Technical Basis for the EPA's Development of the Significant Impact Thresholds for PM2.5 and Ozone

EPA-454/R-18-001
April 2018

Technical Basis for the EPA's Development of the Significant Impact Thresholds for PM2.5 and Ozone

U.S. Environmental Protection Agency
Office of Air Quality Planning and Standards
Air Quality Assessment Division
Research Triangle Park, NC

Contents

List of Figures

1 Bootstrap examples

Bootstrap examples from selects $PM_{2.5}$ sites for the 2008-2010 DV period. Top left, top right, and middle left plots show the distribution of daily PM concentrations for 2008, 2009, and 2010, respectively. The vertical red line shows the annual mean and the vertical blue line shows the annual 98th percentile. Middle left plots show sample distributions of resampled data from 2008, along with the annual mean and the 98th percentile from each resample. The bottom left plots show the distribution of the annual DVs from the 20,000 resampled DV periods (2008-2010). The bottom right plots show the distribution of the 24-hr DVs from the 20,000 resampled DV periods (2008-2010)

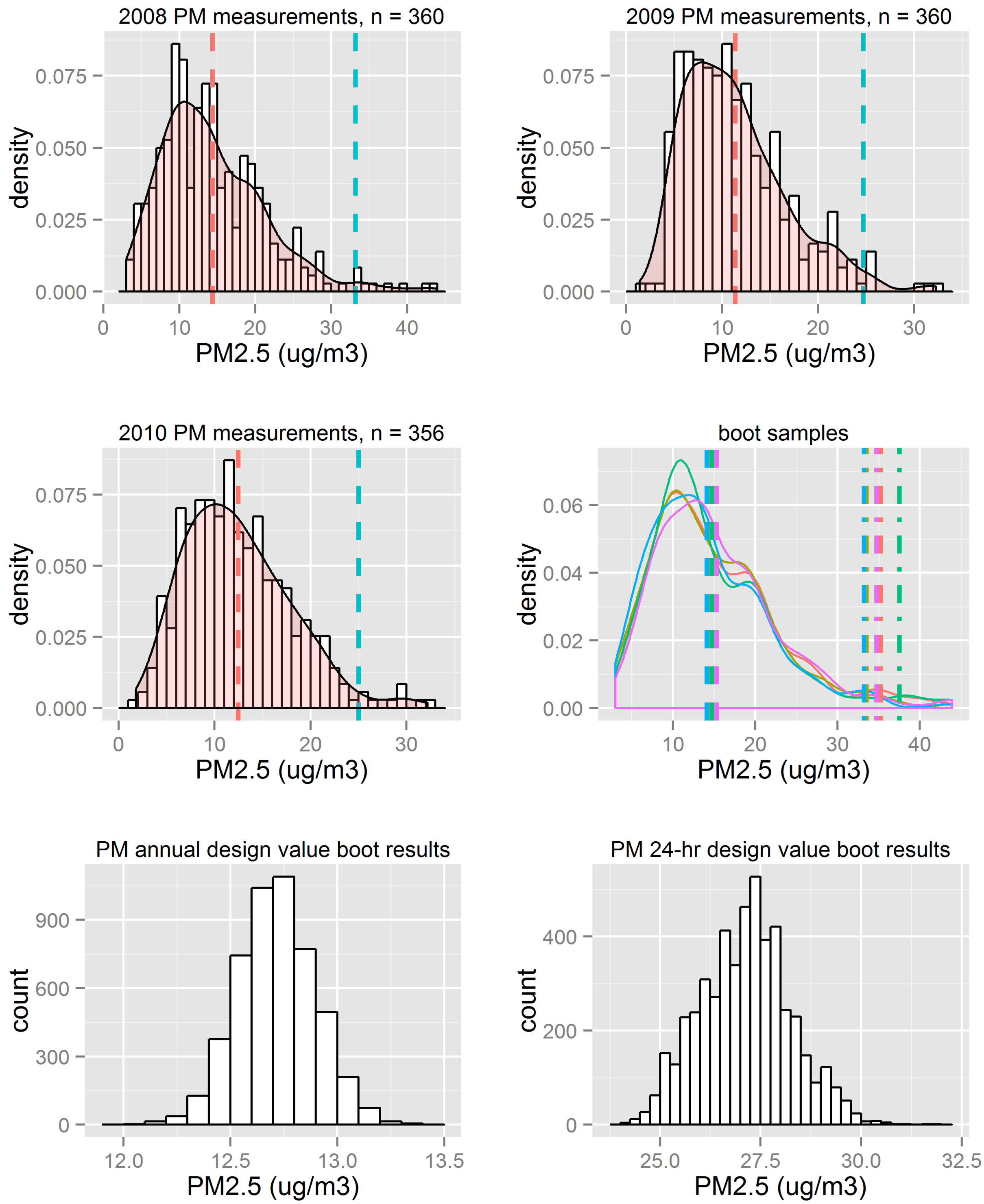


Figure 1: Example from site 10732003.

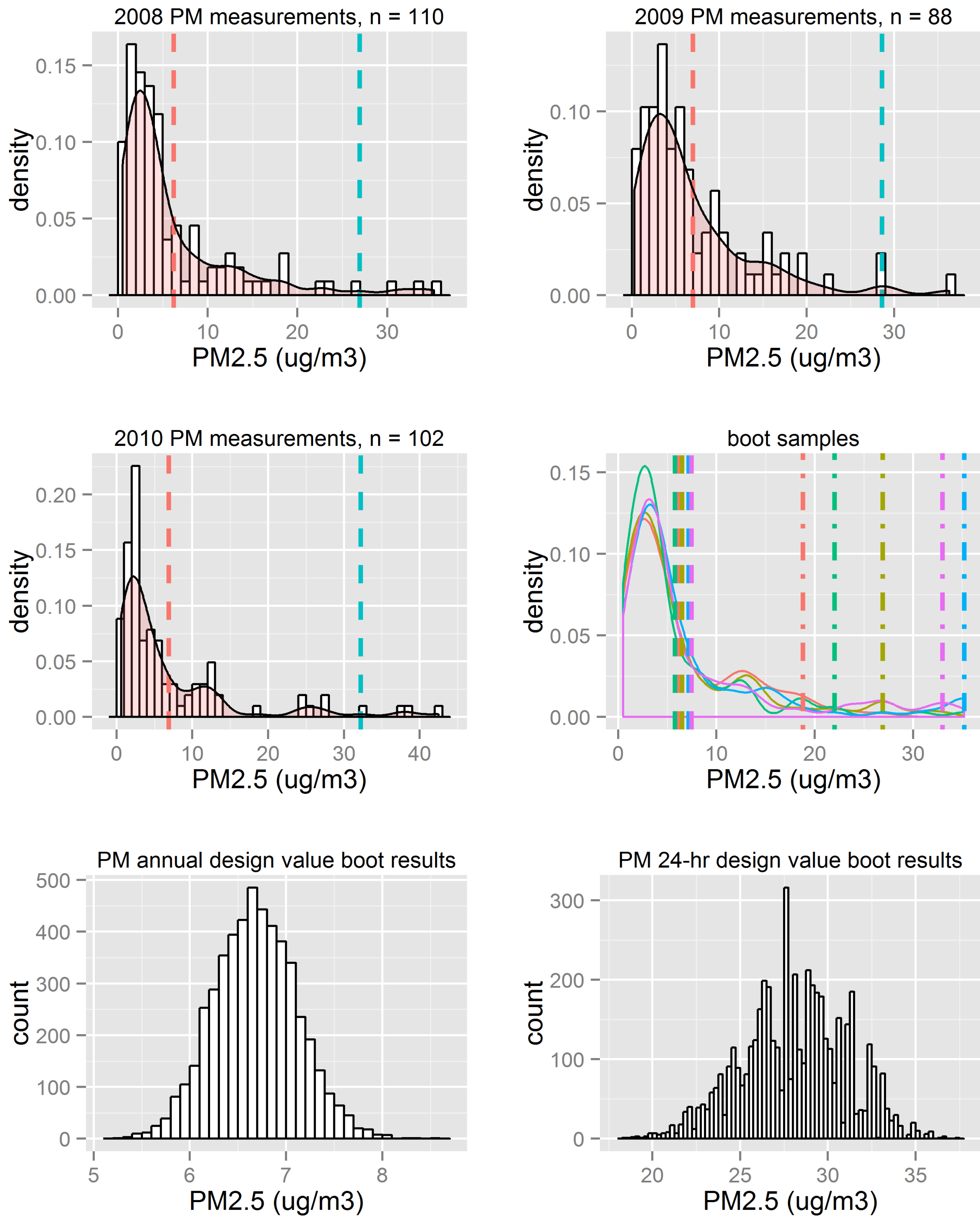


Figure 2: Example from site 21700008.

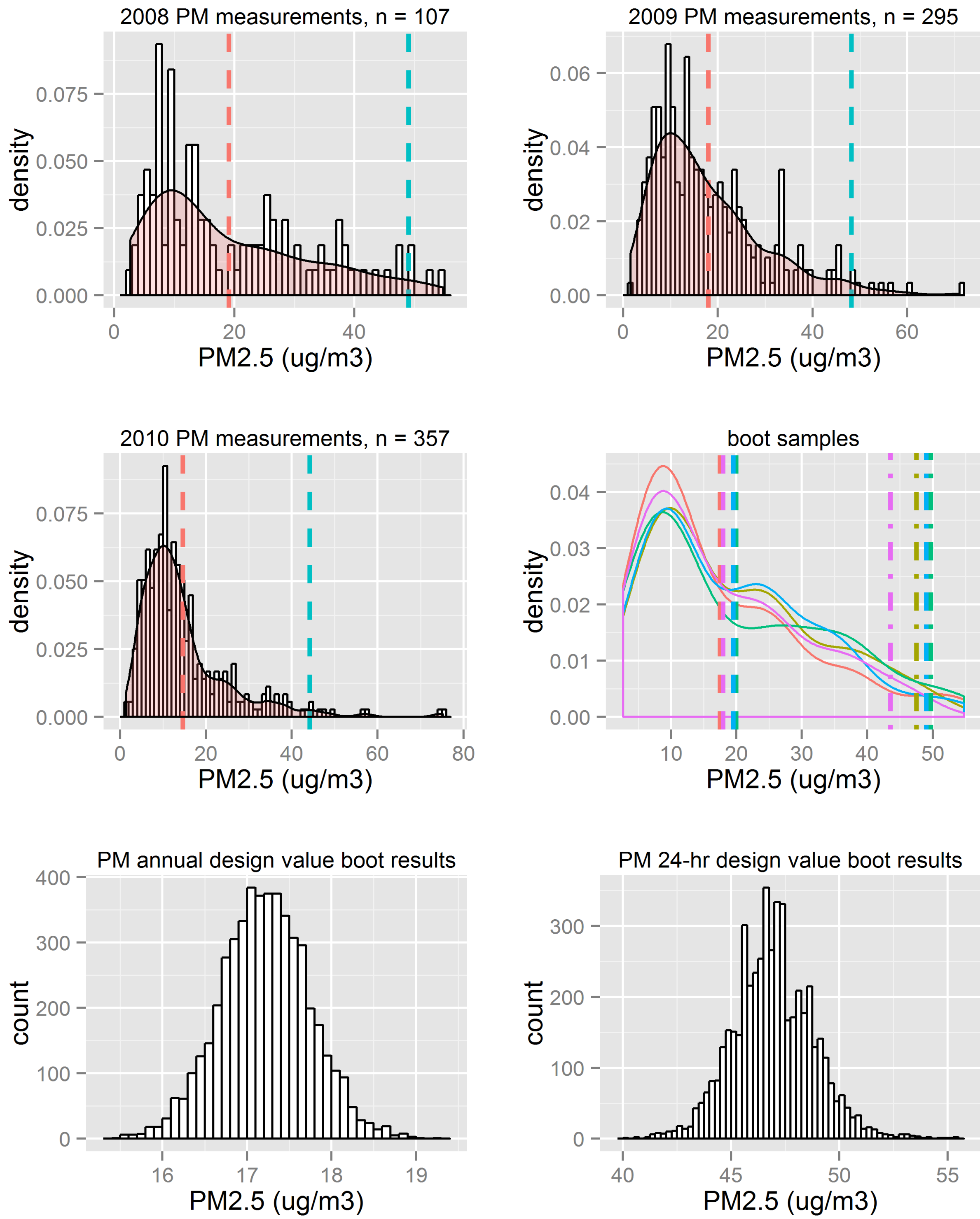


Figure 3: Example from site 60195001.

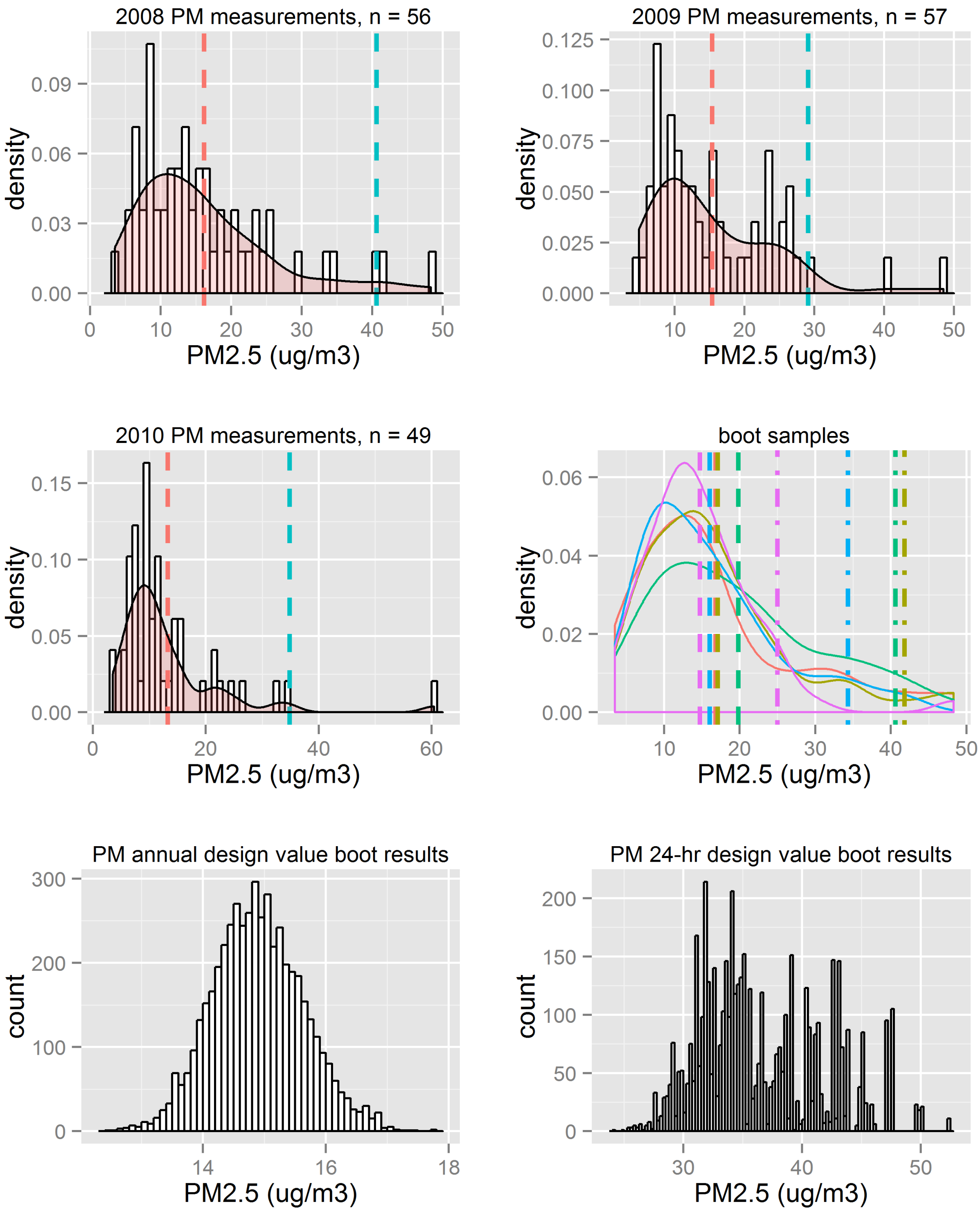


Figure 4: Example from site 481410053.

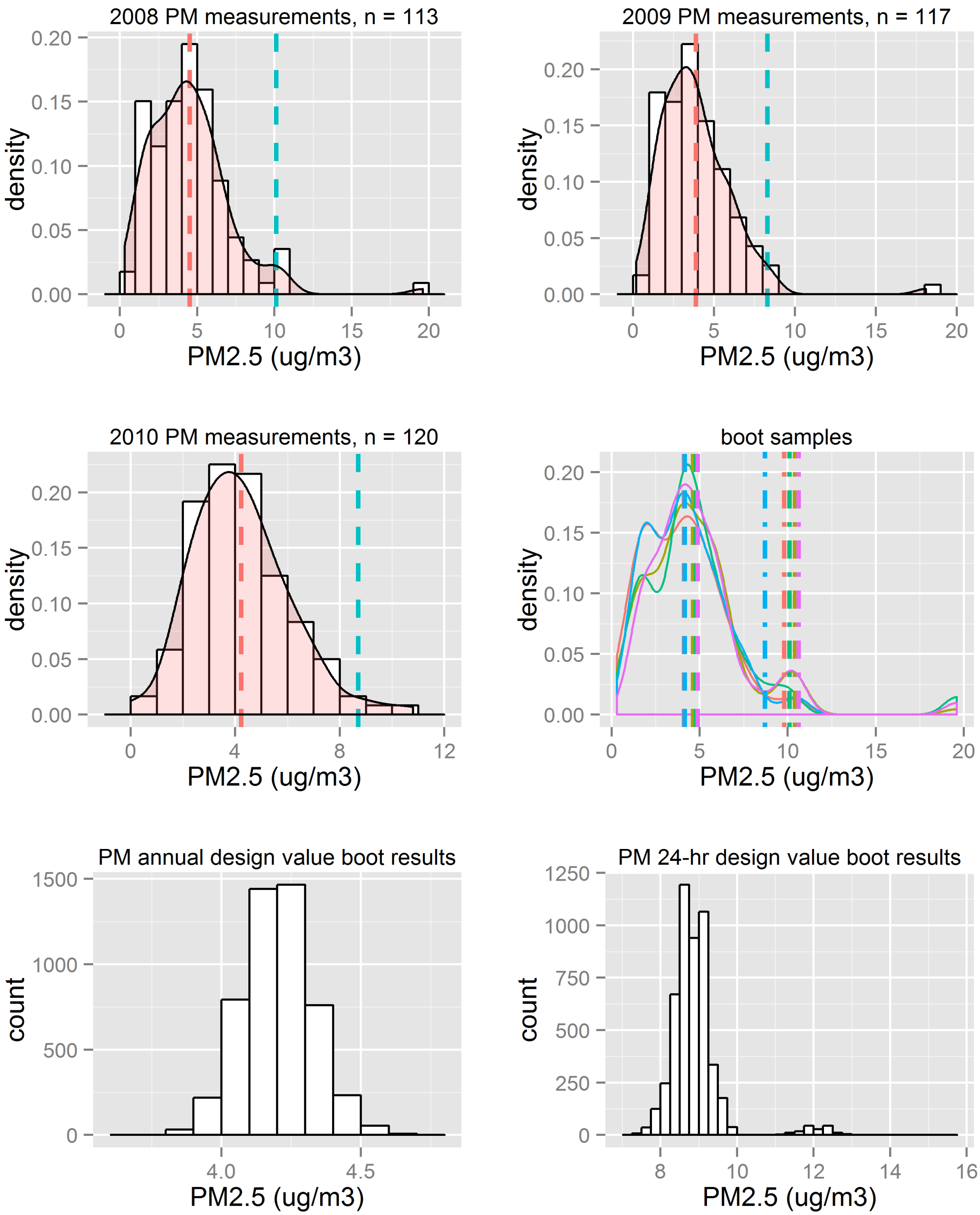


Figure 5: Example from site 560210001.

2 Ozone results

Bootstrap results for ozone data from the years 2000-2013. Each section containts a single DV period,*e.g.*, the results for 2015 include data from 2013-2015.

2.1 2013-2015 ozone bootstrap results

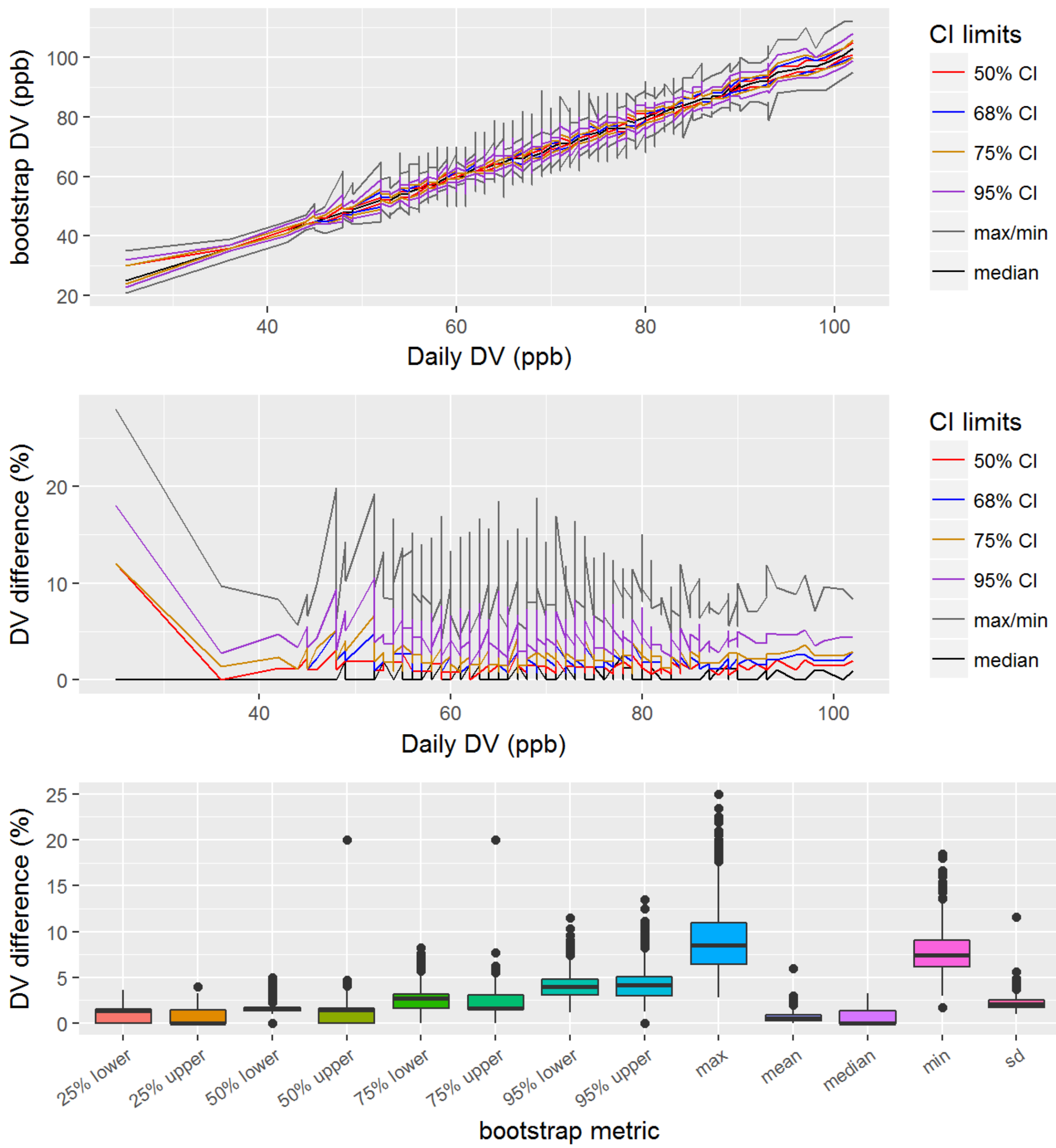


Figure 6: Bootstrap results for the ozone 2015 DVs, showing the 50%, 68%, 75%, and 95% CIs, along with the mean and median bootstrap DVs. The top panel shows the DVs at the various CIs, the middle panel shows the relative difference between the CI and the actual DV, and the bottom panel shows the distribution of the relative differences between the CI and the actual DV.

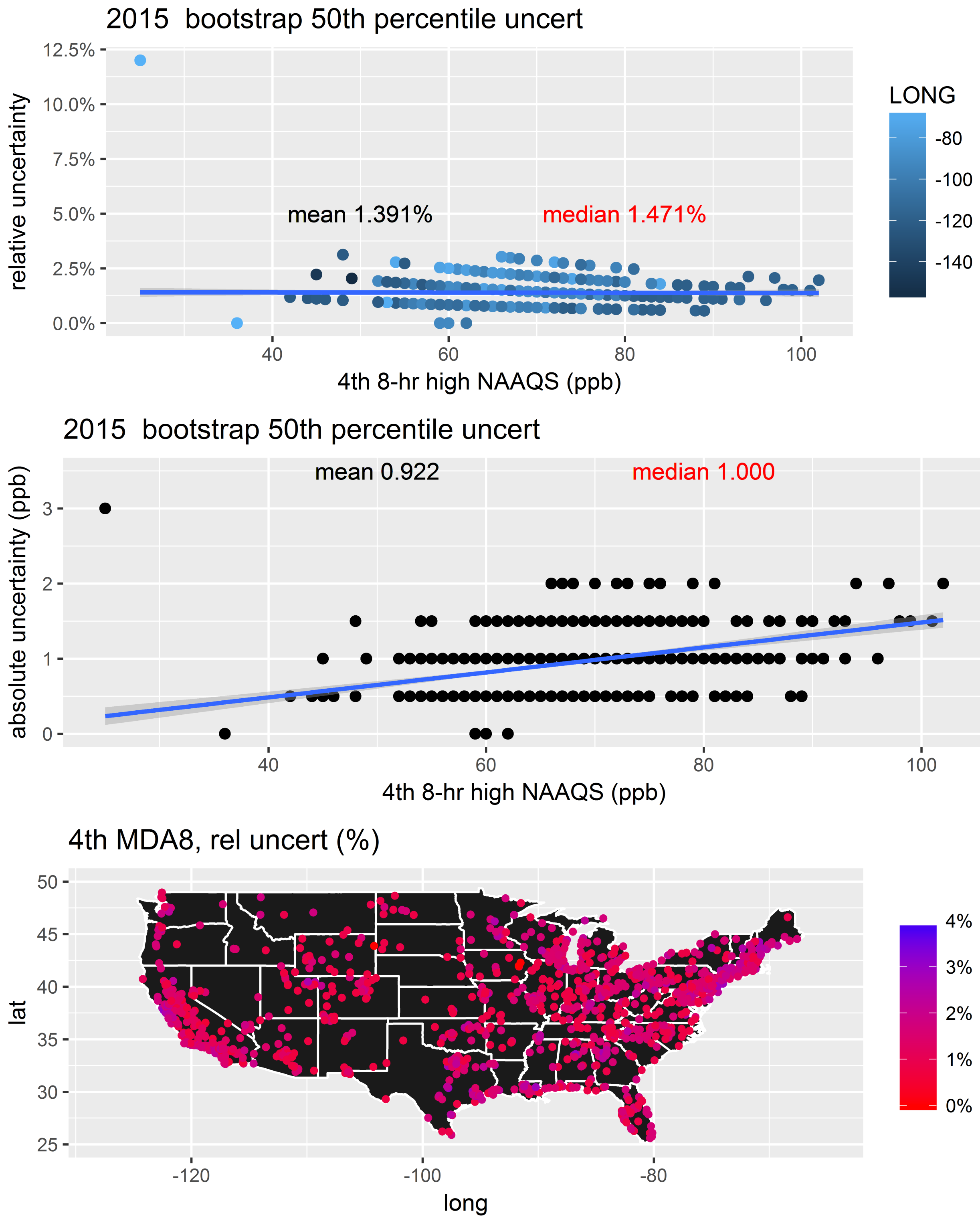


Figure 7: Bootstrap results from the 50% CIs for the 2015 ozone DVs. The top panel shows the relative difference between the CI and the actual DV, the middle panel shows the absolute difference between the values for the DVs at each site and the CI, and the bottom panel shows the spatial distribution of the relative difference between the 50% CIs for the 2015 ozone DV at each site.

2.2 2012-2014 ozone bootstrap results

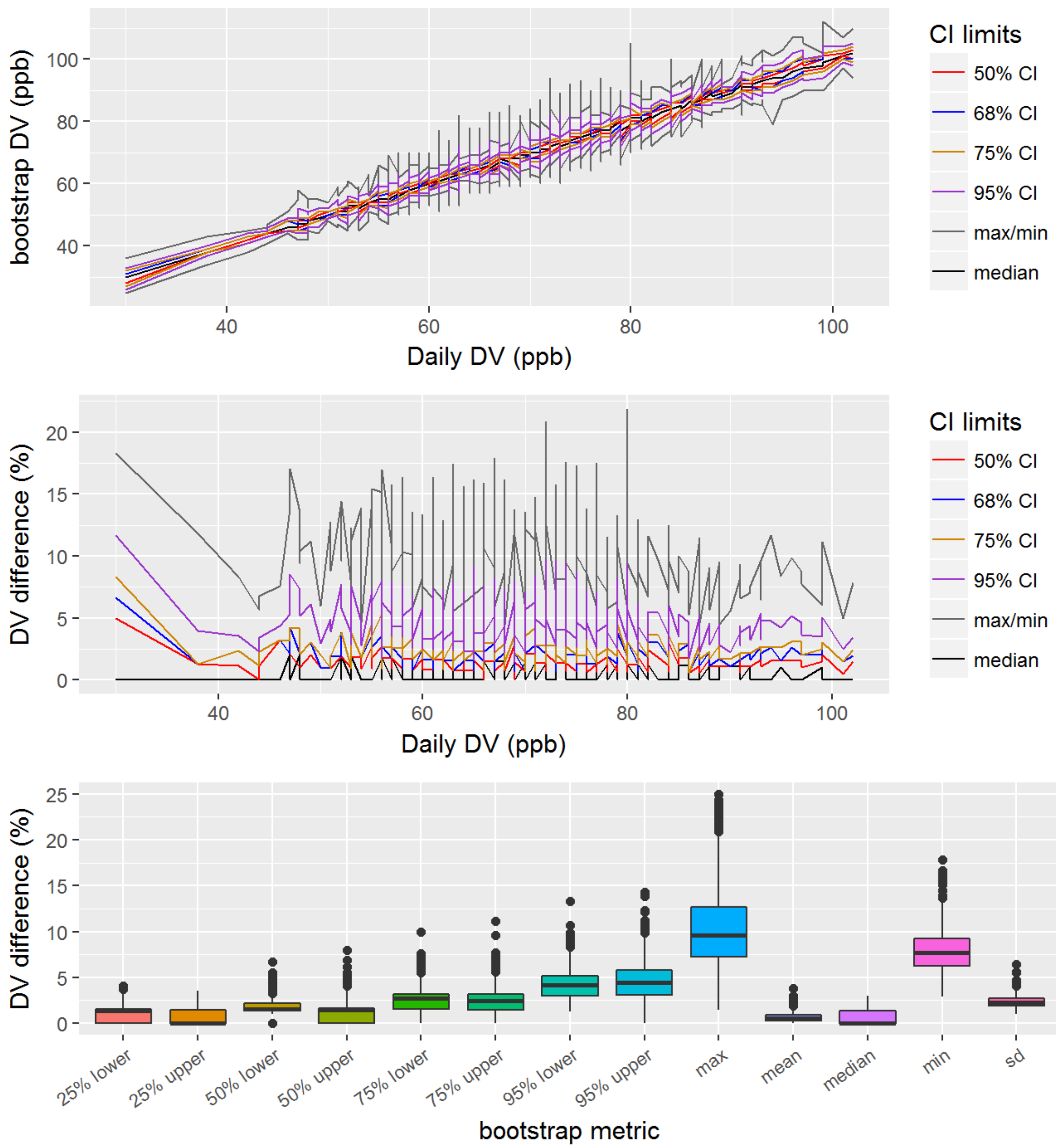


Figure 8: Bootstrap results for the ozone 2014 DVs, showing the 50%, 68%, 75%, and 95% CIs, along with the mean and median bootstrap DVs. The top panel shows the DVs at the various CIs, the middle panel shows the relative difference between the CI and the actual DV, and the bottom panel shows the distribution of the relative differences between the CI and the actual DV.

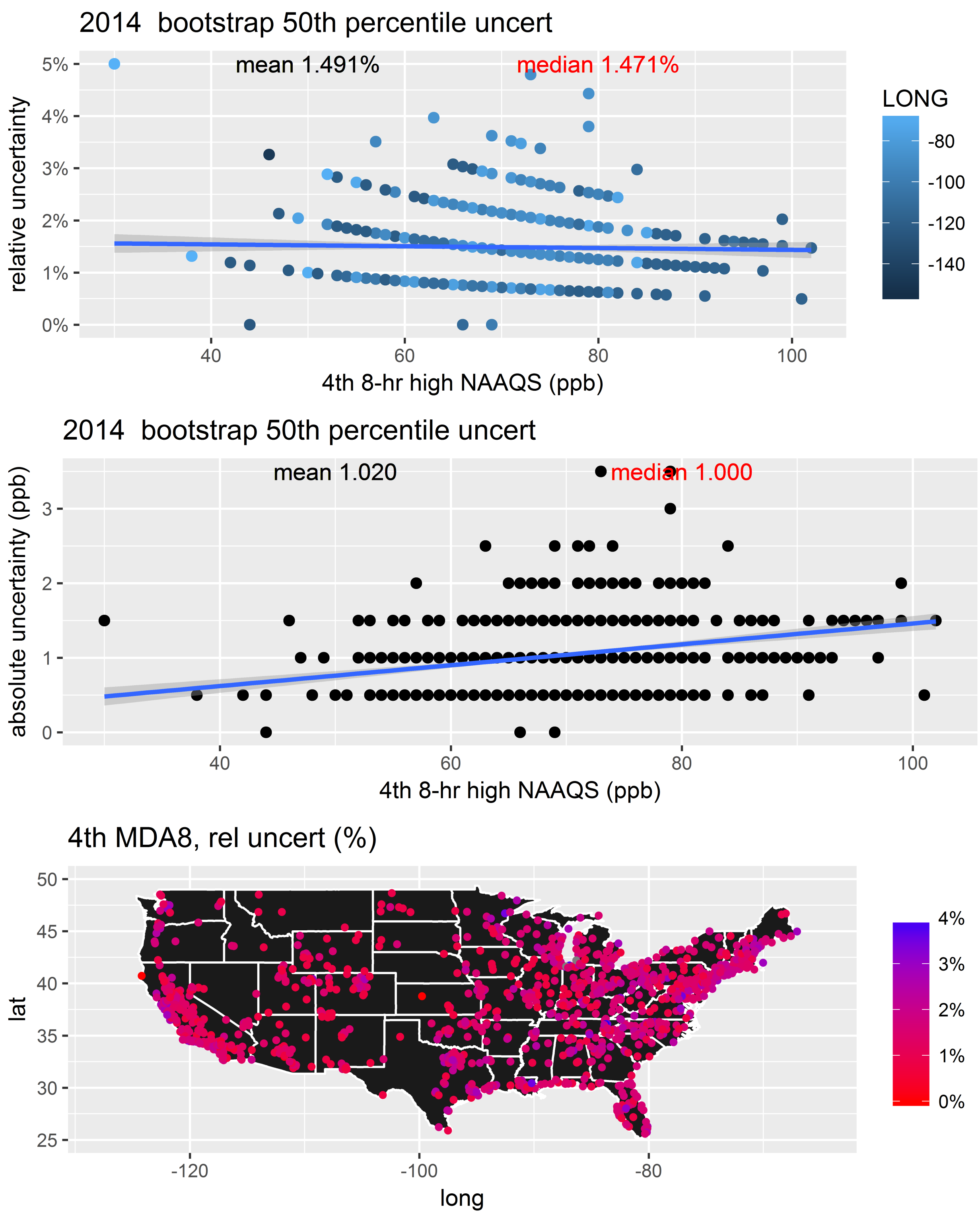


Figure 9: Bootstrap results from the 50% CIs for the 2014 ozone DVs. The top panel shows the relative difference between the CI and the actual DV, the middle panel shows the absolute difference between the values for the DVs at each site and the CI, and the bottom panel shows the spatial distribution of the relative difference between the 50% CIs for the 2014 ozone DV at each site.

2.3 2011-2013 ozone bootstrap results

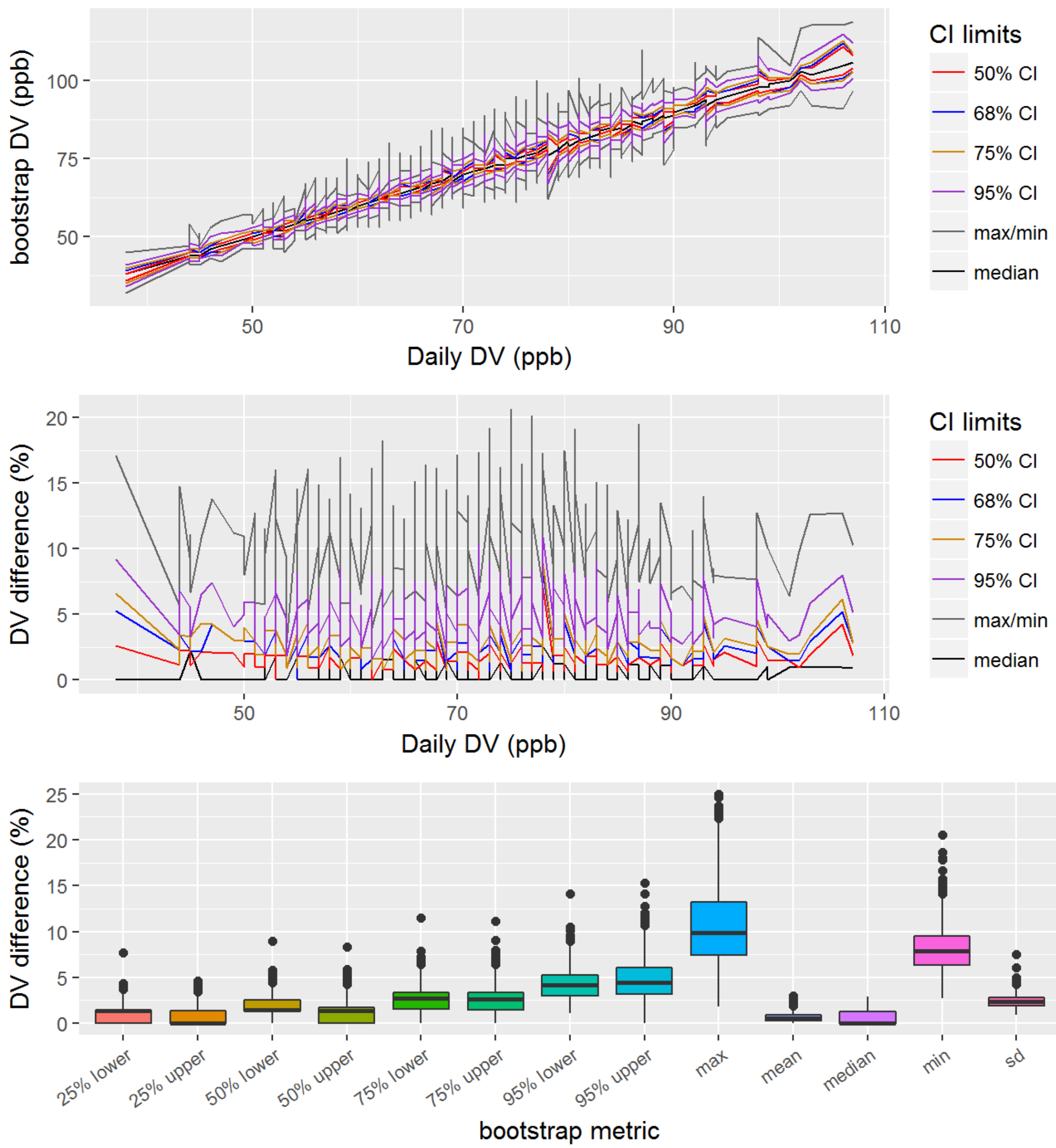


Figure 10: Bootstrap results for the ozone 2013 DVs, showing the 50%, 68%, 75%, and 95% CIs, along with the mean and median bootstrap DVs. The top panel shows the DVs at the various CIs, the middle panel shows the relative difference between the CI and the actual DV, and the bottom panel shows the distribution of the relative differences between the CI and the actual DV.

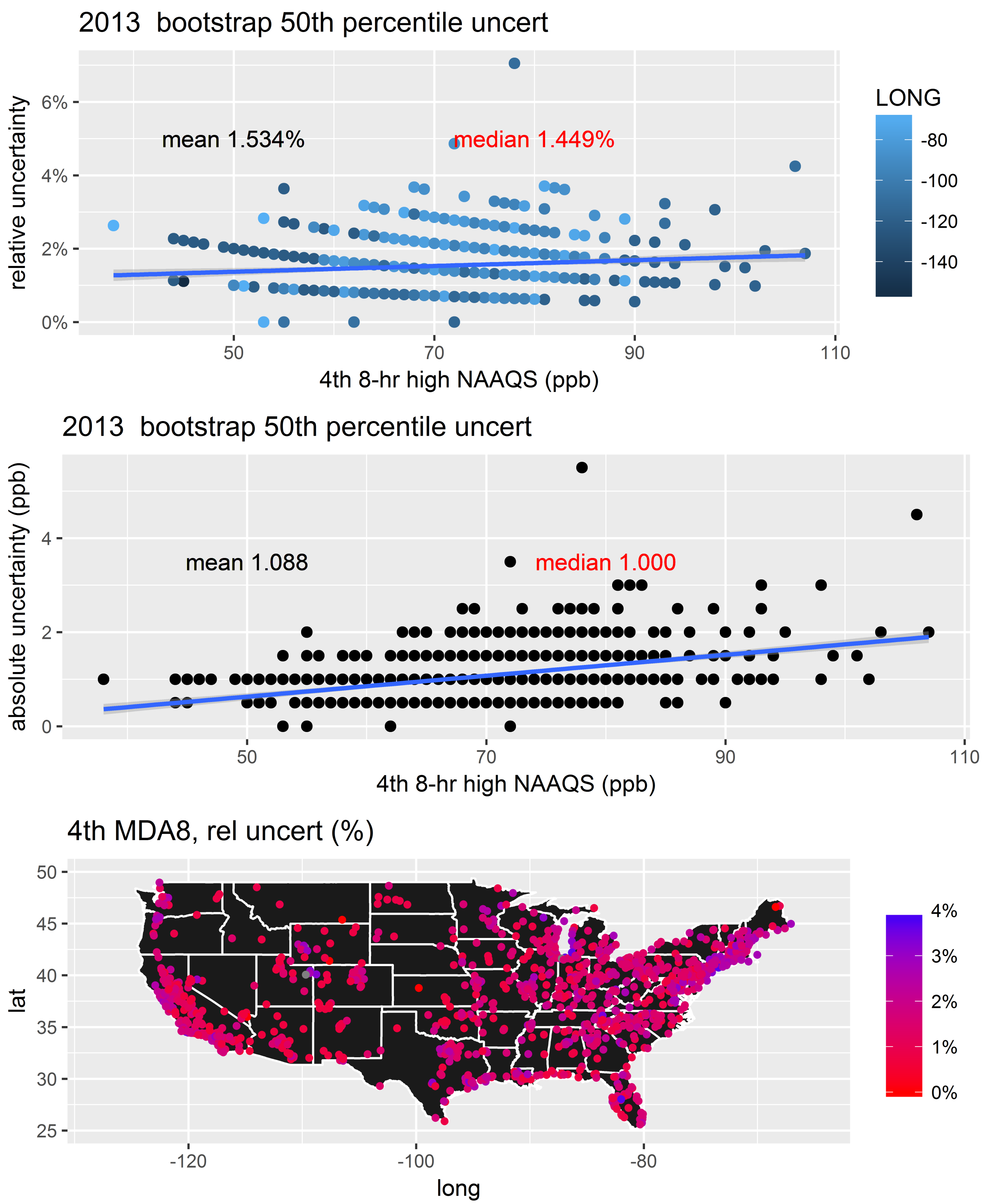


Figure 11: Bootstrap results from the 50% CIs for the 2013 ozone DVs. The top panel shows the relative difference between the CI and the actual DV, the middle panel shows the absolute difference between the values for the DVs at each site and the CI, and the bottom panel shows the spatial distribution of the relative difference between the 50% CIs for the 2013 ozone DV at each site.

2.4 2010-2012 ozone bootstrap results

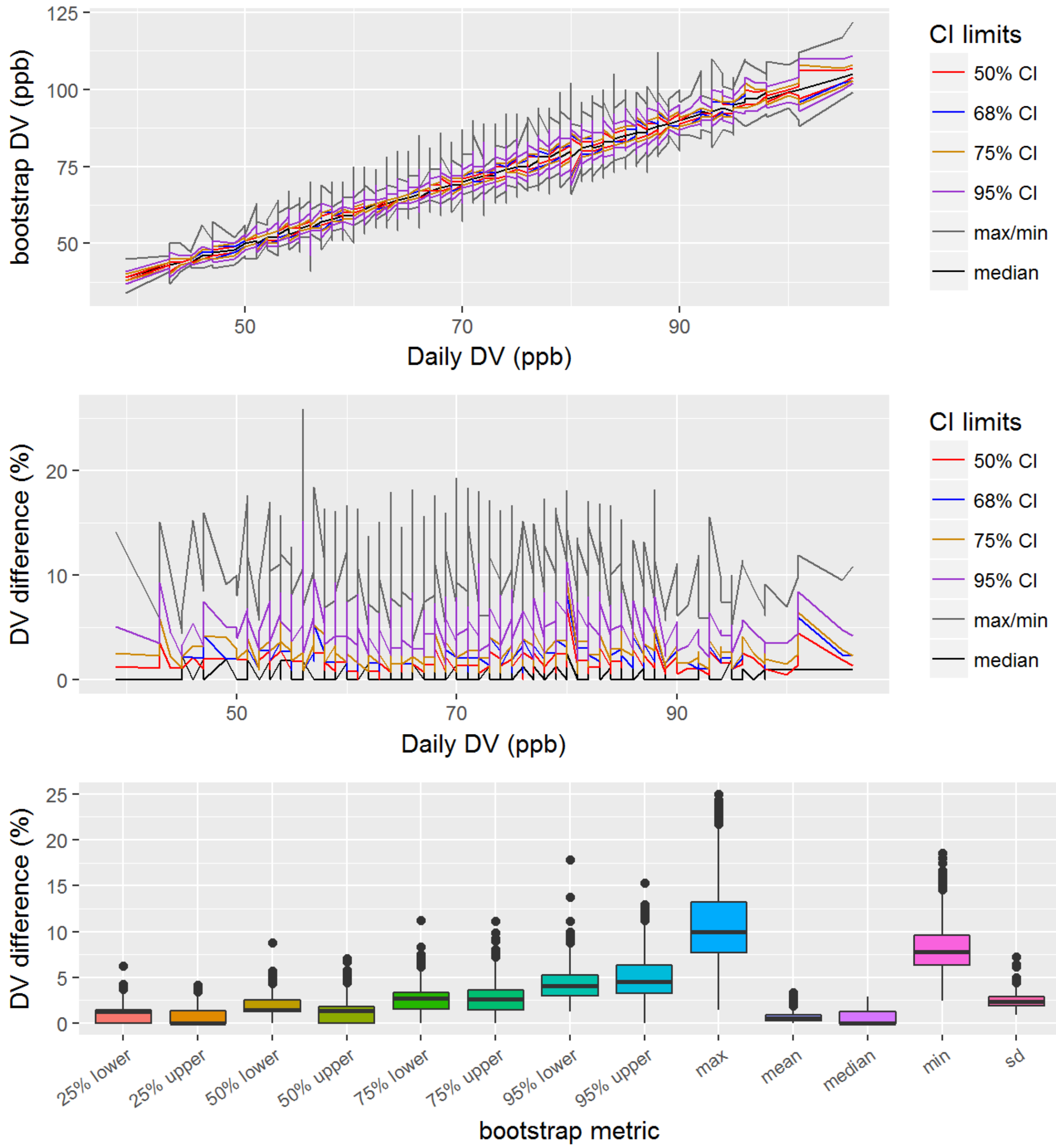


Figure 12: Bootstrap results for the ozone 2012 DVs, showing the 50%, 68%, 75%, and 95% CIs, along with the mean and median bootstrap DVs. The top panel shows the DVs at the various CIs, the middle panel shows the relative difference between the CI and the actual DV, and the bottom panel shows the distribution of the relative differences between the CI and the actual DV.

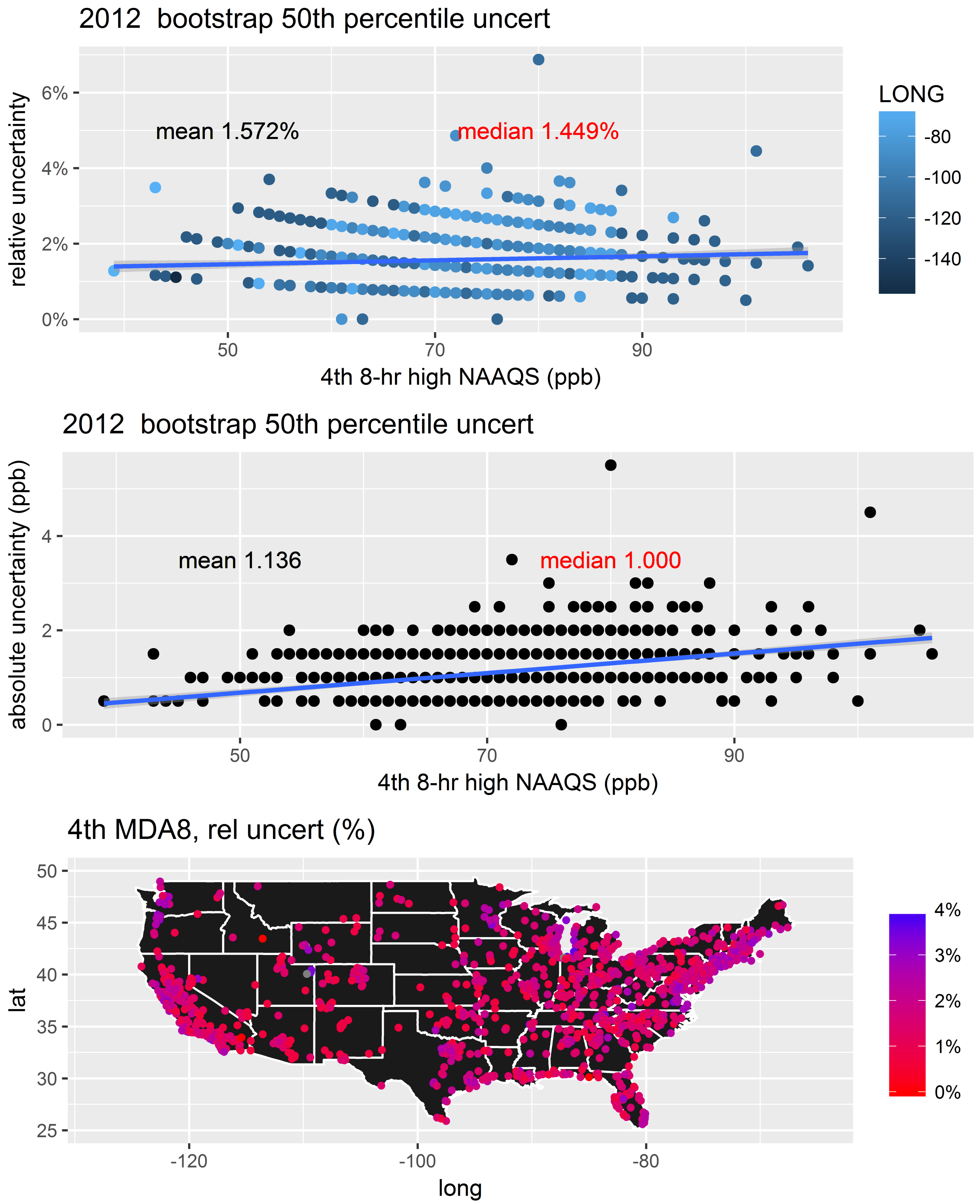


Figure 13: Bootstrap results from the 50% CIs for the 2012 ozone DVs. The top panel shows the relative difference between the CI and the actual DV, the middle panel shows the absolute difference between the values for the DVs at each site and the CI, and the bottom panel shows the spatial distribution of the relative difference between the 50% CIs for the 2012 ozone DV at each site.

2.5 2009-2011 ozone bootstrap results

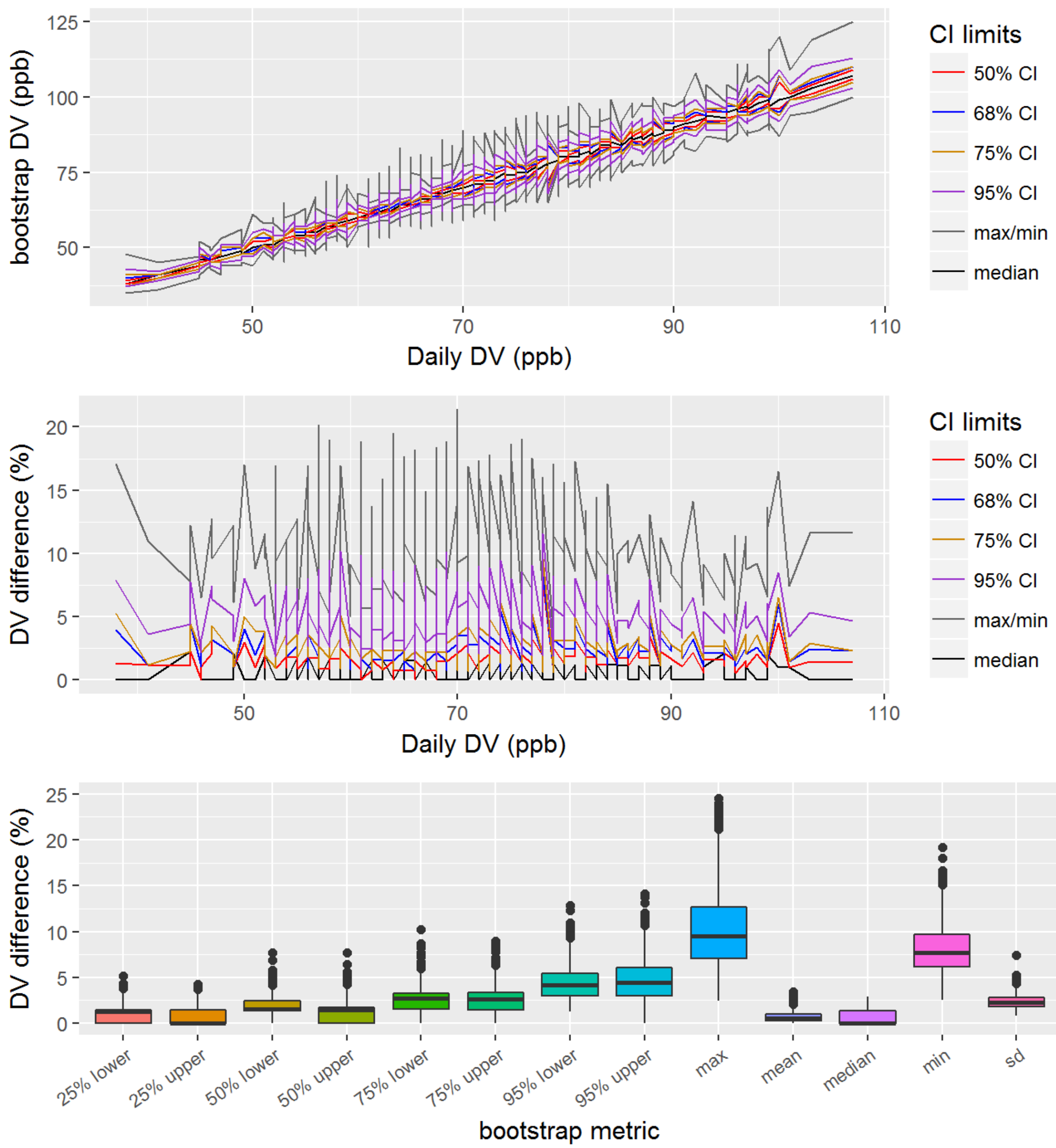


Figure 14: Bootstrap results for the ozone 2011 DVs, showing the 50%, 68%, 75%, and 95% CIs, along with the mean and median bootstrap DVs. The top panel shows the DVs at the various CIs, the middle panel shows the relative difference between the CI and the actual DV, and the bottom panel shows the distribution of the relative differences between the CI and the actual DV.

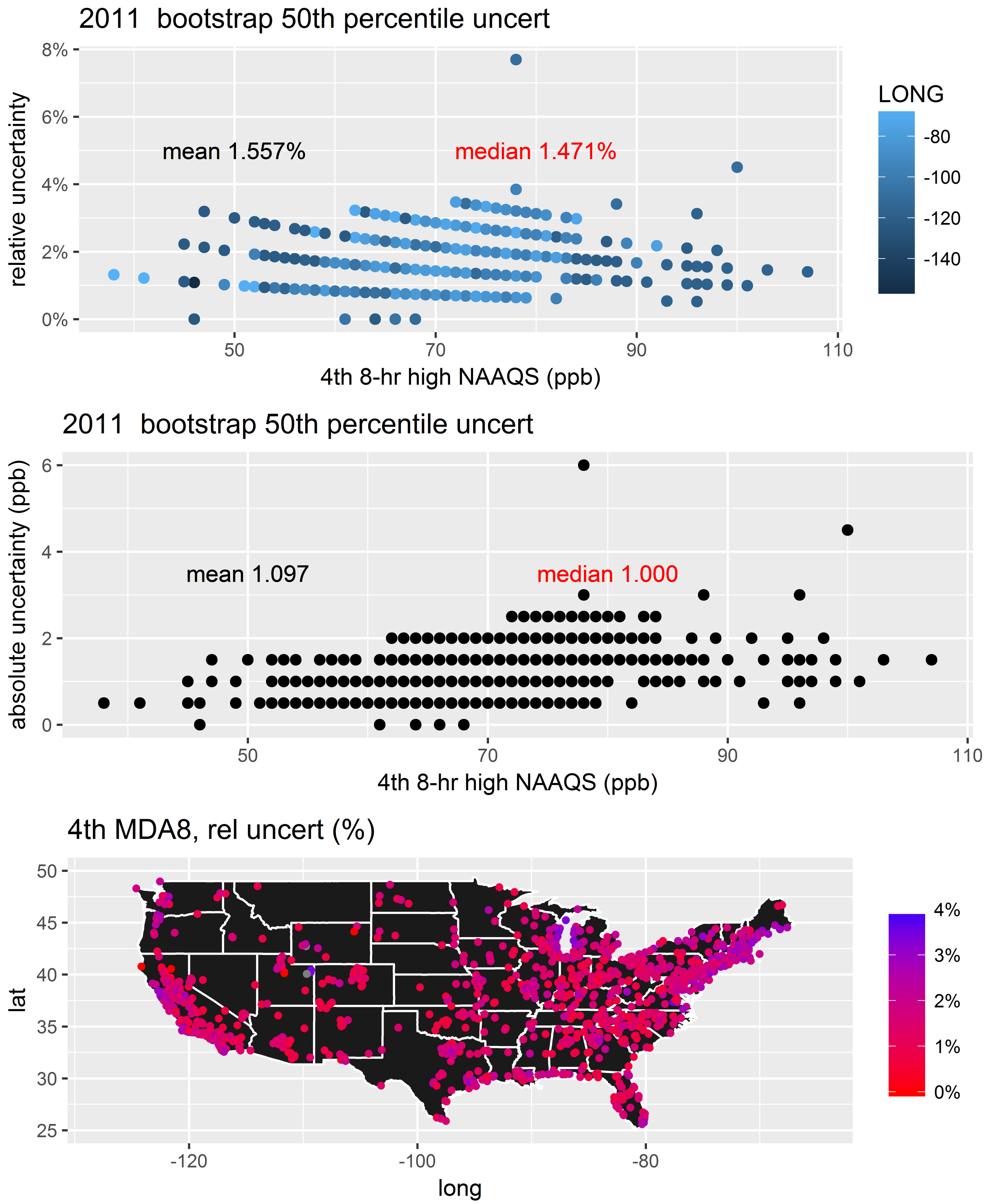


Figure 15: Bootstrap results from the 50% CIs for the 2011 ozone DVs. The top panel shows the relative difference between the CI and the actual DV, the middle panel shows the absolute difference between the values for the DVs at each site and the CI, and the bottom panel shows the spatial distribution of the relative difference between the 50% CIs for the 2011 ozone DV at each site.

2.6 2008-2010 ozone bootstrap results

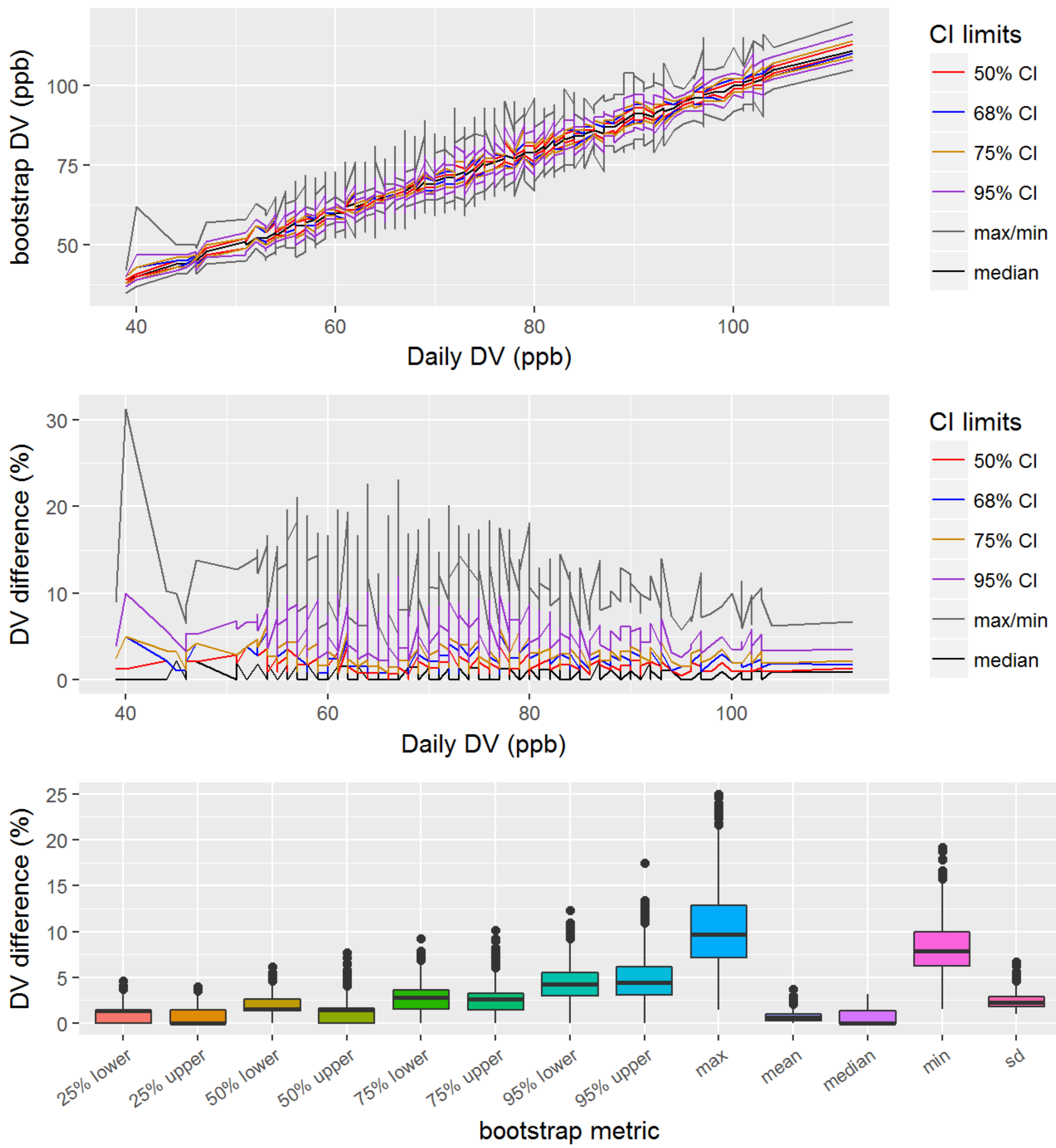


Figure 16: Bootstrap results for the ozone 2010 DVs, showing the 50%, 68%, 75%, and 95% CIs, along with the mean and median bootstrap DVs. The top panel shows the DVs at the various CIs, the middle panel shows the relative difference between the CI and the actual DV, and the bottom panel shows the distribution of the relative differences between the CI and the actual DV.

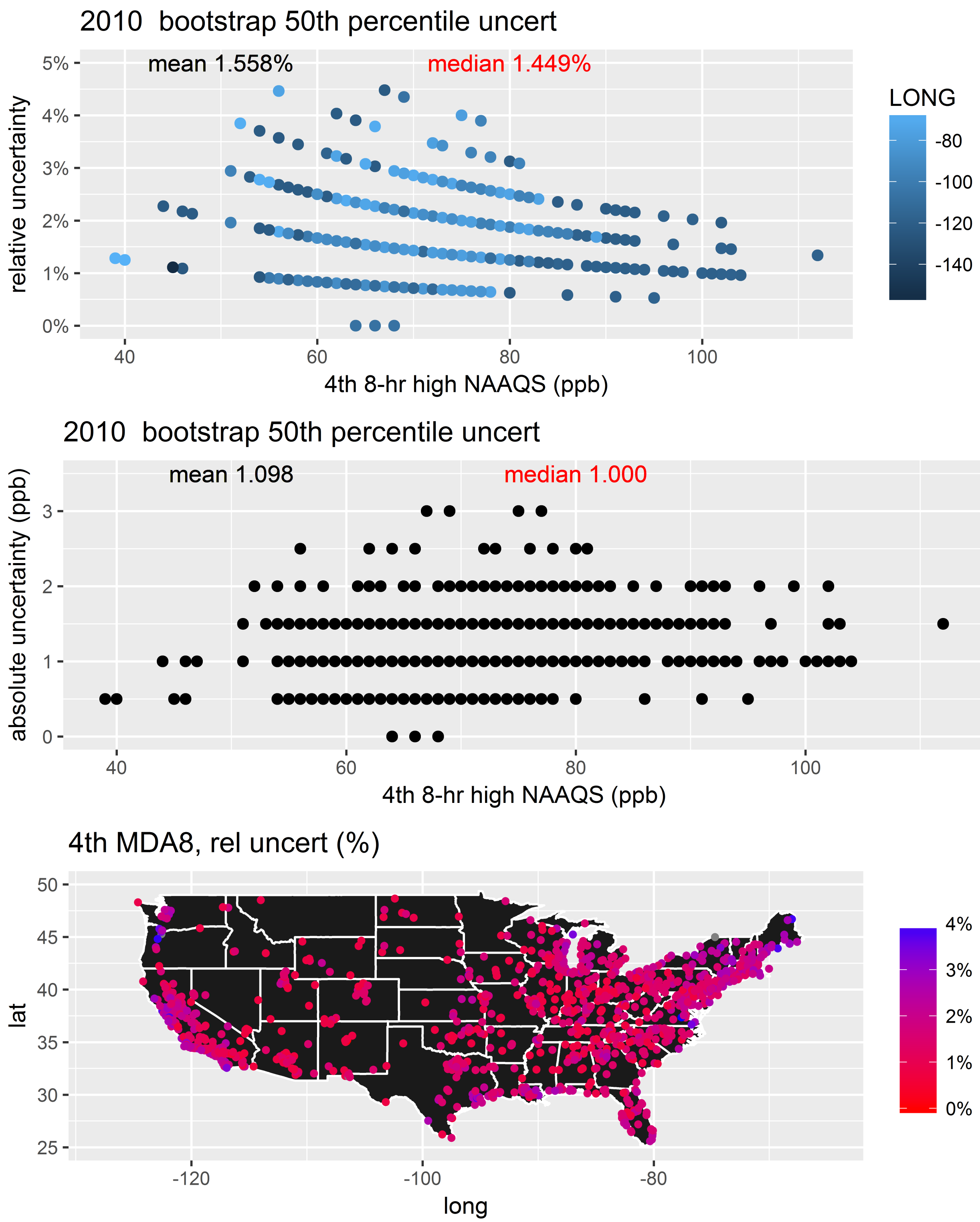


Figure 17: Bootstrap results from the 50% CIs for the 2010 ozone DVs. The top panel shows the relative difference between the CI and the actual DV, the middle panel shows the absolute difference between the values for the DVs at each site and the CI, and the bottom panel shows the spatial distribution of the relative difference between the 50% CIs for the 2010 ozone DV at each site.

2.7 2007-2009 ozone bootstrap results

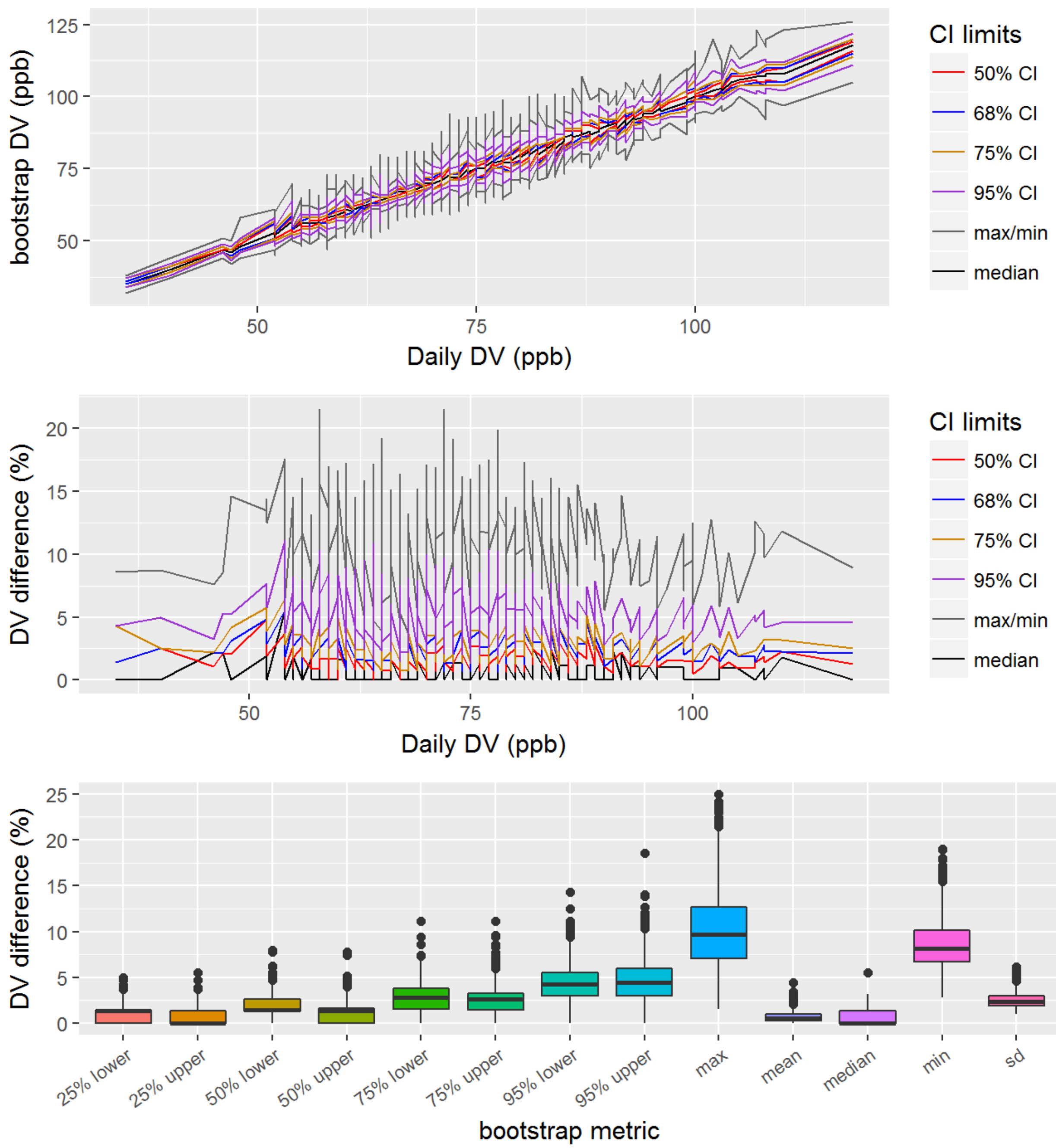


Figure 18: Bootstrap results for the ozone 2009 DVs, showing the 50%, 68%, 75%, and 95% CIs, along with the mean and median bootstrap DVs. The top panel shows the DVs at the various CIs, the middle panel shows the relative difference between the CI and the actual DV, and the bottom panel shows the distribution of the relative differences between the CI and the actual DV.

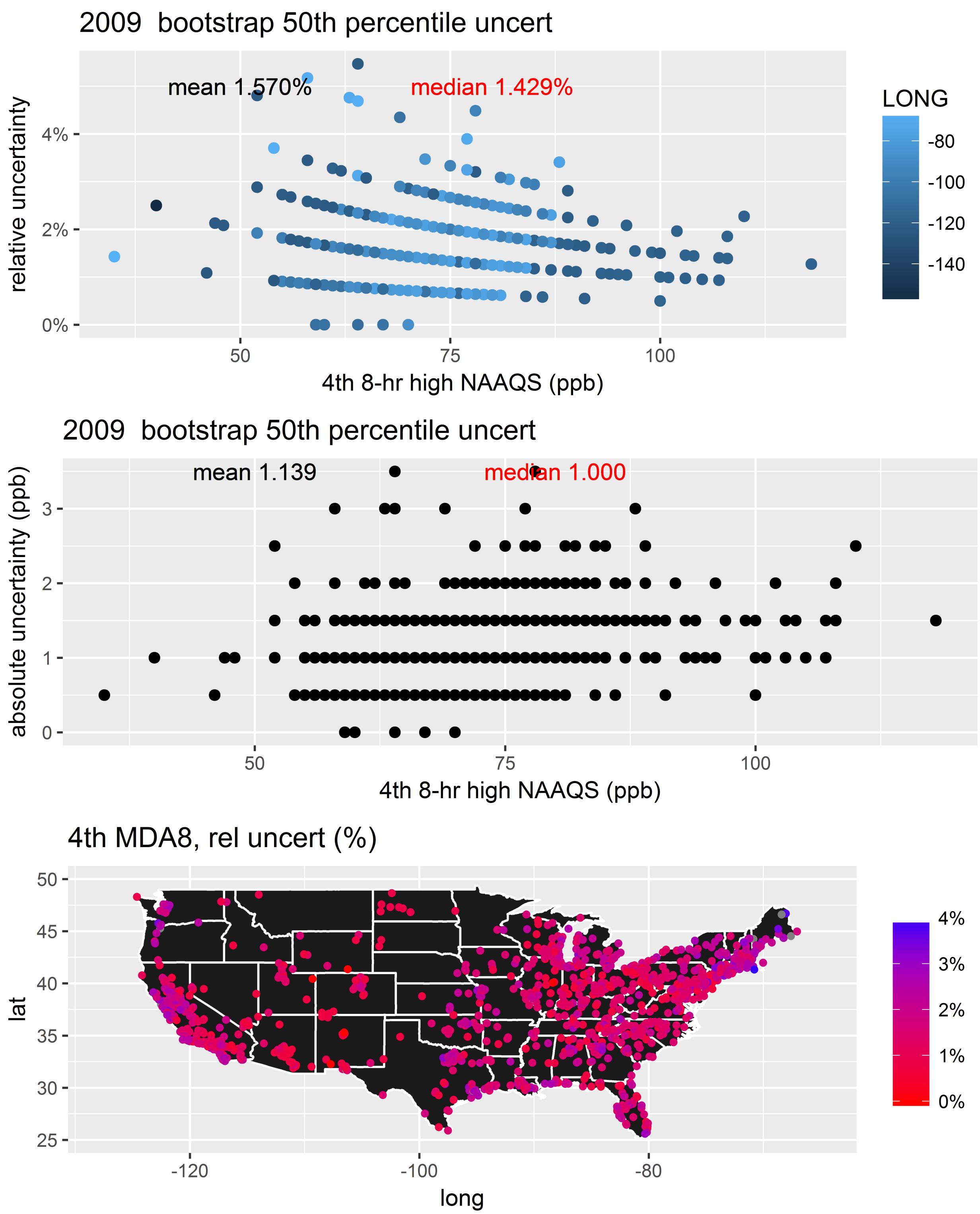


Figure 19: Bootstrap results from the 50% CIs for the 2009 ozone DVs. The top panel shows the relative difference between the CI and the actual DV, the middle panel shows the absolute difference between the values for the DVs at each site and the CI, and the bottom panel shows the spatial distribution of the relative difference between the 50% CIs for the 2009 ozone DV at each site.

2.8 2006-2008 ozone bootstrap results

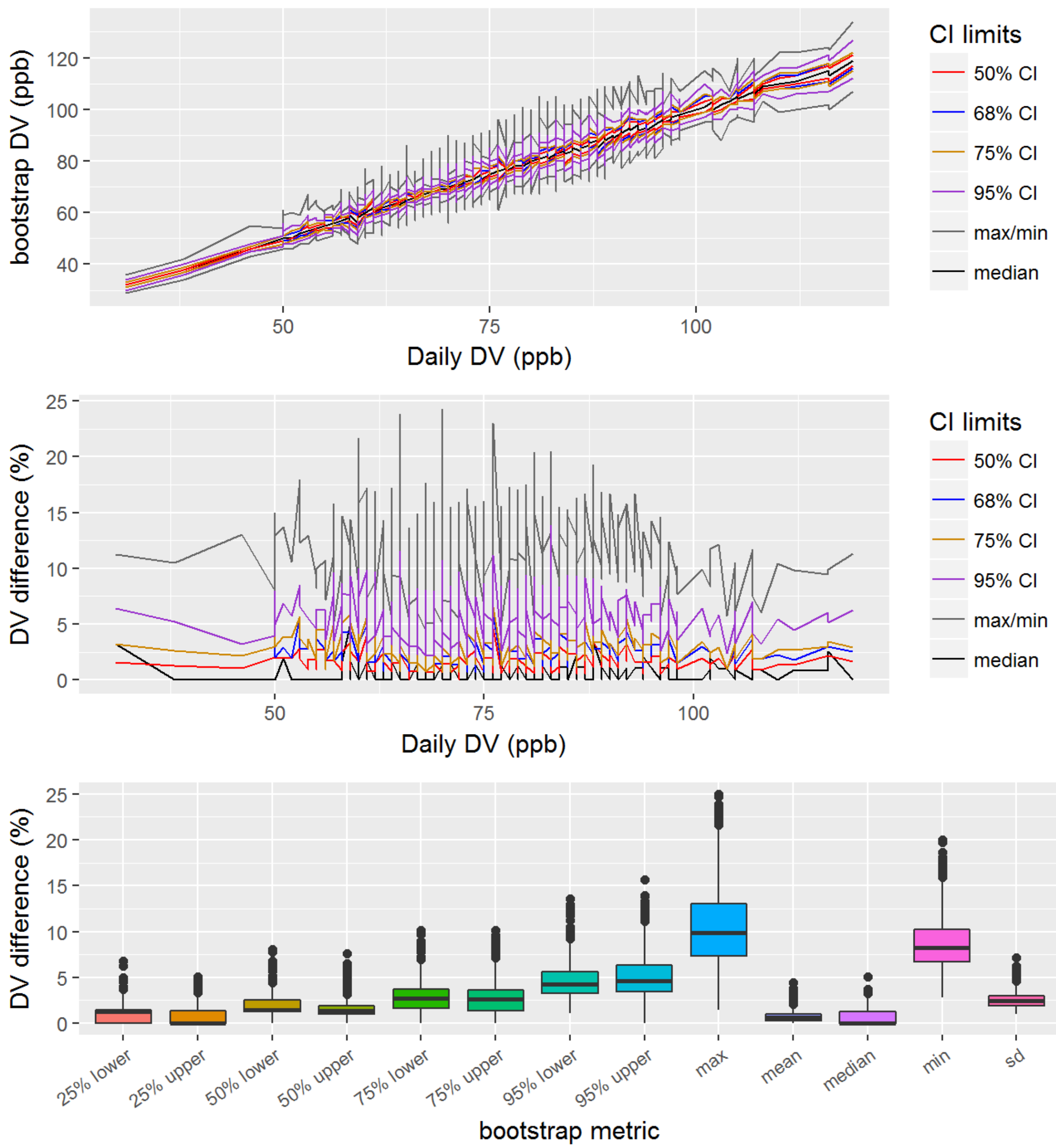


Figure 20: Bootstrap results for the ozone 2008 DVs, showing the 50%, 68%, 75%, and 95% CIs, along with the mean and median bootstrap DVs. The top panel shows the DVs at the various CIs, the middle panel shows the relative difference between the CI and the actual DV, and the bottom panel shows the distribution of the relative differences between the CI and the actual DV.

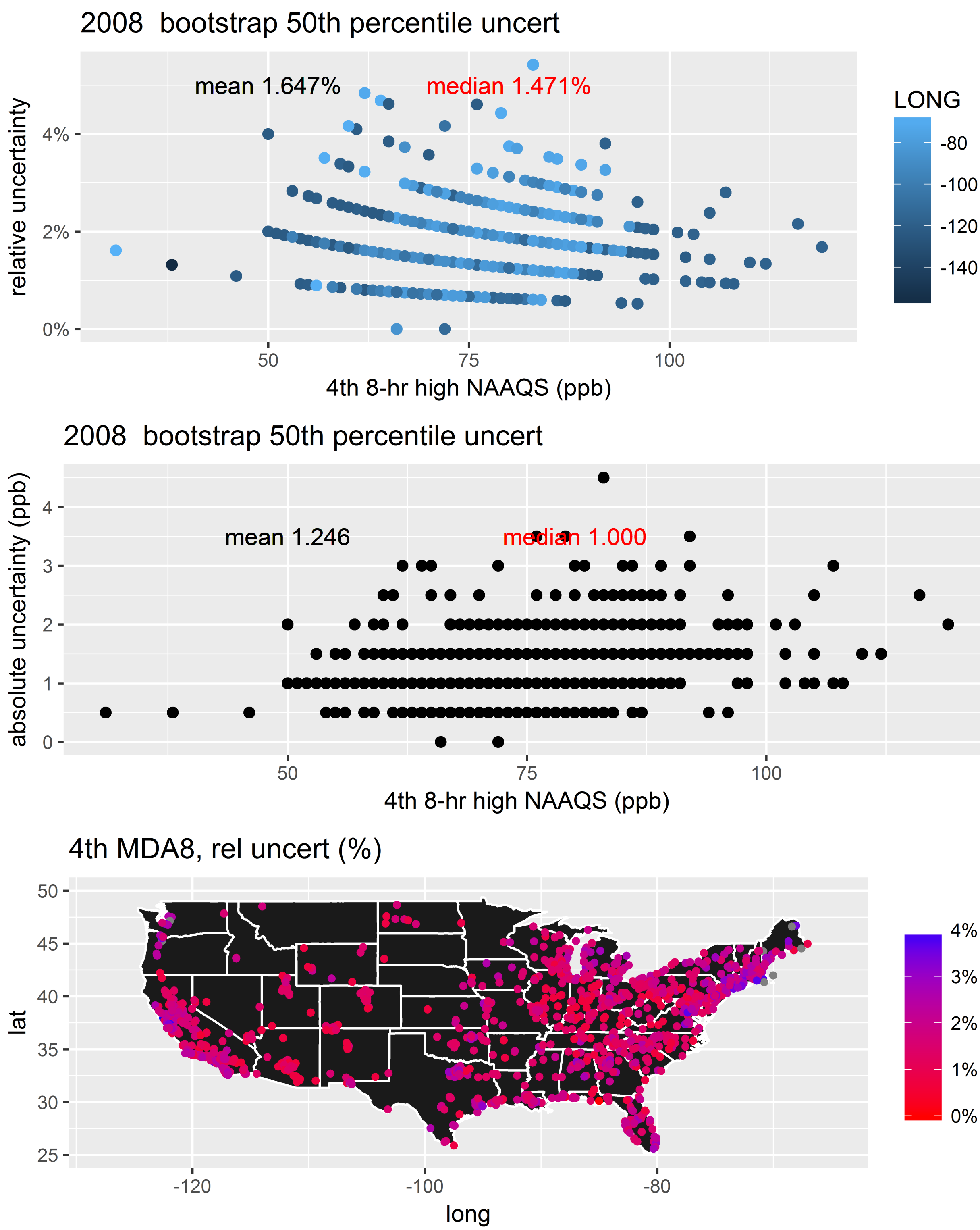


Figure 21: Bootstrap results from the 50% CIs for the 2008 ozone DVs. The top panel shows the relative difference between the CI and the actual DV, the middle panel shows the absolute difference between the values for the DVs at each site and the CI, and the bottom panel shows the spatial distribution of the relative difference between the 50% CIs for the 2008 ozone DV at each site.

2.9 2005-2007 ozone bootstrap results

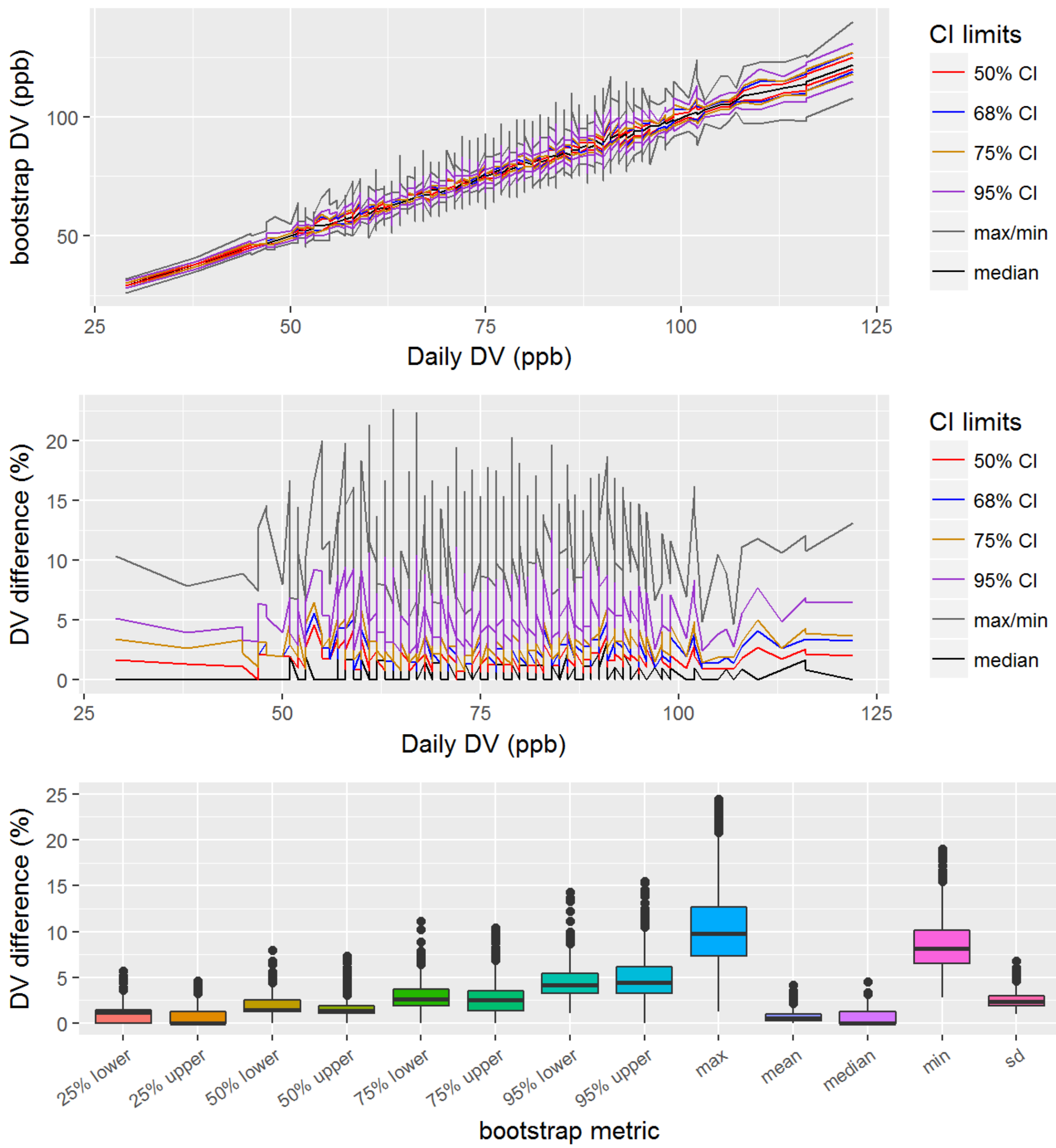


Figure 22: Bootstrap results for the ozone 2007 DVs, showing the 50%, 68%, 75%, and 95% CIs, along with the mean and median bootstrap DVs. The top panel shows the DVs at the various CIs, the middle panel shows the relative difference between the CI and the actual DV, and the bottom panel shows the distribution of the relative differences between the CI and the actual DV.

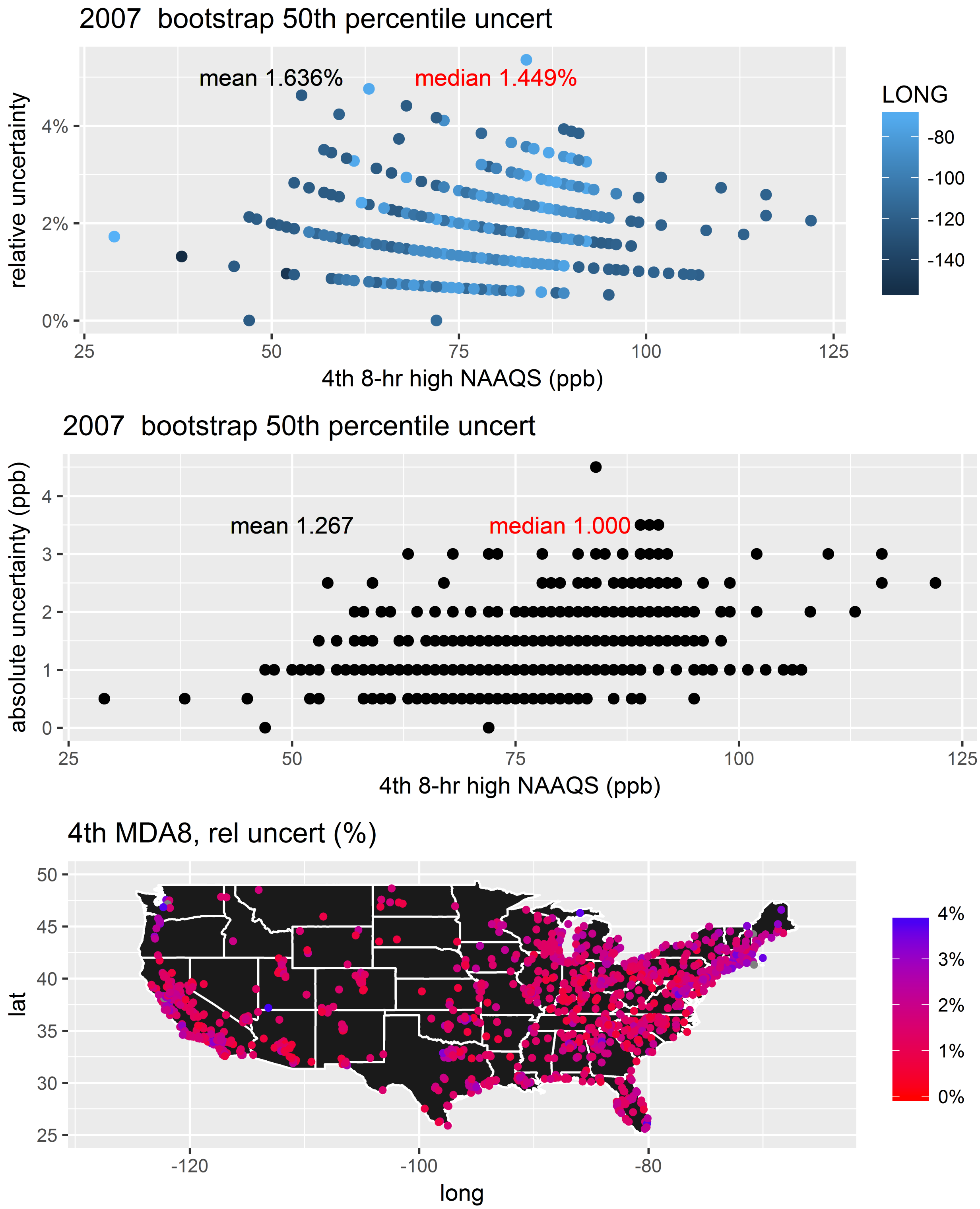


Figure 23: Bootstrap results from the 50% CIs for the 2007 ozone DVs. The top panel shows the relative difference between the CI and the actual DV, the middle panel shows the absolute difference between the values for the DVs at each site and the CI, and the bottom panel shows the spatial distribution of the relative difference between the 50% CIs for the 2007 ozone DV at each site.

2.10 2004-2006 ozone bootstrap results

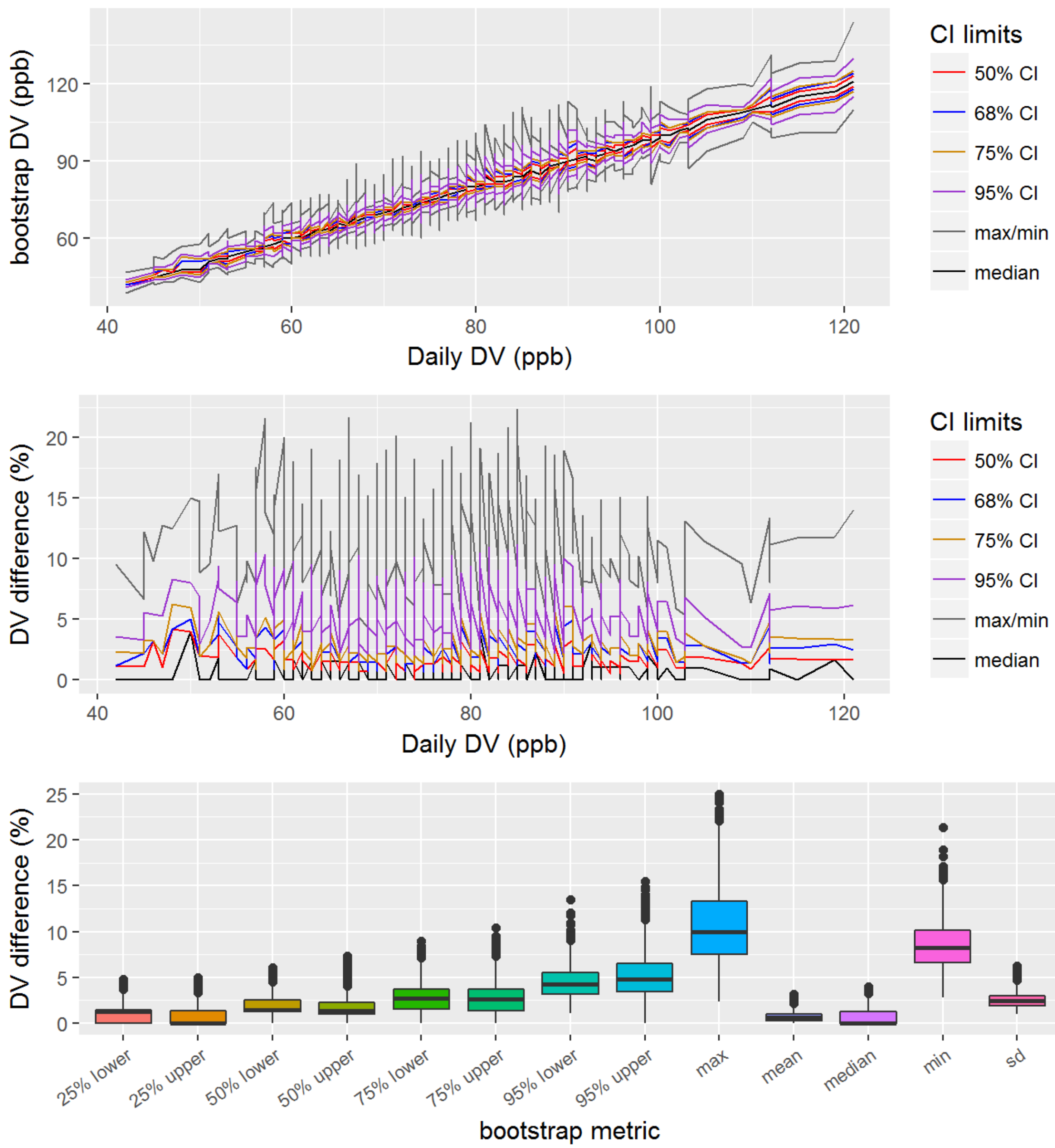


Figure 24: Bootstrap results for the ozone 2006 DVs, showing the 50%, 68%, 75%, and 95% CIs, along with the mean and median bootstrap DVs. The top panel shows the DVs at the various CIs, the middle panel shows the relative difference between the CI and the actual DV, and the bottom panel shows the distribution of the relative differences between the CI and the actual DV.

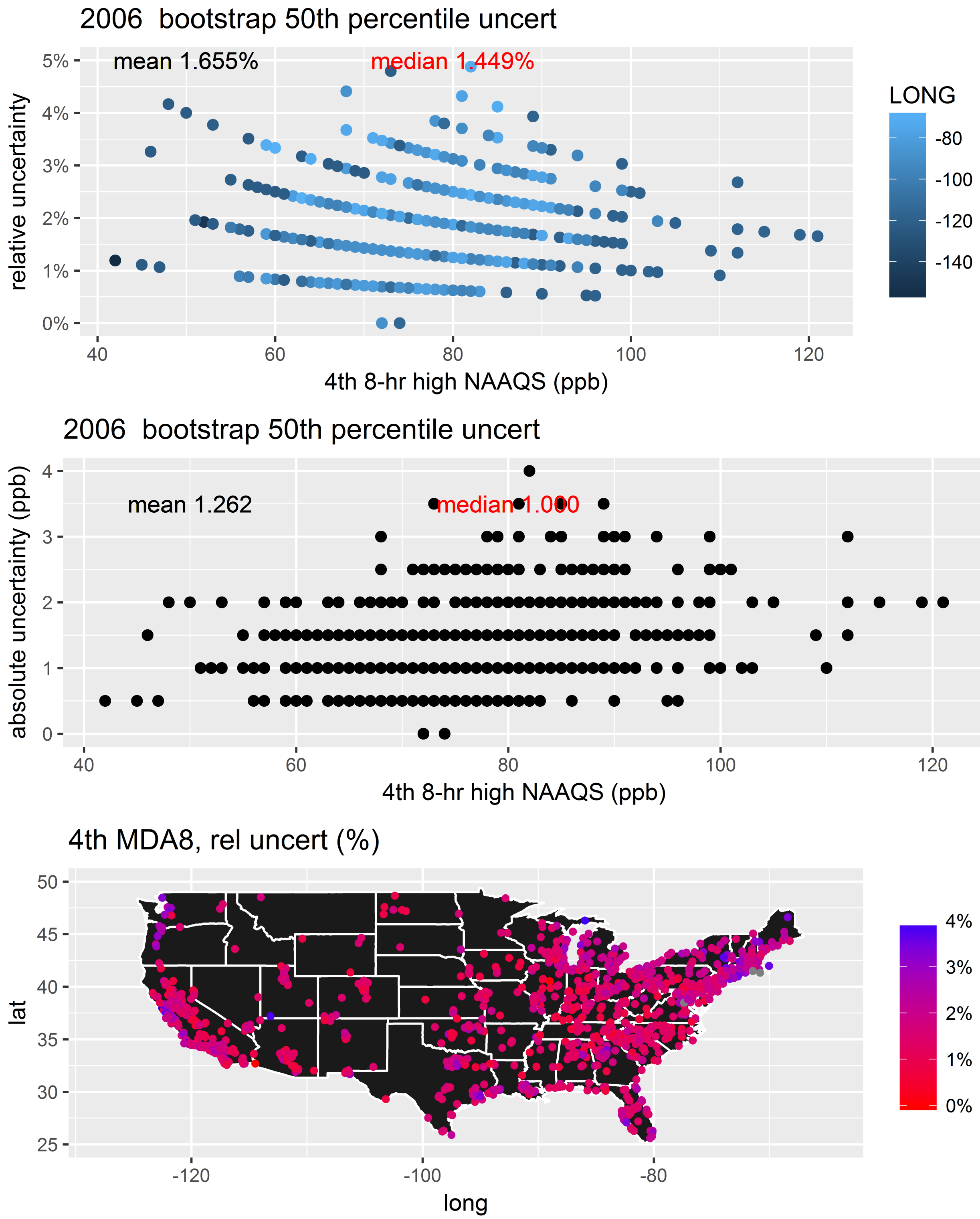


Figure 25: Bootstrap results from the 50% CIs for the 2006 ozone DVs. The top panel shows the relative difference between the CI and the actual DV, the middle panel shows the absolute difference between the values for the DVs at each site and the CI, and the bottom panel shows the spatial distribution of the relative difference between the 50% CIs for the 2006 ozone DV at each site.

2.11 2003-2005 ozone bootstrap results

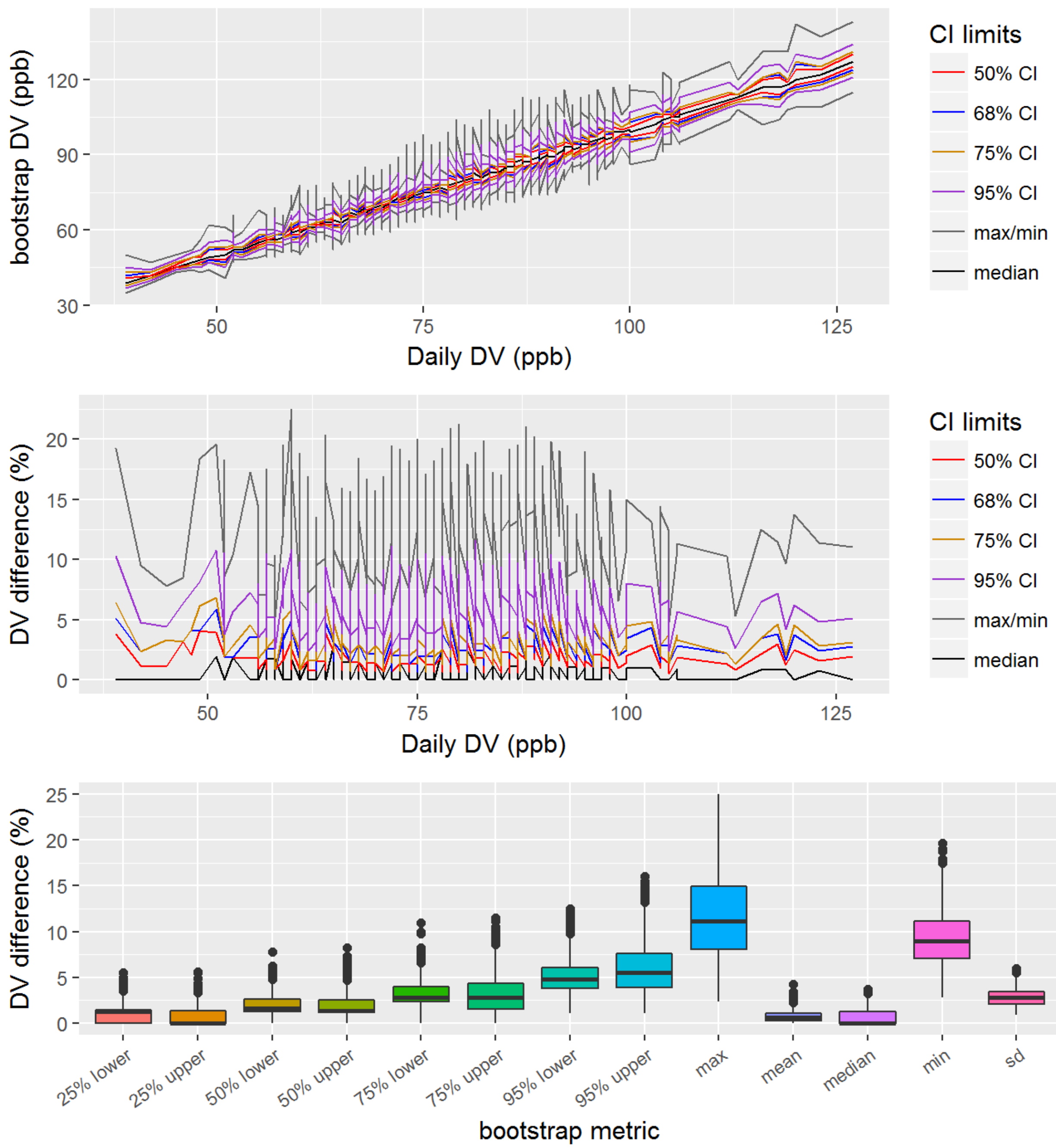


Figure 26: Bootstrap results for the ozone 2005 DVs, showing the 50%, 68%, 75%, and 95% CIs, along with the mean and median bootstrap DVs. The top panel shows the DVs at the various CIs, the middle panel shows the relative difference between the CI and the actual DV, and the bottom panel shows the distribution of the relative differences between the CI and the actual DV.

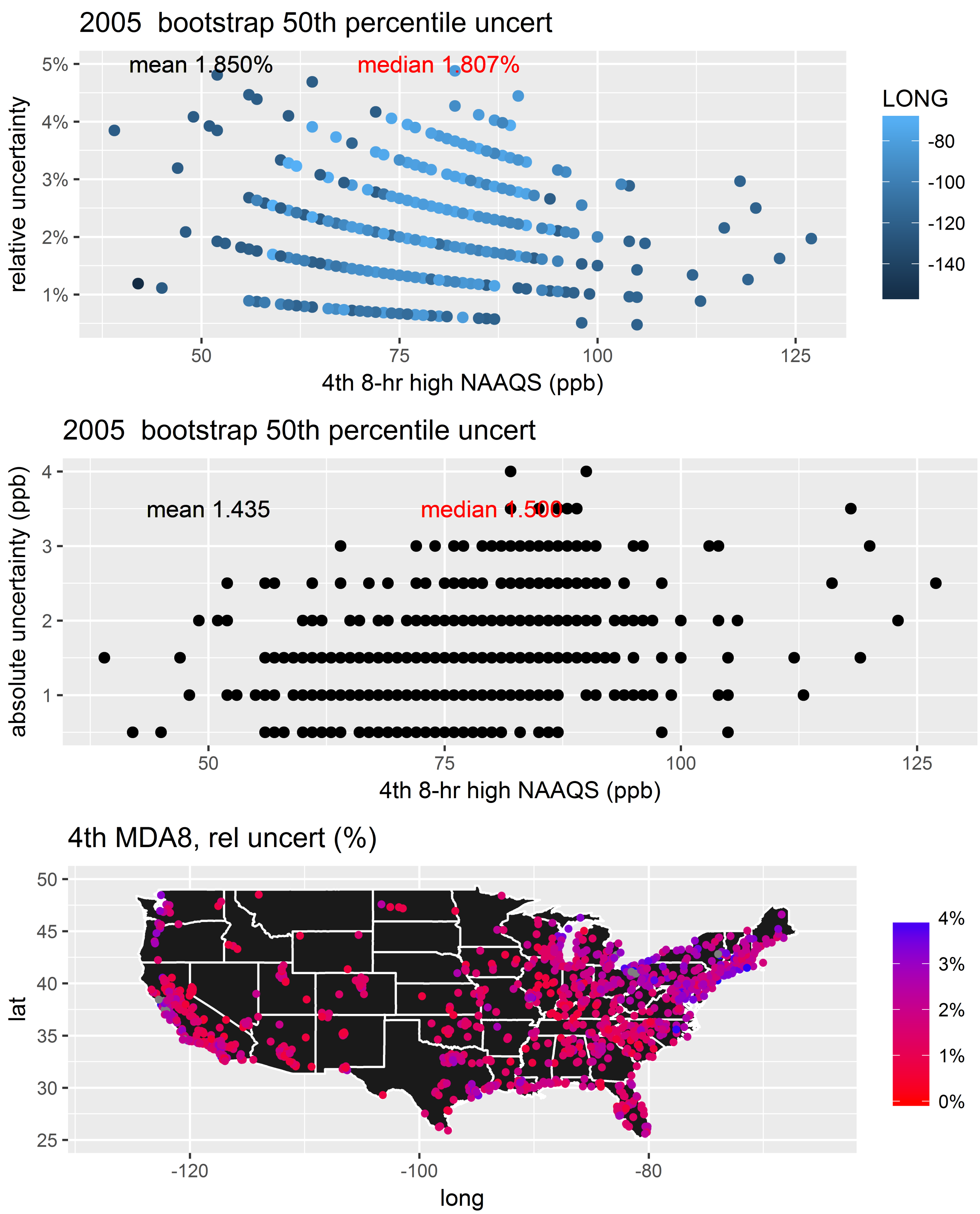


Figure 27: Bootstrap results from the 50% CIs for the 2005 ozone DVs. The top panel shows the relative difference between the CI and the actual DV, the middle panel shows the absolute difference between the values for the DVs at each site and the CI, and the bottom panel shows the spatial distribution of the relative difference between the 50% CIs for the 2005 ozone DV at each site.

2.12 2002-2004 ozone bootstrap results

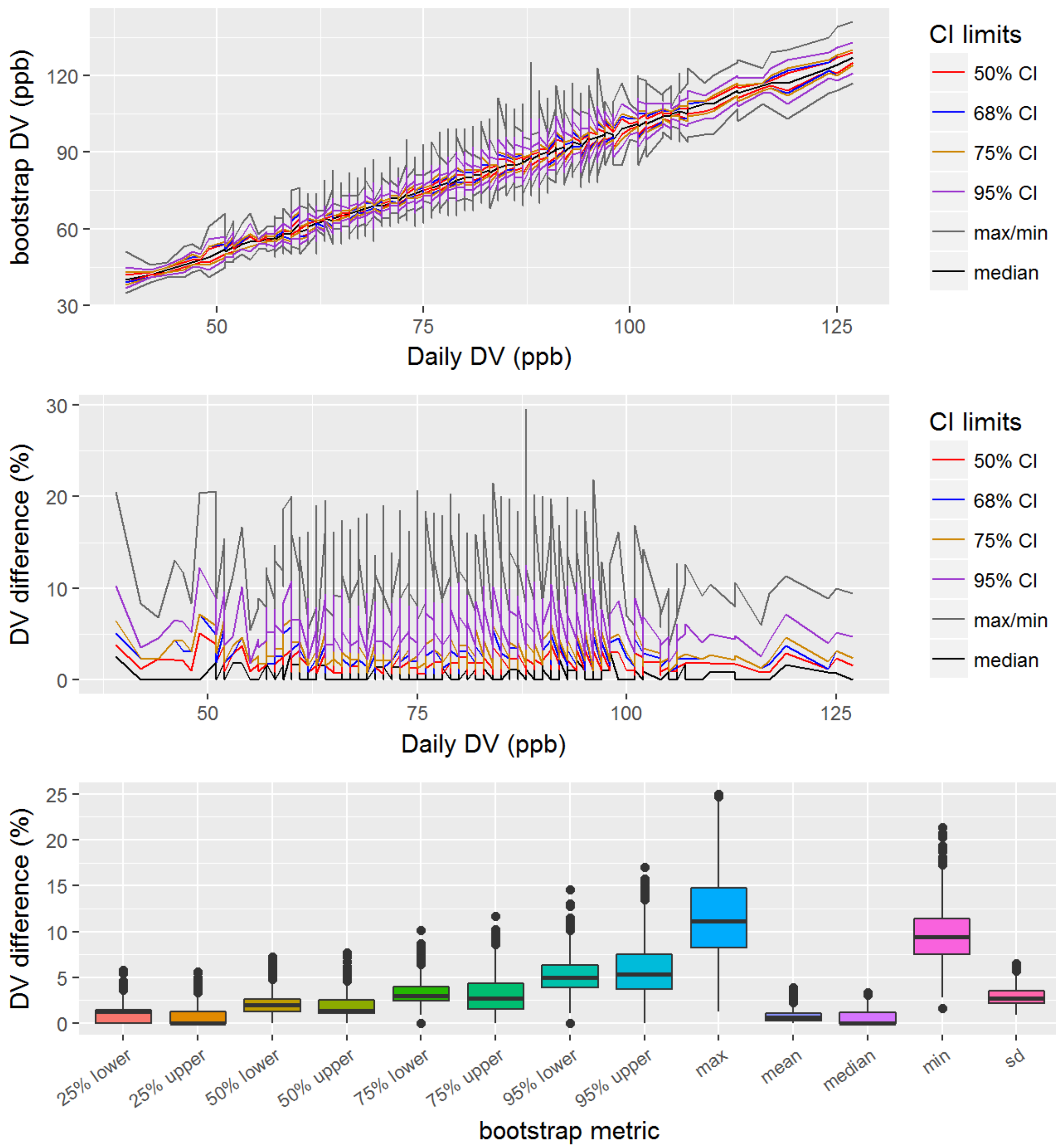


Figure 28: Bootstrap results for the ozone 2004 DVs, showing the 50%, 68%, 75%, and 95% CIs, along with the mean and median bootstrap DVs. The top panel shows the DVs at the various CIs, the middle panel shows the relative difference between the CI and the actual DV, and the bottom panel shows the distribution of the relative differences between the CI and the actual DV.

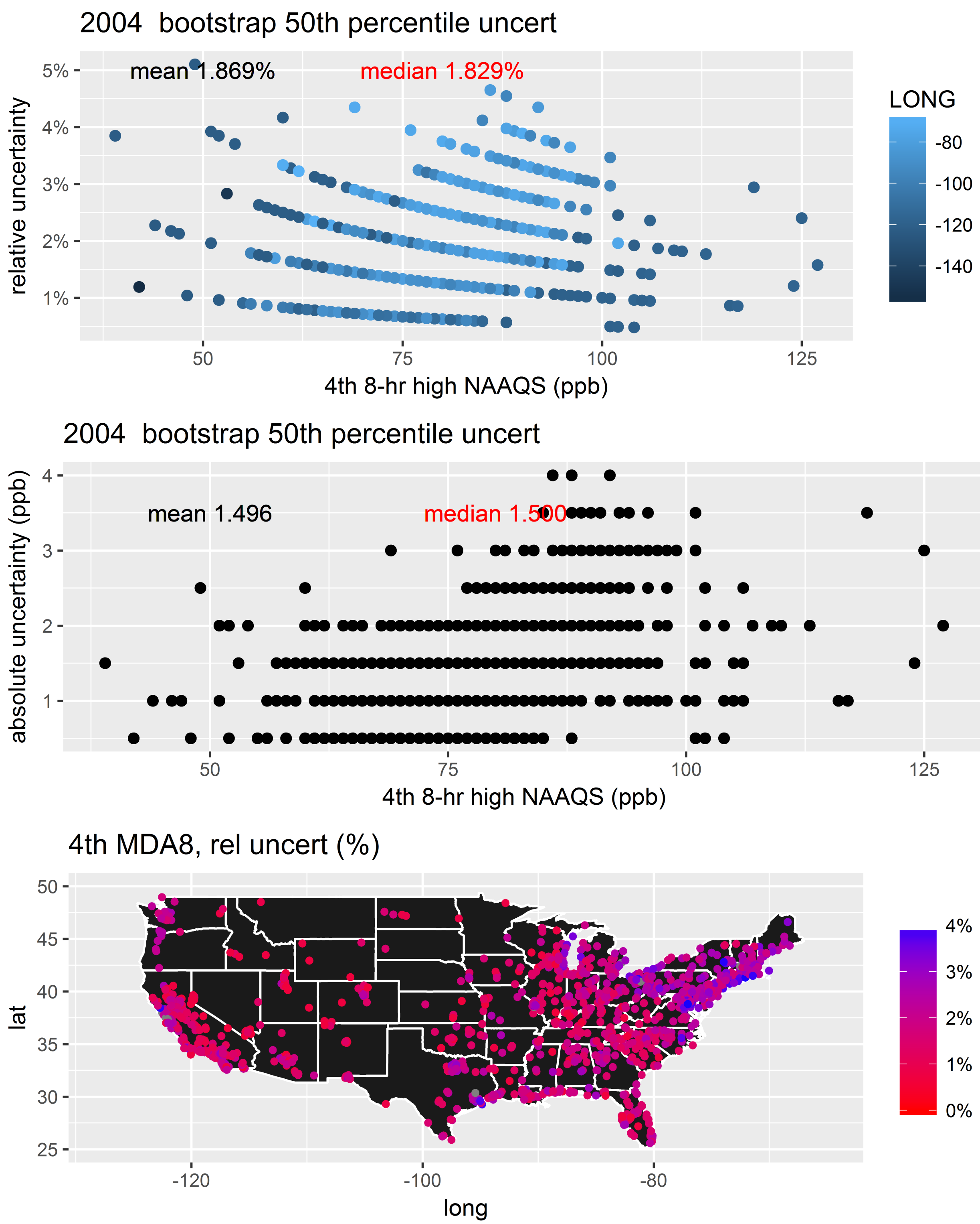


Figure 29: Bootstrap results from the 50% CIs for the 2004 ozone DVs. The top panel shows the relative difference between the CI and the actual DV, the middle panel shows the absolute difference between the values for the DVs at each site and the CI, and the bottom panel shows the spatial distribution of the relative difference between the 50% CIs for the 2004 ozone DV at each site.

2.13 2001-2003 ozone bootstrap results

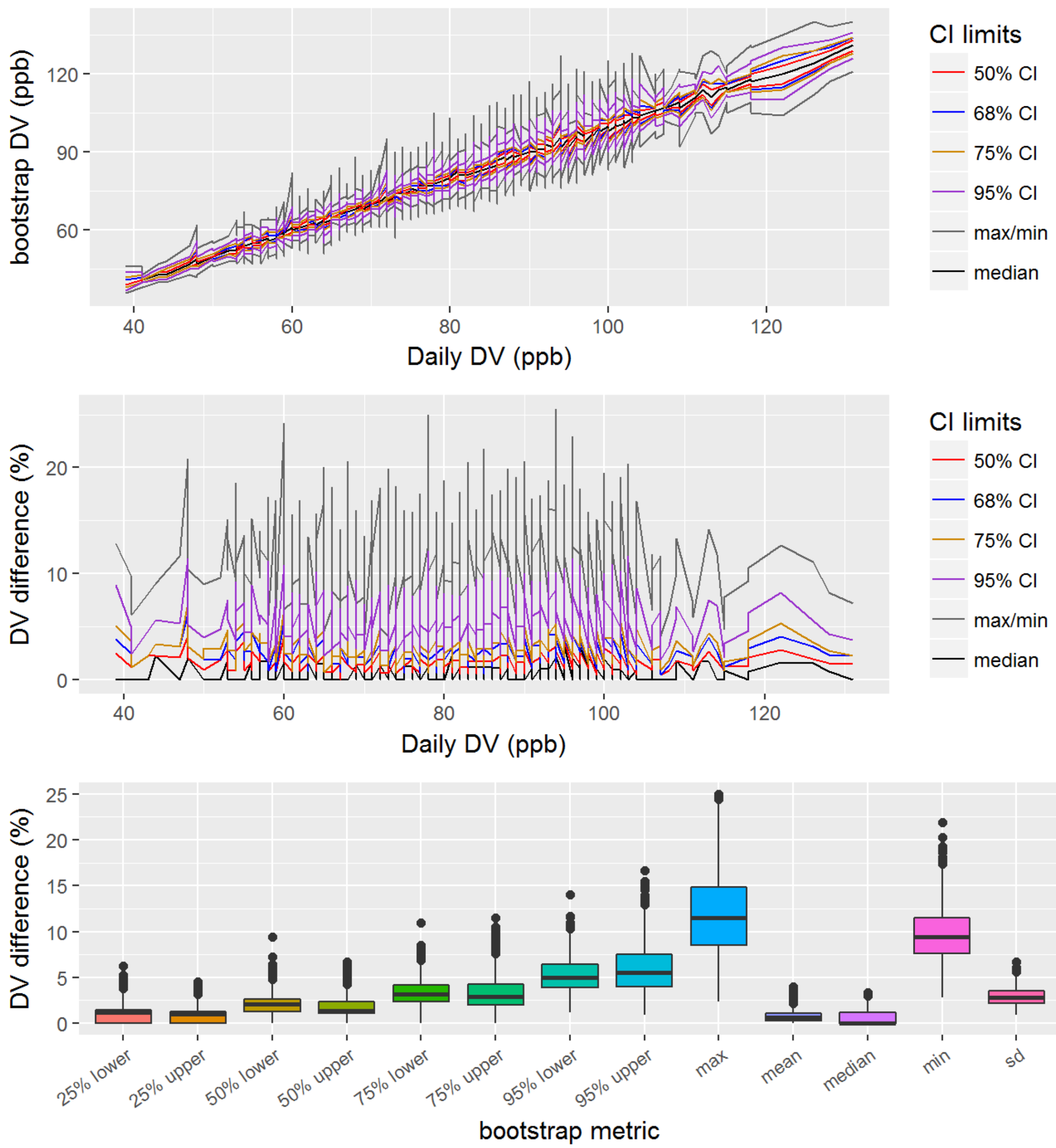


Figure 30: Bootstrap results for the ozone 2003 DVs, showing the 50%, 68%, 75%, and 95% CIs, along with the mean and median bootstrap DVs. The top panel shows the DVs at the various CIs, the middle panel shows the relative difference between the CI and the actual DV, and the bottom panel shows the distribution of the relative differences between the CI and the actual DV.

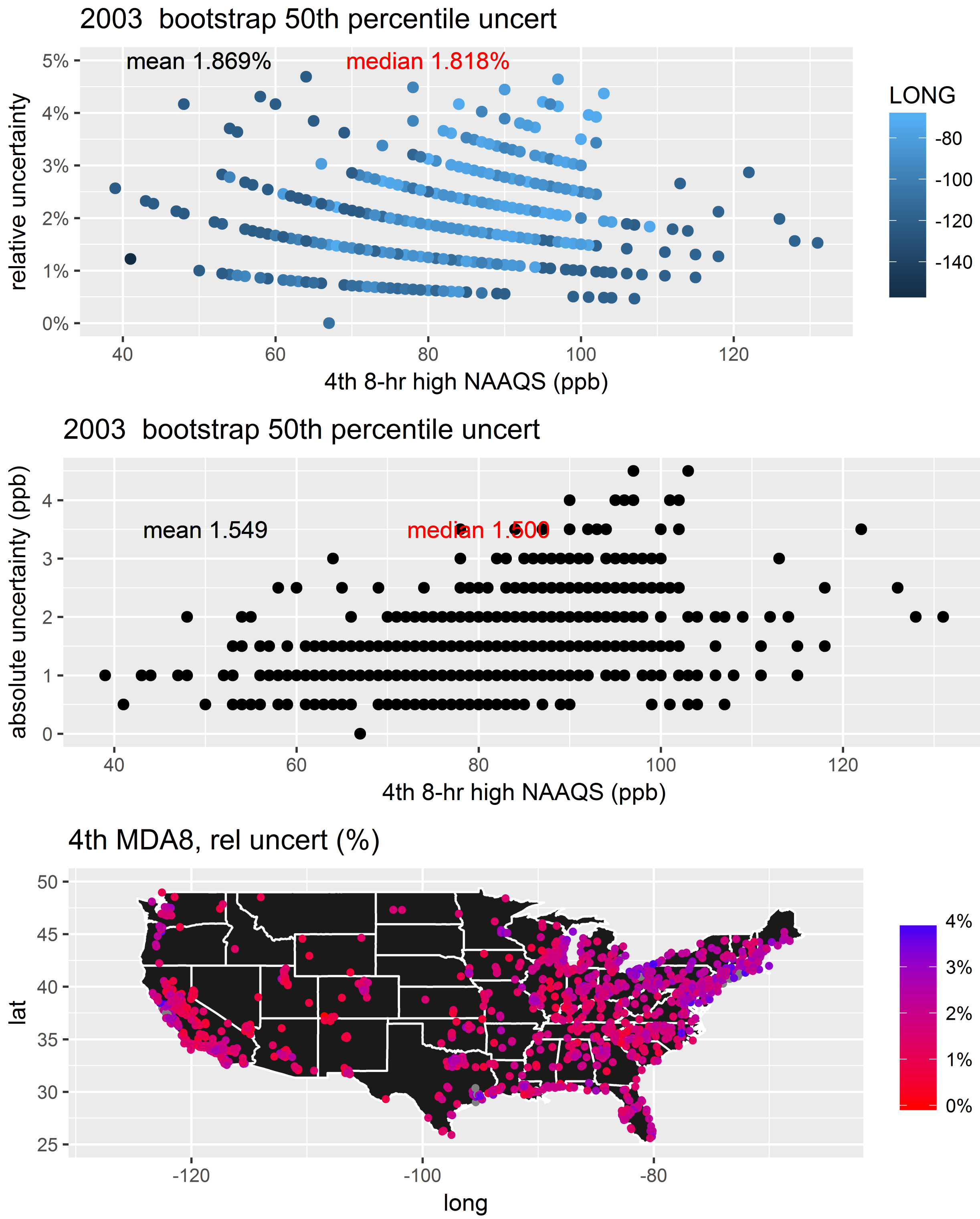


Figure 31: Bootstrap results from the 50% CIs for the 2003 ozone DVs. The top panel shows the relative difference between the CI and the actual DV, the middle panel shows the absolute difference between the values for the DVs at each site and the CI, and the bottom panel shows the spatial distribution of the relative difference between the 50% CIs for the 2003 ozone DV at each site.

2.14 2000-2002 ozone bootstrap results

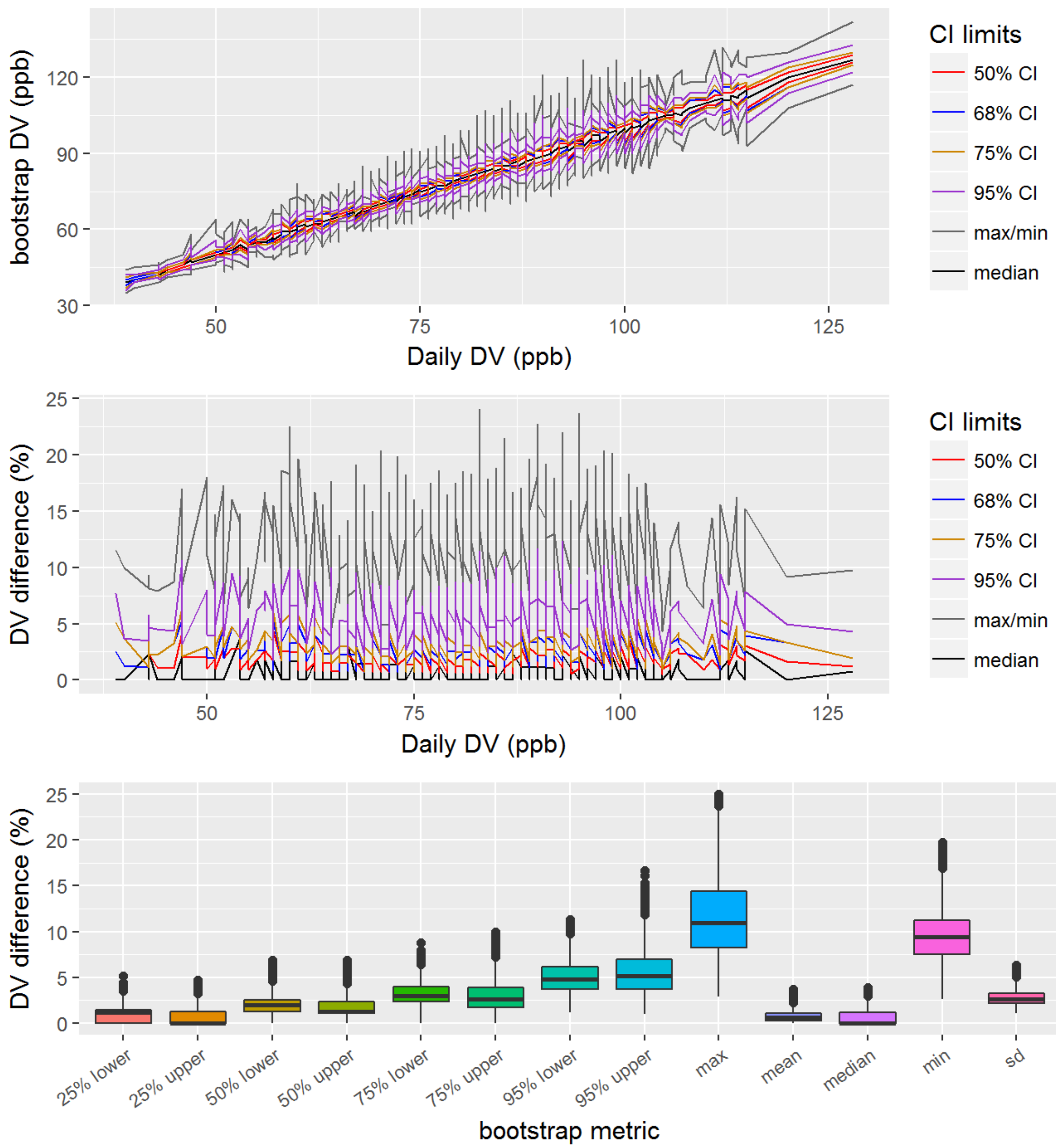


Figure 32: Bootstrap results for the ozone 2002 DVs, showing the 50%, 68%, 75%, and 95% CIs, along with the mean and median bootstrap DVs. The top panel shows the DVs at the various CIs, the middle panel shows the relative difference between the CI and the actual DV, and the bottom panel shows the distribution of the relative differences between the CI and the actual DV.

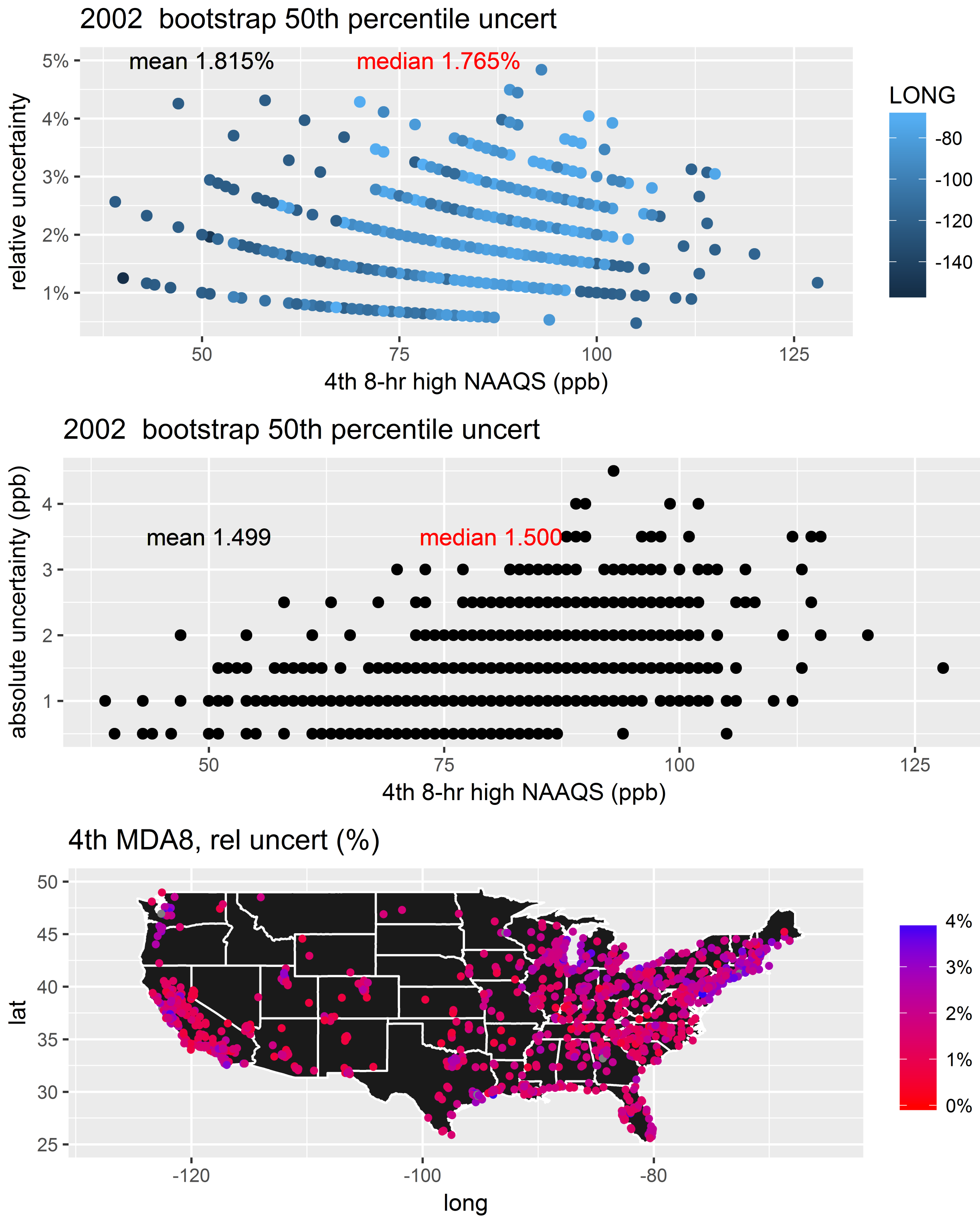


Figure 33: Bootstrap results from the 50% CIs for the 2002 ozone DVs. The top panel shows the relative difference between the CI and the actual DV, the middle panel shows the absolute difference between the values for the DVs at each site and the CI, and the bottom panel shows the spatial distribution of the relative difference between the 50% CIs for the 2002 ozone DV at each site.

3 Air quality variability results for years 2002-2013 for $PM_{2.5}$

Bootstrap results for $PM_{2.5}$ data from the years 2000-2015. Each section containts a single DV period,*e.g.*, the results for 2015 include data from 2013-2015.

3.1 2013-2015 $PM_{2.5}$ bootstrap results

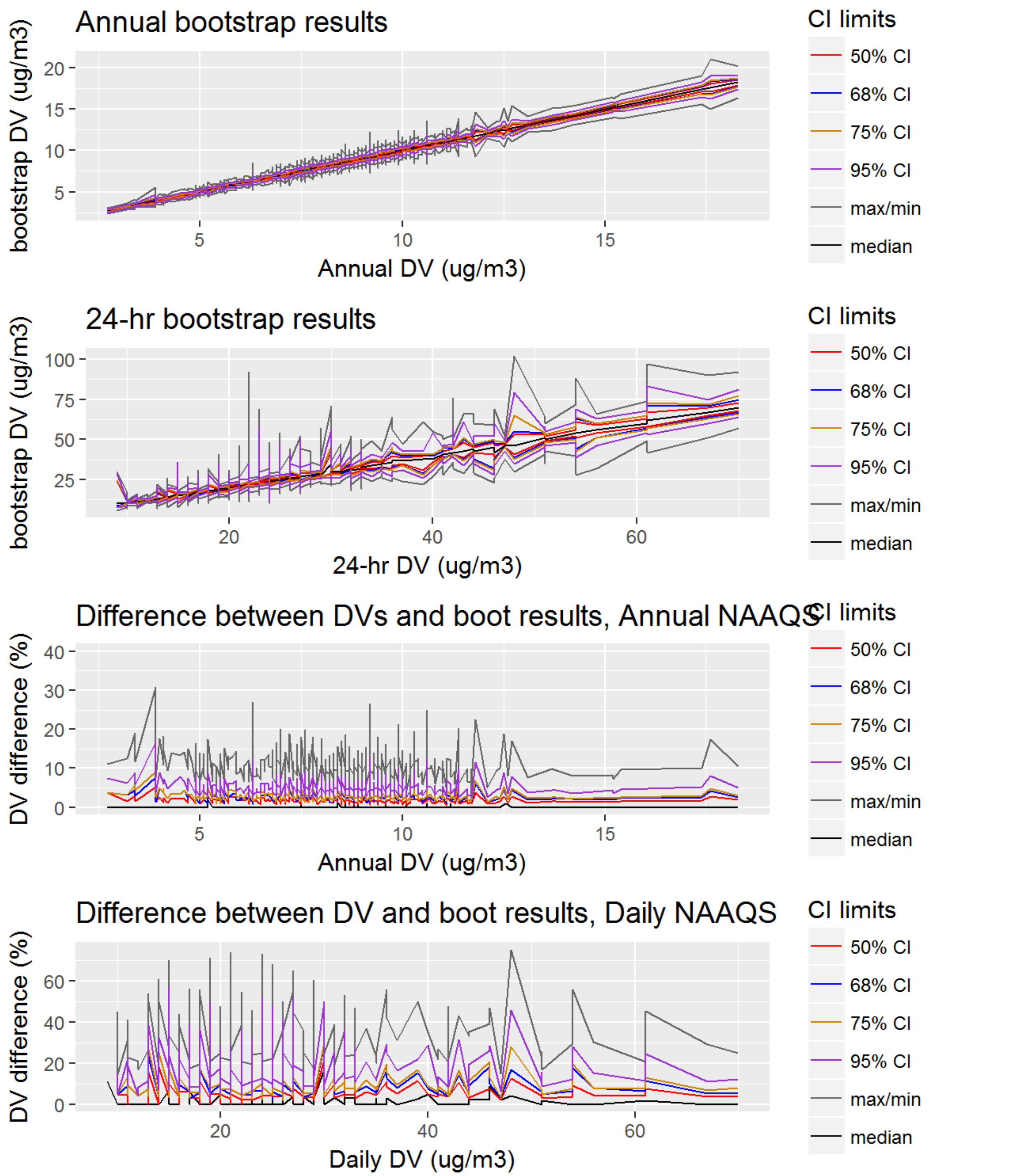


Figure 34: Bootstrap results for the 2015 $PM_{2.5}$ DVs, showing the 50%, 68%, 75%, and 95% CIs, along with the mean and median bootstrap DVs. The top two panels show the values for the DVs at the various CIs, while the bottom two panels show the relative difference between the CI and the actual DV.

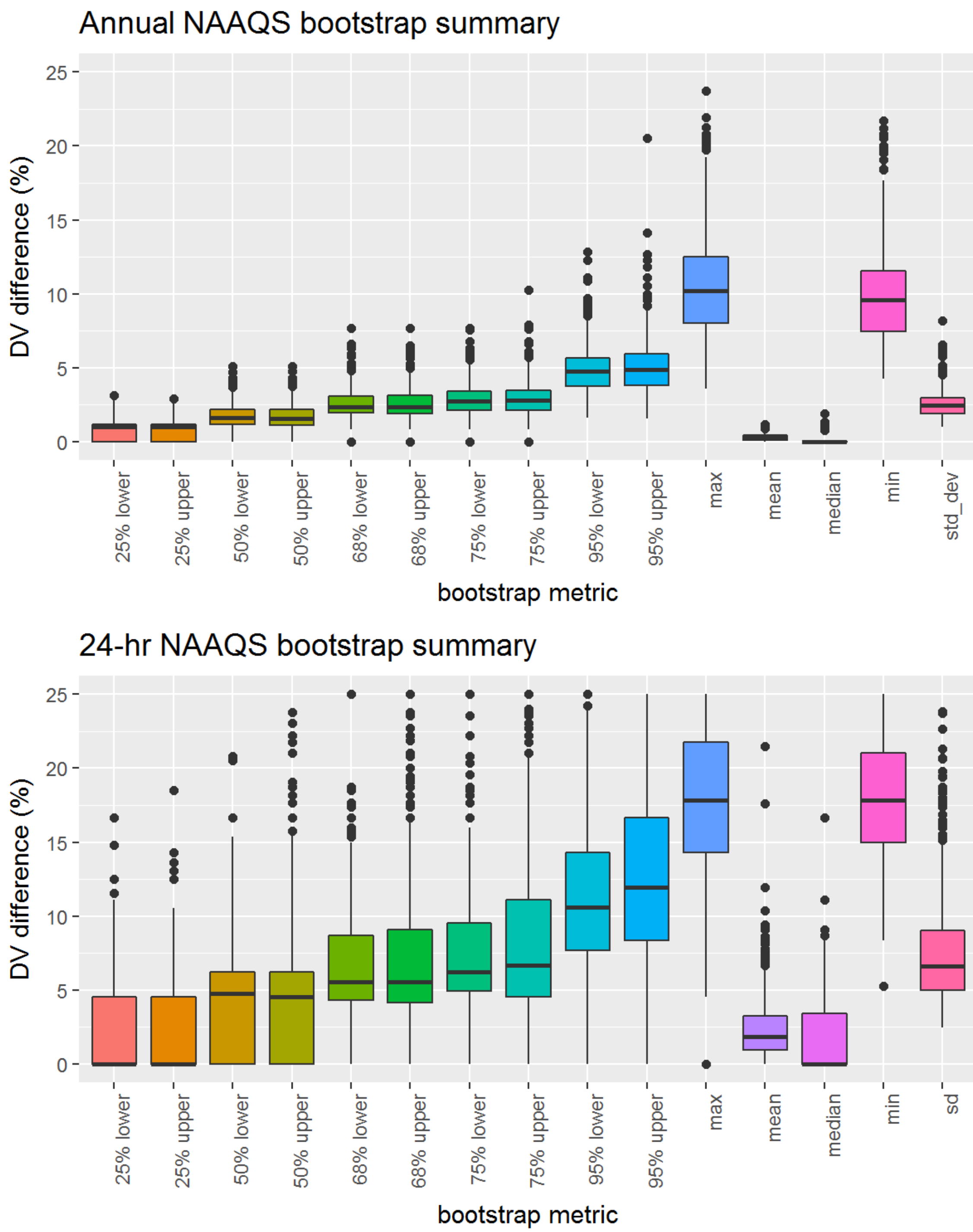


Figure 35: Bootstrap results for the 2015 $PM_{2.5}$ DVs, showing distribution of the relative differences between the bootstrap DVs and the actual DV at the 50%, 68%, 75%, and 95% CIs, along with the mean, median, maximum, minimum, standard deviations of the relative differences.

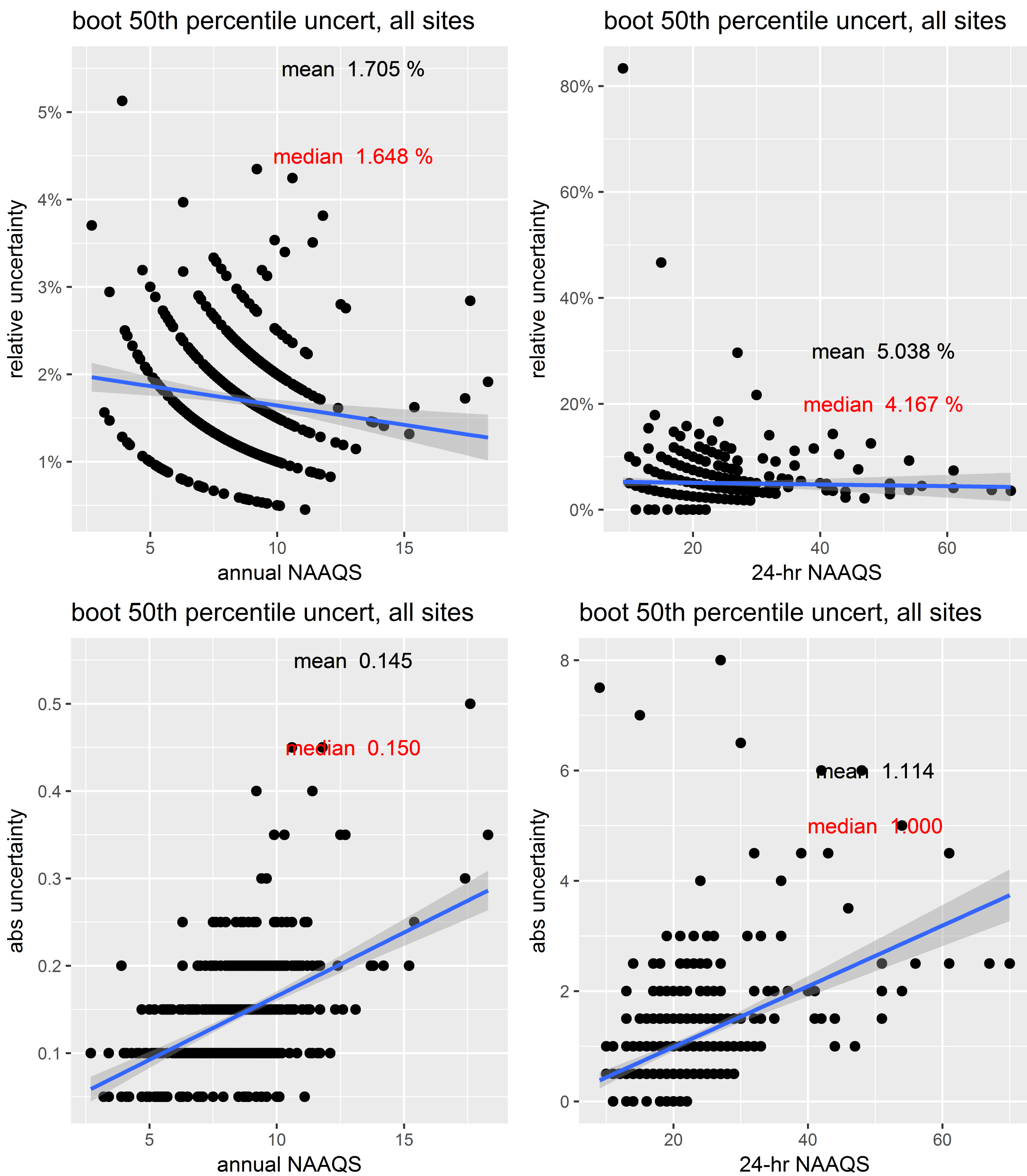
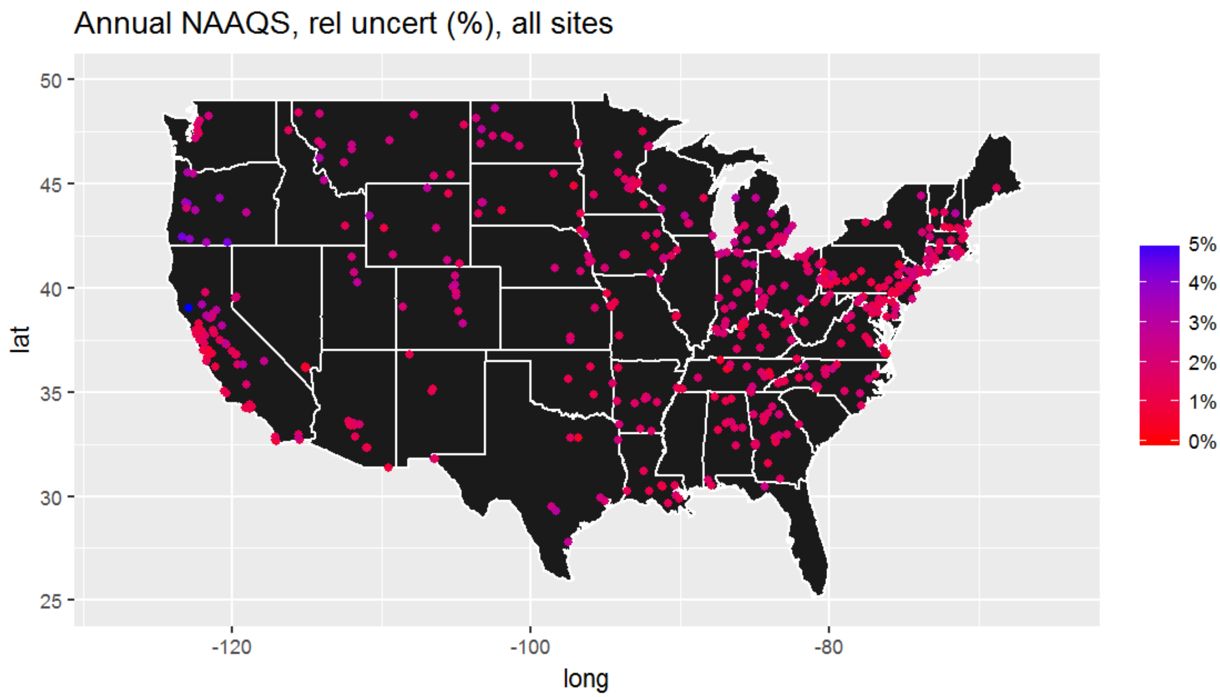


Figure 36: Bootstrap results from the 50% CIs for $PM_{2.5}$ DVs. The top two panels show the relative difference between the CI and the actual DV and the bottom two panels show the absolute difference between the values for the DVs at each site and the CI.



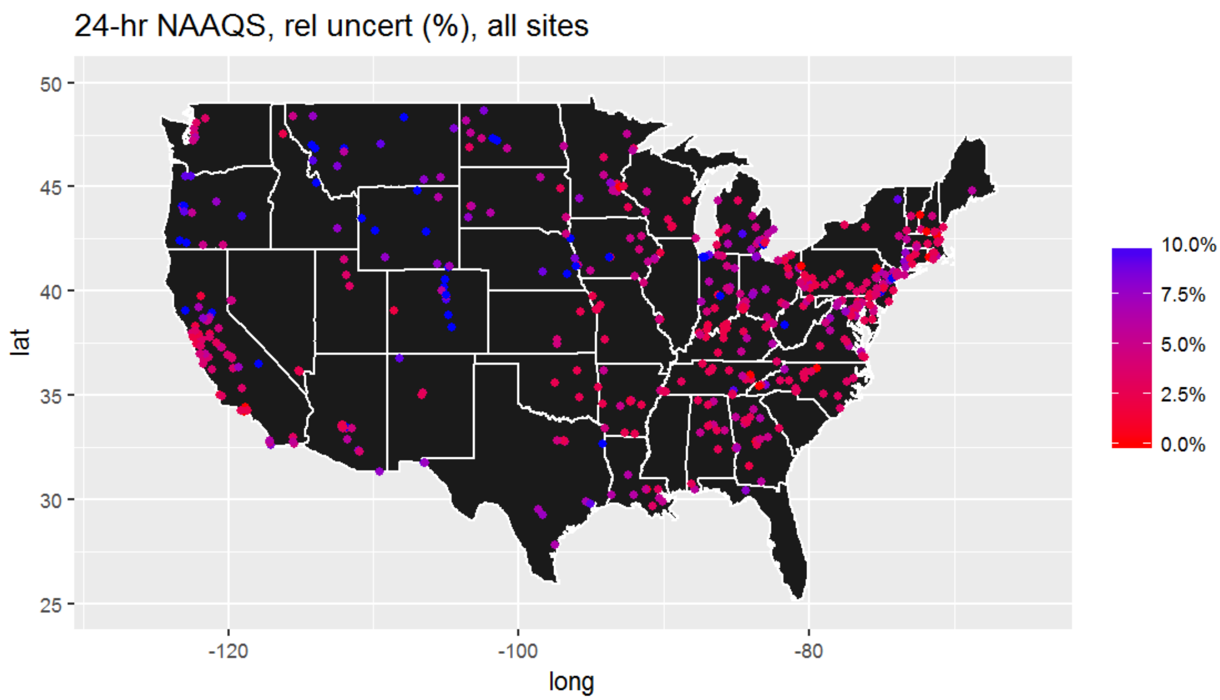


Figure 37: Spatial distribution of the relative difference between the CI and the actual DV from the 50% CIs for the 2015 $PM_{2.5}$ DVs.

3.2 2012-2014 $PM_{2.5}$ bootstrap results

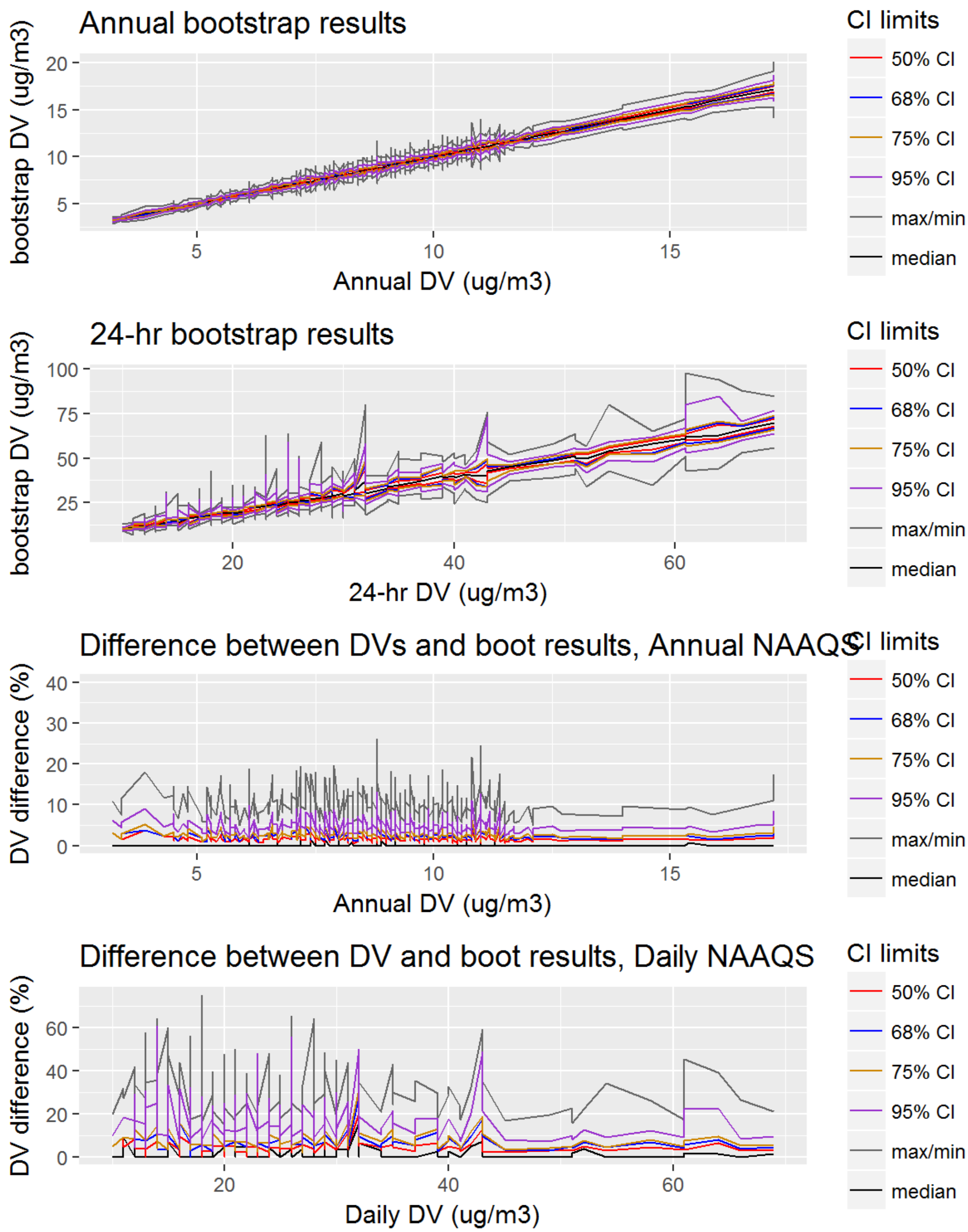


Figure 38: Bootstrap results for the 2014 $PM_{2.5}$ DVs, showing the 50%, 68%, 75%, and 95% CIs, along with the mean and median bootstrap DVs. The top two panels show the values for the DVs at the various CIs, while the bottom two panels show the relative difference between the CI and the actual DV.

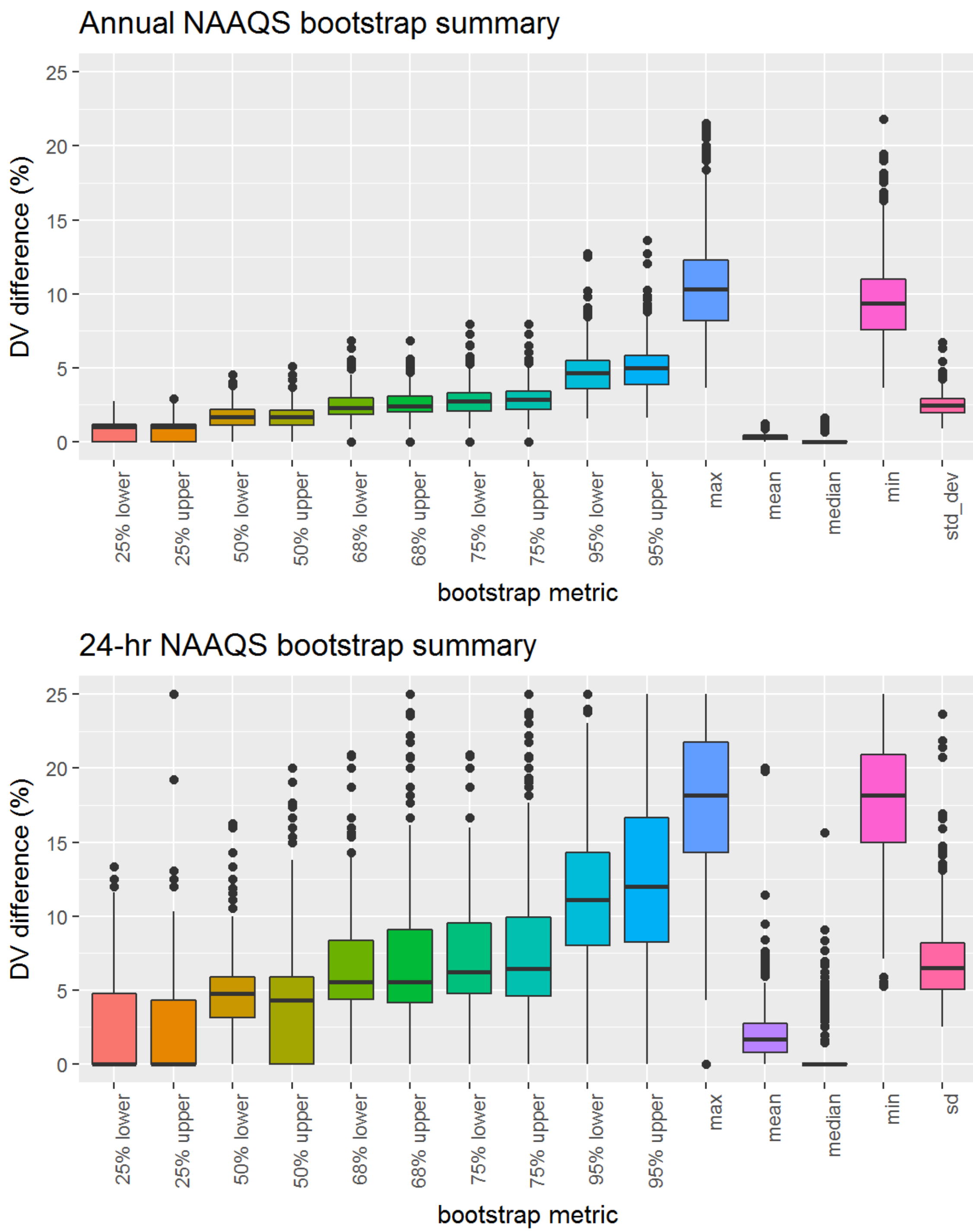


Figure 39: Bootstrap results for the 2014 $PM_{2.5}$ DVs, showing distribution of the relative differences between the bootstrap DVs and the actual DV at the 50%, 68%, 75%, and 95% CIs, along with the mean, median, maximum, minimum, standard deviations of the relative differences.

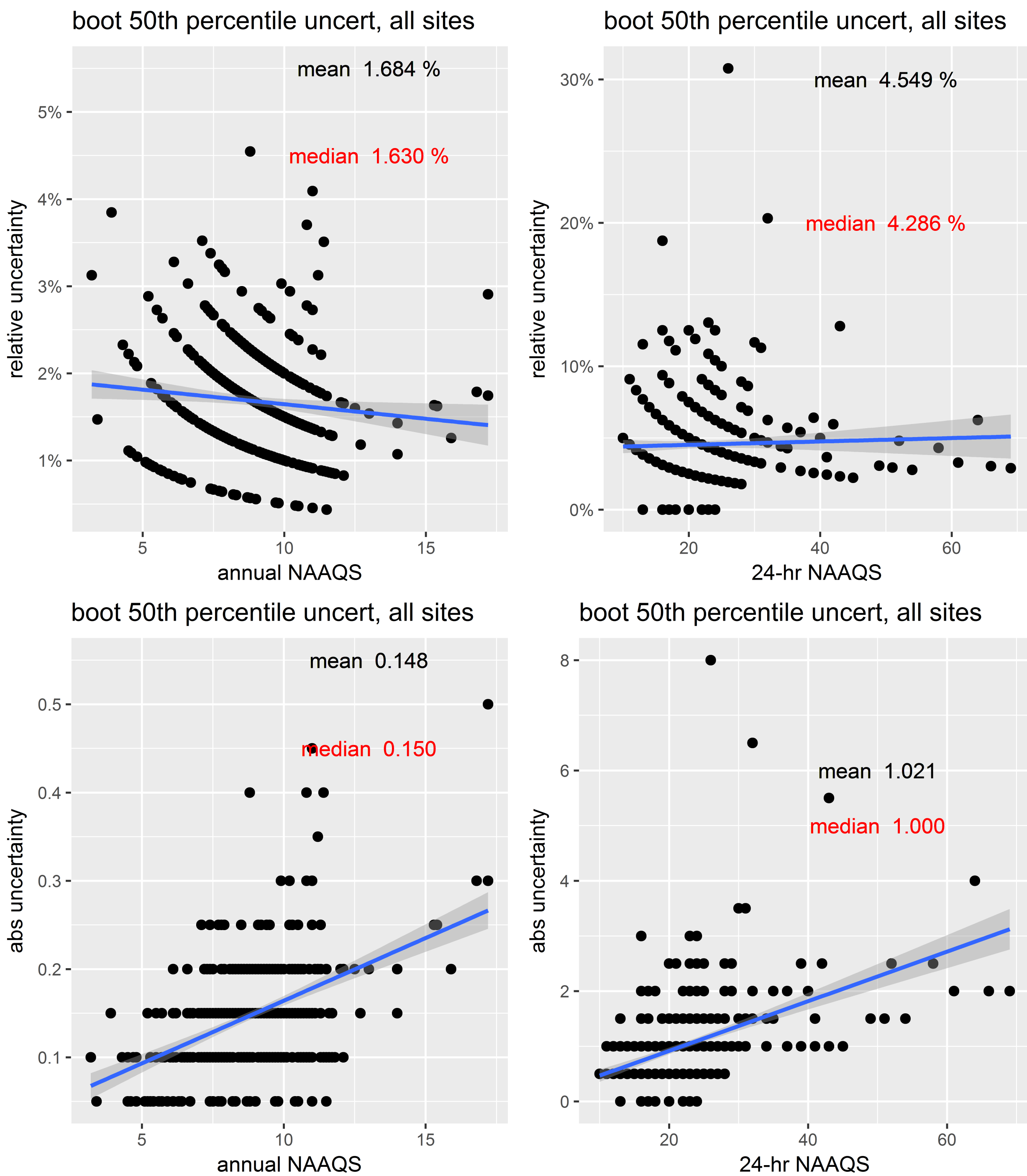


Figure 40: Bootstrap results from the 50% CIs for $PM_{2.5}$ DVs. The top two panels show the relative difference between the CI and the actual DV and the bottom two panels show the absolute difference between the values for the DVs at each site and the CI.

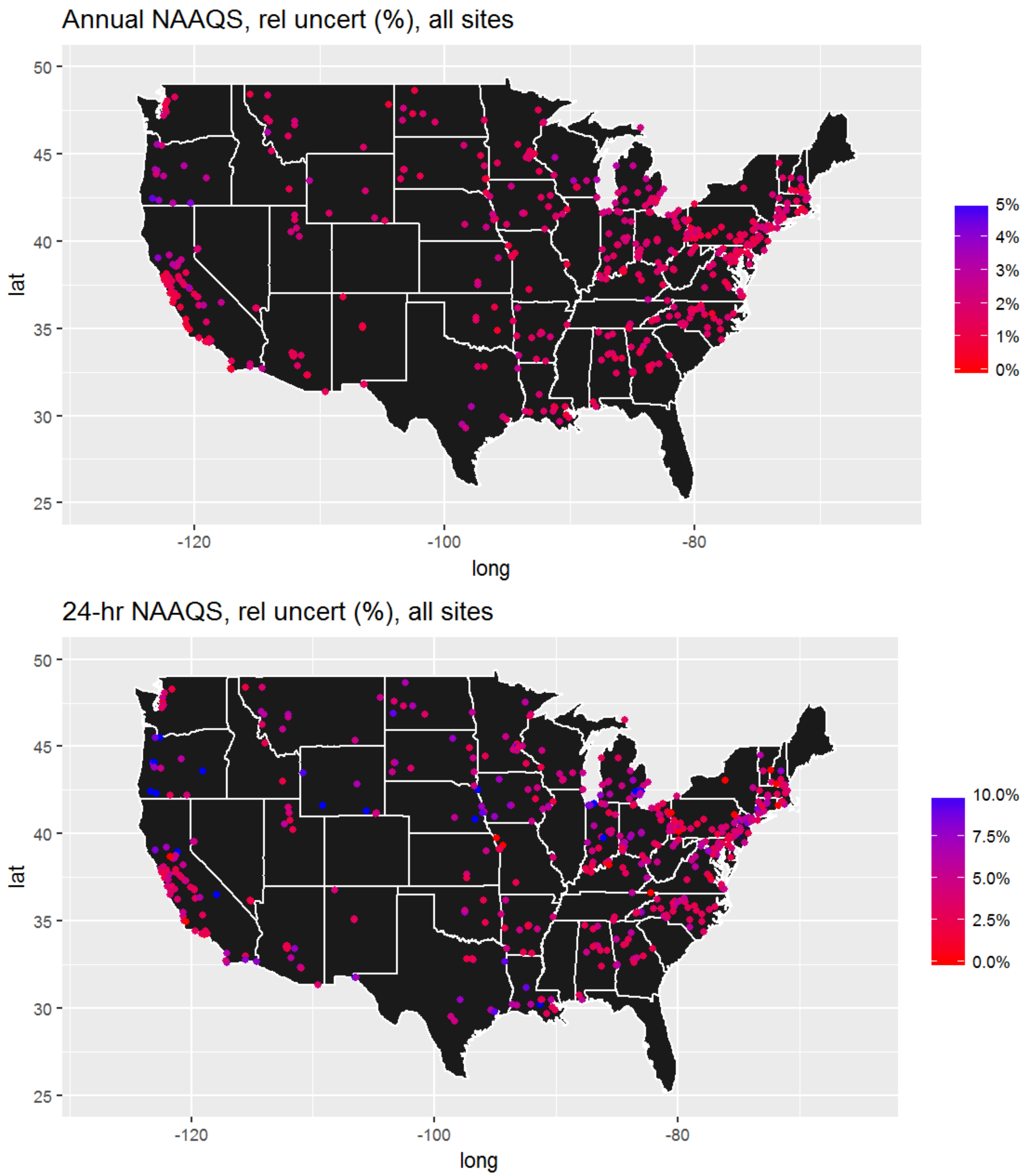


Figure 41: Spatial distribution of the relative difference between the CI and the actual DV from the 50% CIs for the 2014 $PM_{2.5}$ DVs.

3.3 2011-2013 $\mathbf{PM_{2.5}}$ bootstrap results

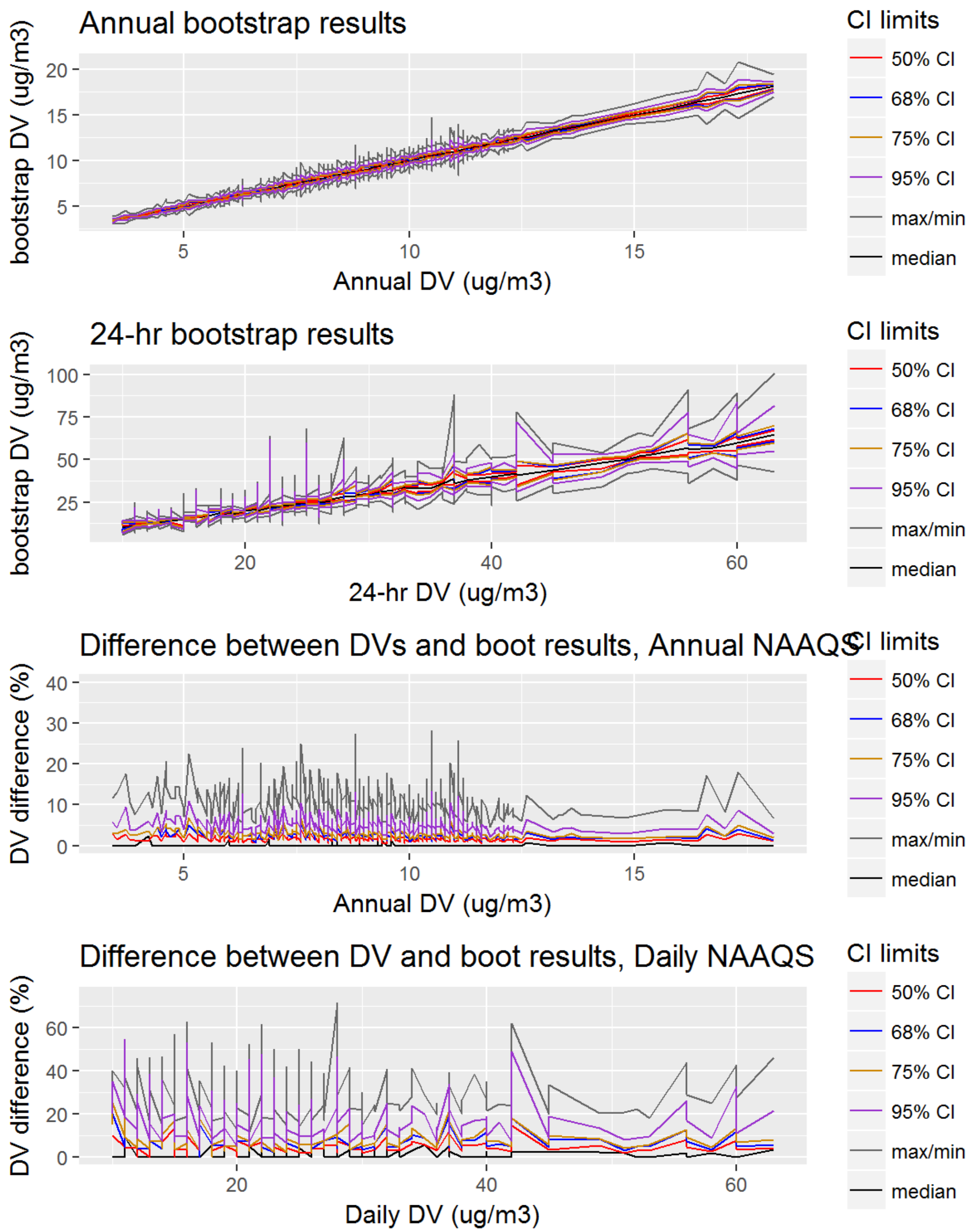


Figure 42: Bootstrap results for the 2013 $PM_{2.5}$ DVs, showing the 50%, 68%, 75%, and 95% CIs, along with the mean and median bootstrap DVs. The top two panels show the values for the DVs at the various CIs, while the bottom two panels show the relative difference between the CI and the actual DV.

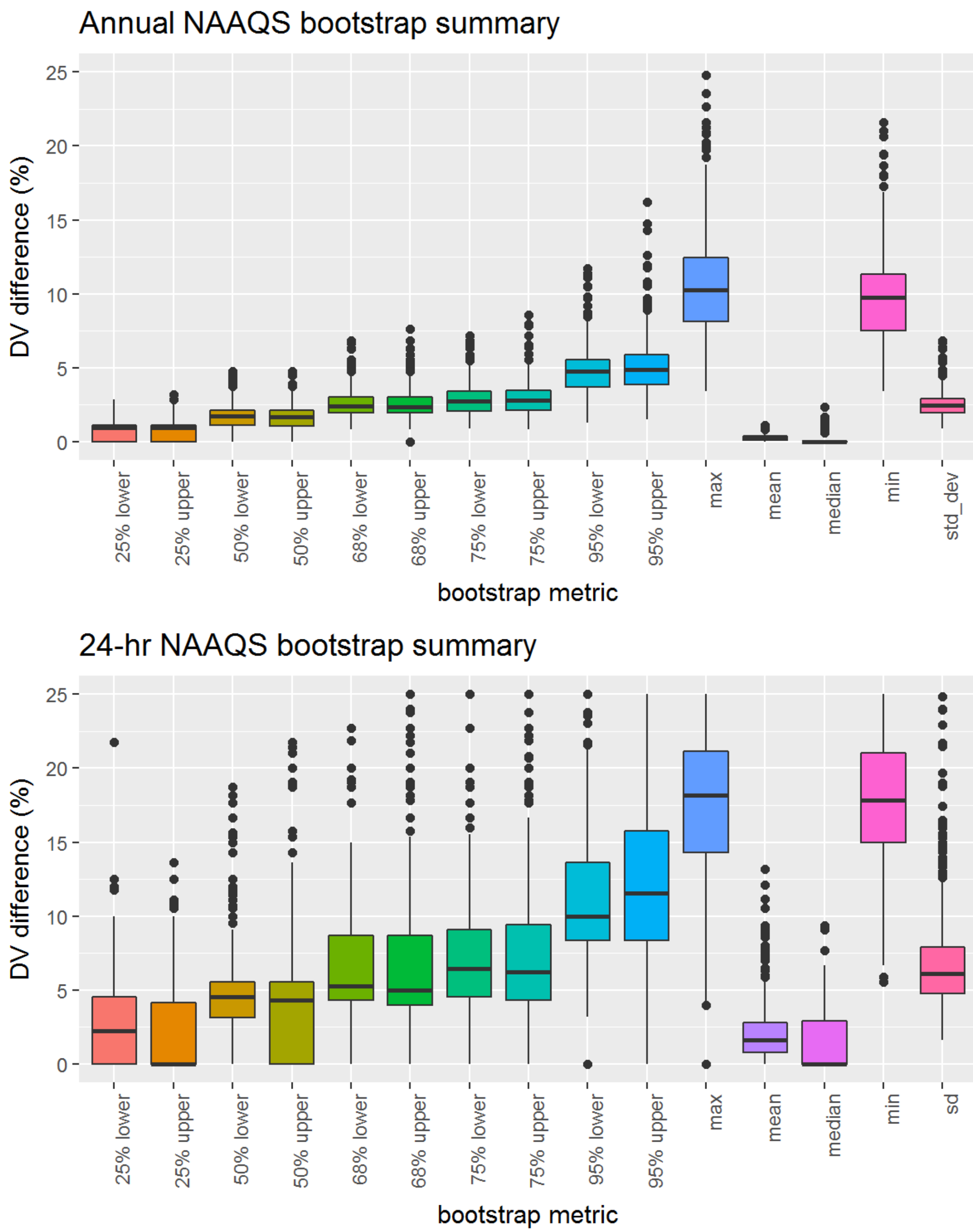


Figure 43: Bootstrap results for the 2013 $PM_{2.5}$ DVs, showing distribution of the relative differences between the bootstrap DVs and the actual DV at the 50%, 68%, 75%, and 95% CIs, along with the mean, median, maximum, minimum, standard deviations of the relative differences.

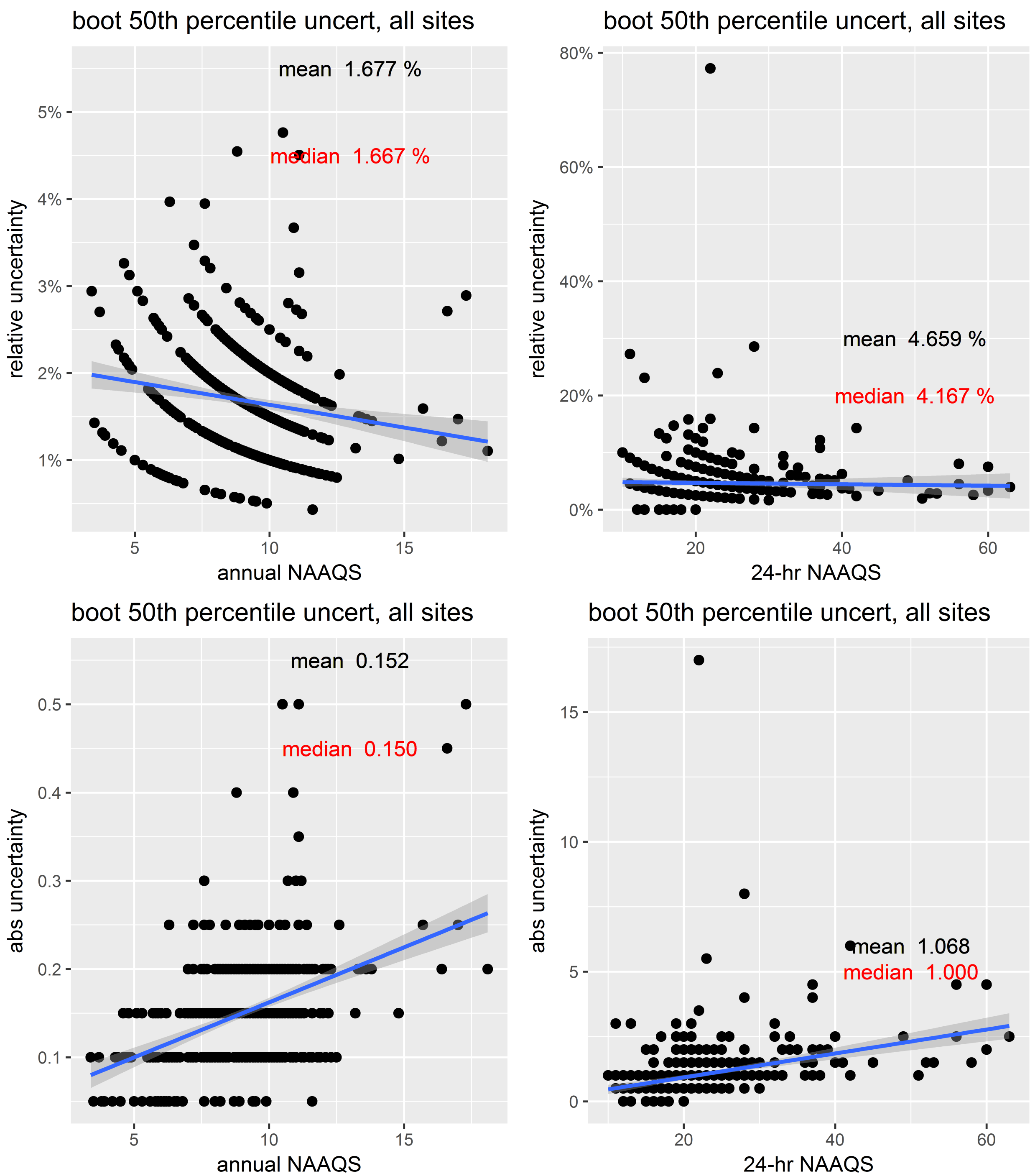


Figure 44: Bootstrap results from the 50% CIs for $PM_{2.5}$ DVs. The top two panels show the relative difference between the CI and the actual DV and the bottom two panels show the absolute difference between the values for the DVs at each site and the CI.

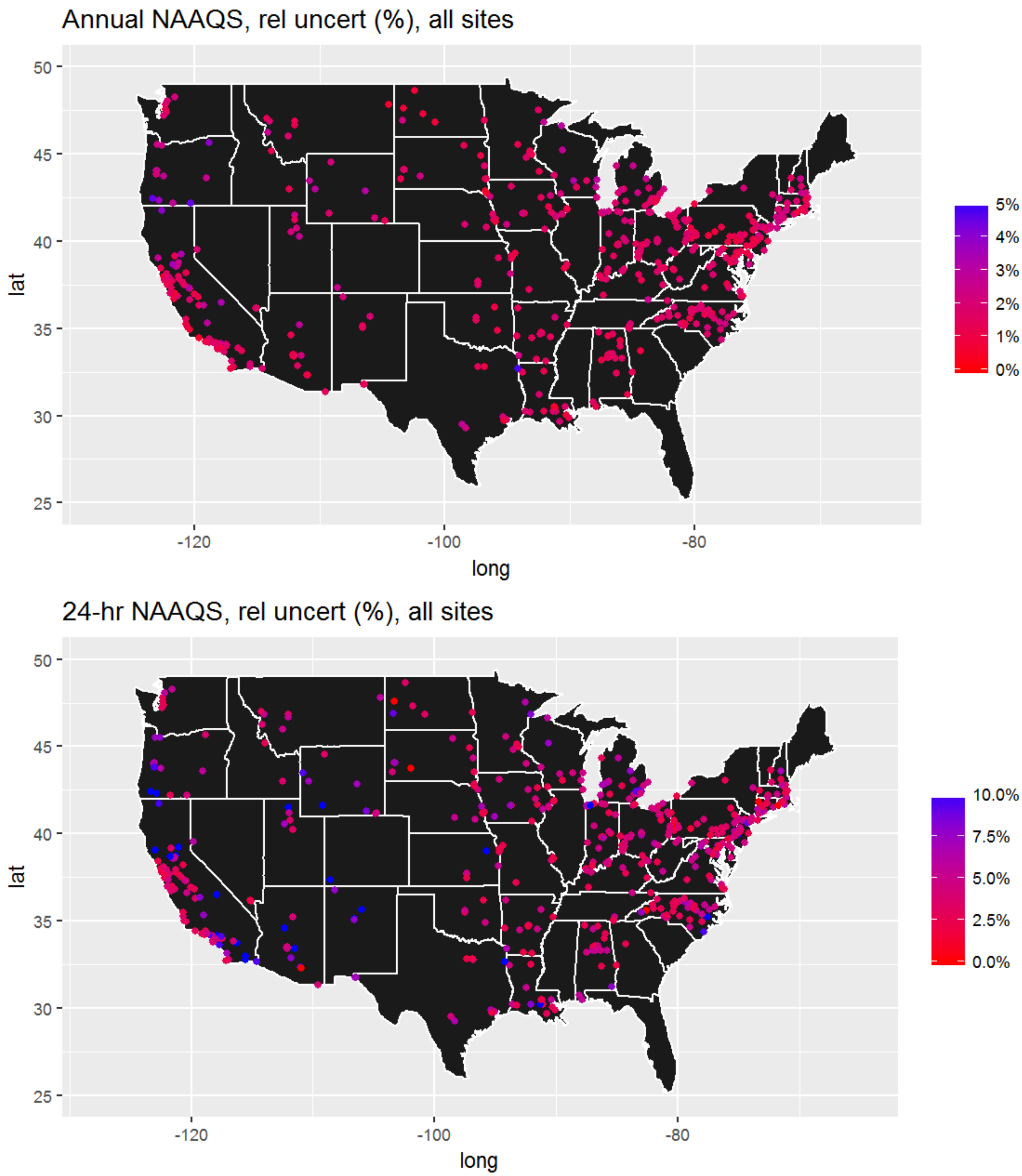


Figure 45: Spatial distribution of the relative difference between the CI and the actual DV from the 50% CIs for the 2013 $PM_{2.5}$ DVs.

3.4 2010-2012 $PM_{2.5}$ bootstrap results

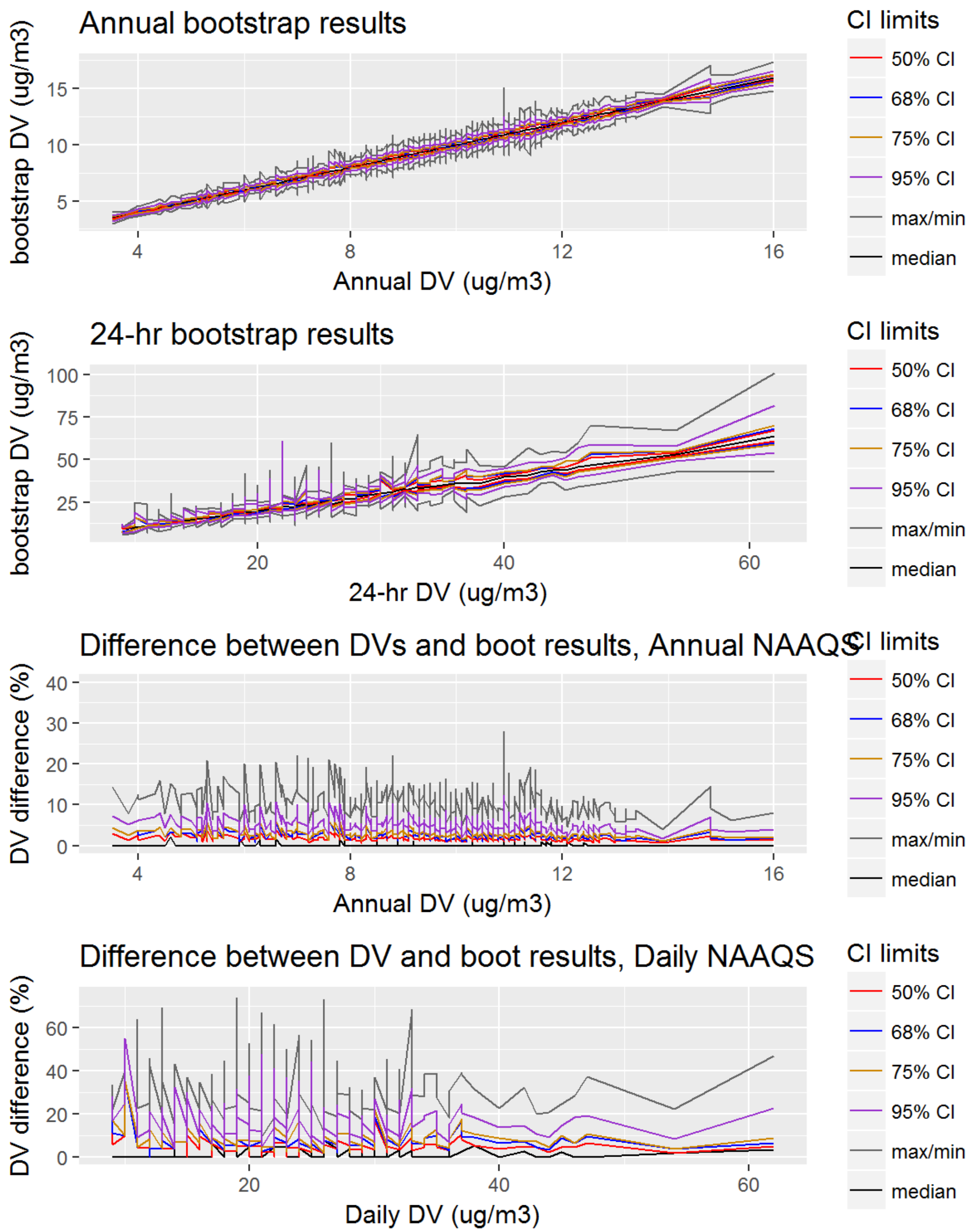


Figure 46: Bootstrap results for the 2012 $PM_{2.5}$ DVs, showing the 50%, 68%, 75%, and 95% CIs, along with the mean and median bootstrap DVs. The top two panels show the values for the DVs at the various CIs, while the bottom two panels show the relative difference between the CI and the actual DV.

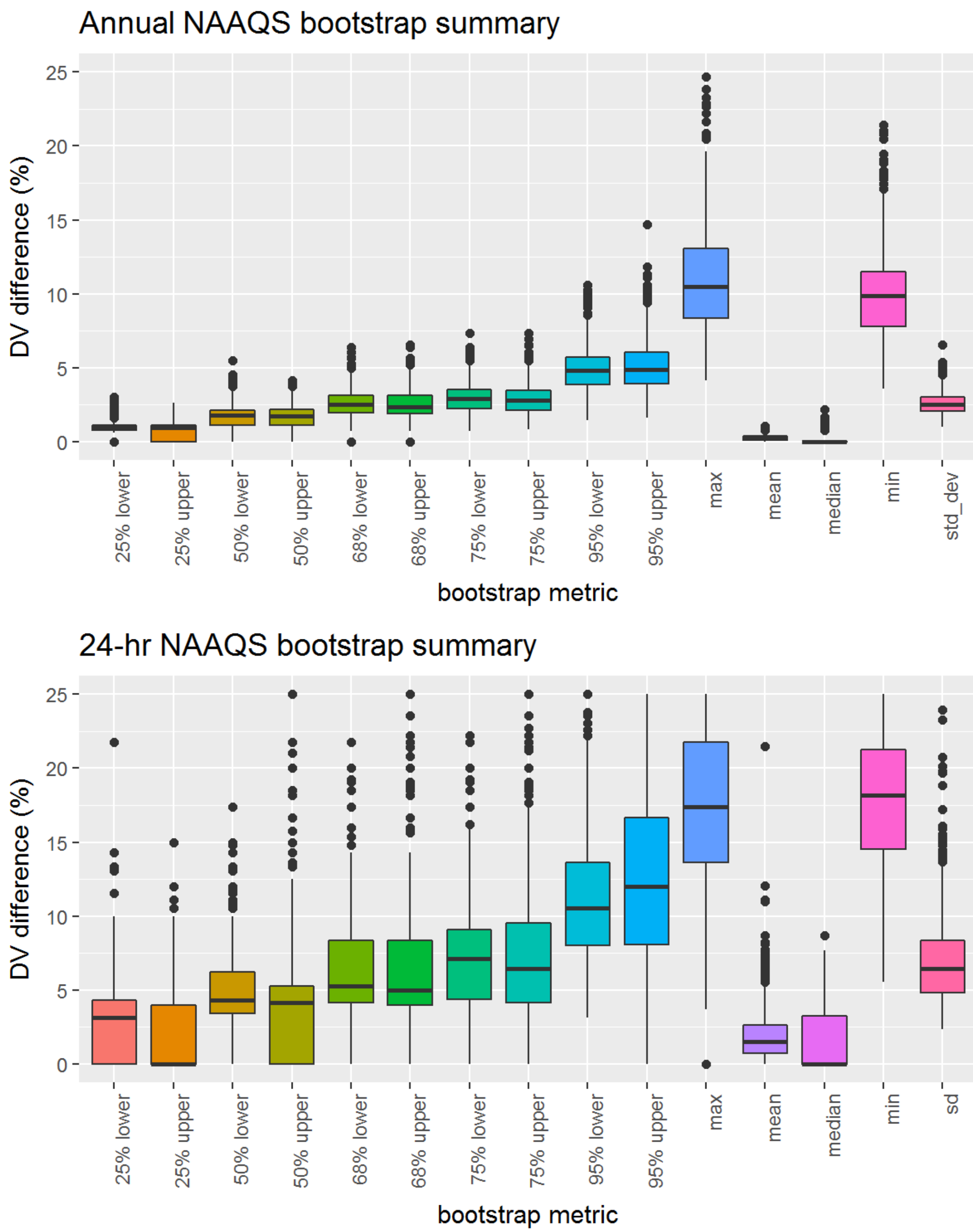


Figure 47: Bootstrap results for the 2012 $PM_{2.5}$ DVs, showing distribution of the relative differences between the bootstrap DVs and the actual DV at the 50%, 68%, 75%, and 95% CIs, along with the mean, median, maximum, minimum, standard deviations of the relative differences.

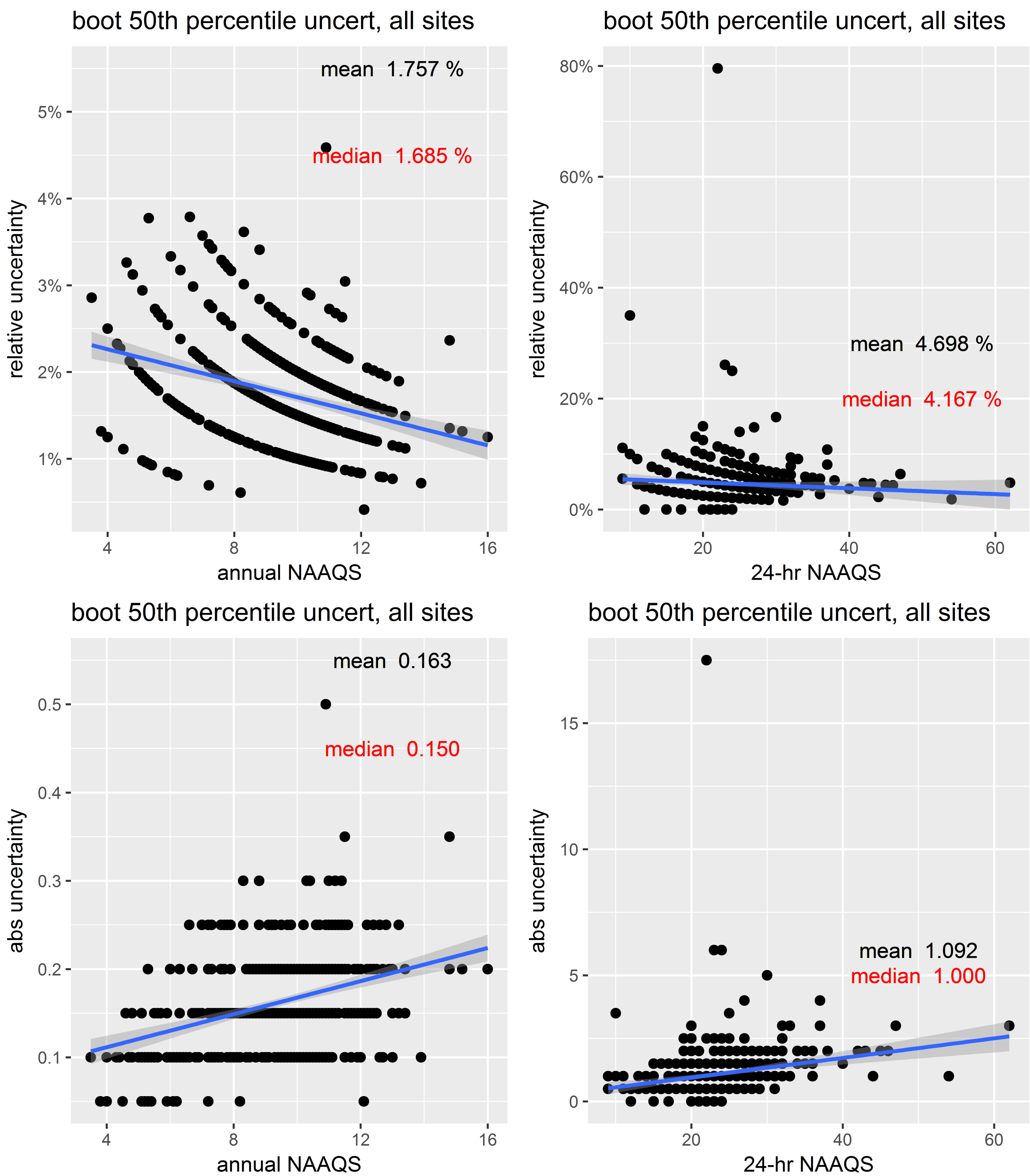
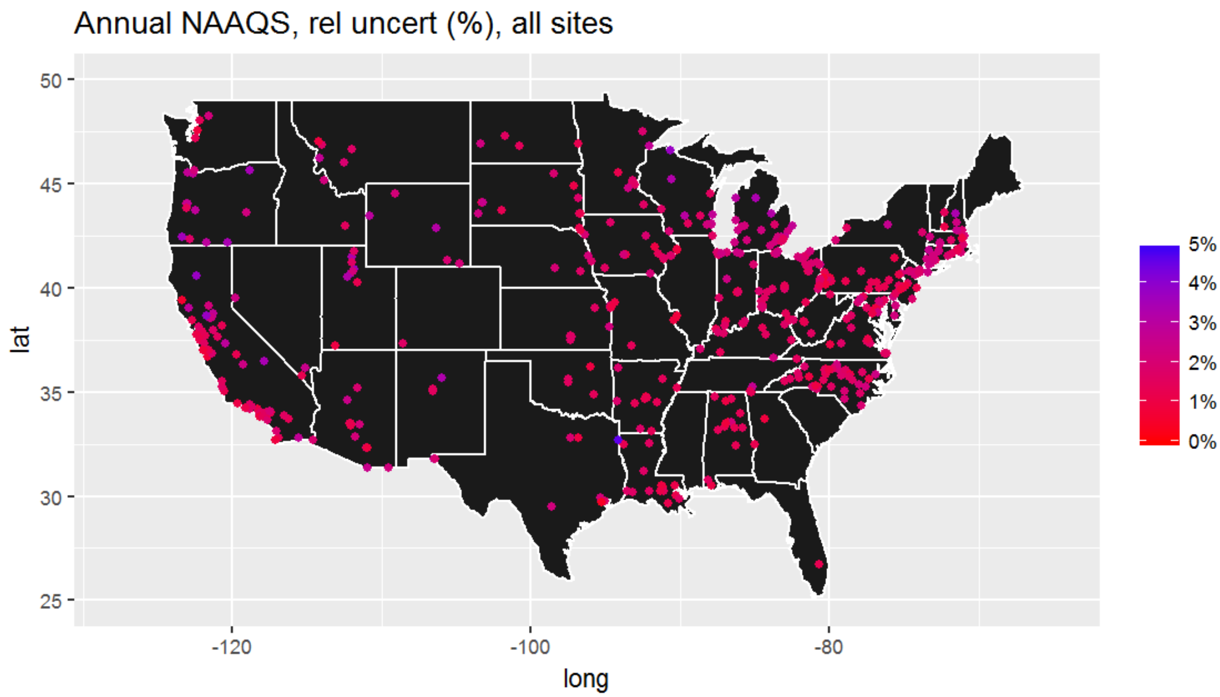


Figure 48: Bootstrap results from the 50% CIs for $PM_{2.5}$ DVs. The top two panels show the relative difference between the CI and the actual DV and the bottom two panels show the absolute difference between the values for the DVs at each site and the CI.



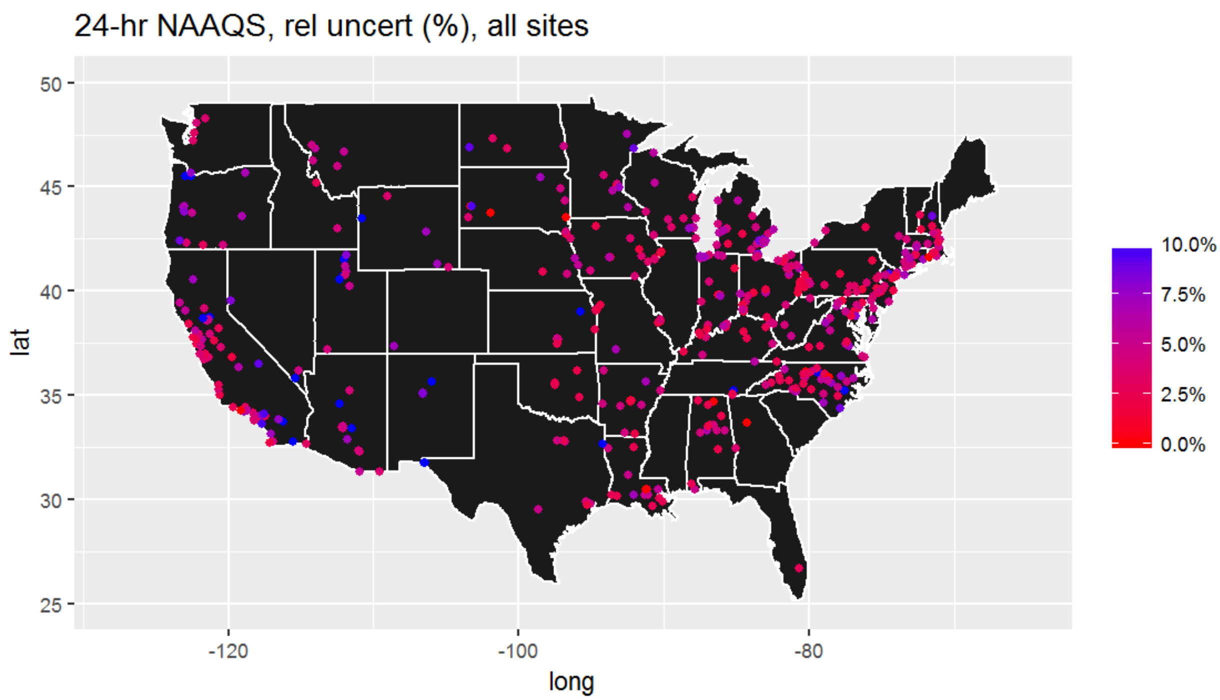


Figure 49: Spatial distribution of the relative difference between the CI and the actual DV from the 50% CIs for the 2012 $PM_{2.5}$ DVs.

3.5 2009-2011 $PM_{2.5}$ bootstrap results

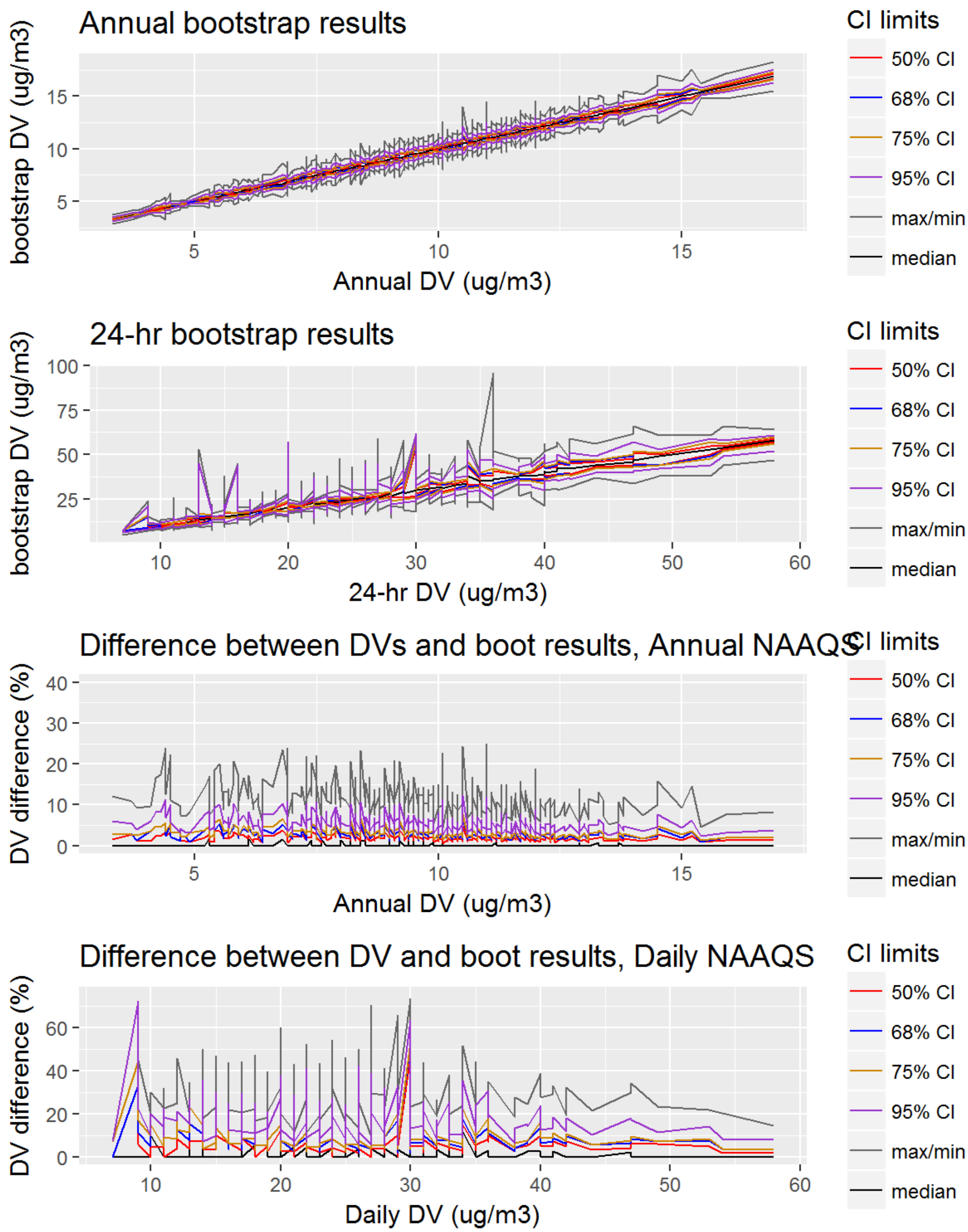


Figure 50: Bootstrap results for the 2011 $PM_{2.5}$ DVs, showing the 50%, 68%, 75%, and 95% CIs, along with the mean and median bootstrap DVs. The top two panels show the values for the DVs at the various CIs, while the bottom two panels show the relative difference between the CI and the actual DV.

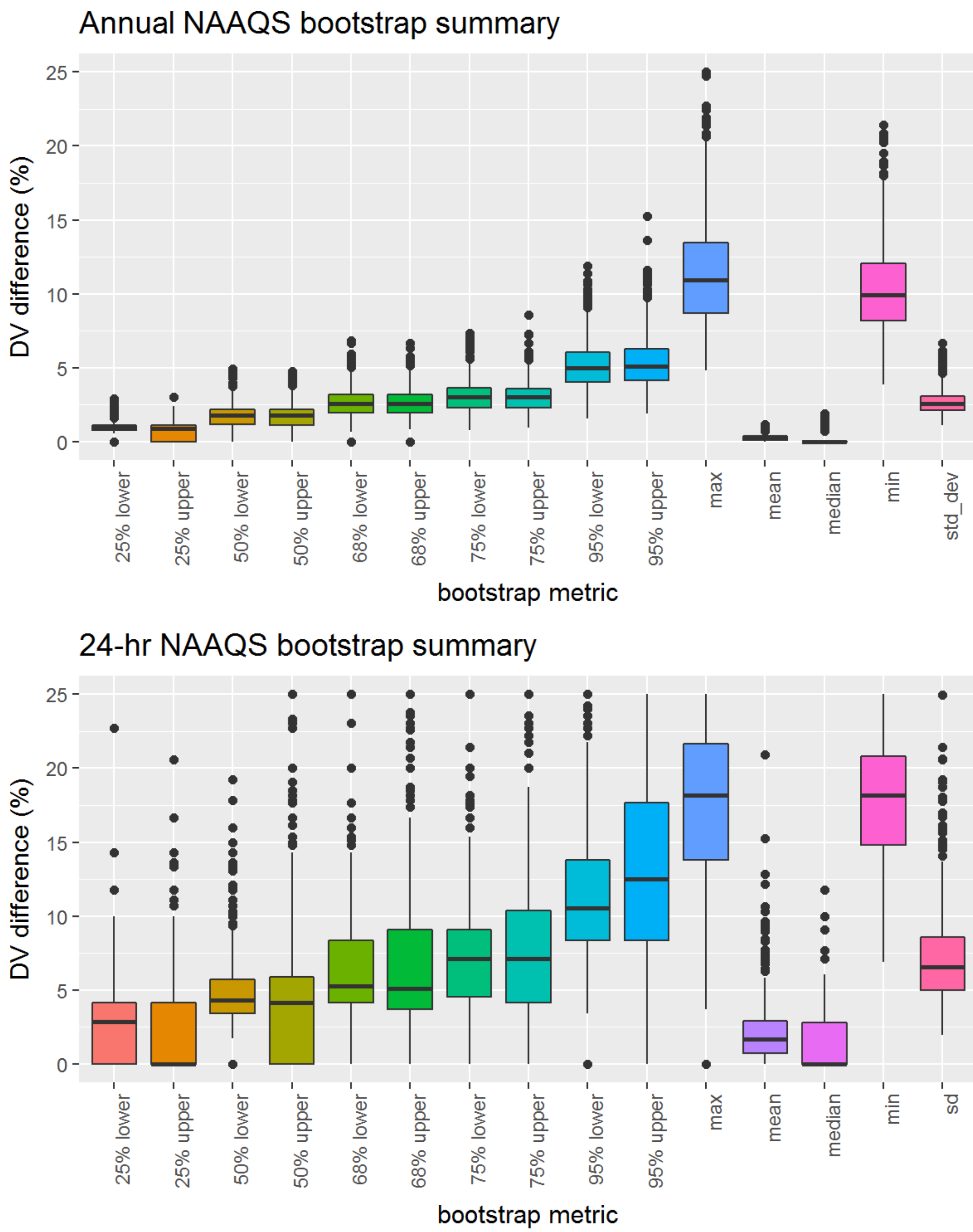


Figure 51: Bootstrap results for the 2011 $PM_{2.5}$ DVs, showing distribution of the relative differences between the bootstrap DVs and the actual DV at the 50%, 68%, 75%, and 95% CIs, along with the mean, median, maximum, minimum, standard deviations of the relative differences.

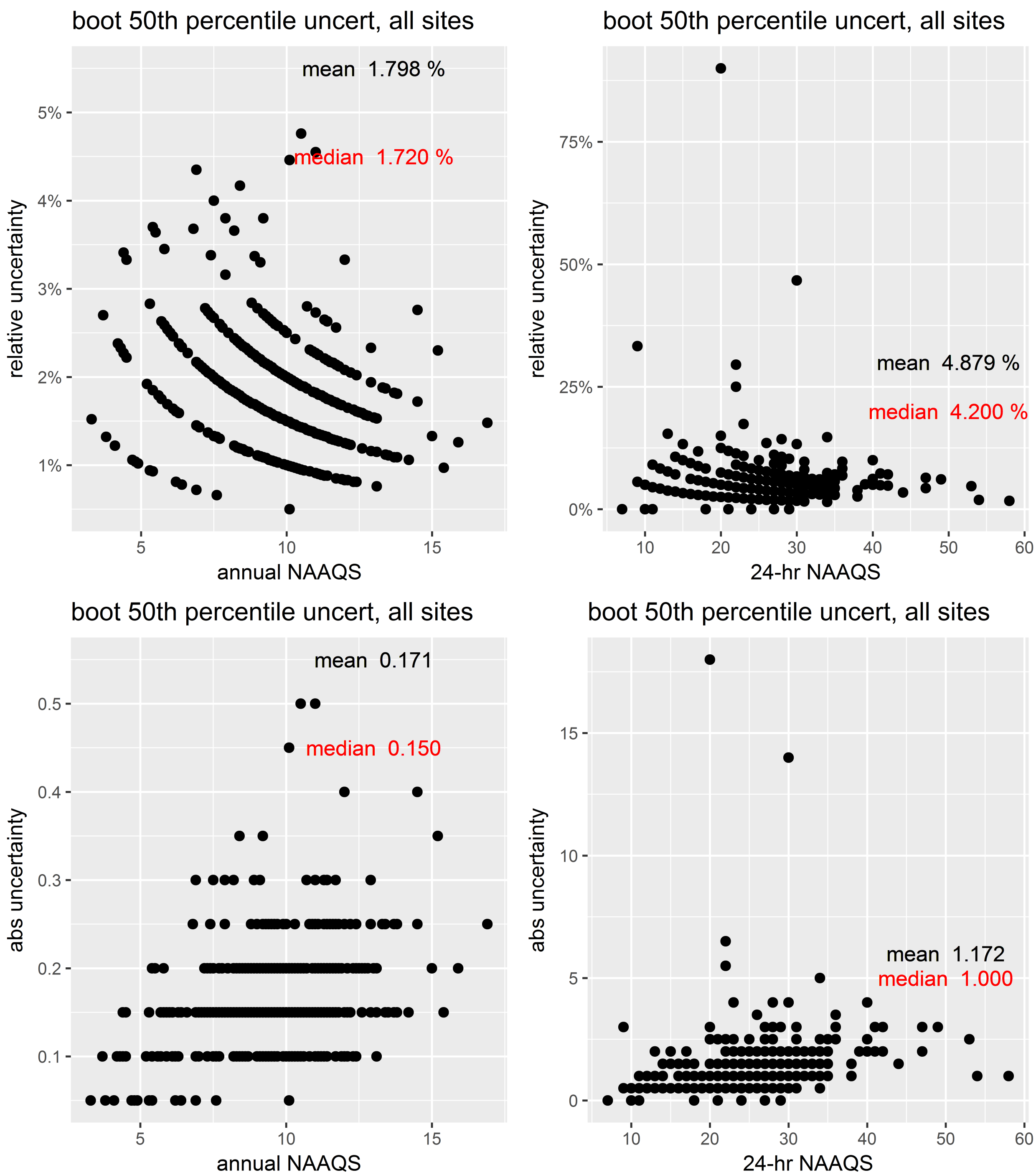


Figure 52: Bootstrap results from the 50% CIs for $PM_{2.5}$ DVs. The top two panels show the relative difference between the CI and the actual DV and the bottom two panels show the absolute difference between the values for the DVs at each site and the CI.

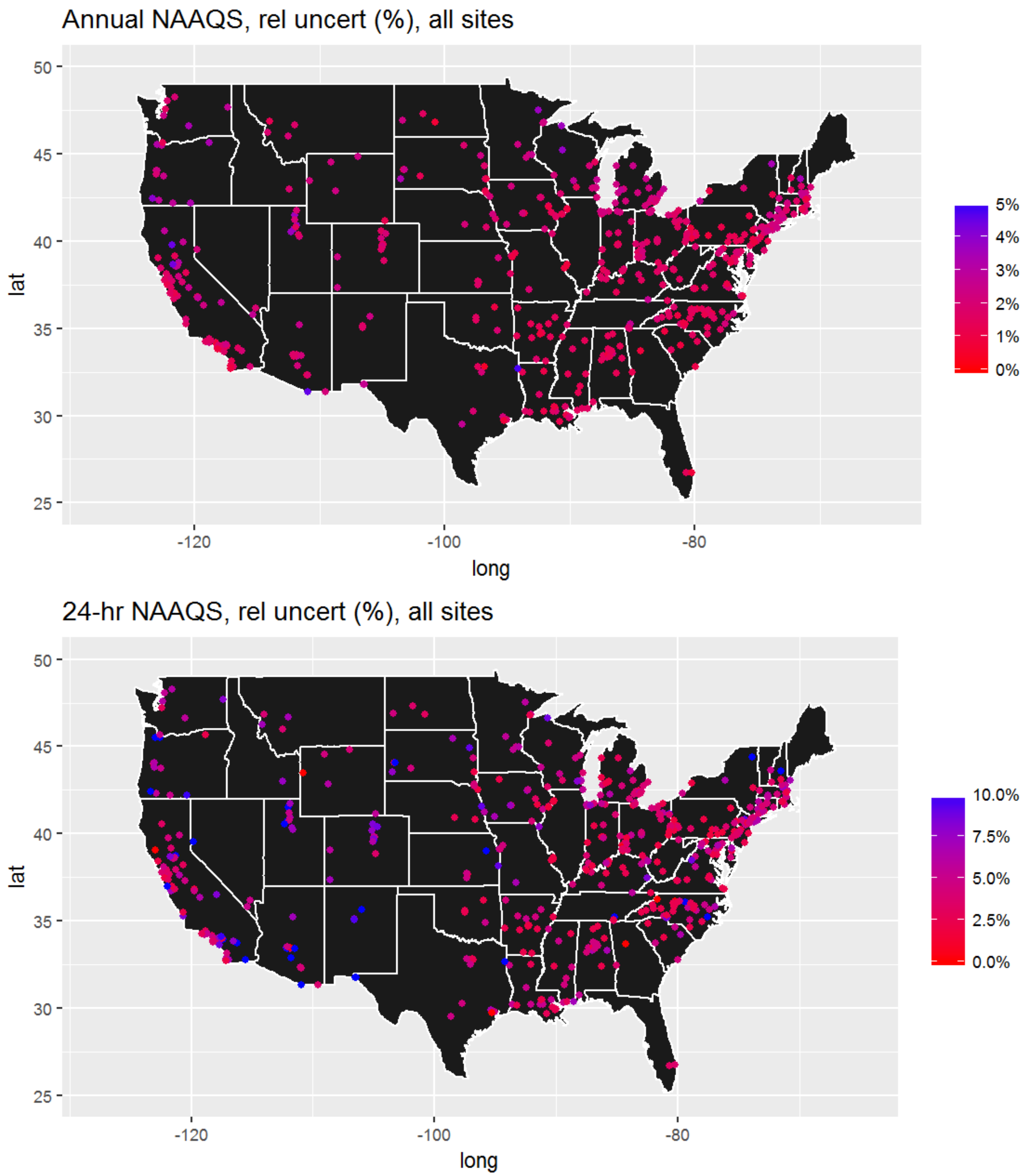


Figure 53: Spatial distribution of the relative difference between the CI and the actual DV from the 50% CIs for the 2011 $PM_{2.5}$ DVs.

3.6 2008-2010 $PM_{2.5}$ bootstrap results

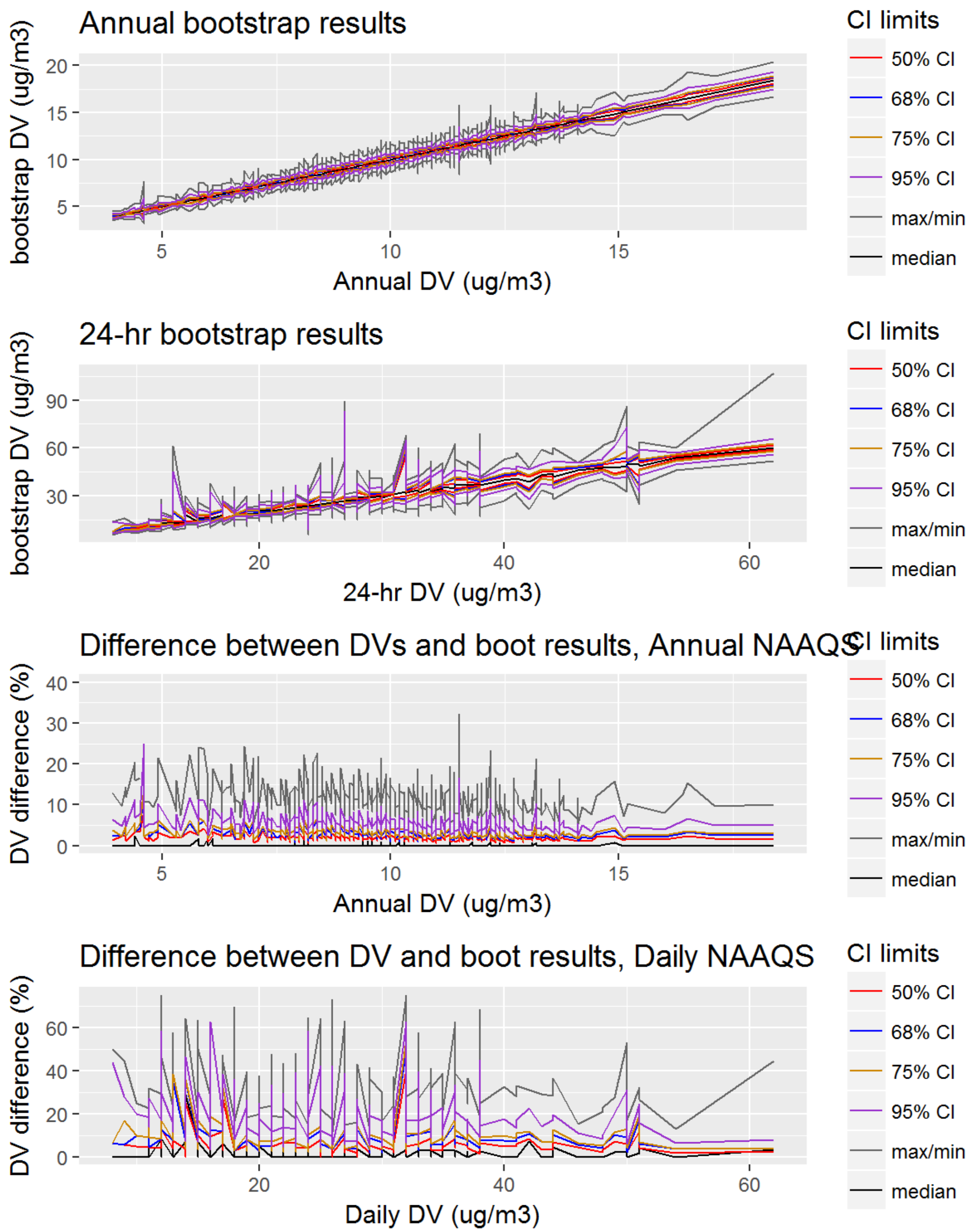


Figure 54: Bootstrap results for the 2010 $PM_{2.5}$ DVs, showing the 50%, 68%, 75%, and 95% CIs, along with the mean and median bootstrap DVs. The top two panels show the values for the DVs at the various CIs, while the bottom two panels show the relative difference between the CI and the actual DV.

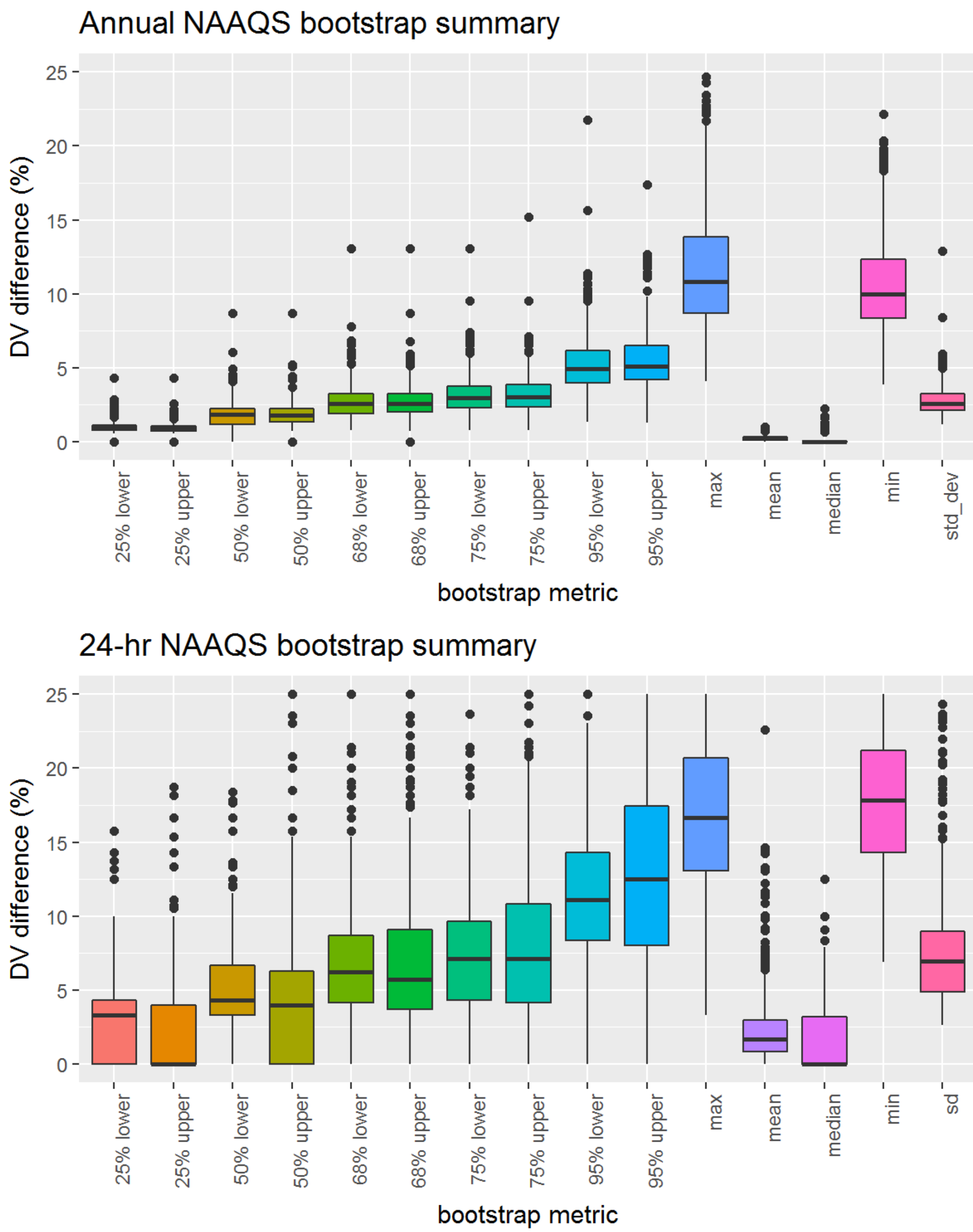


Figure 55: Bootstrap results for the 2010 $PM_{2.5}$ DVs, showing distribution of the relative differences between the bootstrap DVs and the actual DV at the 50%, 68%, 75%, and 95% CIs, along with the mean, median, maximum, minimum, standard deviations of the relative differences.

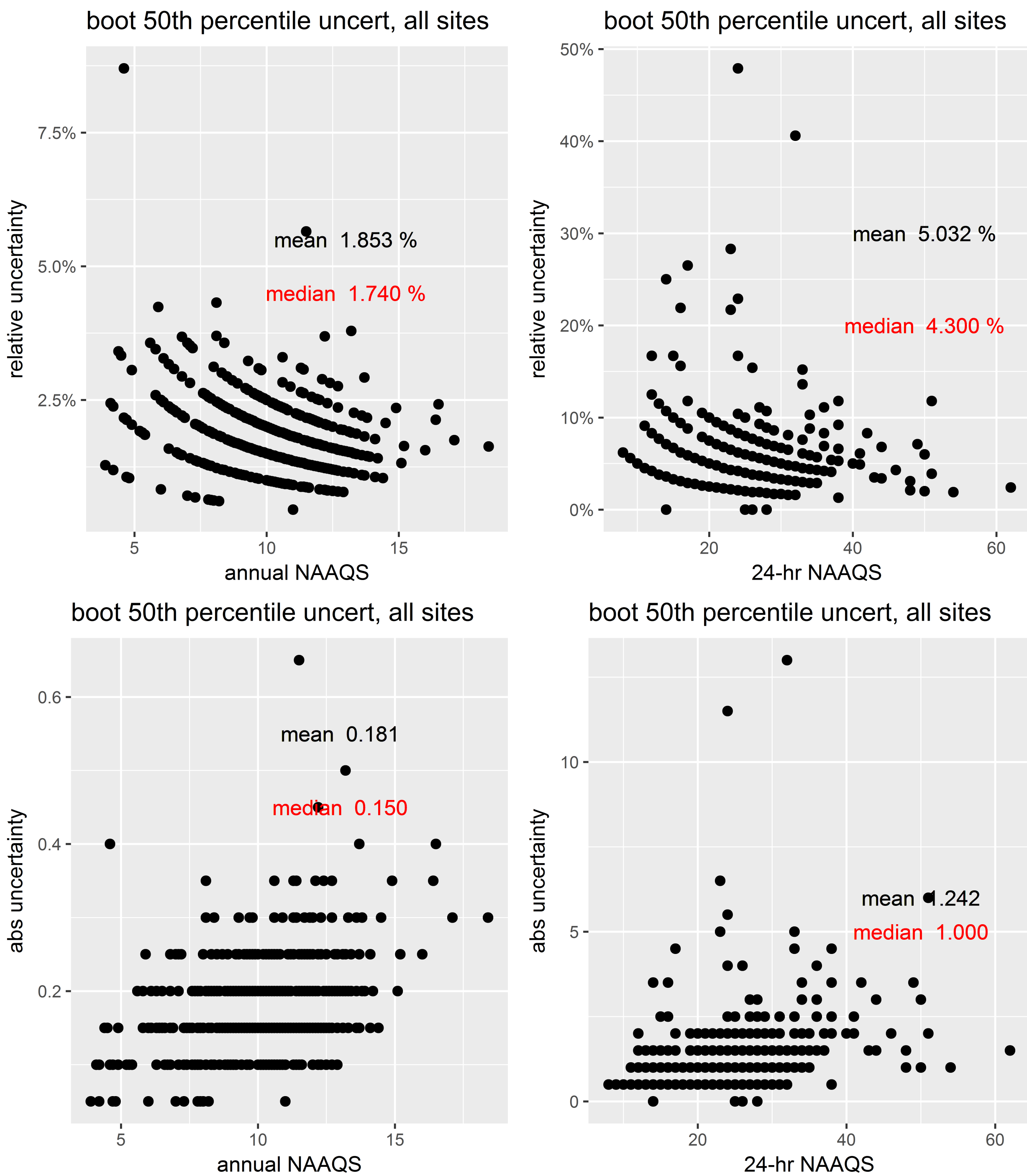
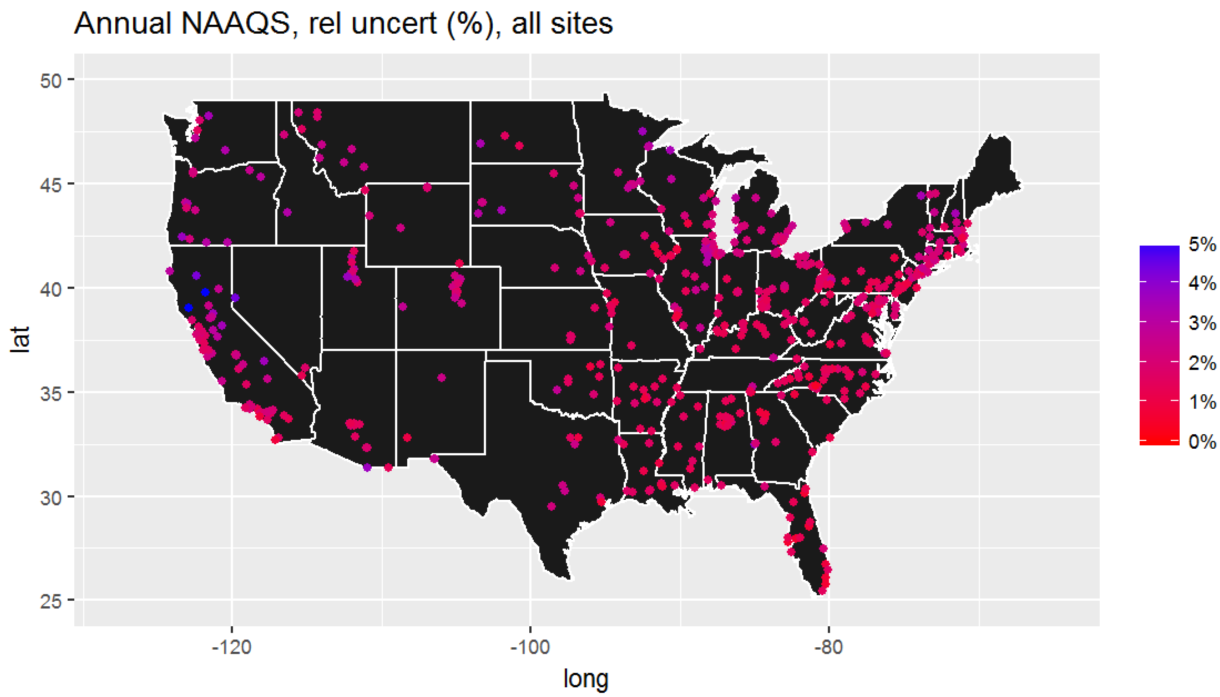


Figure 56: Bootstrap results from the 50% CIs for $PM_{2.5}$ DVs. The top two panels show the relative difference between the CI and the actual DV and the bottom two panels show the absolute difference between the values for the DVs at each site and the CI.



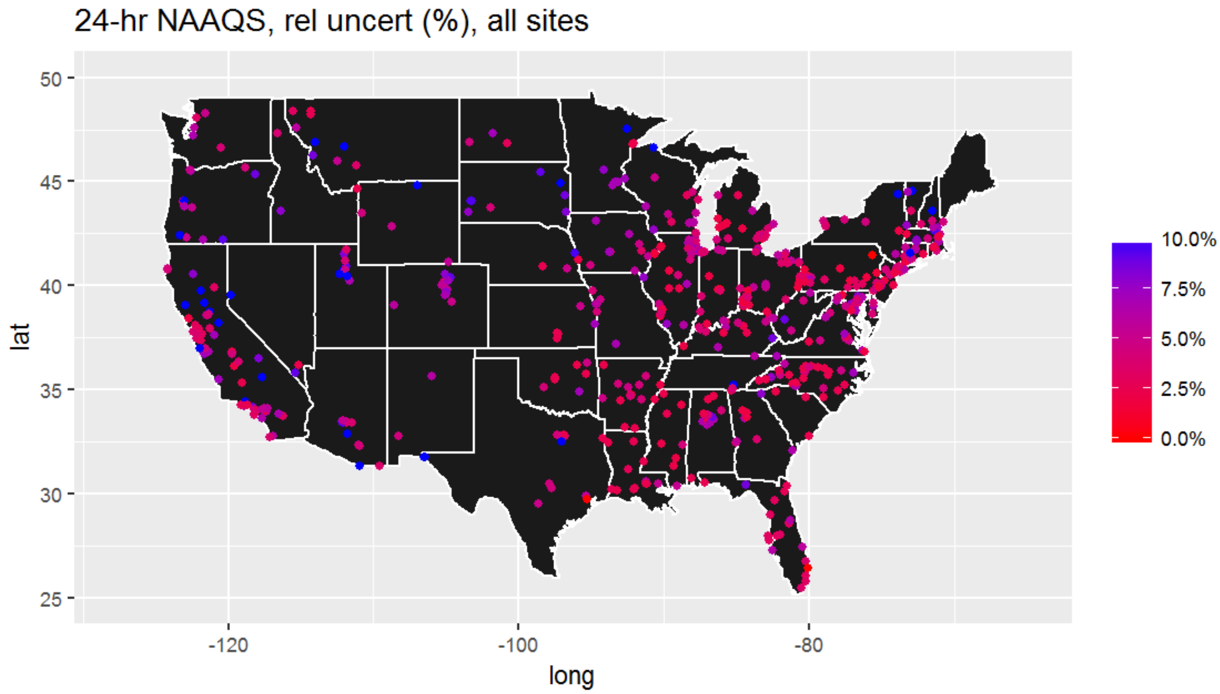


Figure 57: Spatial distribution of the relative difference between the CI and the actual DV from the 50% CIs for the 2010 $PM_{2.5}$ DVs.

3.7 2007-2009 $PM_{2.5}$ bootstrap results

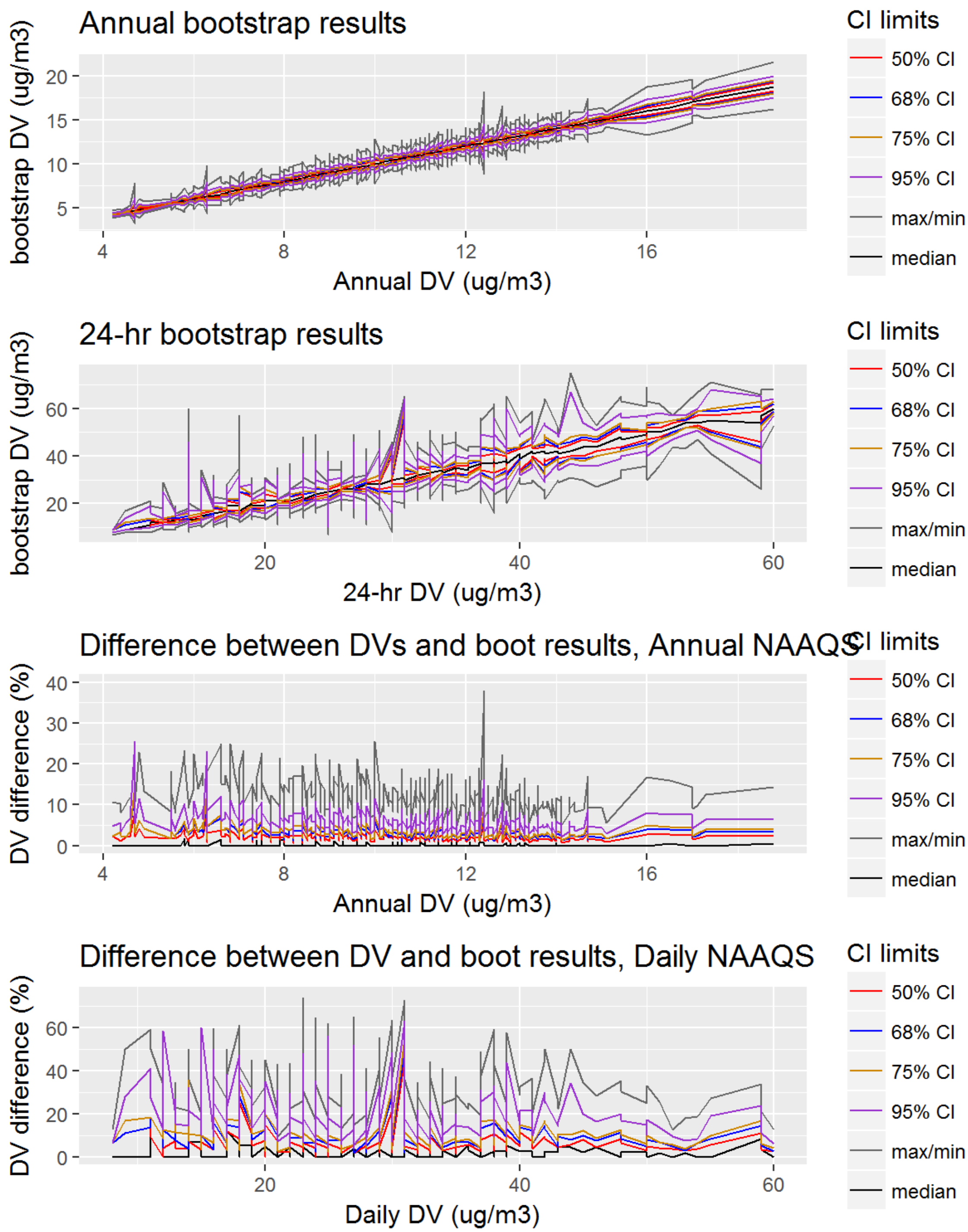


Figure 58: Bootstrap results for the 2009 $PM_{2.5}$ DVs, showing the 50%, 68%, 75%, and 95% CIs, along with the mean and median bootstrap DVs. The top two panels show the values for the DVs at the various CIs, while the bottom two panels show the relative difference between the CI and the actual DV.

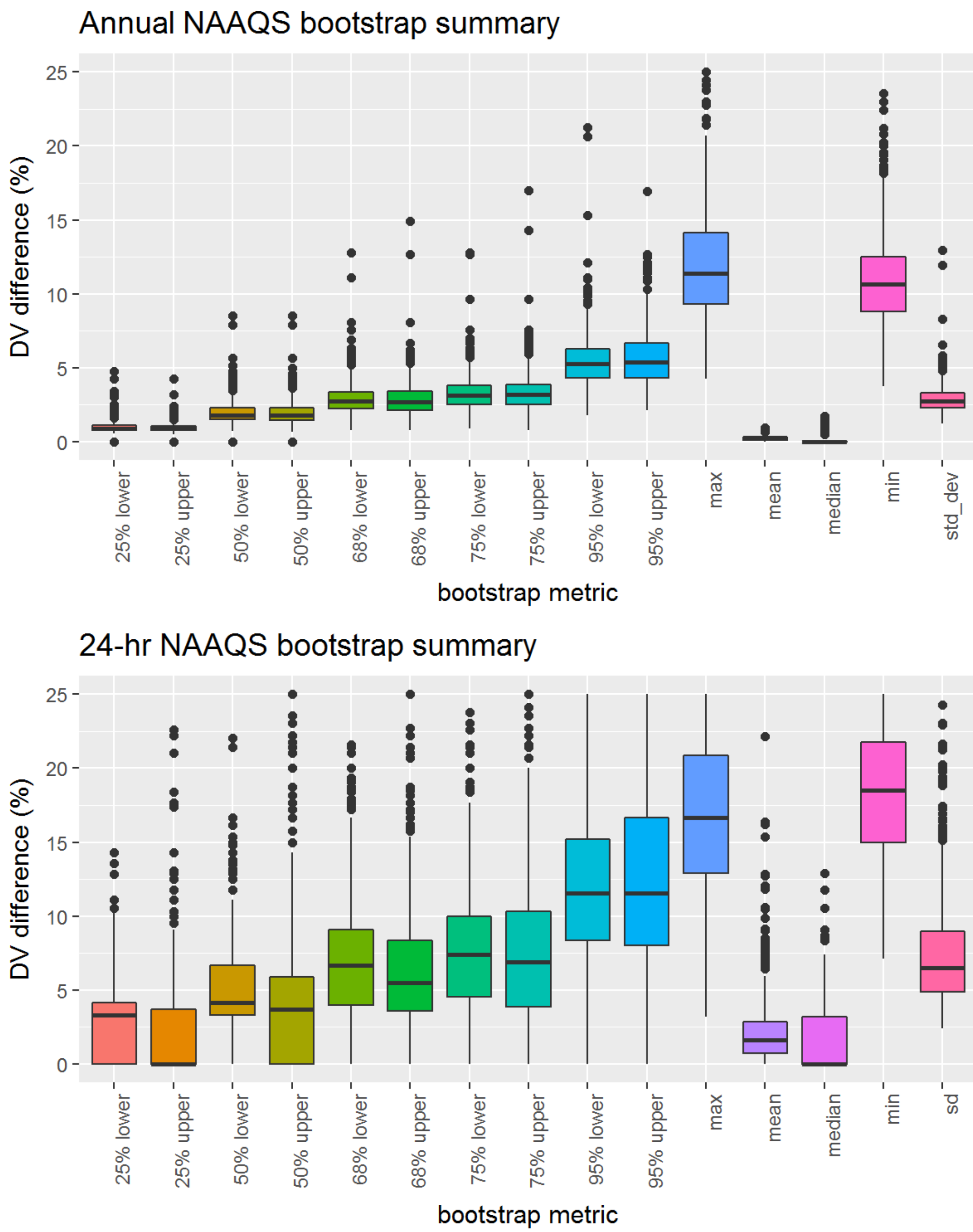


Figure 59: Bootstrap results for the 2009 $PM_{2.5}$ DVs, showing distribution of the relative differences between the bootstrap DVs and the actual DV at the 50%, 68%, 75%, and 95% CIs, along with the mean, median, maximum, minimum, standard deviations of the relative differences.

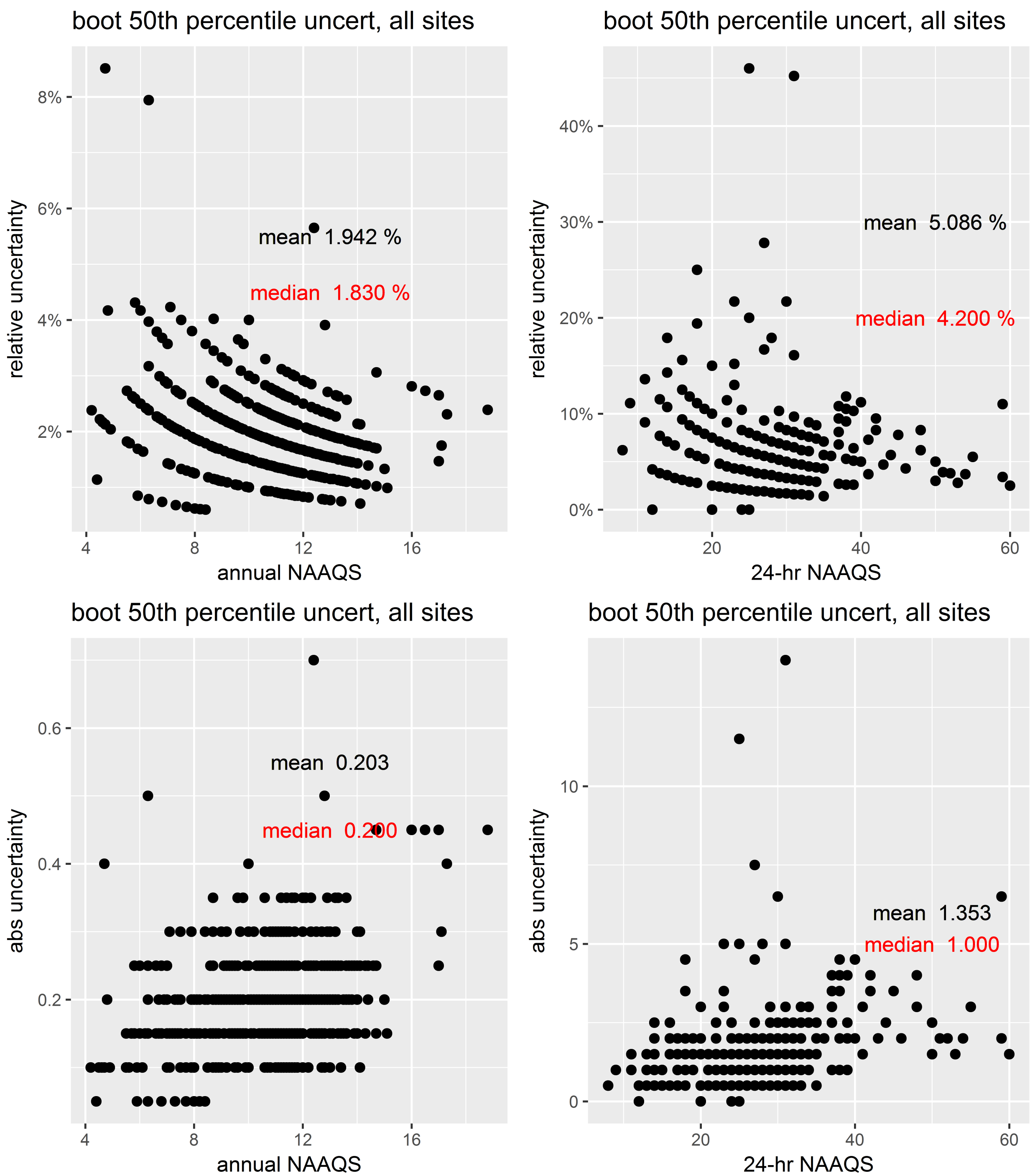


Figure 60: Bootstrap results from the 50% CIs for $PM_{2.5}$ DVs. The top two panels show the relative difference between the CI and the actual DV and the bottom two panels show the absolute difference between the values for the DVs at each site and the CI.

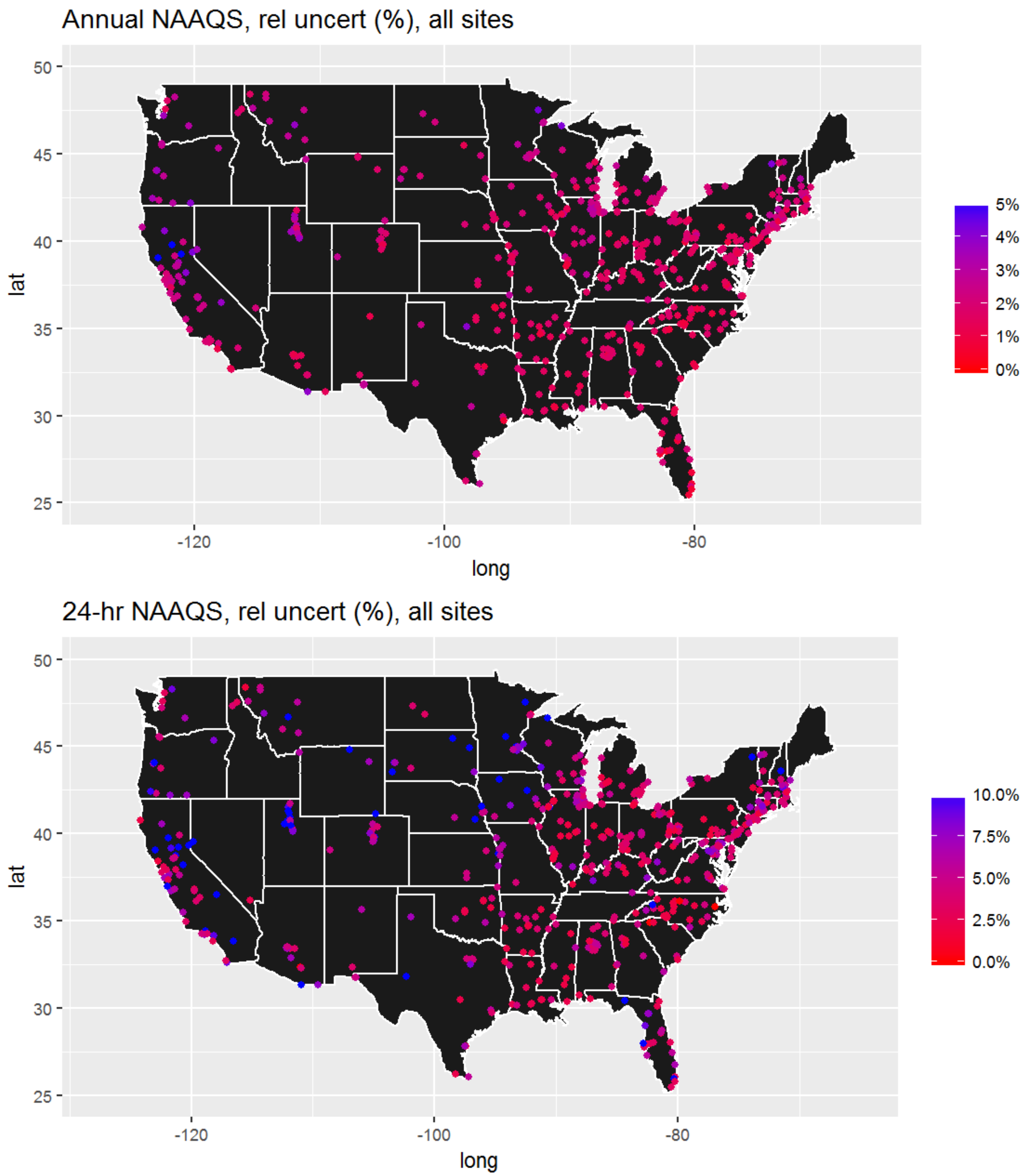


Figure 61: Spatial distribution of the relative difference between the CI and the actual DV from the 50% CIs for the 2009 $PM_{2.5}$ DVs.

3.8 2006-2008 $PM_{2.5}$ bootstrap results

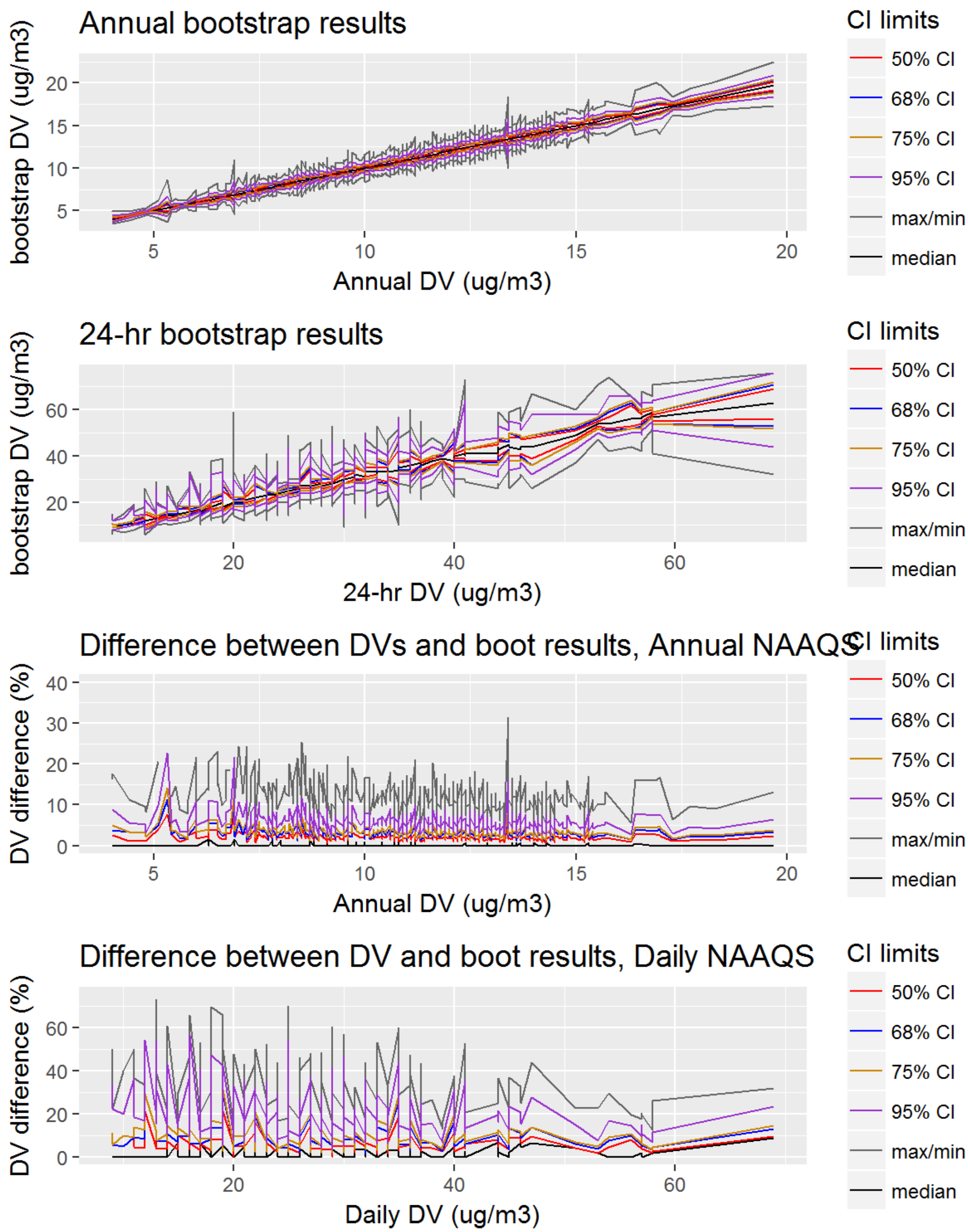


Figure 62: Bootstrap results for the 2008 $PM_{2.5}$ DVs, showing the 50%, 68%, 75%, and 95% CIs, along with the mean and median bootstrap DVs. The top two panels show the values for the DVs at the various CIs, while the bottom two panels show the relative difference between the CI and the actual DV.

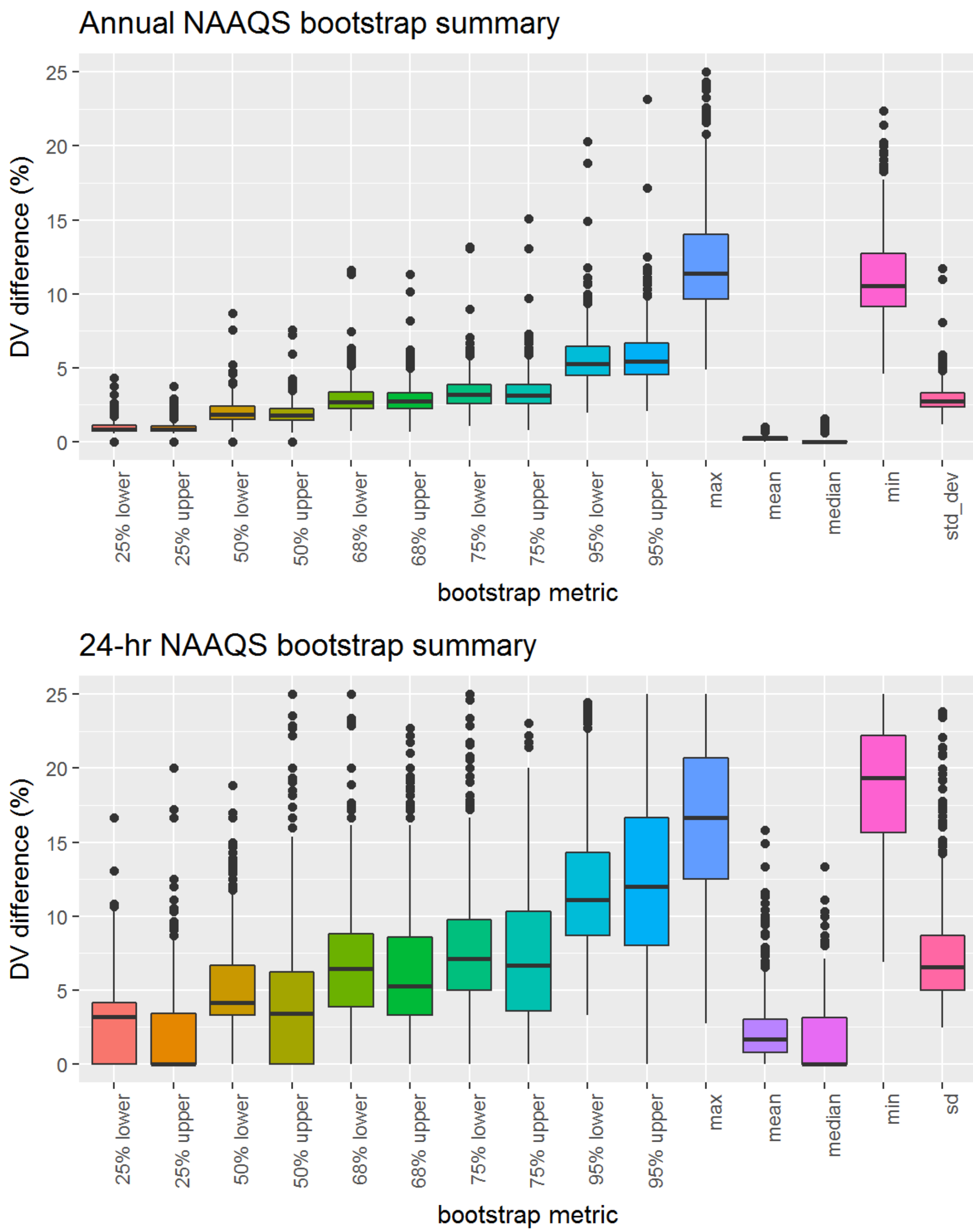


Figure 63: Bootstrap results for the 2008 $PM_{2.5}$ DVs, showing distribution of the relative differences between the bootstrap DVs and the actual DV at the 50%, 68%, 75%, and 95% CIs, along with the mean, median, maximum, minimum, standard deviations of the relative differences.

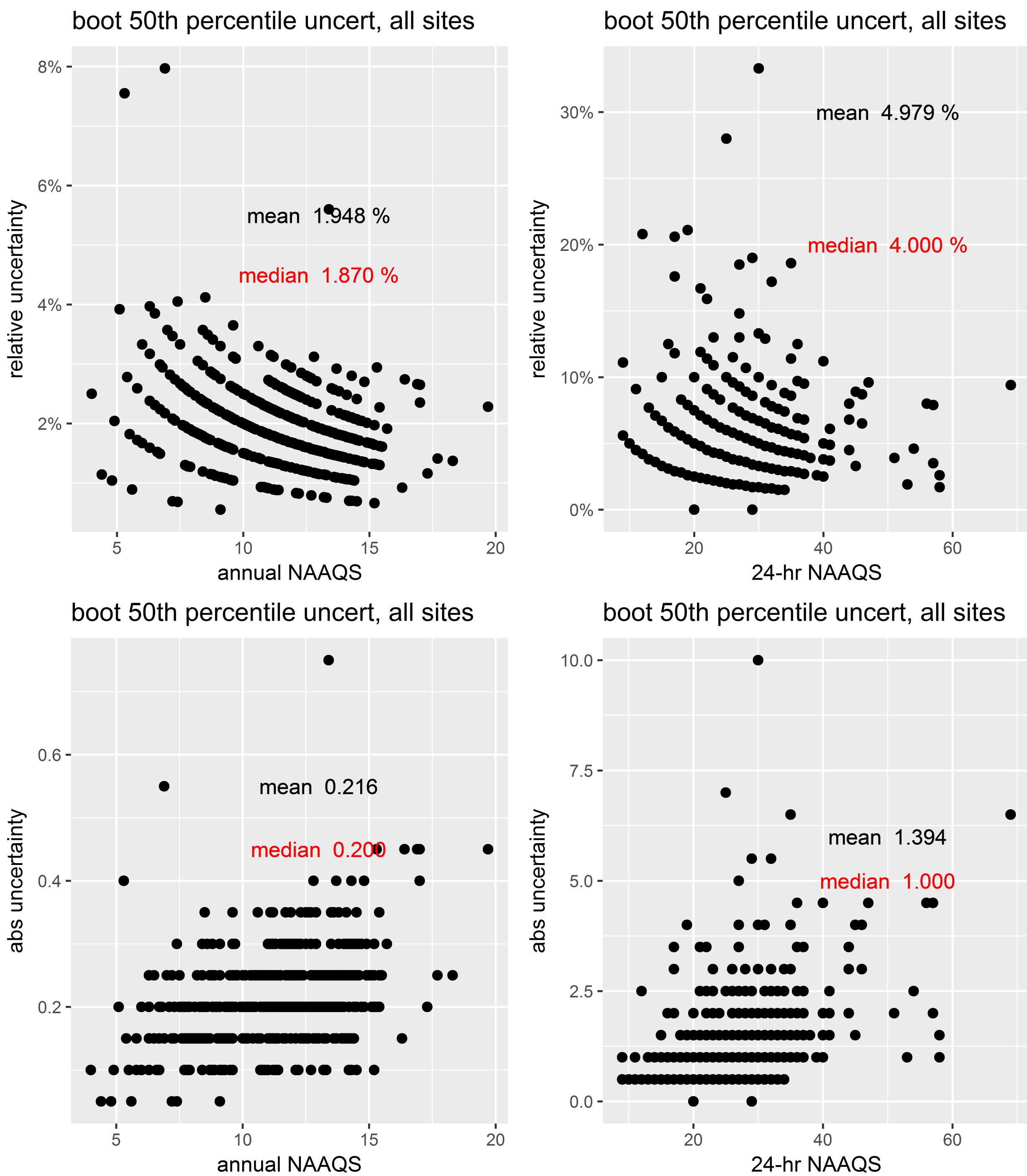


Figure 64: Bootstrap results from the 50% CIs for $PM_{2.5}$ DVs. The top two panels show the relative difference between the CI and the actual DV and the bottom two panels show the absolute difference between the values for the DVs at each site and the CI.

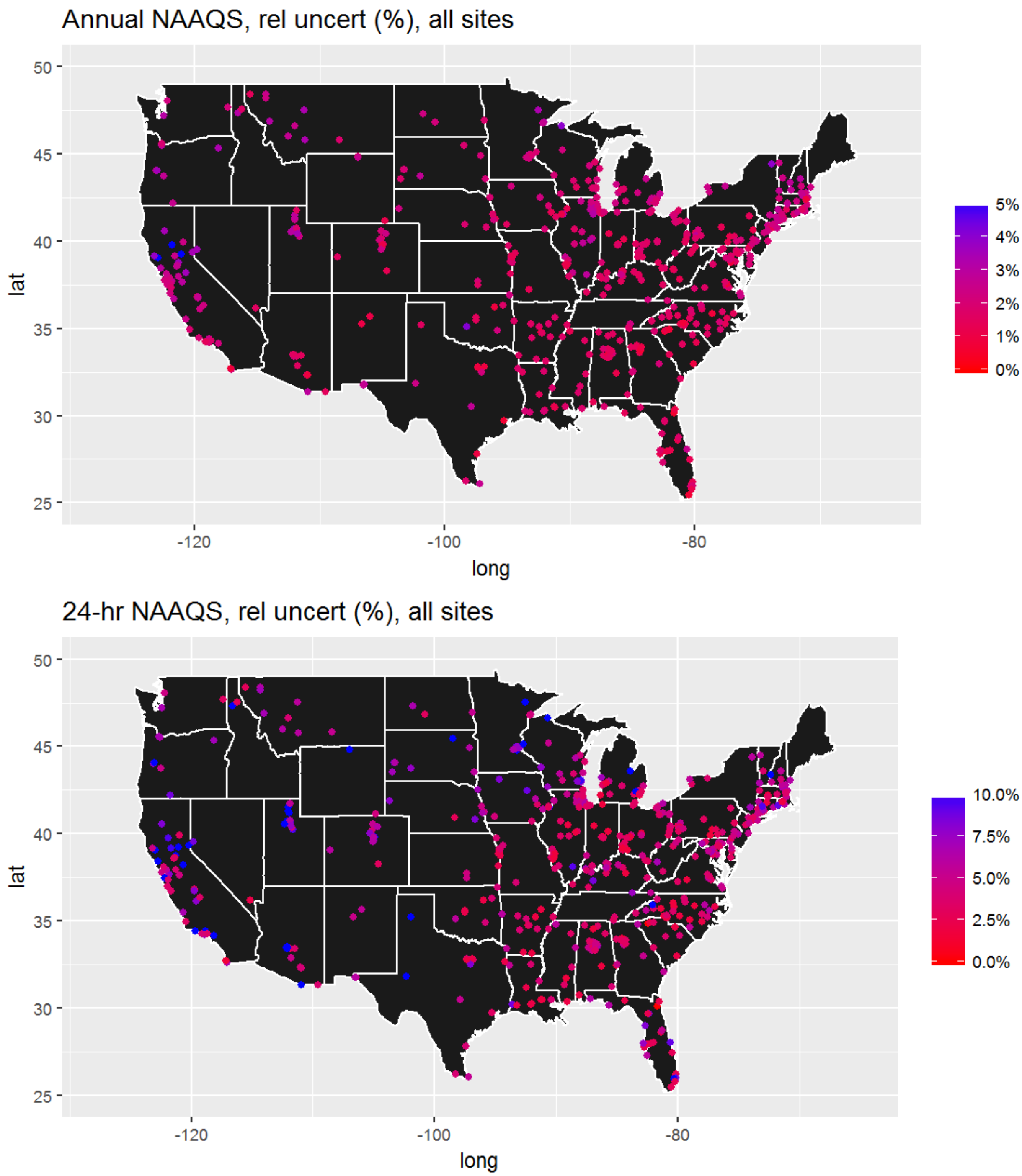


Figure 65: Spatial distribution of the relative difference between the CI and the actual DV from the 50% CIs for the 2008 $PM_{2.5}$ DVs.

3.9 2005-2007 $PM_{2.5}$ bootstrap results

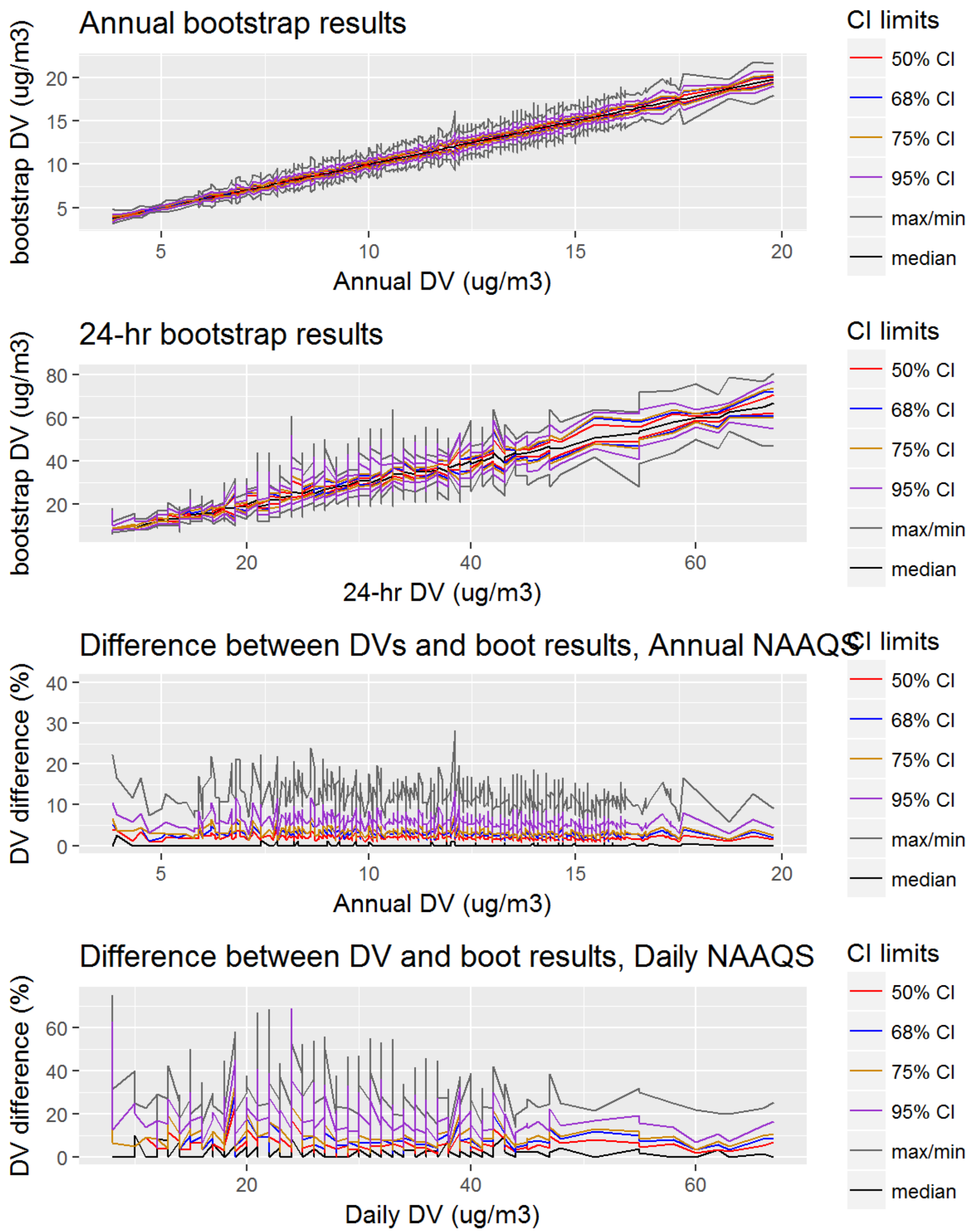


Figure 66: Bootstrap results for the 2007 $PM_{2.5}$ DVs, showing the 50%, 68%, 75%, and 95% CIs, along with the mean and median bootstrap DVs. The top two panels show the values for the DVs at the various CIs, while the bottom two panels show the relative difference between the CI and the actual DV.

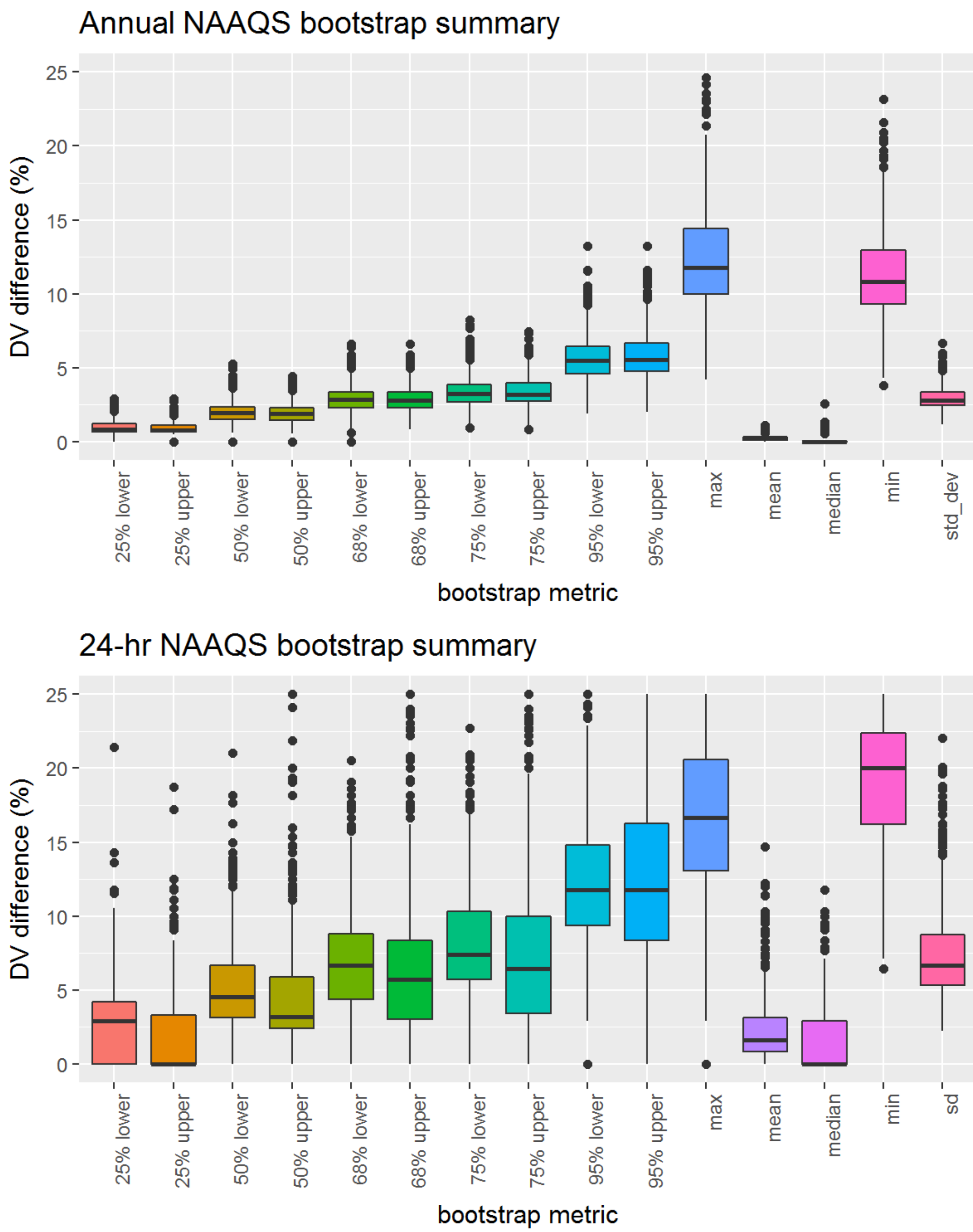


Figure 67: Bootstrap results for the 2007 $PM_{2.5}$ DVs, showing distribution of the relative differences between the bootstrap DVs and the actual DV at the 50%, 68%, 75%, and 95% CIs, along with the mean, median, maximum, minimum, standard deviations of the relative differences.

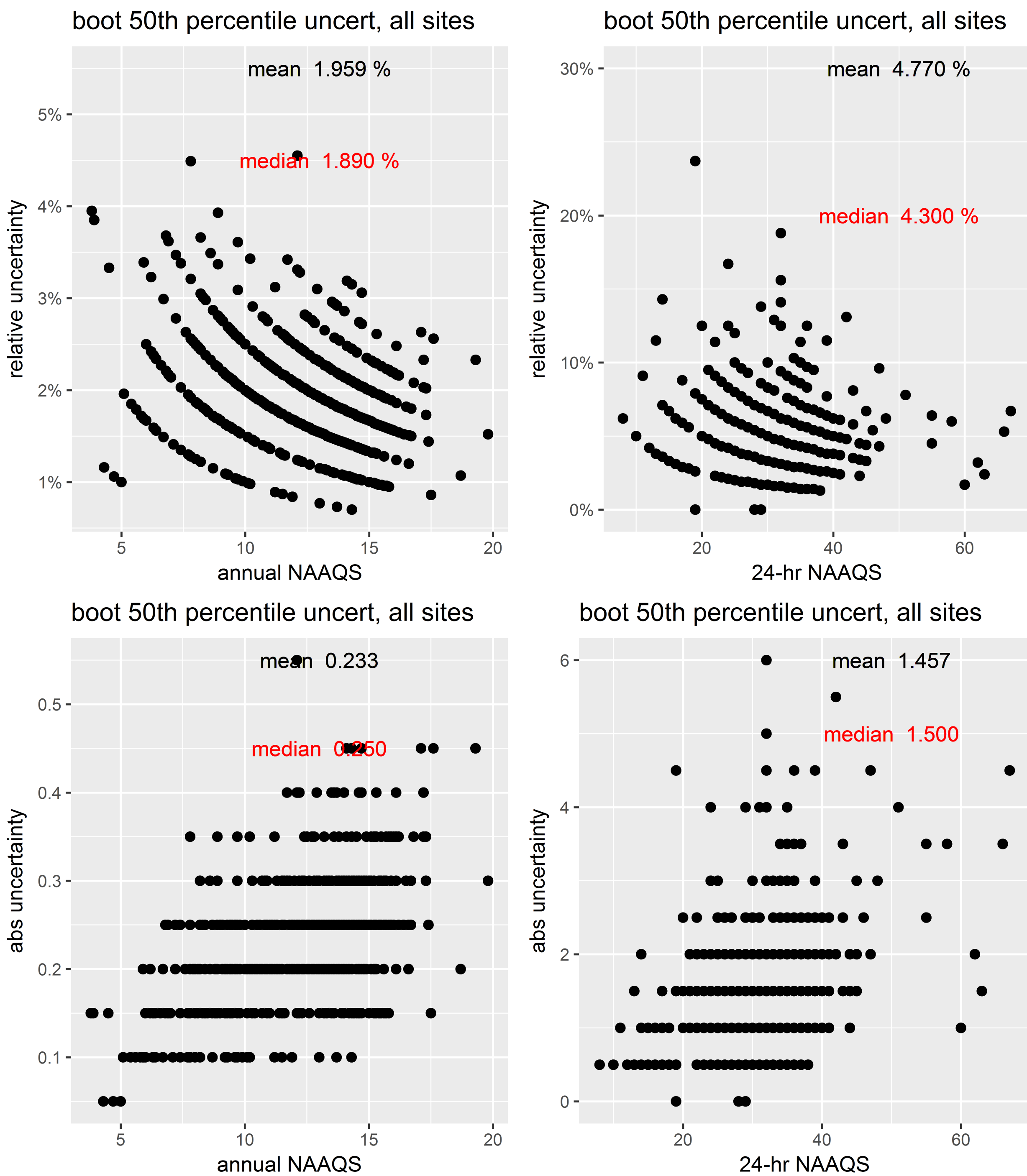


Figure 68: Bootstrap results from the 50% CIs for $PM_{2.5}$ DVs. The top two panels show the relative difference between the CI and the actual DV and the bottom two panels show the absolute difference between the values for the DVs at each site and the CI.

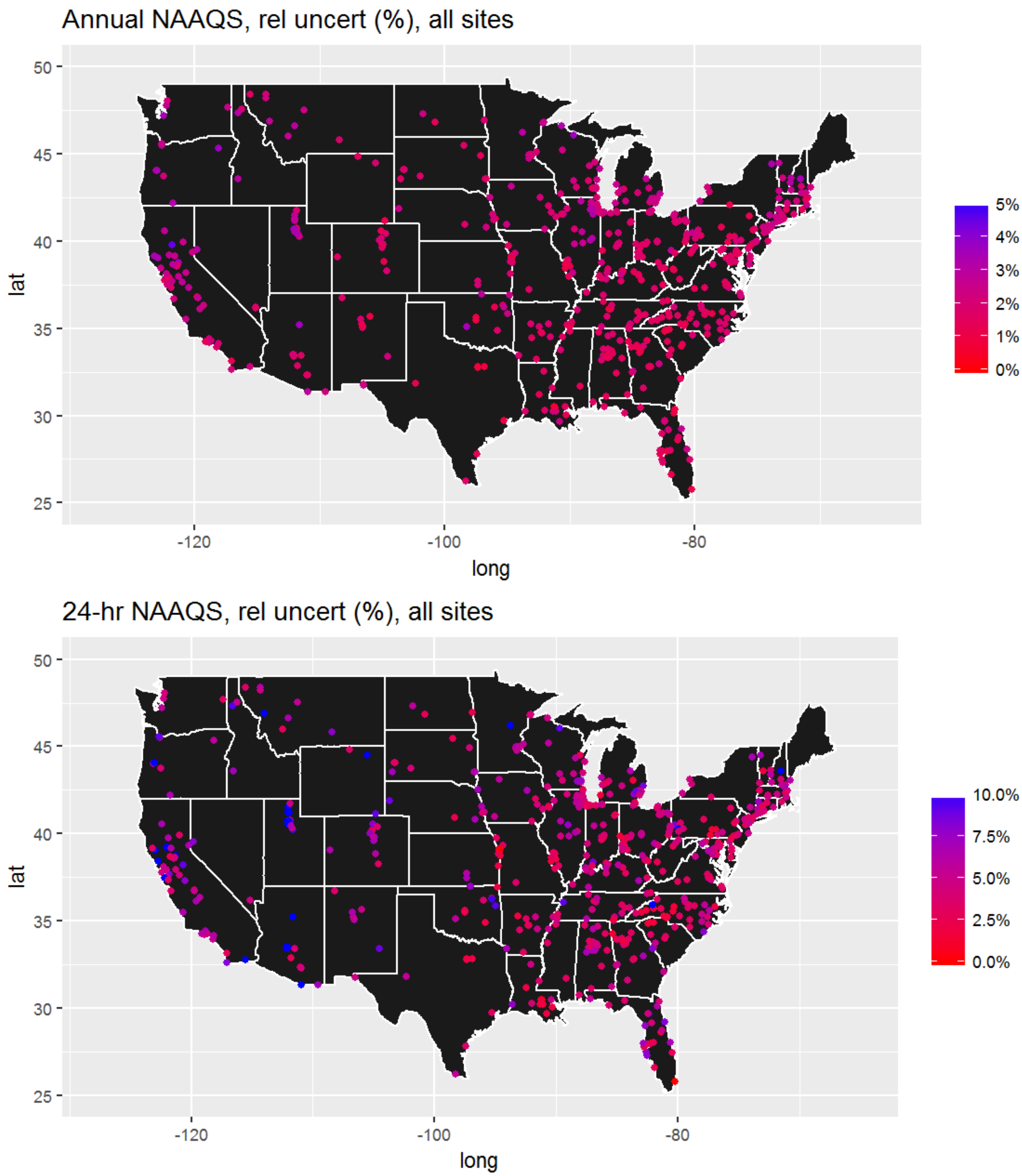


Figure 69: Spatial distribution of the relative difference between the CI and the actual DV from the 50% CIs for the 2007 $PM_{2.5}$ DVs.

3.10 2004-2006 $PM_{2.5}$ bootstrap results

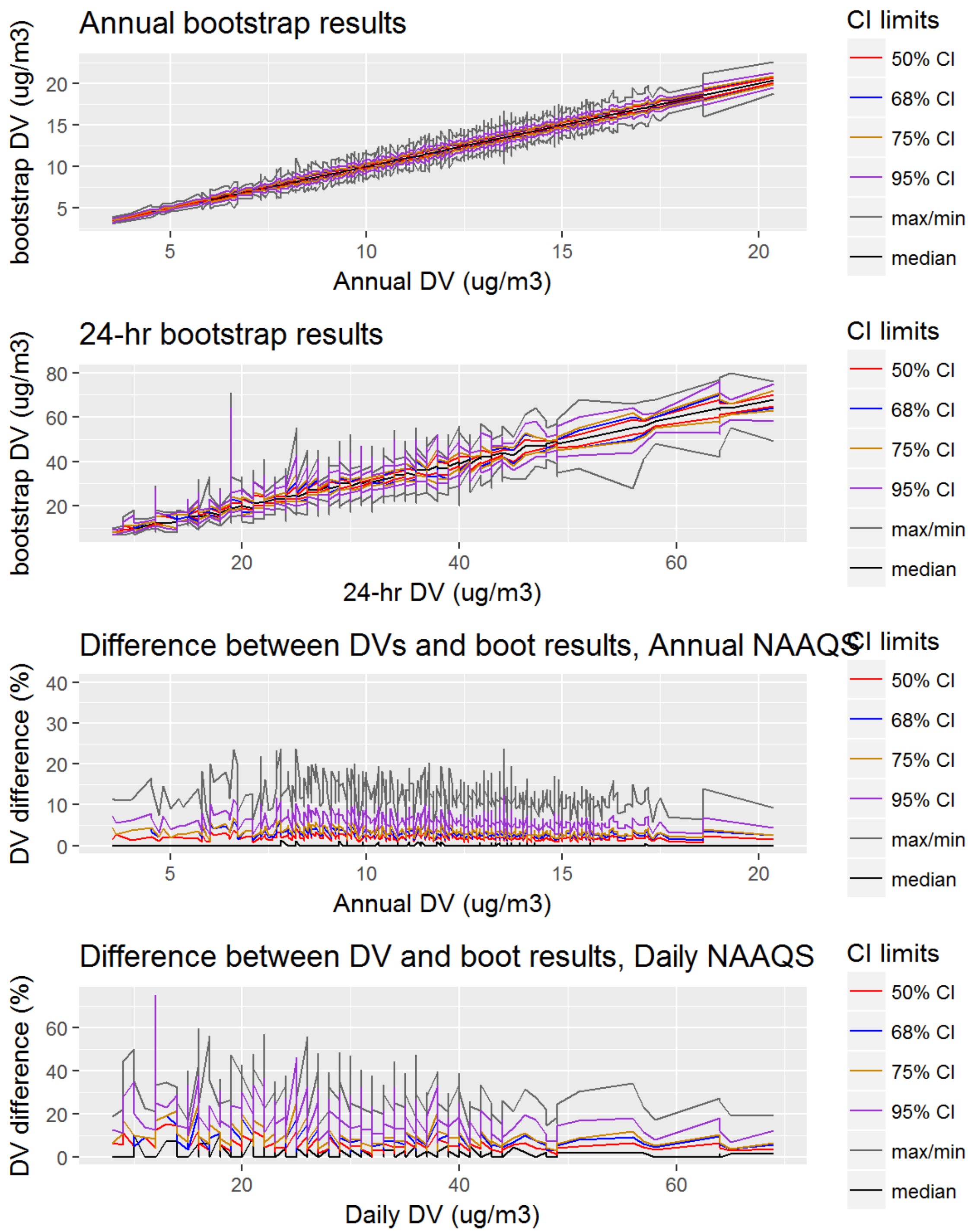


Figure 70: Bootstrap results for the 2006 $PM_{2.5}$ DVs, showing the 50%, 68%, 75%, and 95% CIs, along with the mean and median bootstrap DVs. The top two panels show the values for the DVs at the various CIs, while the bottom two panels show the relative difference between the CI and the actual DV.

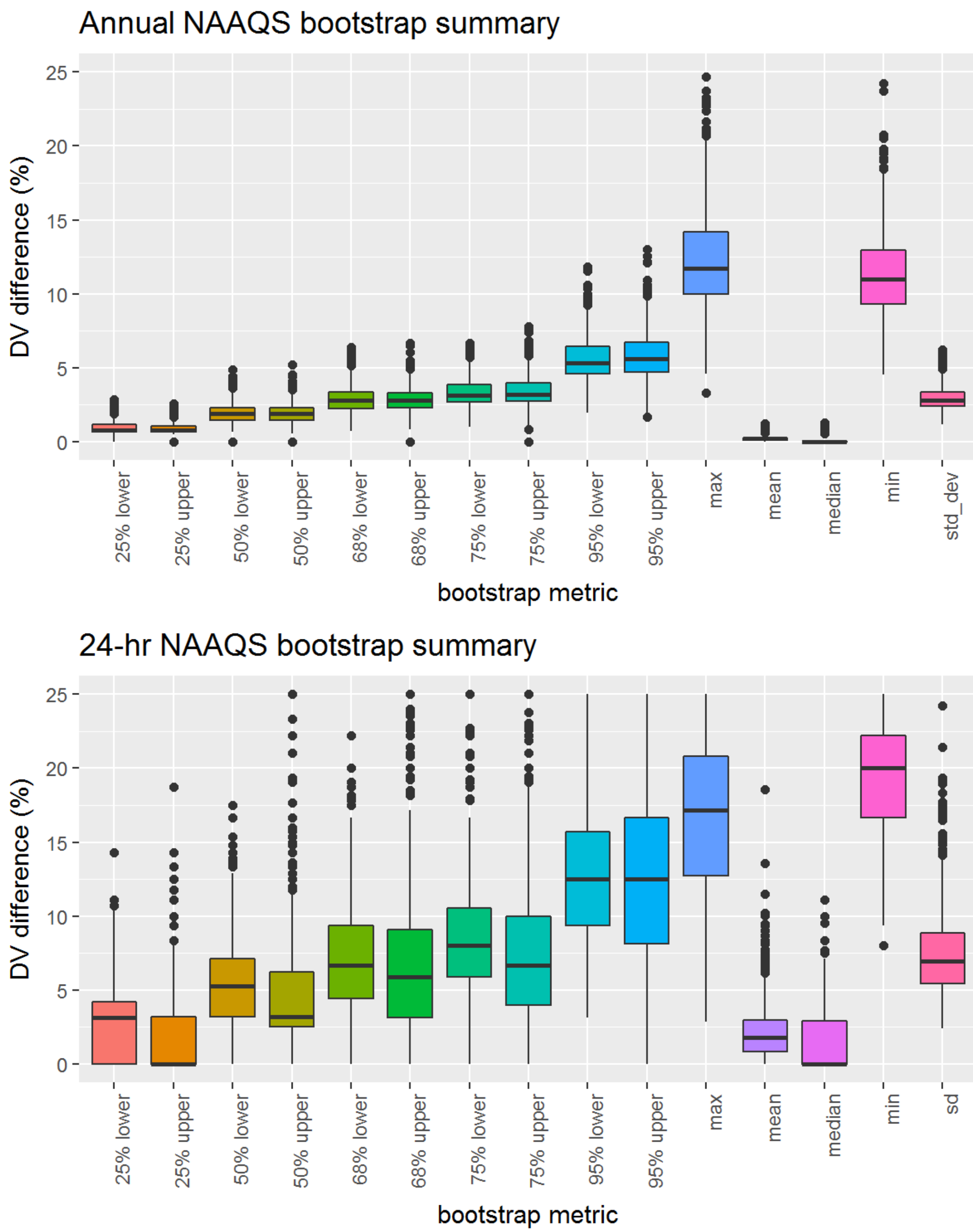


Figure 71: Bootstrap results for the 2006 $PM_{2.5}$ DVs, showing distribution of the relative differences between the bootstrap DVs and the actual DV at the 50%, 68%, 75%, and 95% CIs, along with the mean, median, maximum, minimum, standard deviations of the relative differences.

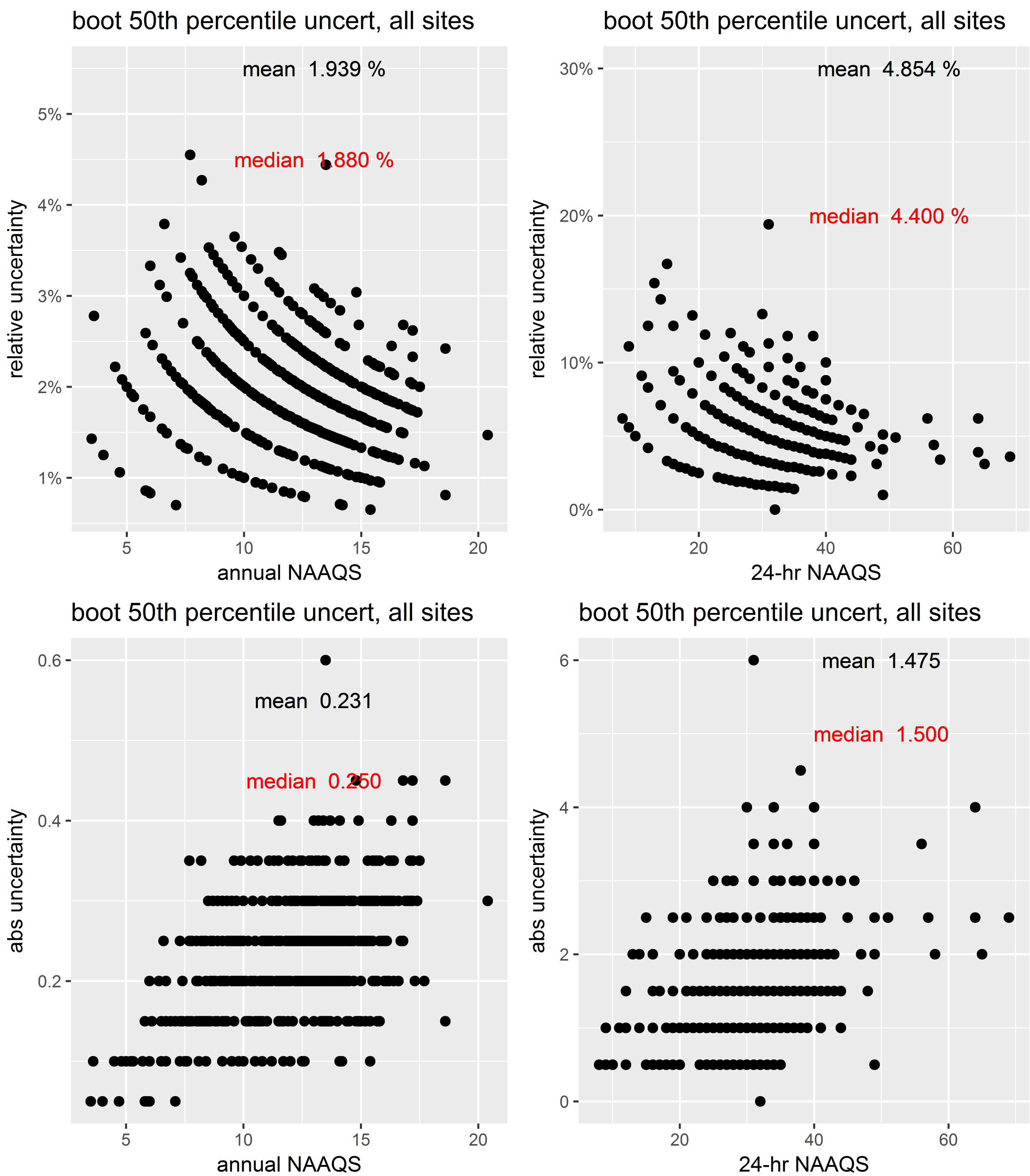


Figure 72: Bootstrap results from the 50% CIs for $PM_{2.5}$ DVs. The top two panels show the relative difference between the CI and the actual DV and the bottom two panels show the absolute difference between the values for the DVs at each site and the CI.

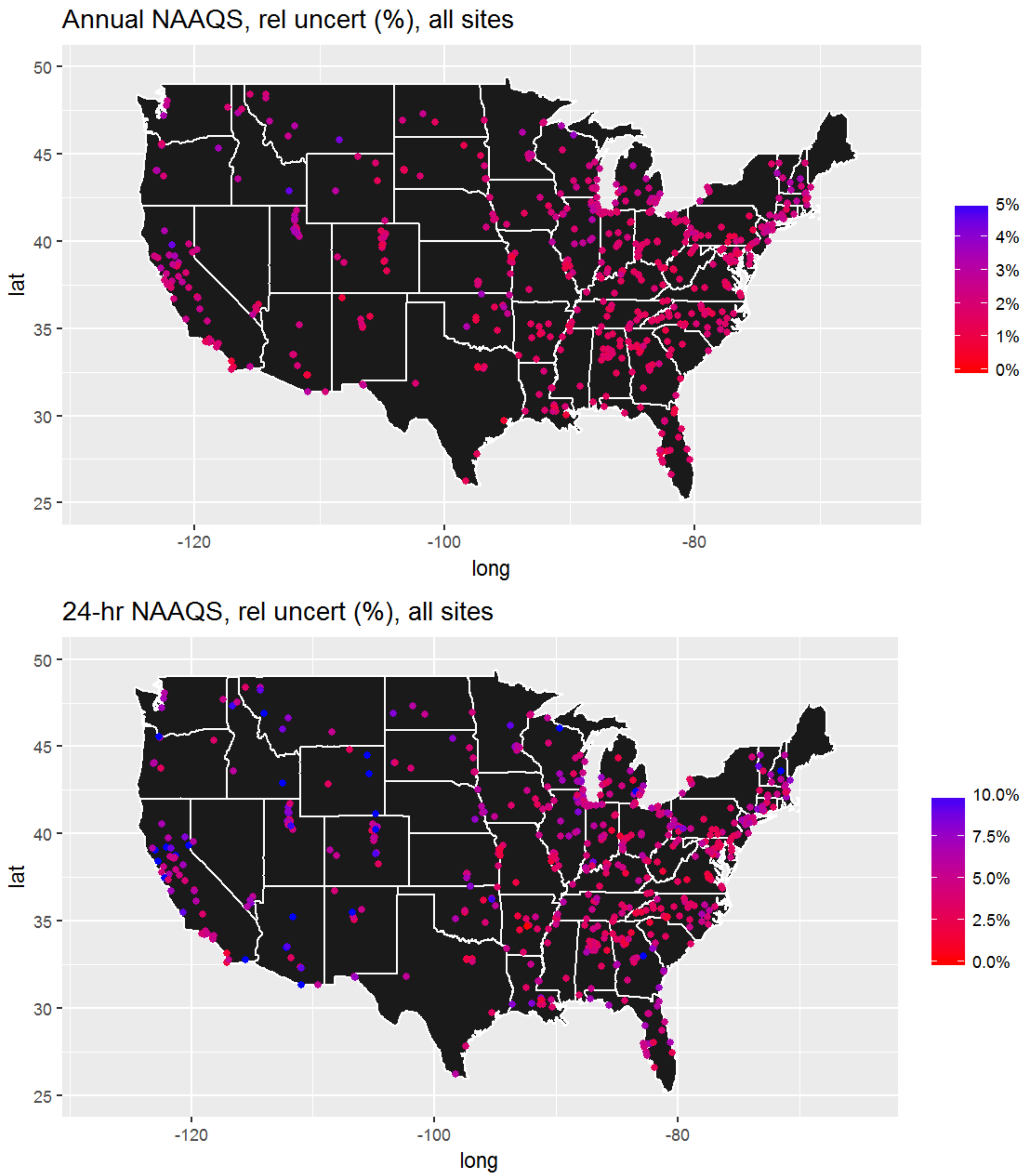


Figure 73: Spatial distribution of the relative difference between the CI and the actual DV from the 50% CIs for the 2006 $PM_{2.5}$ DVs.

3.11 2003-2005 $PM_{2.5}$ bootstrap results

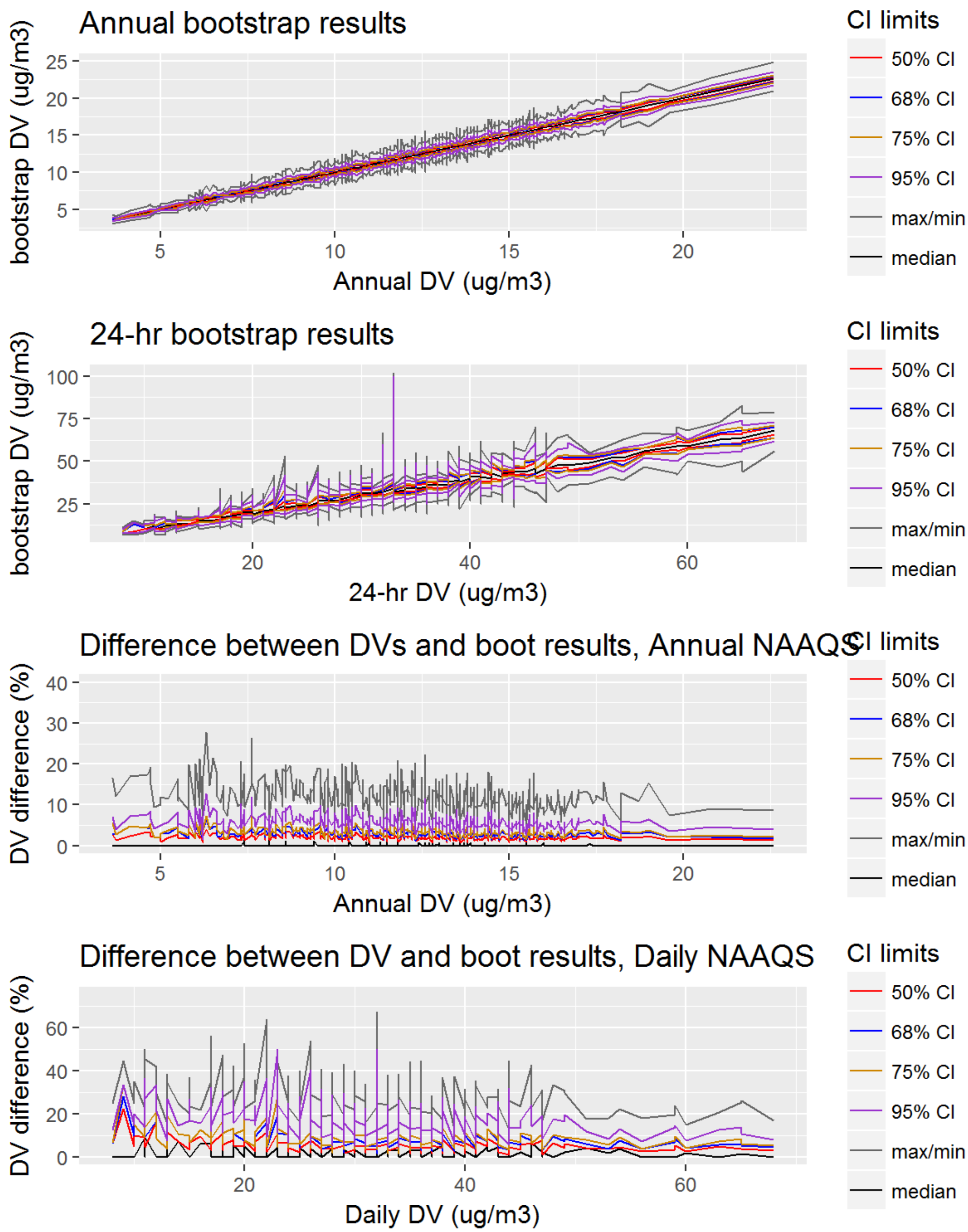


Figure 74: Bootstrap results for the 2005 $PM_{2.5}$ DVs, showing the 50%, 68%, 75%, and 95% CIs, along with the mean and median bootstrap DVs. The top two panels show the values for the DVs at the various CIs, while the bottom two panels show the relative difference between the CI and the actual DV.

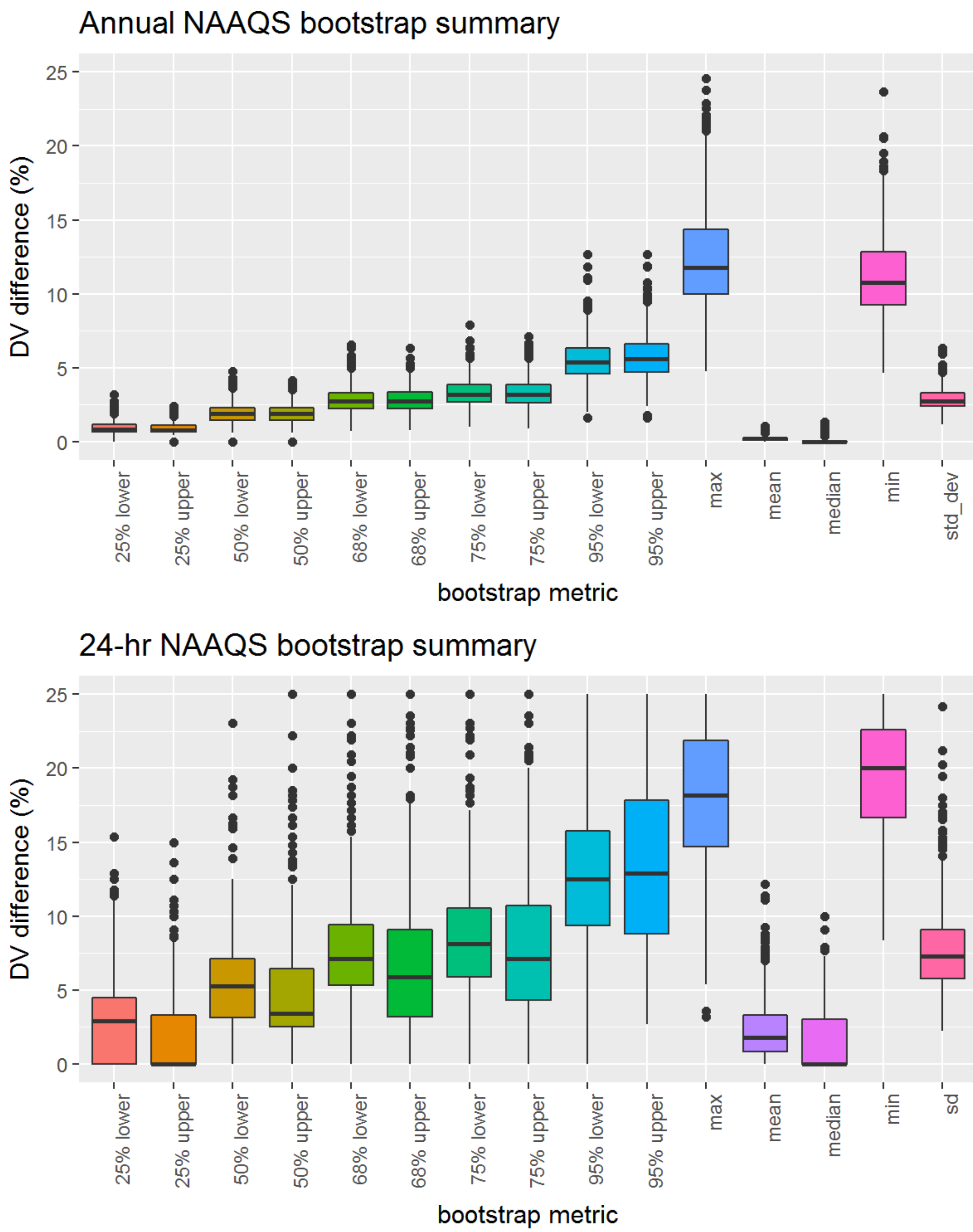


Figure 75: Bootstrap results for the 2005 $PM_{2.5}$ DVs, showing distribution of the relative differences between the bootstrap DVs and the actual DV at the 50%, 68%, 75%, and 95% CIs, along with the mean, median, maximum, minimum, standard deviations of the relative differences.

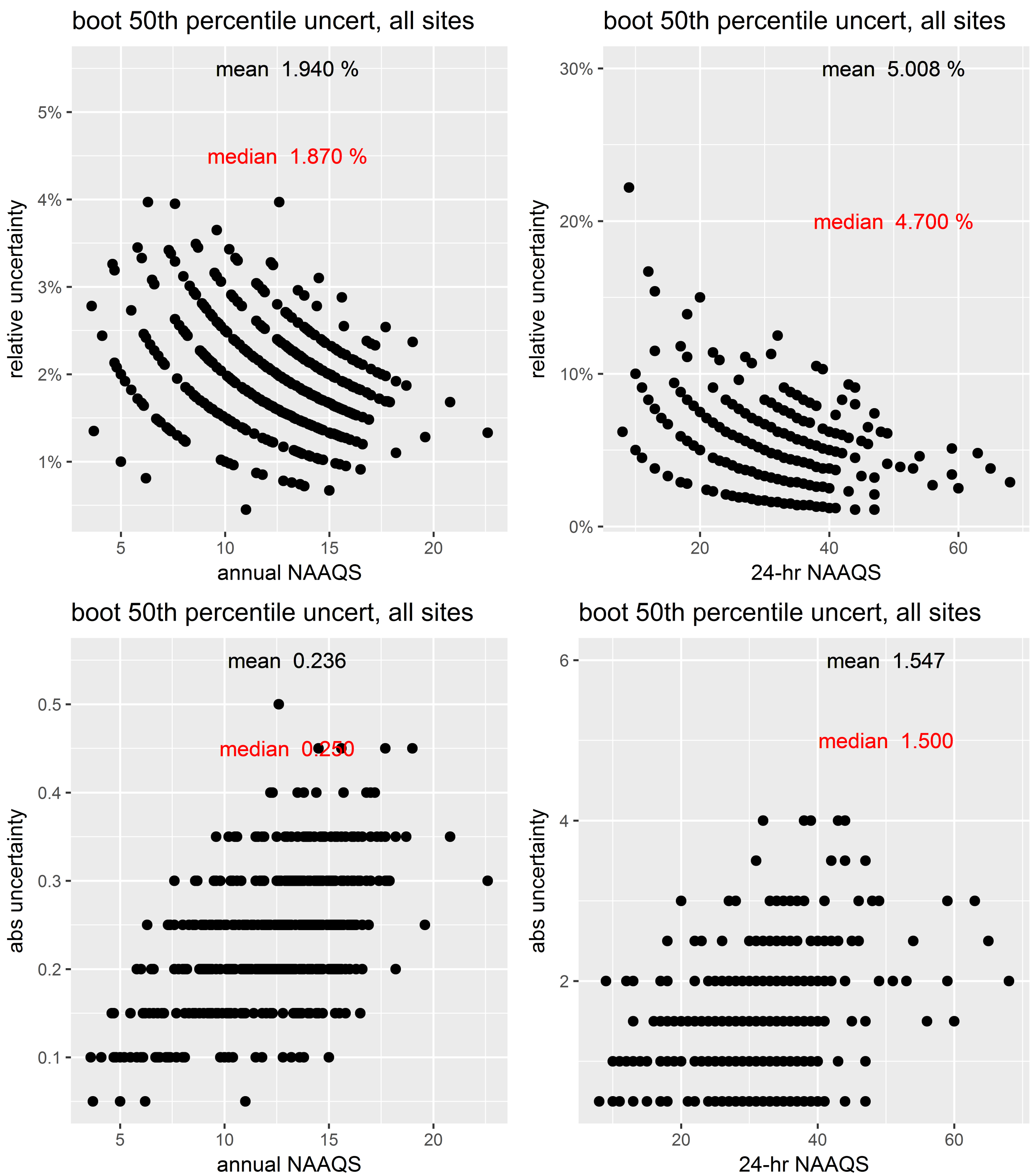


Figure 76: Bootstrap results from the 50% CIs for $PM_{2.5}$ DVs. The top two panels show the relative difference between the CI and the actual DV and the bottom two panels show the absolute difference between the values for the DVs at each site and the CI.

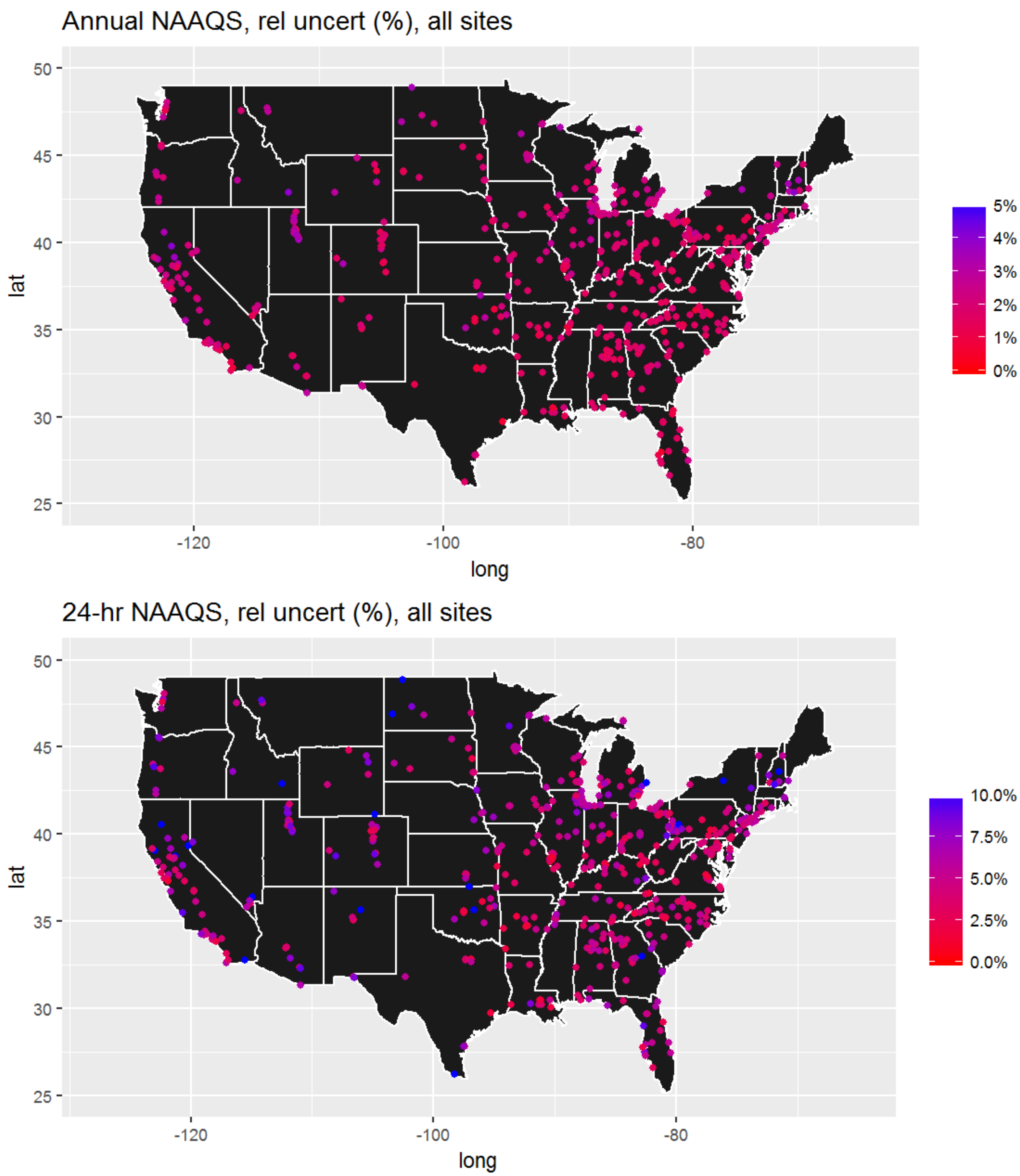


Figure 77: Spatial distribution of the relative difference between the CI and the actual DV from the 50% CIs for the 2005 $PM_{2.5}$ DVs.

3.12 2002-2004 $PM_{2.5}$ bootstrap results

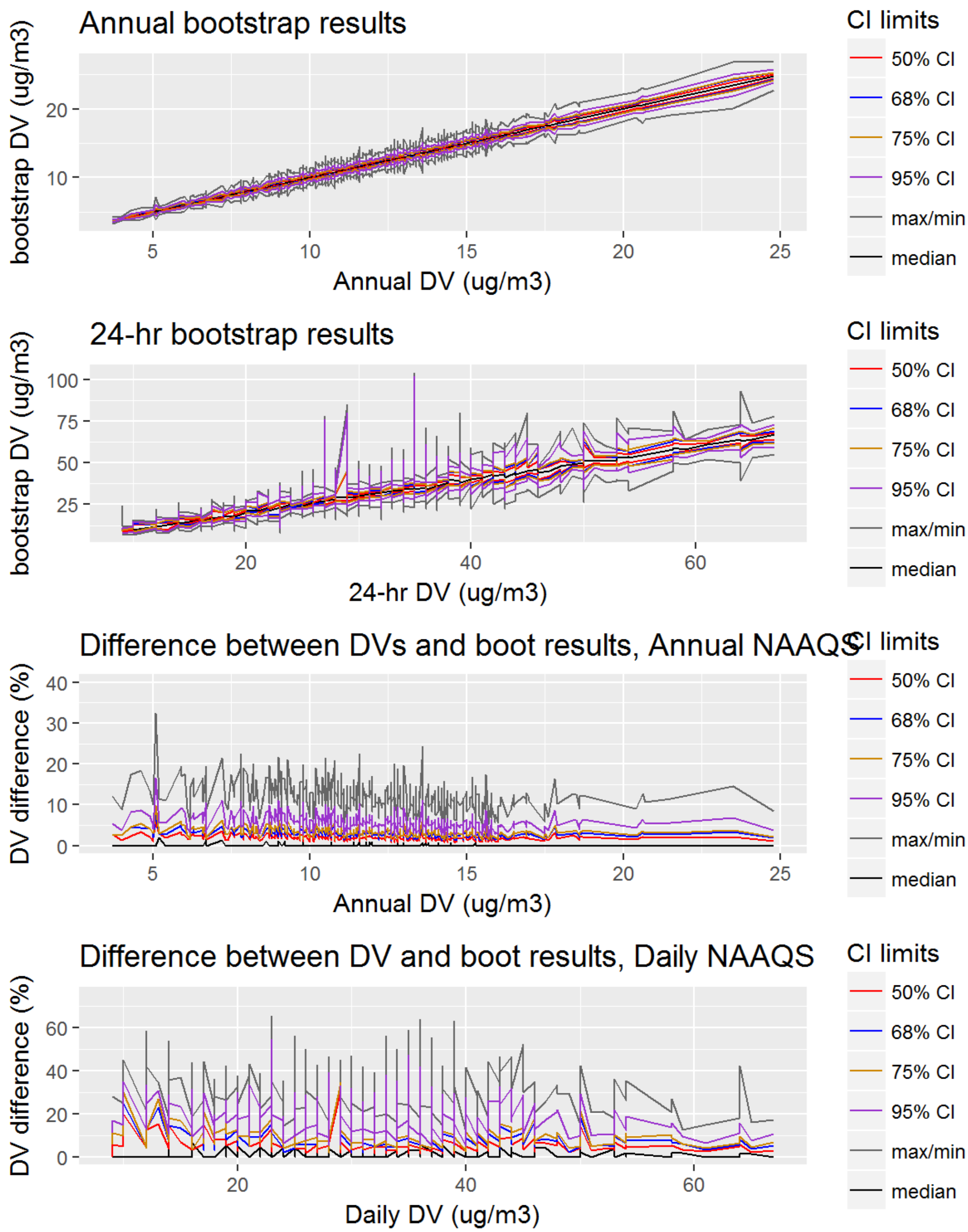


Figure 78: Bootstrap results for the 2004 $PM_{2.5}$ DVs, showing the 50%, 68%, 75%, and 95% CIs, along with the mean and median bootstrap DVs. The top two panels show the values for the DVs at the various CIs, while the bottom two panels show the relative difference between the CI and the actual DV.

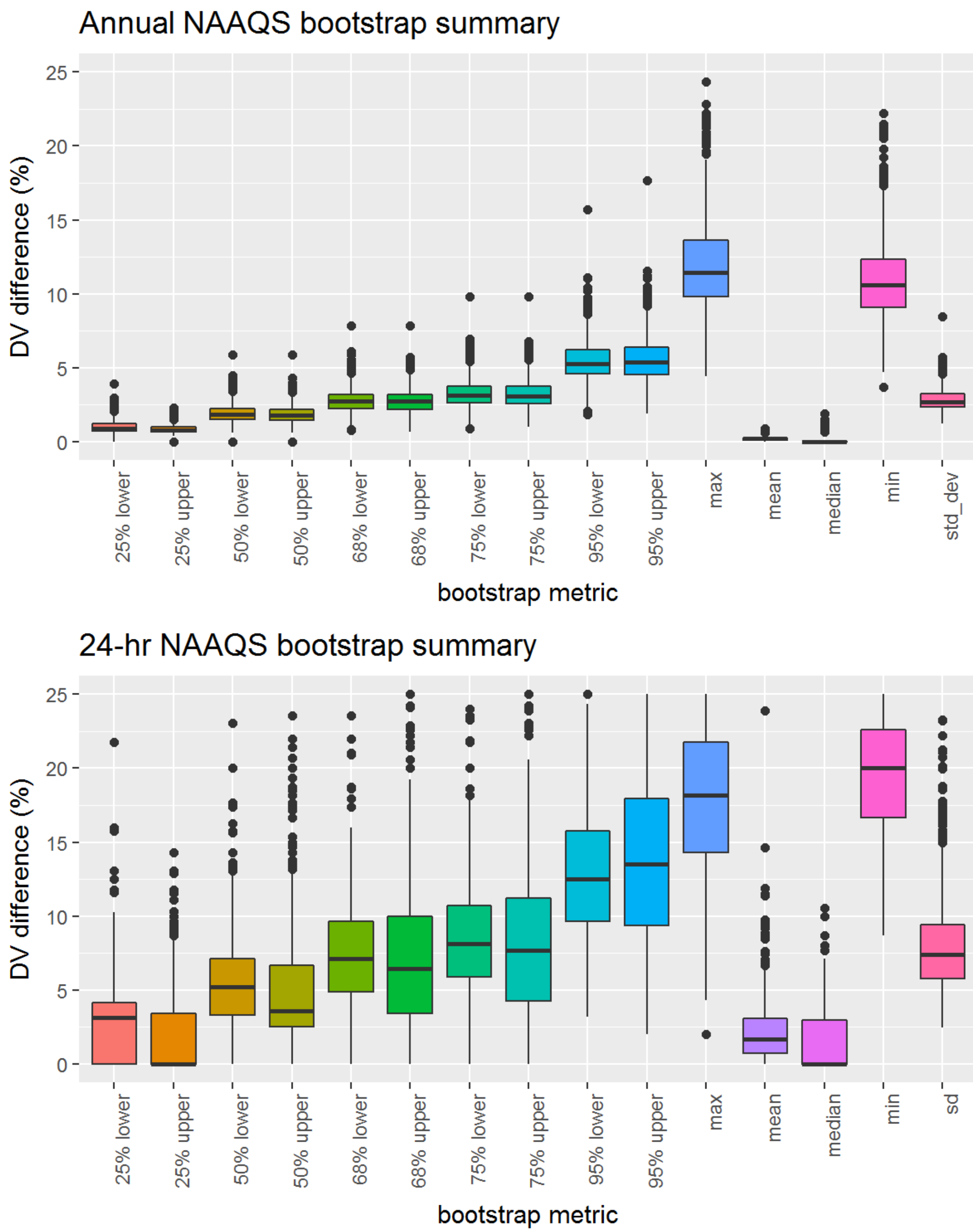


Figure 79: Bootstrap results for the 2004 $PM_{2.5}$ DVs, showing distribution of the relative differences between the bootstrap DVs and the actual DV at the 50%, 68%, 75%, and 95% CIs, along with the mean, median, maximum, minimum, standard deviations of the relative differences.

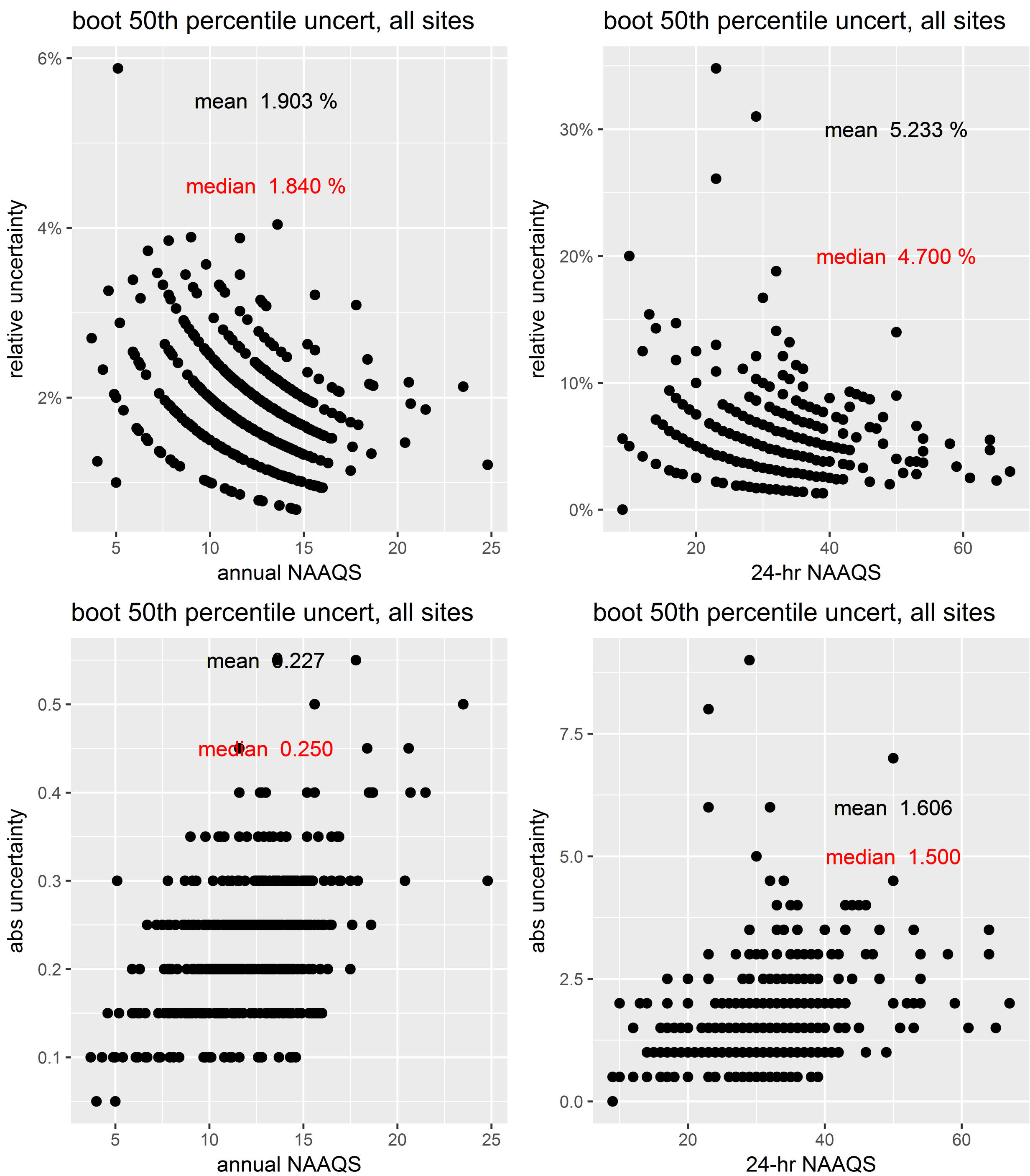


Figure 80: Bootstrap results from the 50% CIs for $PM_{2.5}$ DVs. The top two panels show the relative difference between the CI and the actual DV and the bottom two panels show the absolute difference between the values for the DVs at each site and the CI.

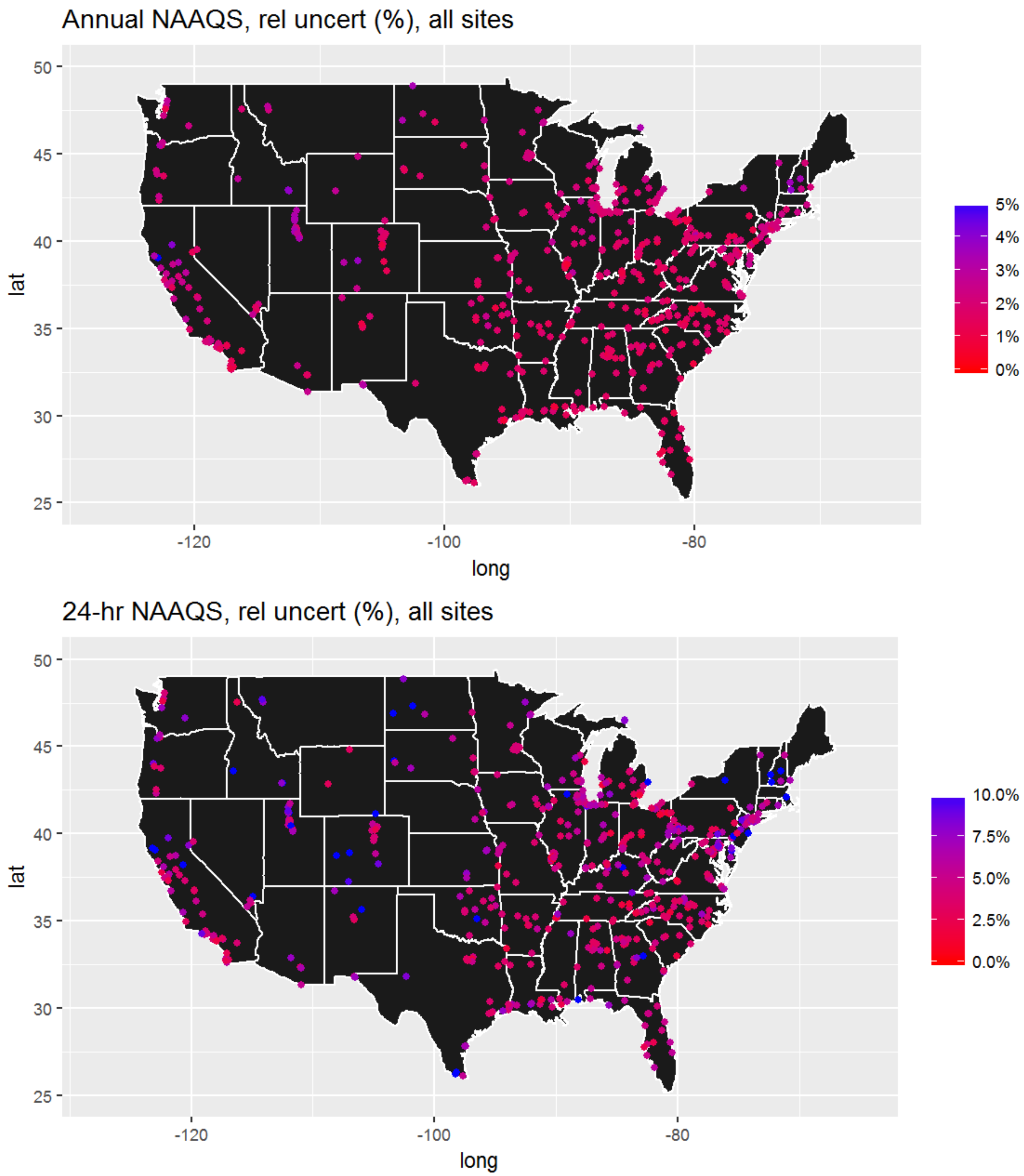


Figure 81: Spatial distribution of the relative difference between the CI and the actual DV from the 50% CIs for the 2004 $PM_{2.5}$ DVs.

3.13 2001-2003 $PM_{2.5}$ bootstrap results

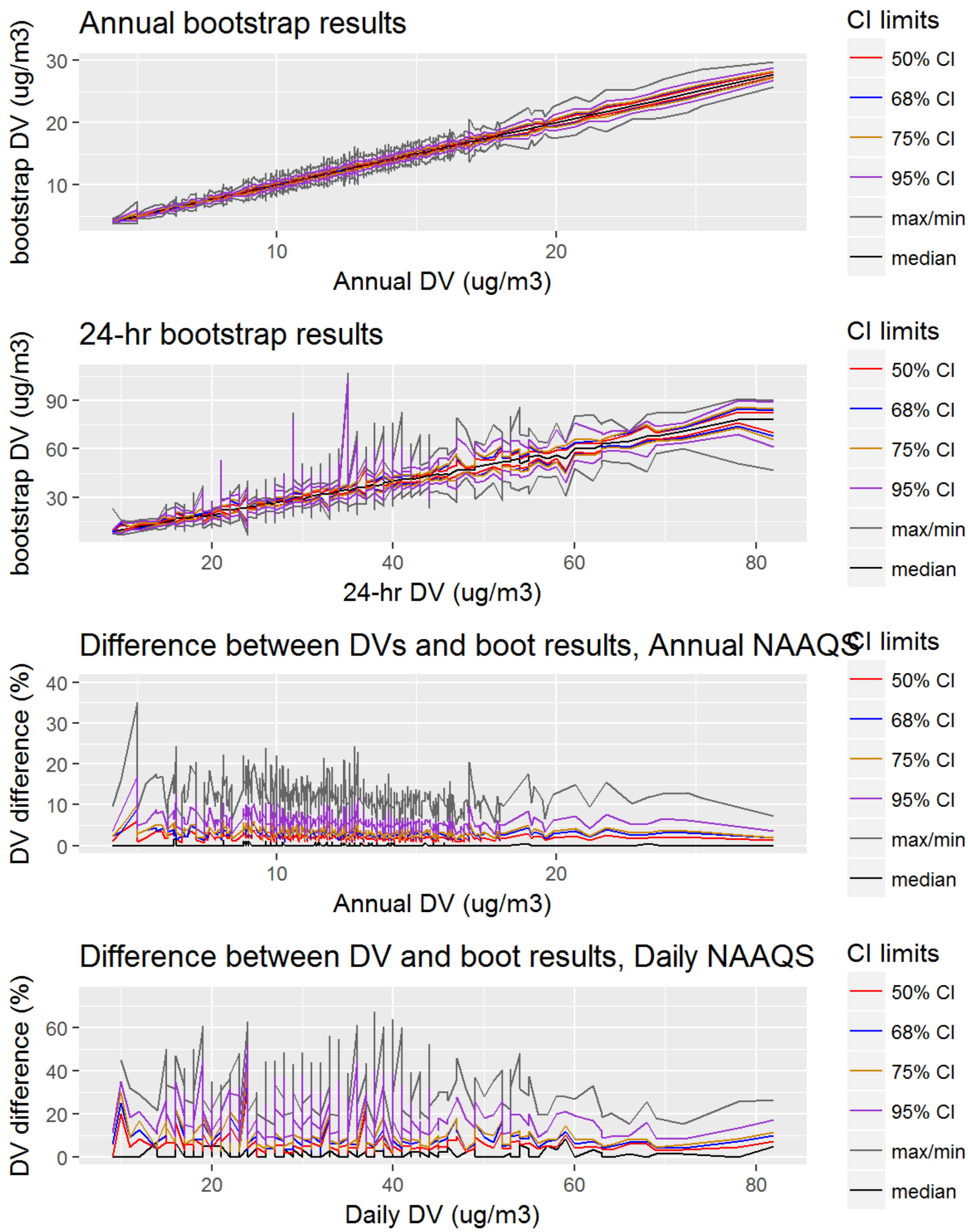


Figure 82: Bootstrap results for the 2003 $PM_{2.5}$ DVs, showing the 50%, 68%, 75%, and 95% CIs, along with the mean and median bootstrap DVs. The top two panels show the values for the DVs at the various CIs, while the bottom two panels show the relative difference between the CI and the actual DV.

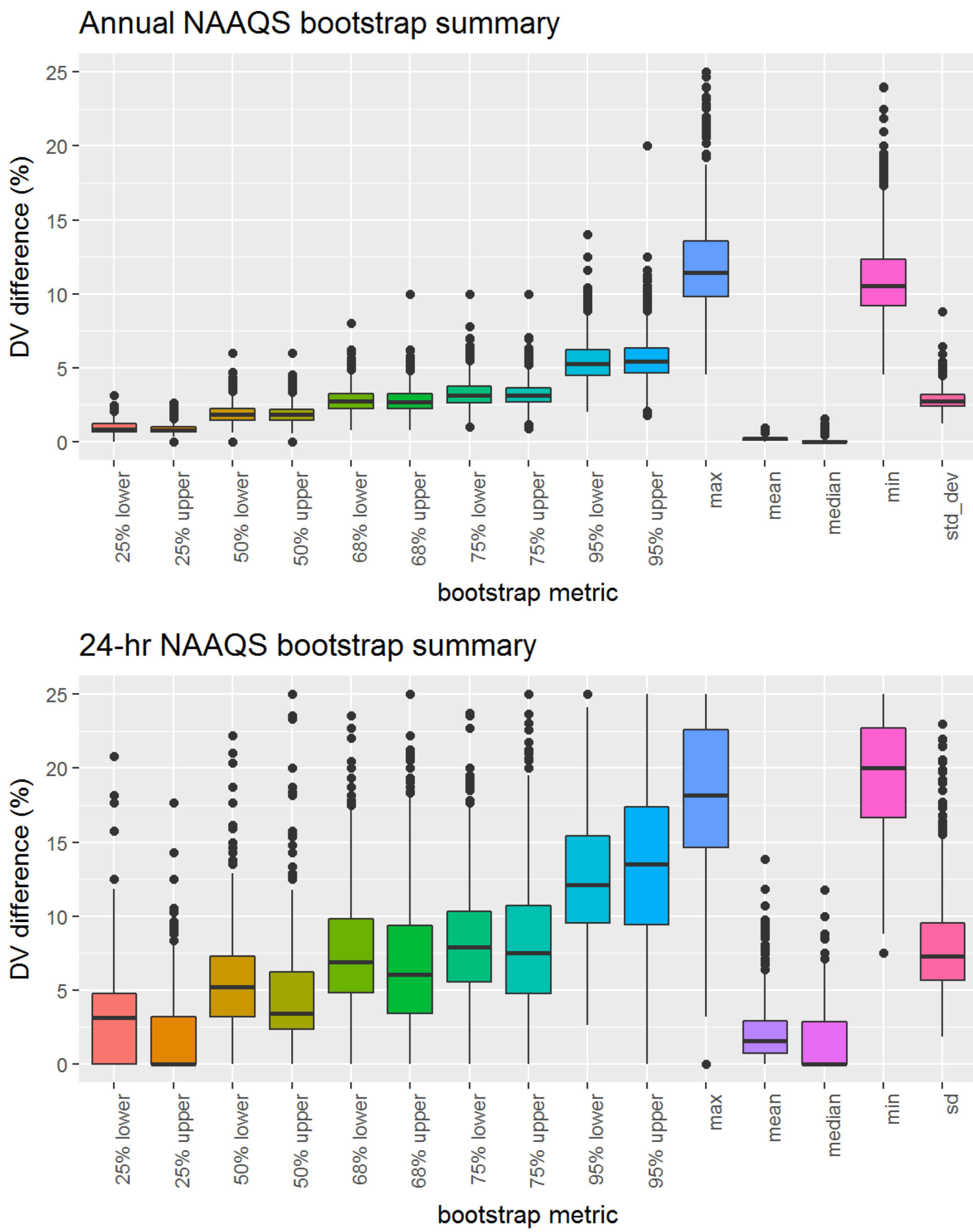


Figure 83: Bootstrap results for the 2003 $PM_{2.5}$ DVs, showing distribution of the relative differences between the bootstrap DVs and the actual DV at the 50%, 68%, 75%, and 95% CIs, along with the mean, median, maximum, minimum, standard deviations of the relative differences.

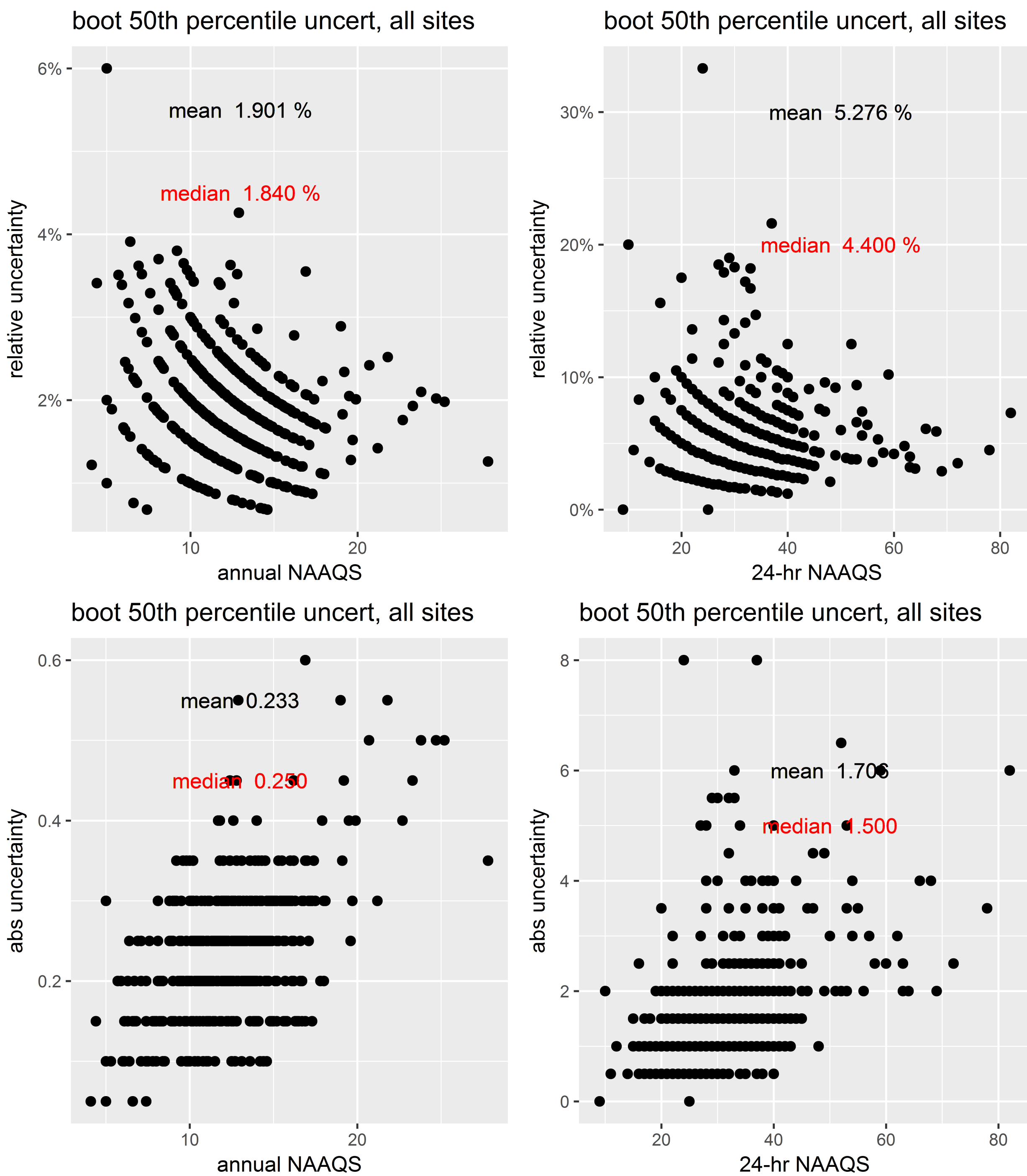


Figure 84: Bootstrap results from the 50% CIs for $PM_{2.5}$ DVs. The top two panels show the relative difference between the CI and the actual DV and the bottom two panels show the absolute difference between the values for the DVs at each site and the CI.

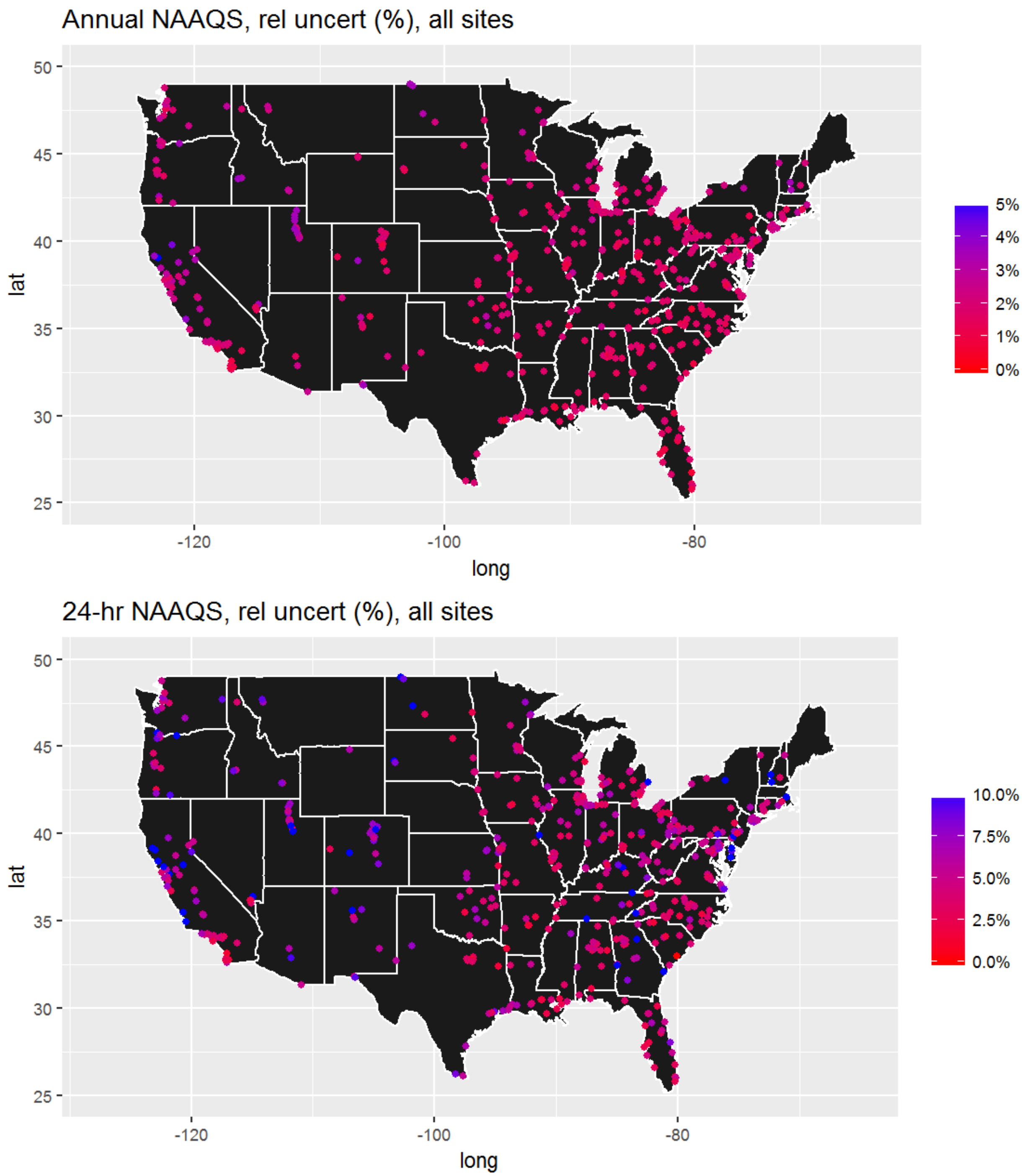


Figure 85: Spatial distribution of the relative difference between the CI and the actual DV from the 50% CIs for the 2003 $PM_{2.5}$ DVs.

3.14 2000-2002 $PM_{2.5}$ bootstrap results

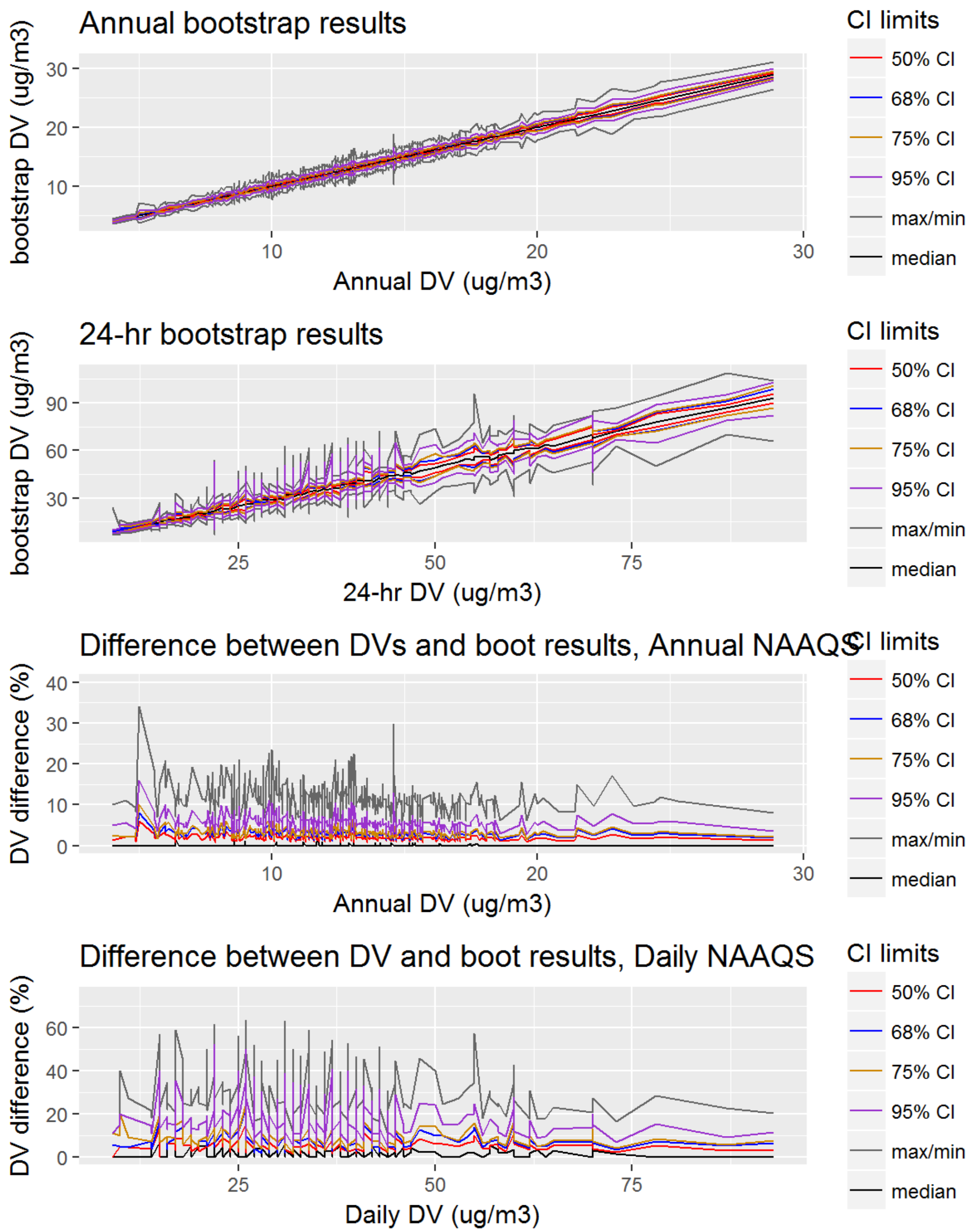


Figure 86: Bootstrap results for the 2002 $PM_{2.5}$ DVs, showing the 50%, 68%, 75%, and 95% CIs, along with the mean and median bootstrap DVs. The top two panels show the values for the DVs at the various CIs, while the bottom two panels show the relative difference between the CI and the actual DV.

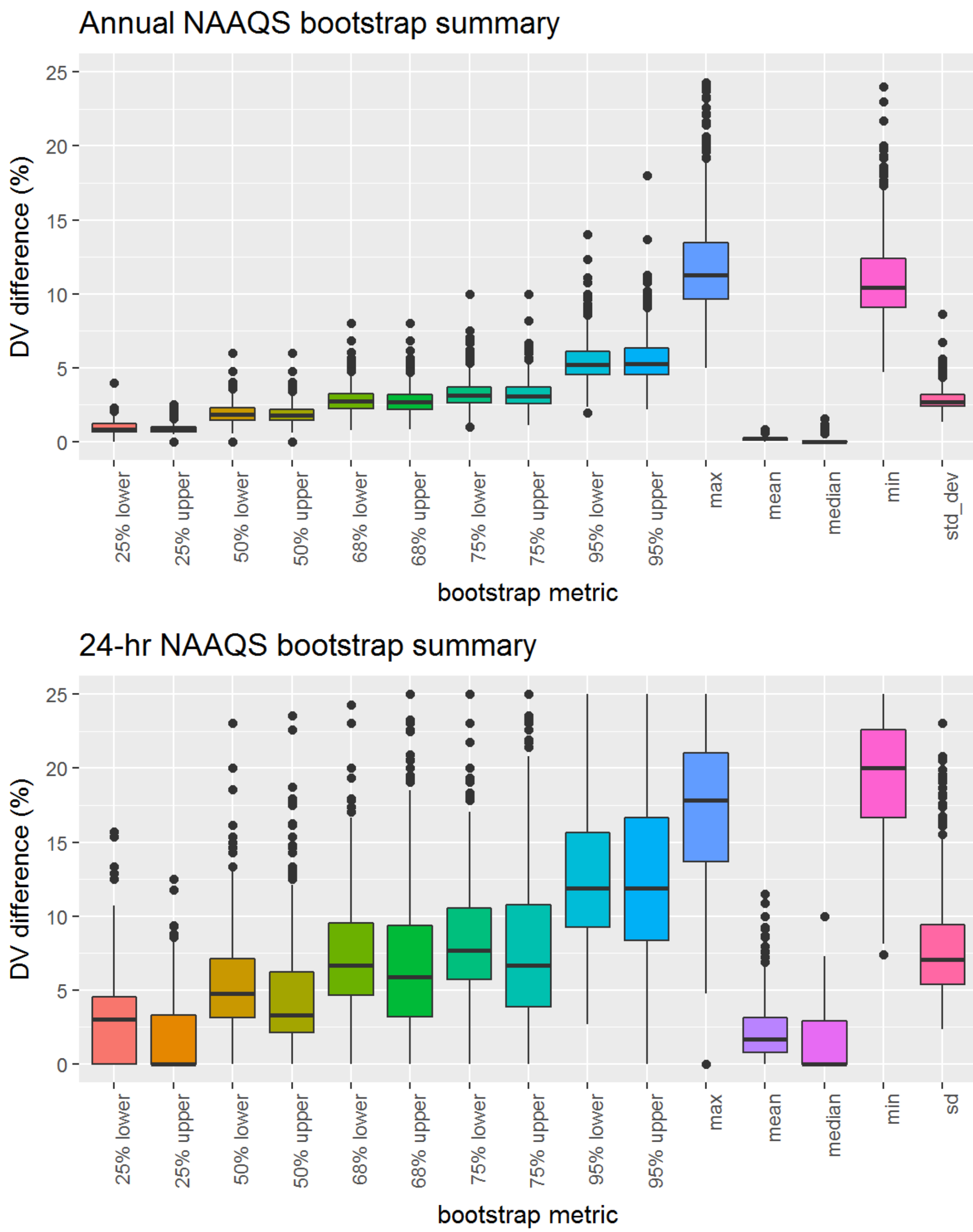


Figure 87: Bootstrap results for the 2002 $PM_{2.5}$ DVs, showing distribution of the relative differences between the bootstrap DVs and the actual DV at the 50%, 68%, 75%, and 95% CIs, along with the mean, median, maximum, minimum, standard deviations of the relative differences.

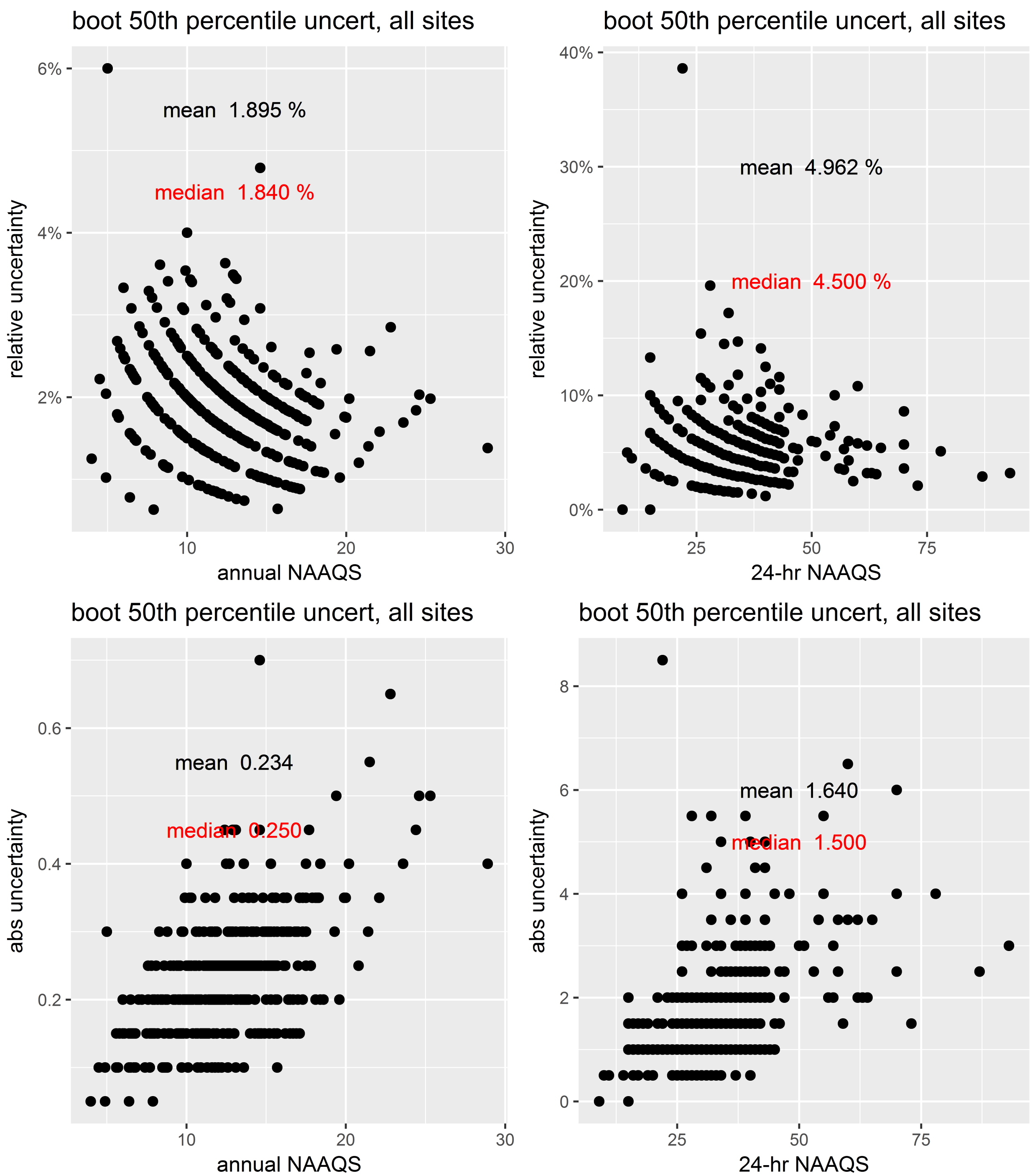


Figure 88: Bootstrap results from the 50% CIs for $PM_{2.5}$ DVs. The top two panels show the relative difference between the CI and the actual DV and the bottom two panels show the absolute difference between the values for the DVs at each site and the CI.

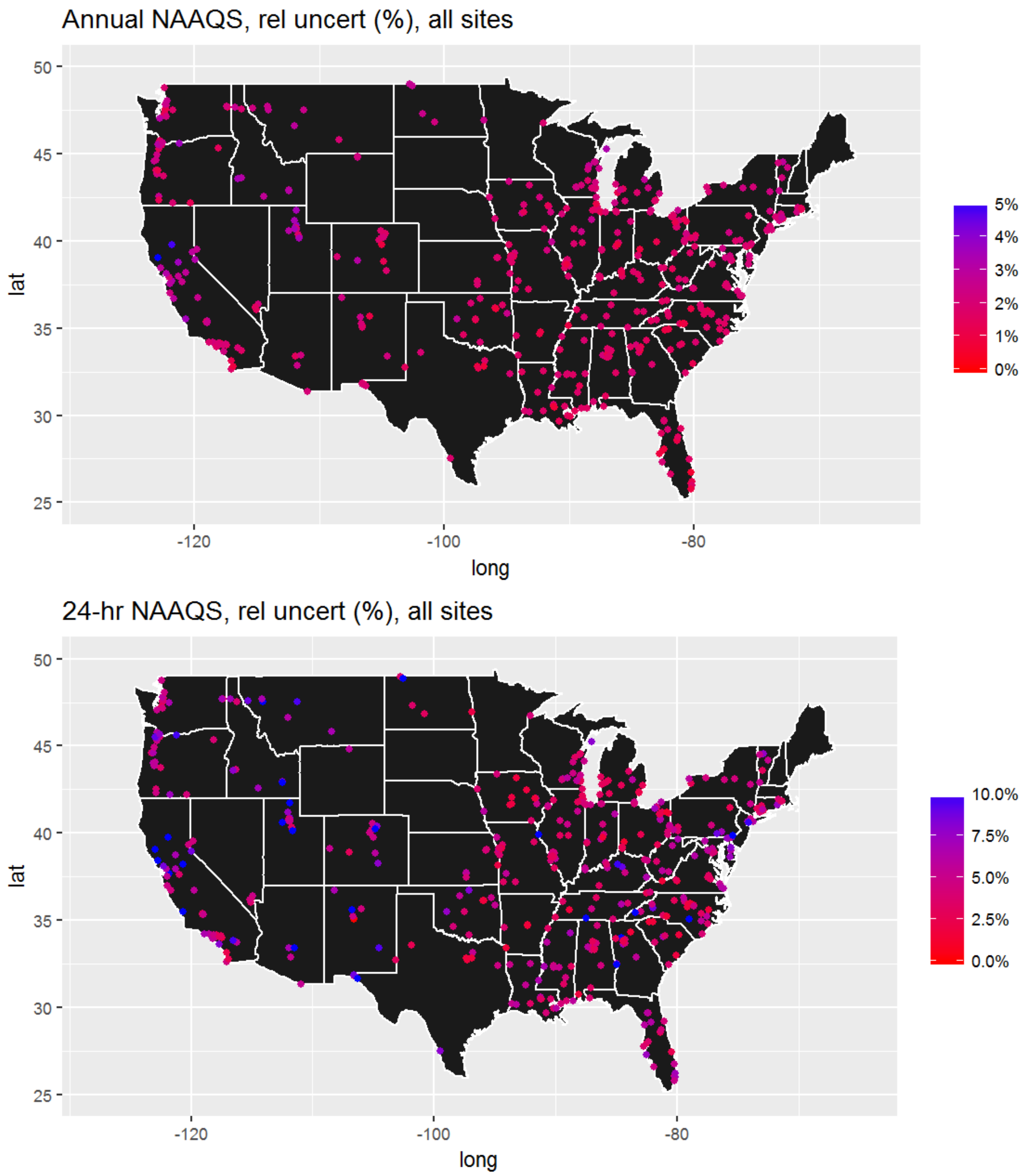


Figure 89: Spatial distribution of the relative difference between the CI and the actual DV from the 50% CIs for the 2002 $PM_{2.5}$ DVs.

4 Comparison plots of nearby sites

Comparison of $PM_{2.5}$ data for paired, nearby sites for the spatial analysis conducted in Section 3.1.2.

Figure 90: Comparison of $PM_{2.5}$ data for sites 150031001 and 150031001. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

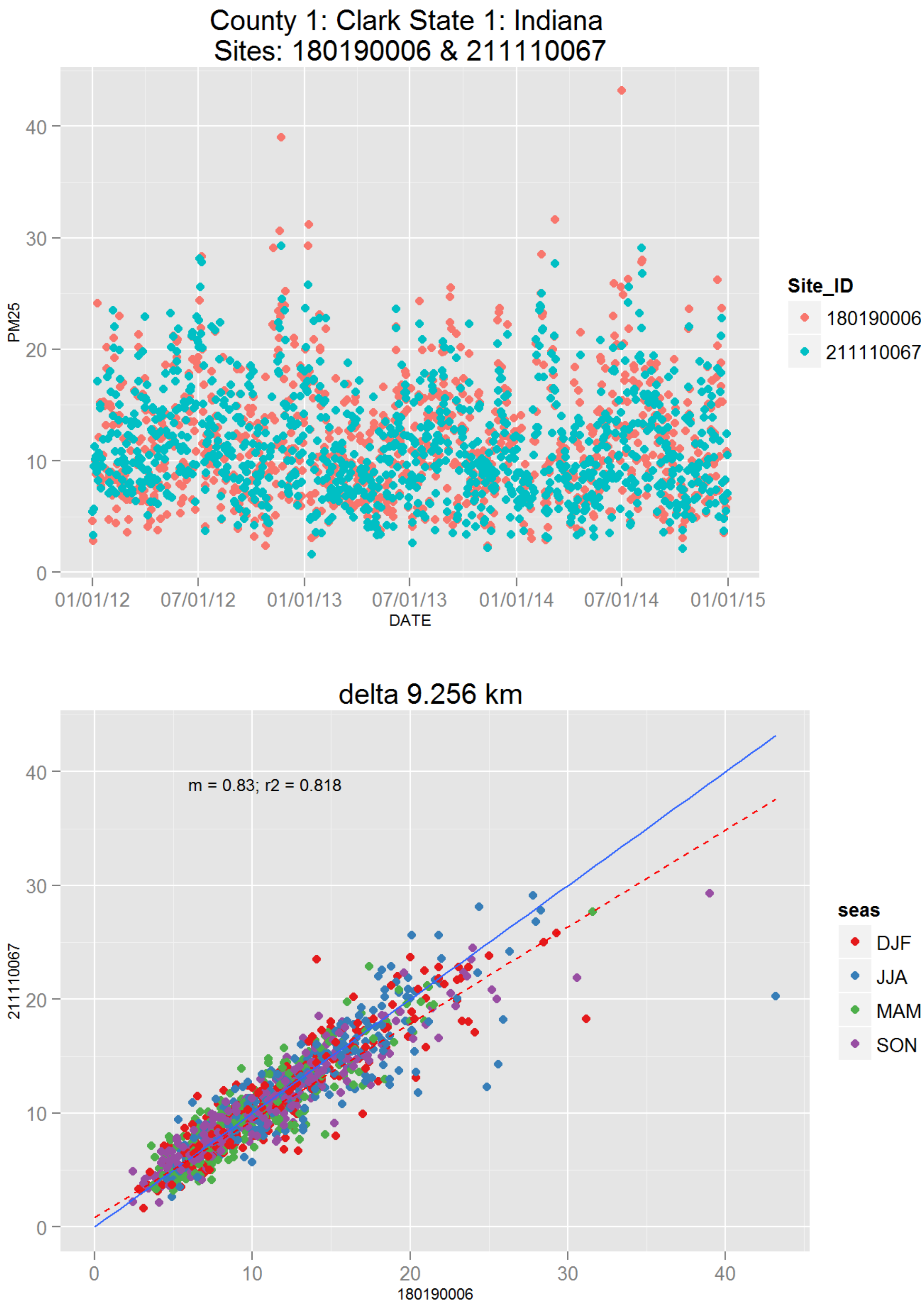


Figure 91: Comparison of $PM_{2.5}$ data for sites 180190006 and 180190006. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

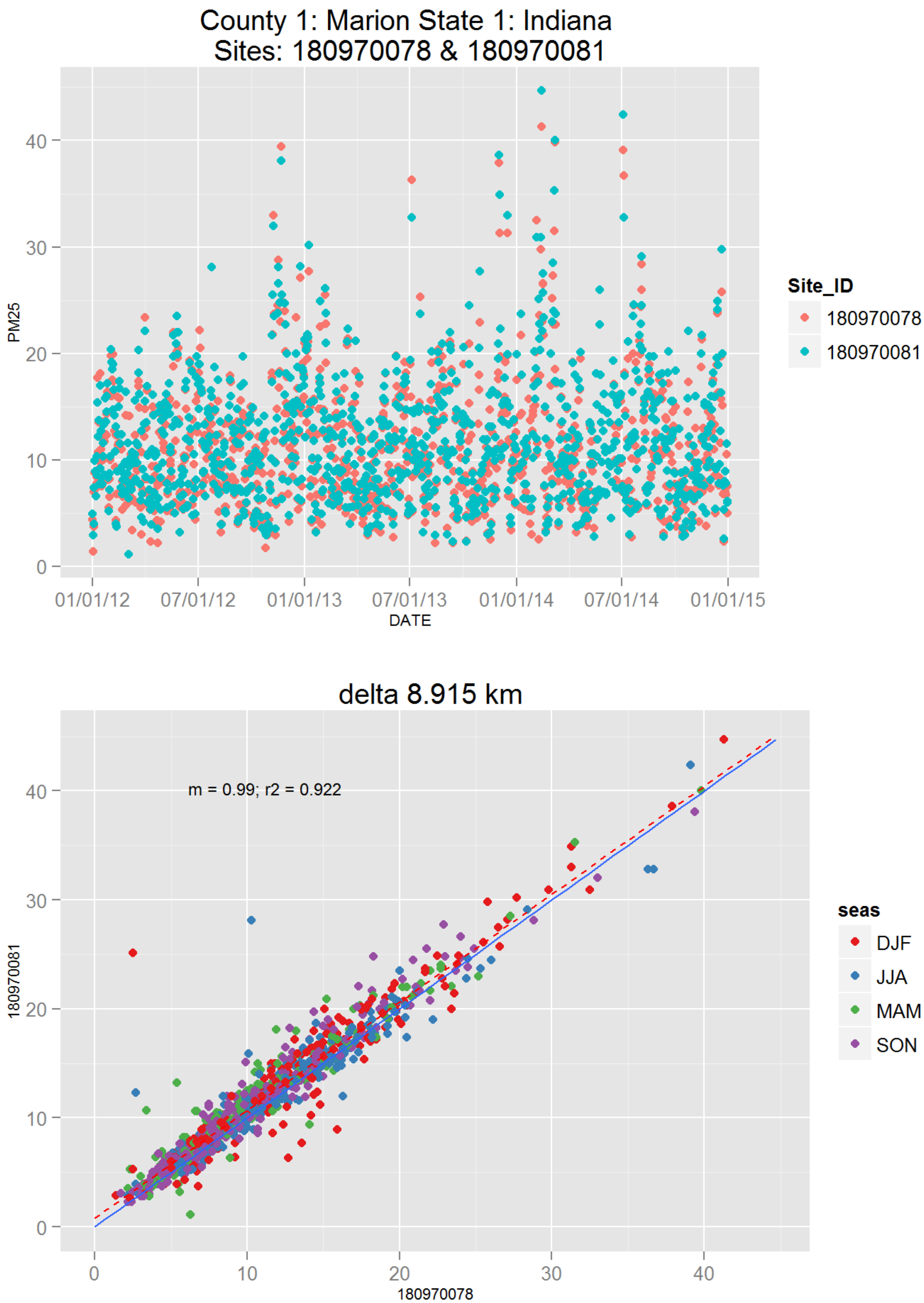
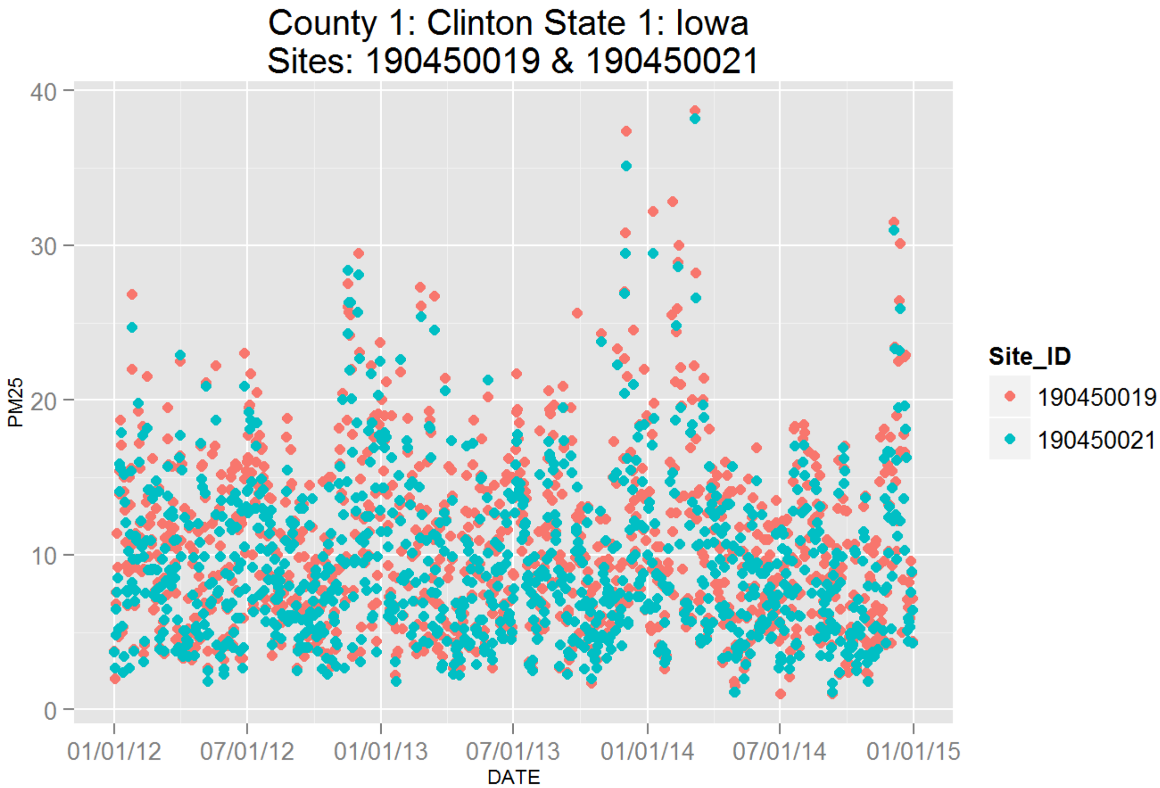


Figure 92: Comparison of $PM_{2.5}$ data for sites 180970078 and 180970078. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.



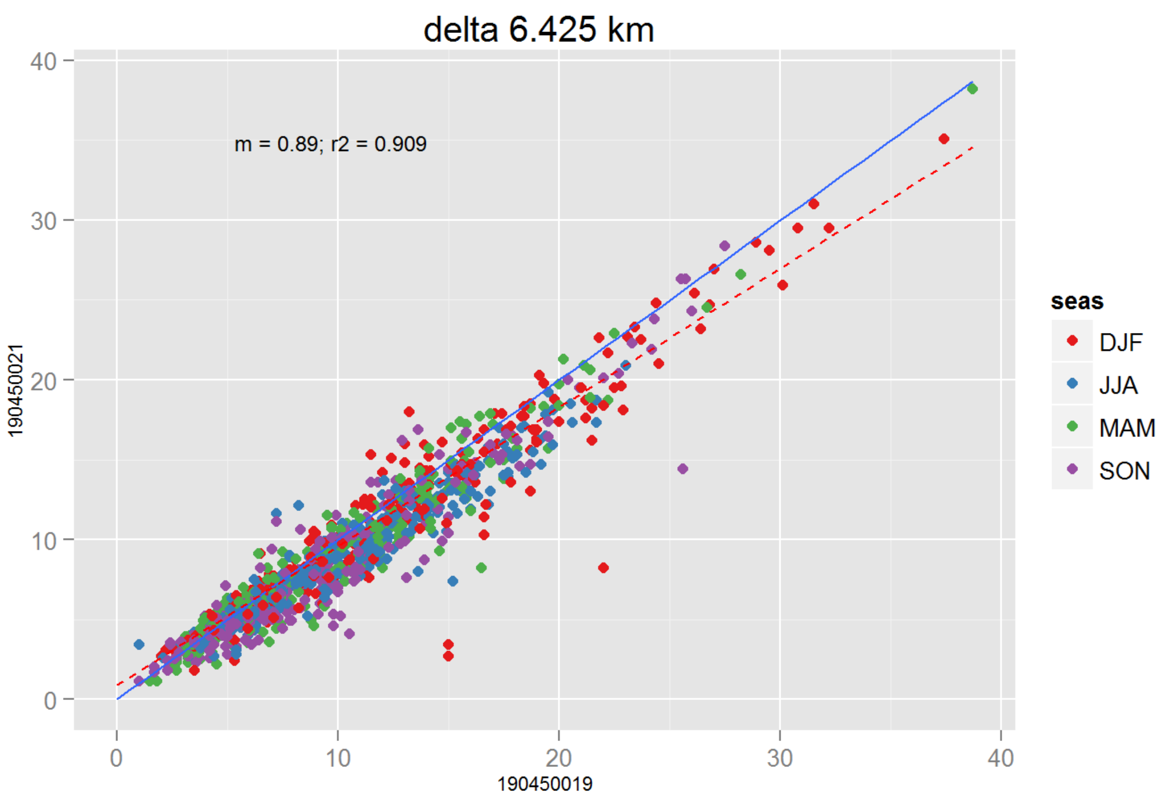


Figure 93: Comparison of $PM_{2.5}$ data for sites 190450019 and 190450019. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

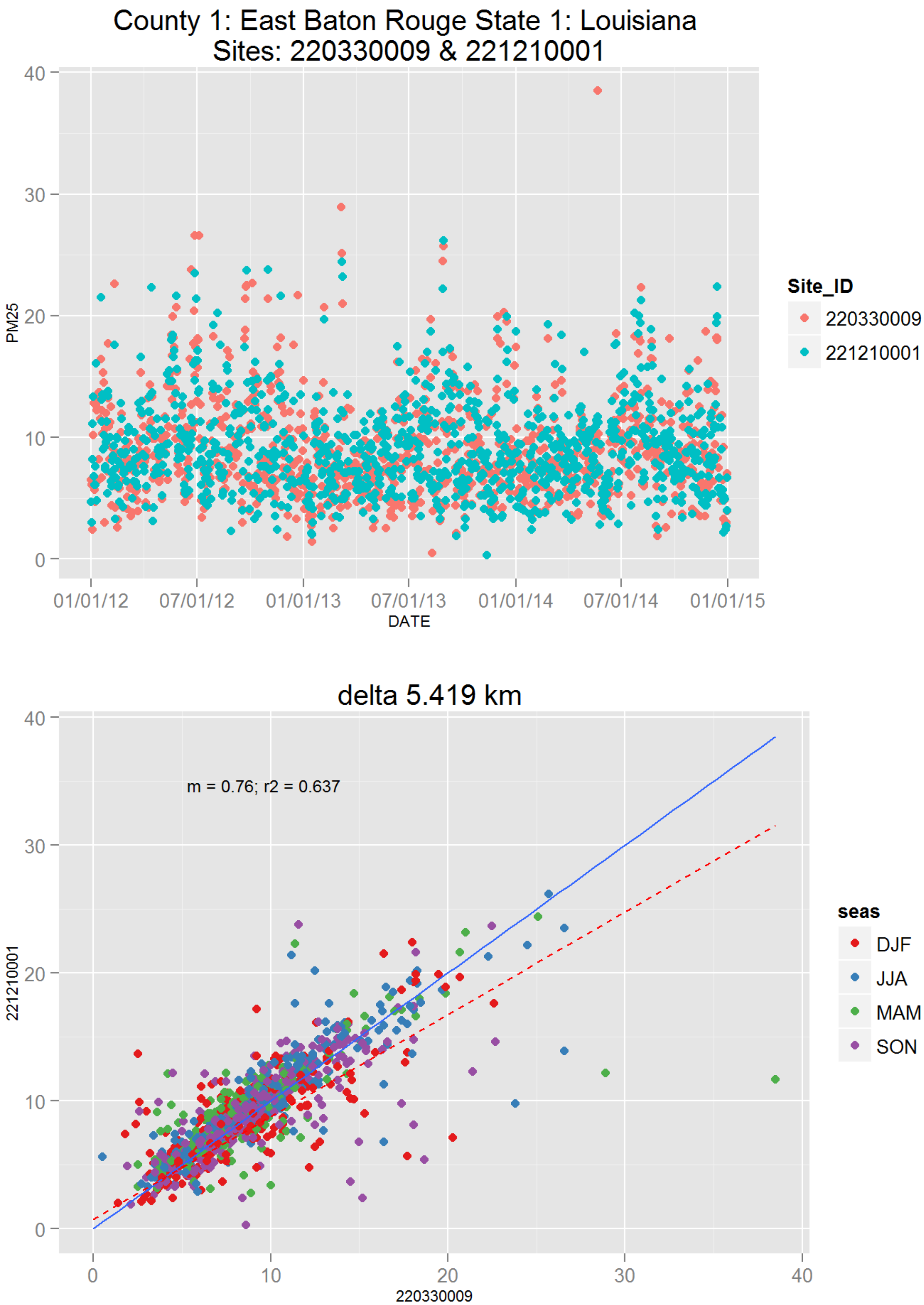


Figure 94: Comparison of $PM_{2.5}$ data for sites 220330009 and 220330009. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

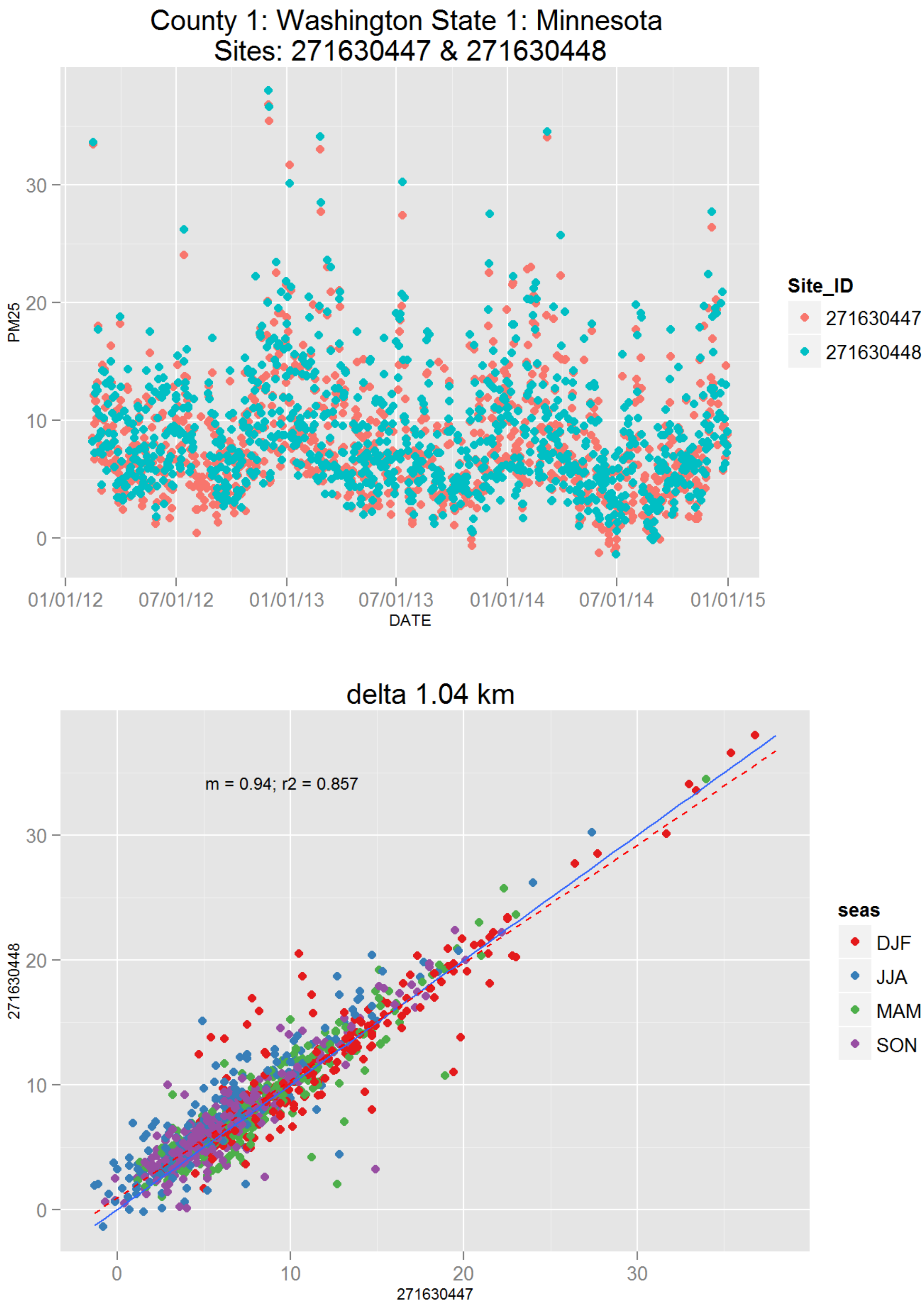


Figure 95: Comparison of $PM_{2.5}$ data for sites 271630447 and 271630447. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

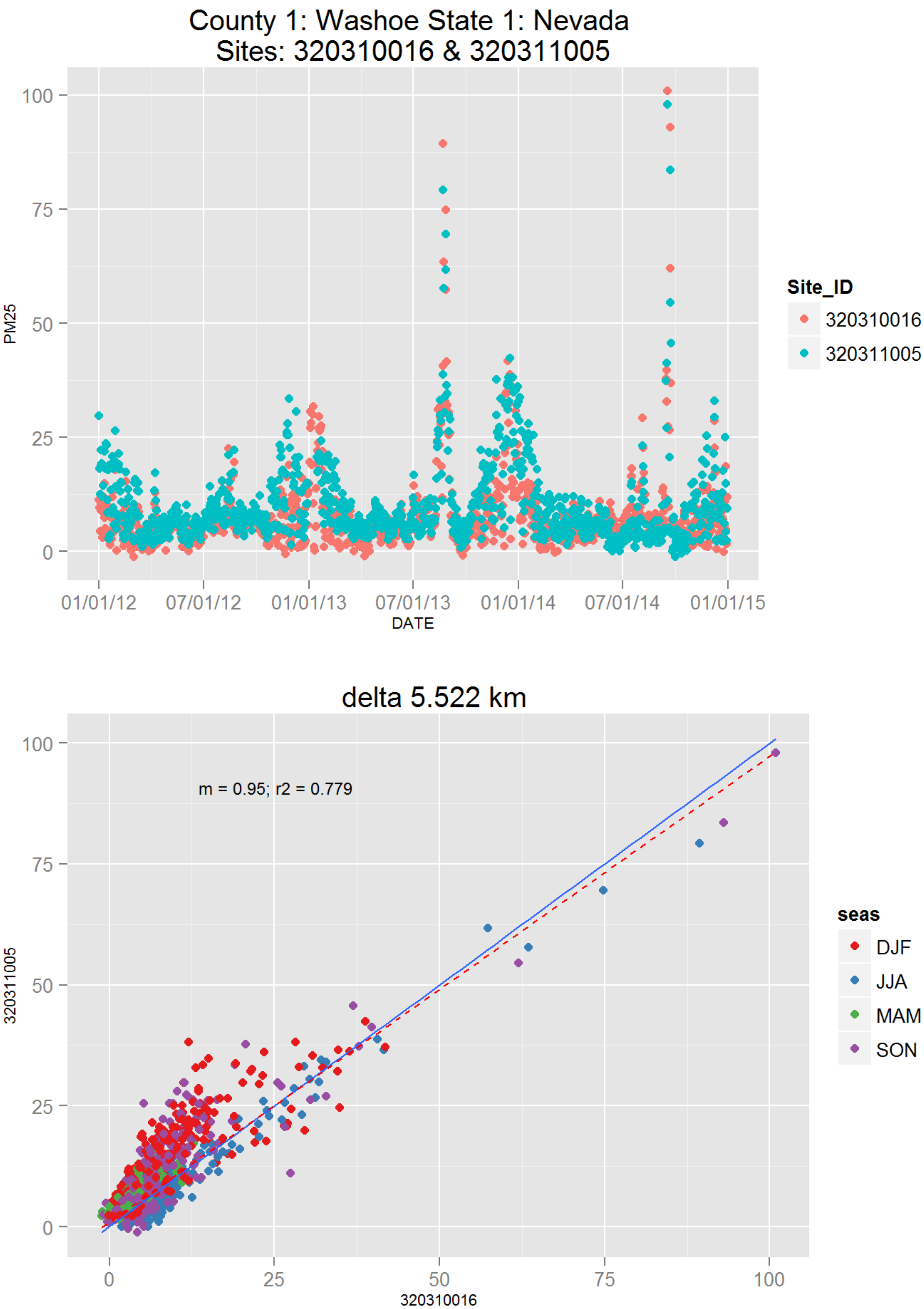


Figure 96: Comparison of $PM_{2.5}$ data for sites 320310016 and 320310016. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

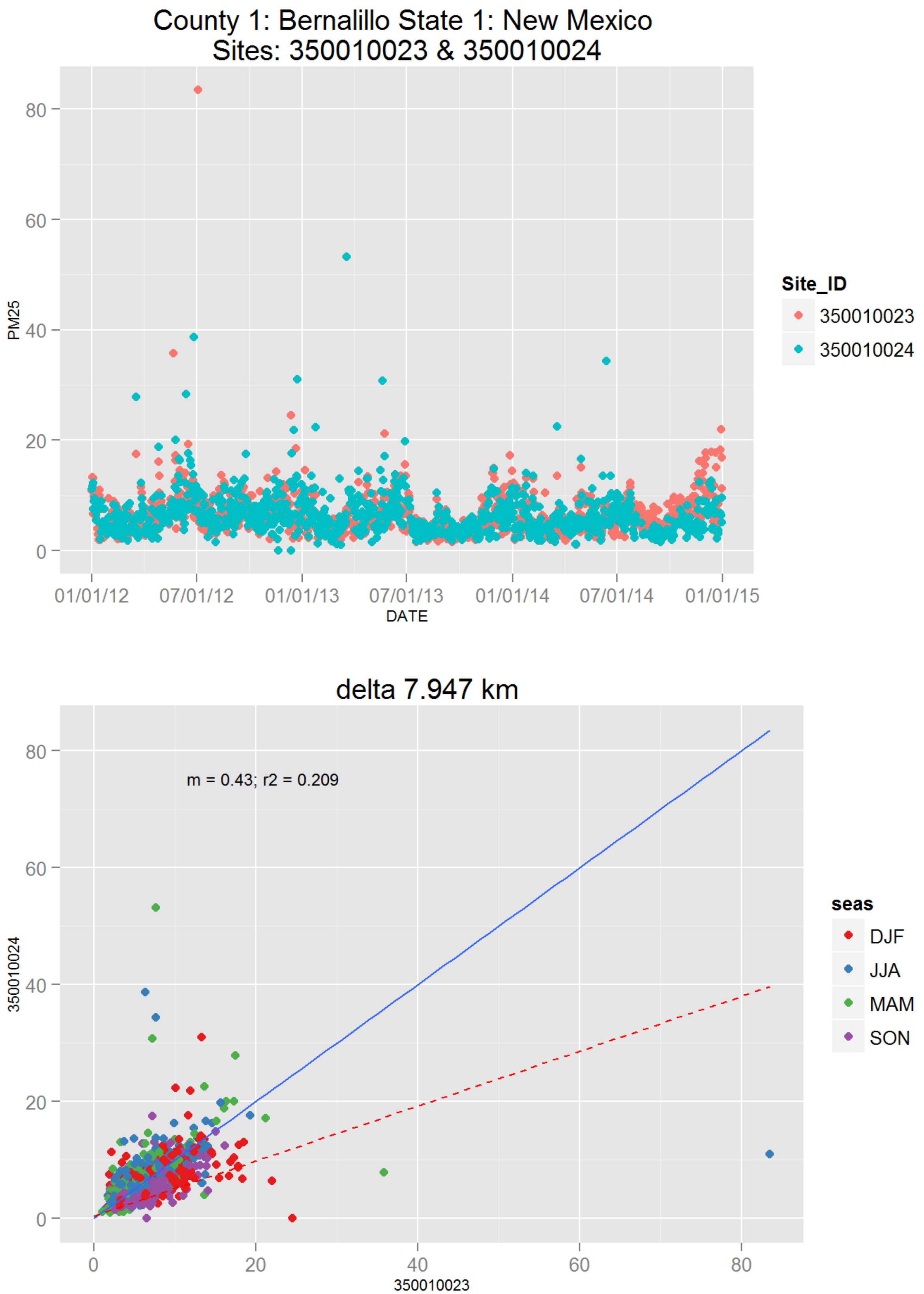


Figure 97: Comparison of $PM_{2.5}$ data for sites 350010023 and 350010023. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

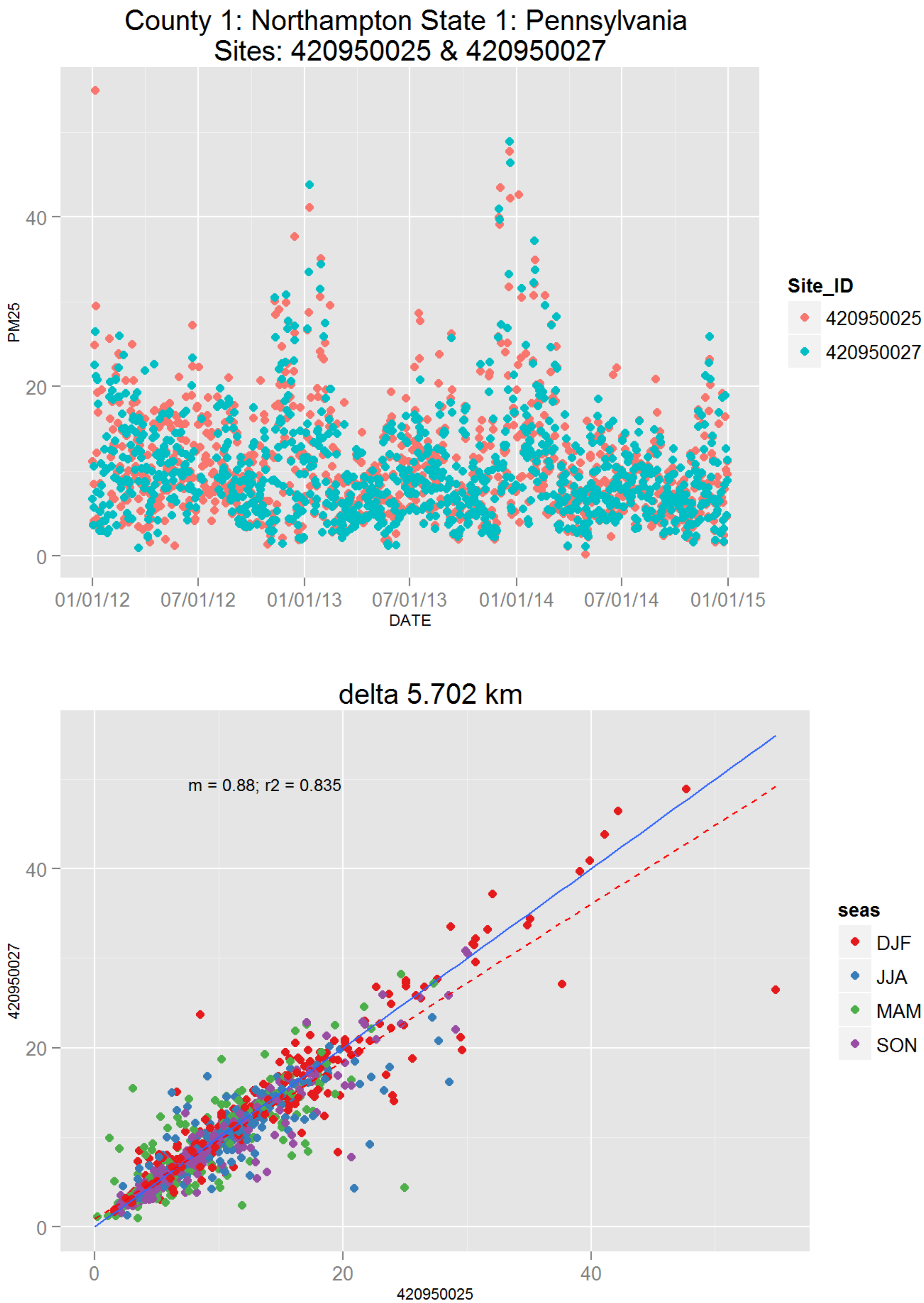


Figure 98: Comparison of $PM_{2.5}$ data for sites 420950025 and 420950025. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

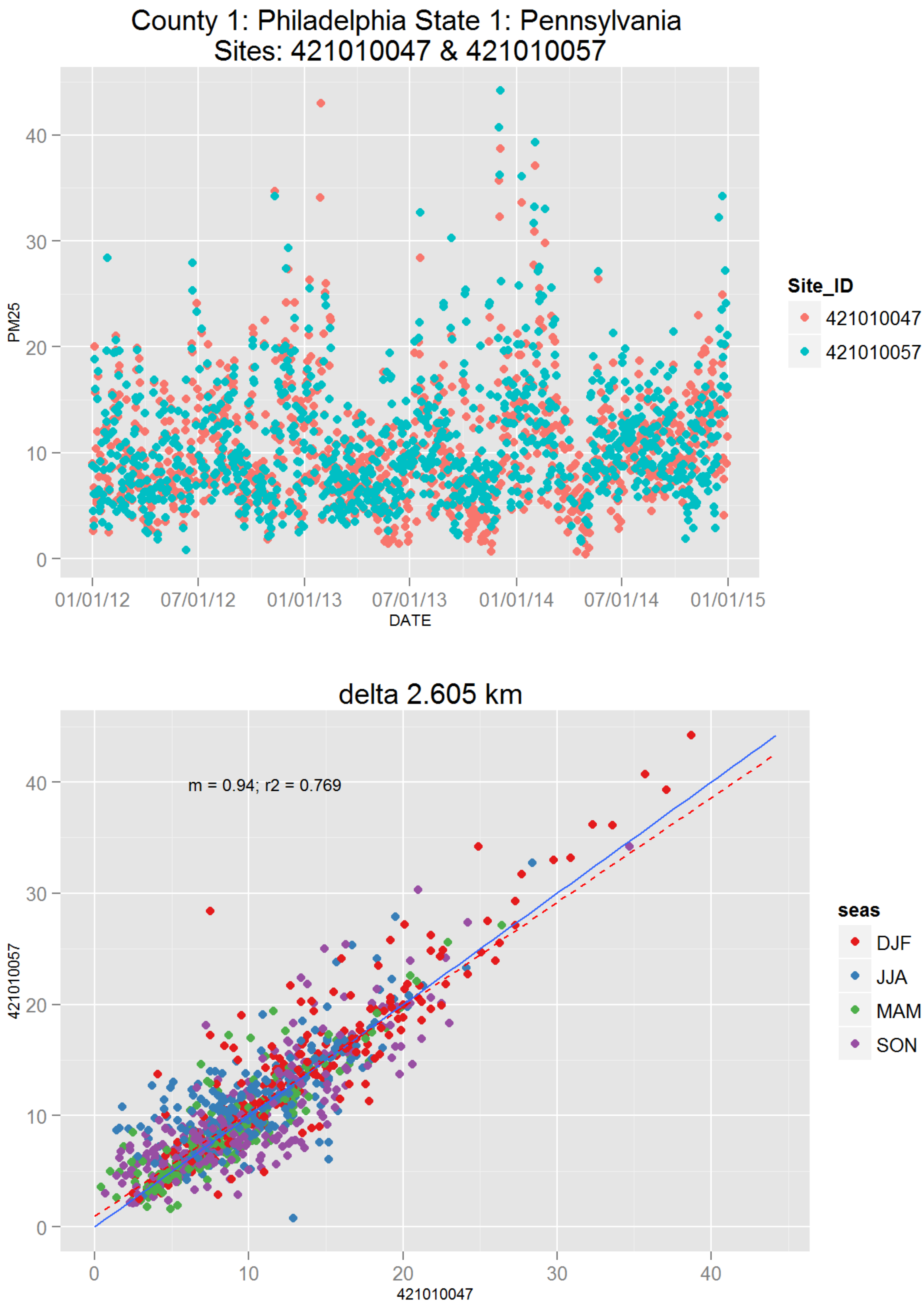


Figure 99: Comparison of $PM_{2.5}$ data for sites 421010047 and 421010047. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

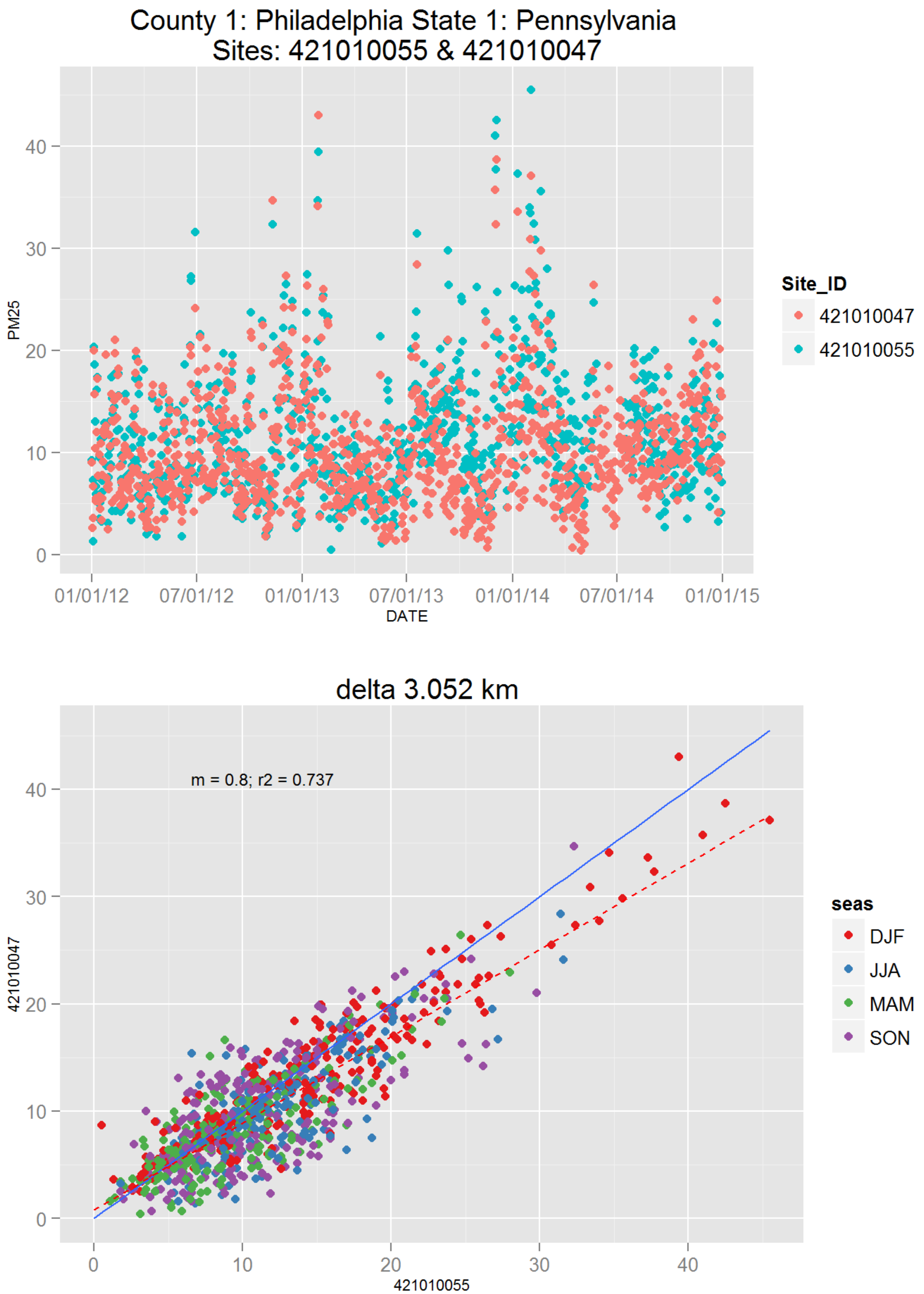
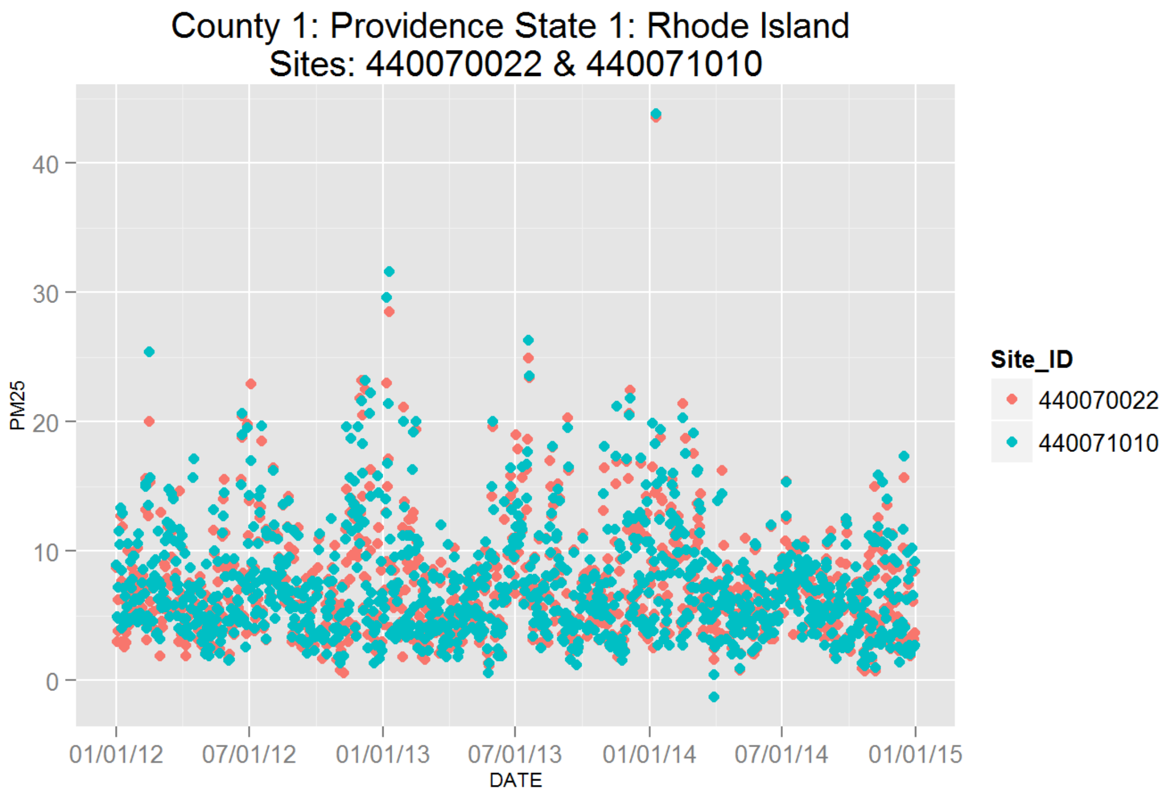


Figure 100: Comparison of $PM_{2.5}$ data for sites 421010055 and 421010055. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.



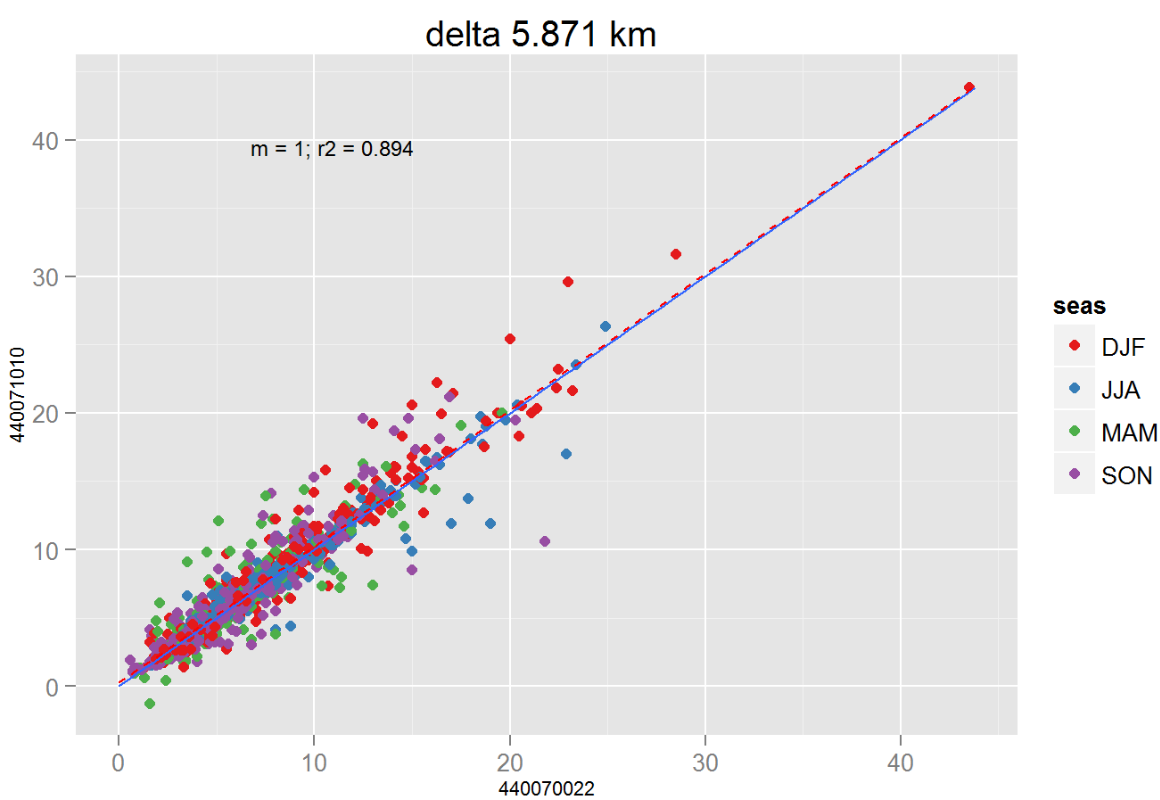
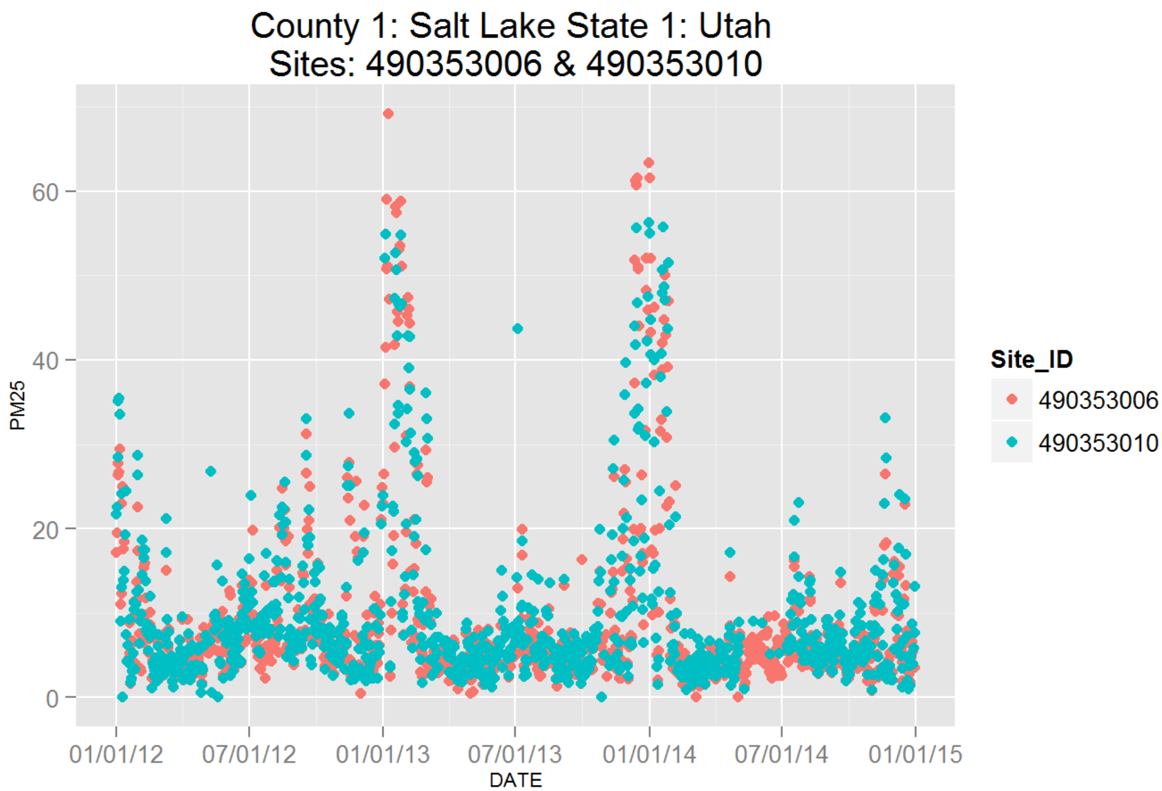


Figure 101: Comparison of $PM_{2.5}$ data for sites 440070022 and 440070022. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.



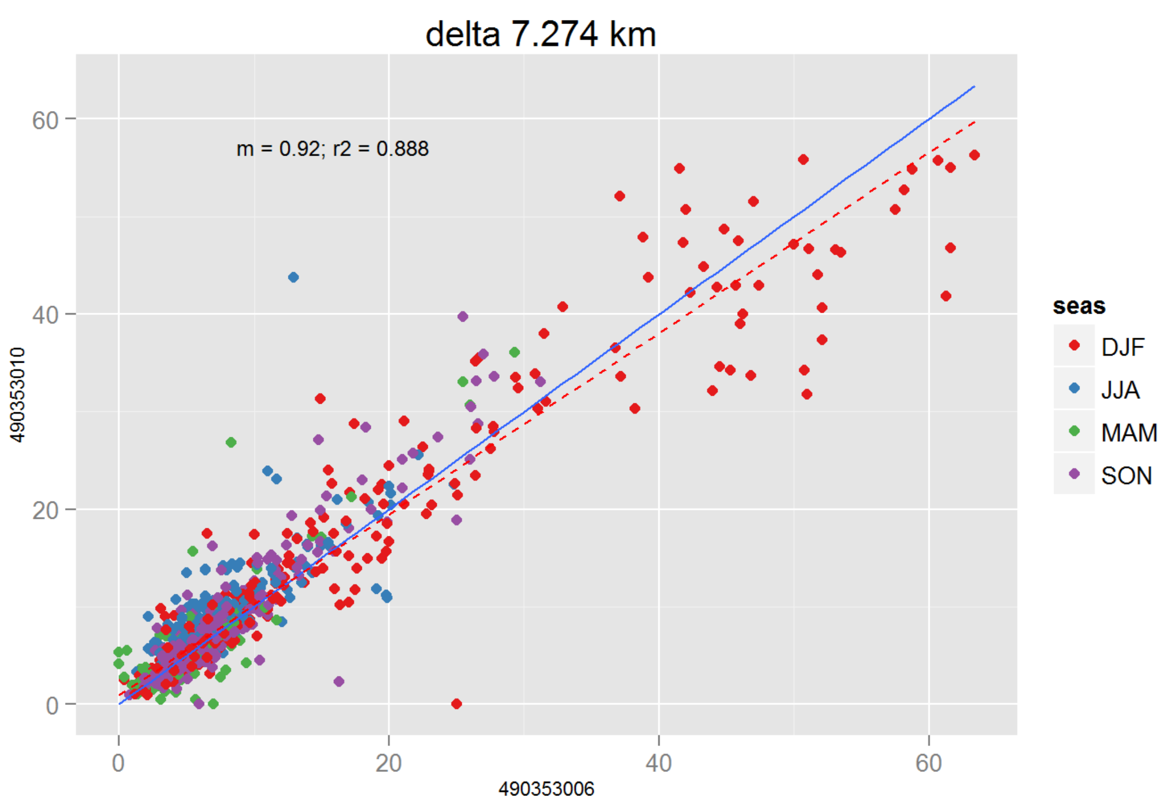


Figure 102: Comparison of $PM_{2.5}$ data for sites 490353006 and 490353006. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

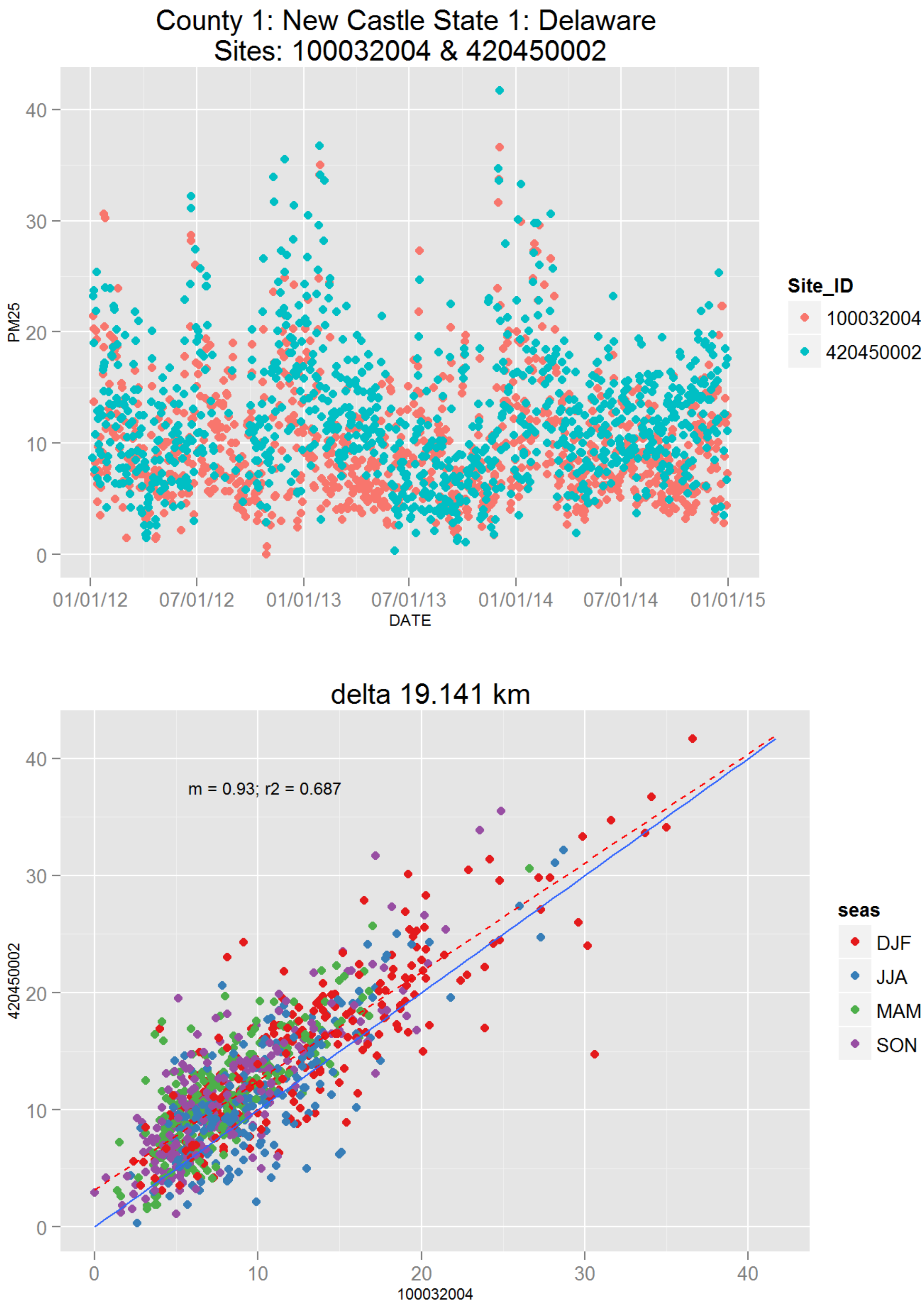


Figure 103: Comparison of $PM_{2.5}$ data for sites 100032004 and 100032004. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

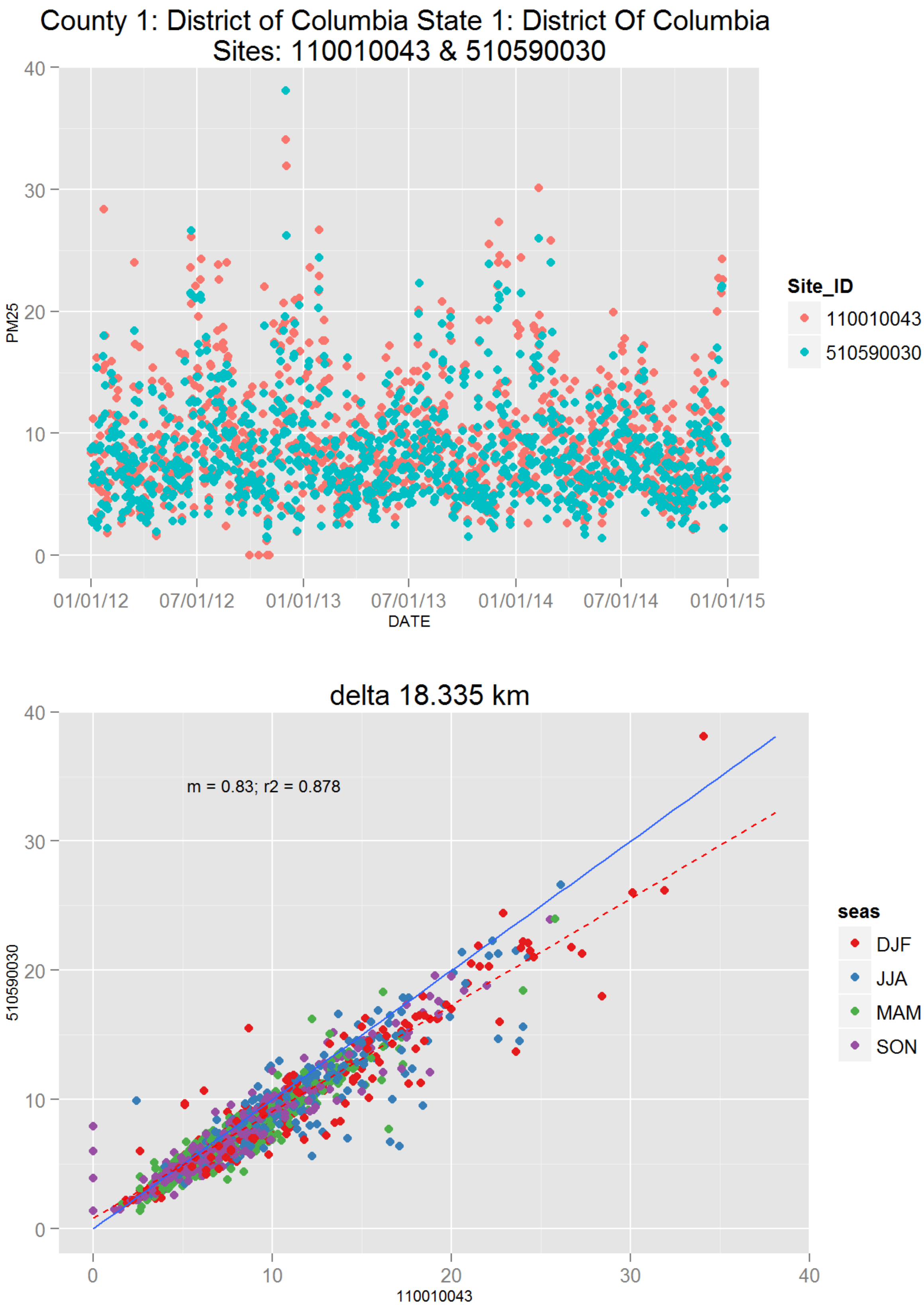
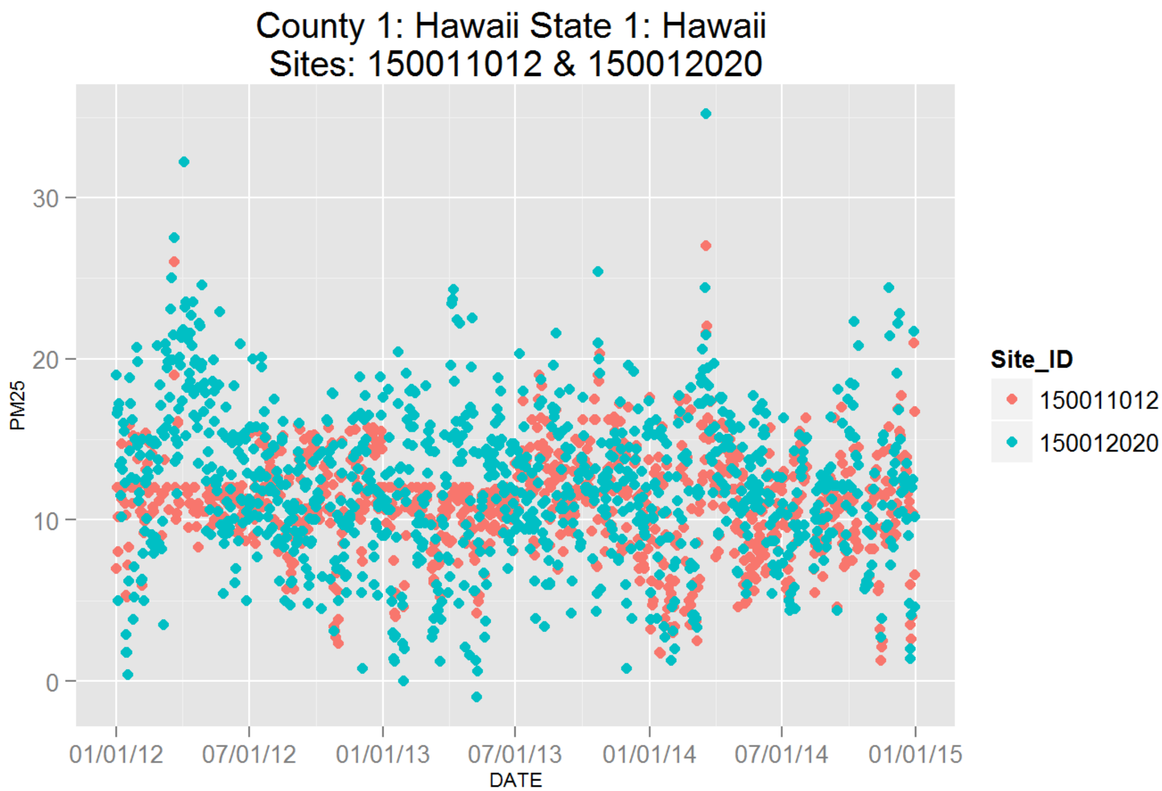


Figure 104: Comparison of $PM_{2.5}$ data for sites 110010043 and 110010043. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

Figure 105: Comparison of $PM_{2.5}$ data for sites 130670003 and 130670003. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

Figure 106: Comparison of $PM_{2.5}$ data for sites 150011006 and 150011006. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.



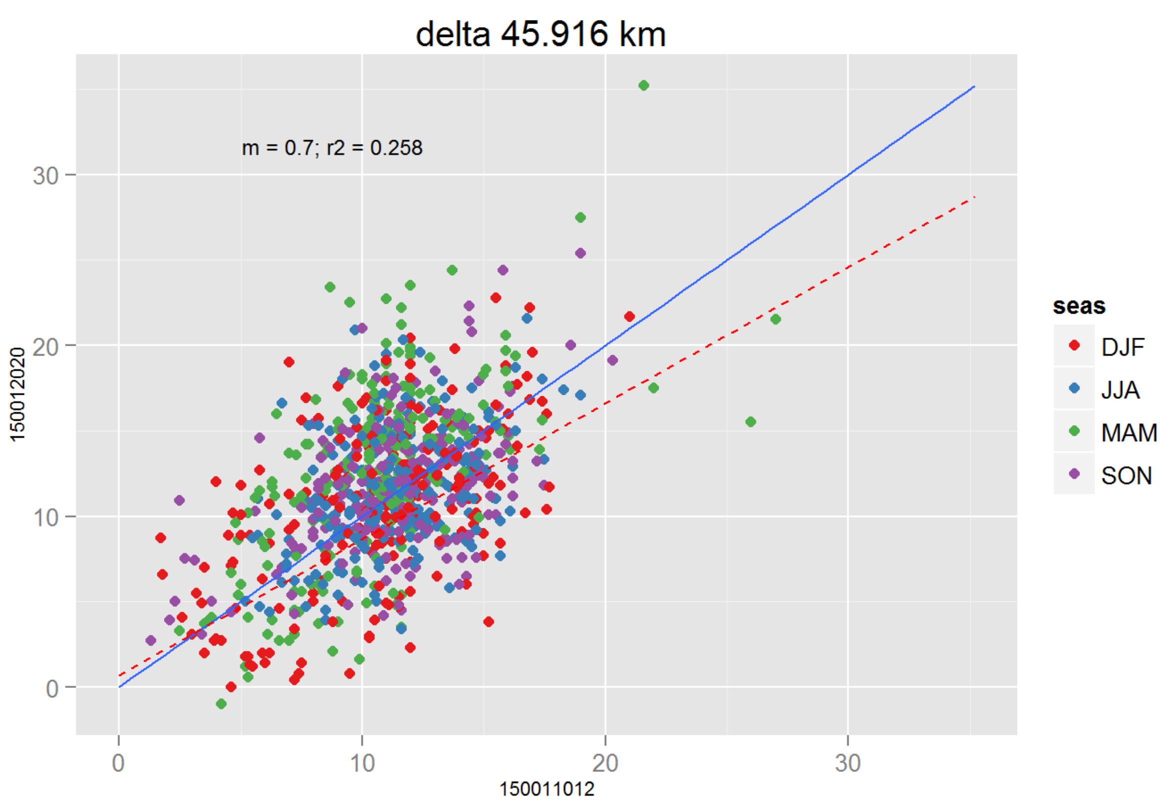


Figure 107: Comparison of $PM_{2.5}$ data for sites 150011012 and 150011012. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

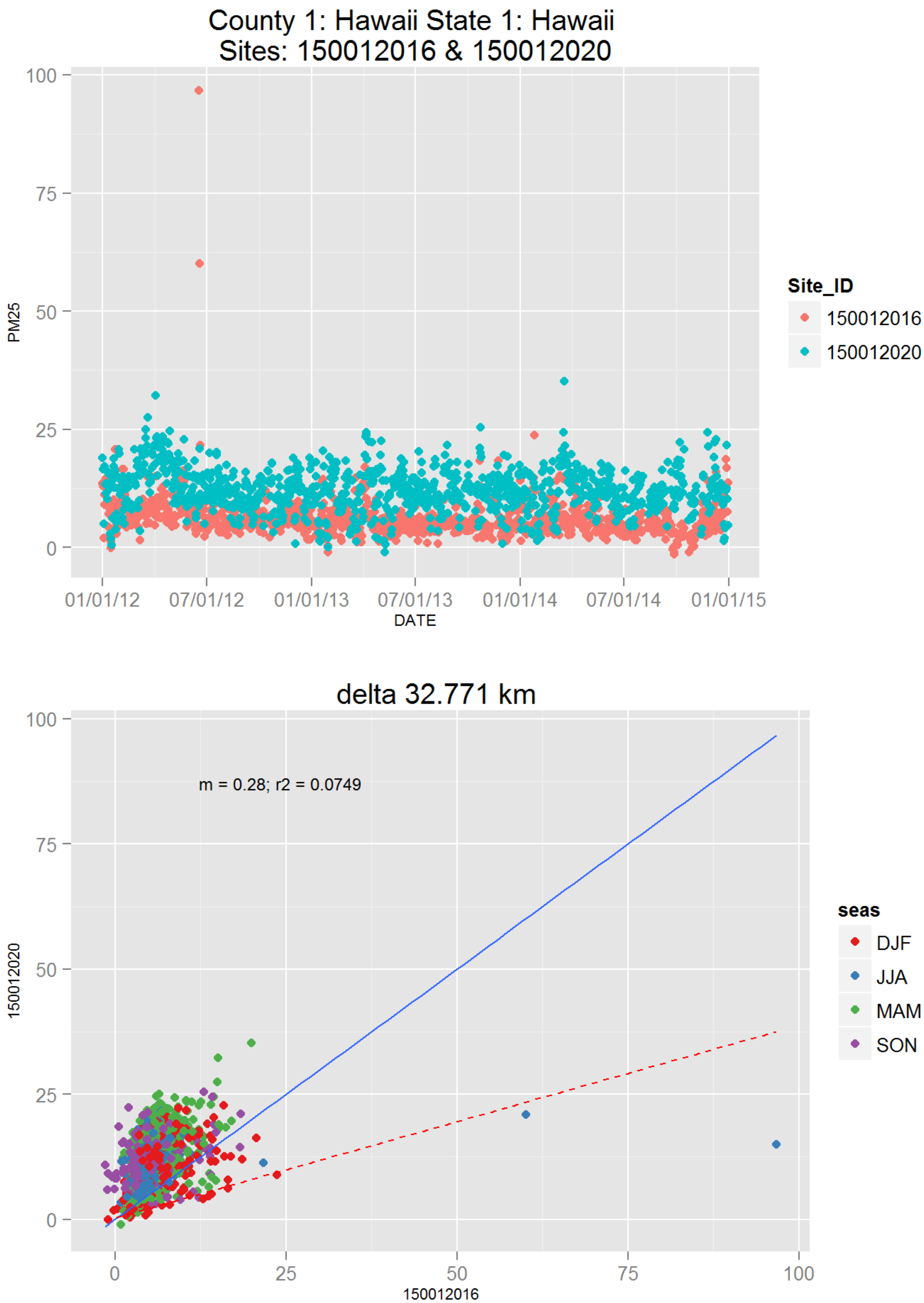


Figure 108: Comparison of $PM_{2.5}$ data for sites 150012016 and 150012014. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

Figure 109: Comparison of $PM_{2.5}$ data for sites 150031001 and 150031001. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

Figure 110: Comparison of $PM_{2.5}$ data for sites 150032004 and 150032004. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

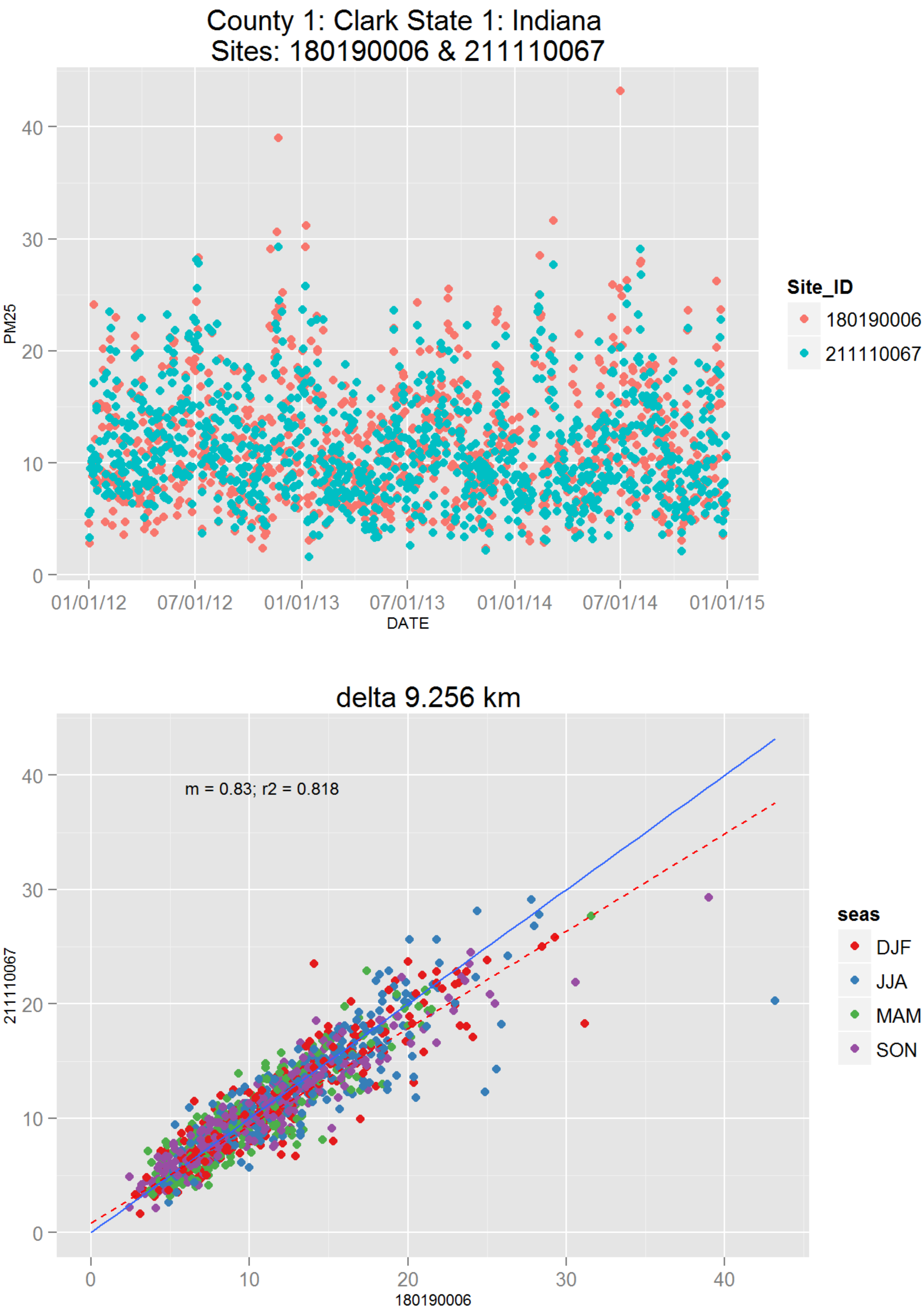


Figure 111: Comparison of $PM_{2.5}$ data for sites 180190006 and 180190006. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

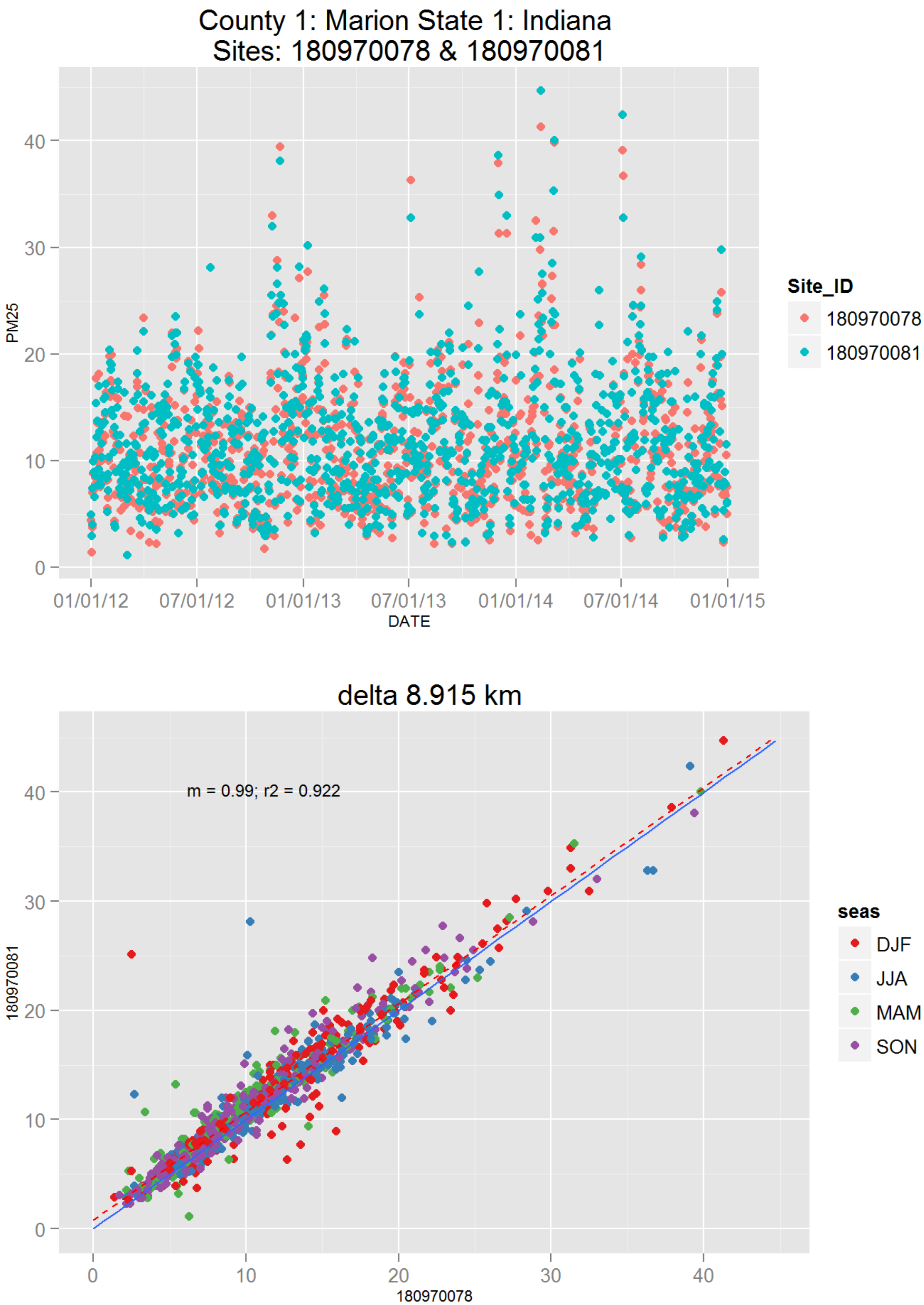
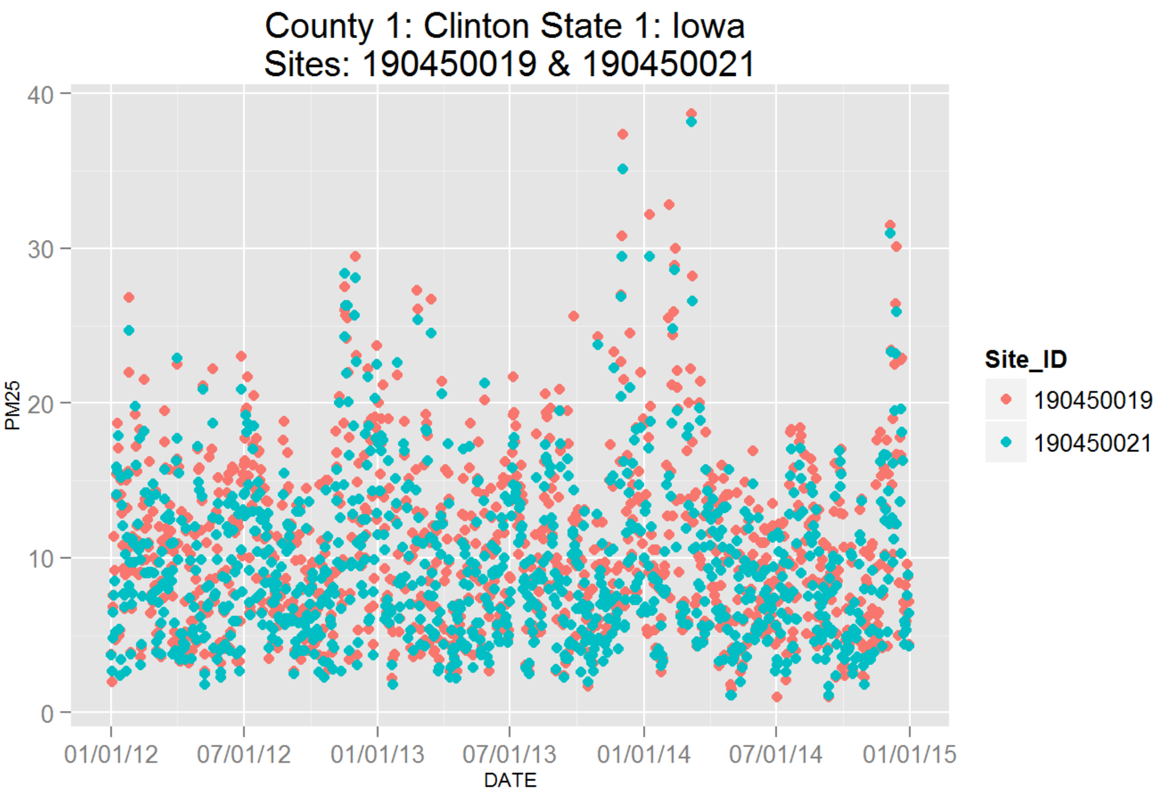


Figure 112: Comparison of $PM_{2.5}$ data for sites 180970078 and 180970078. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.



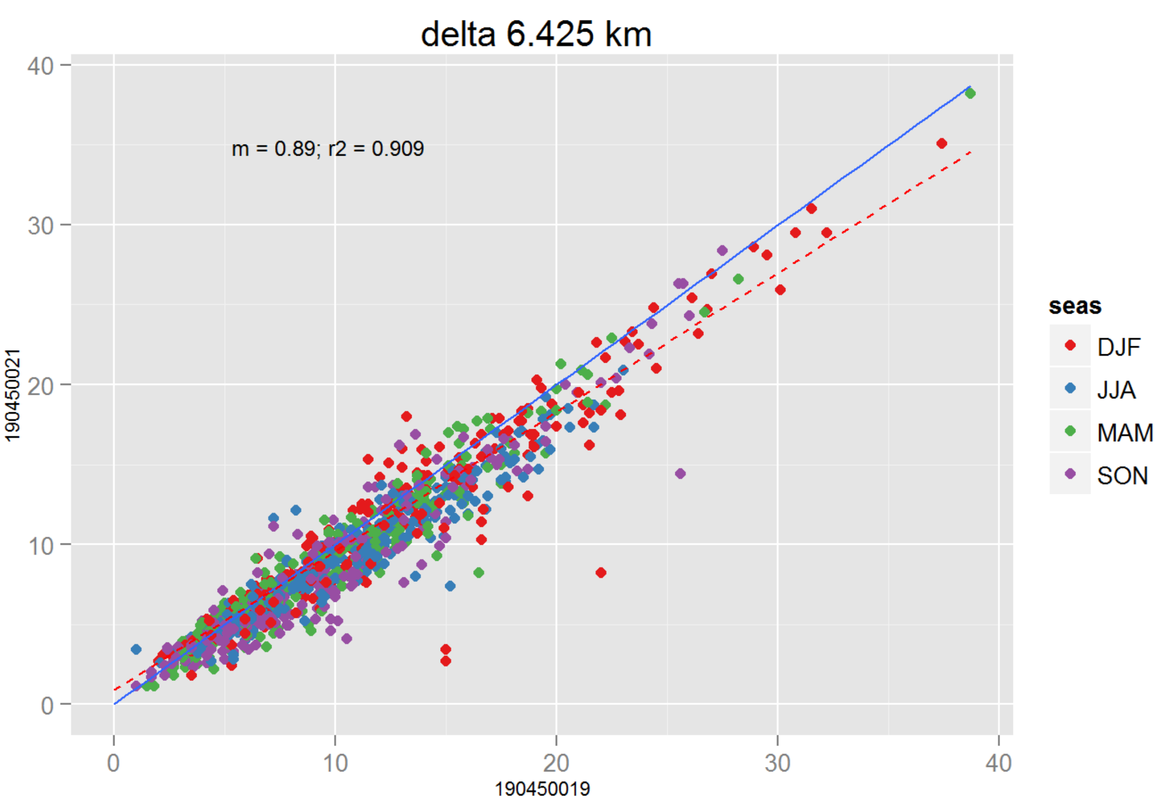


Figure 113: Comparison of $PM_{2.5}$ data for sites 190450019 and 190450019. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

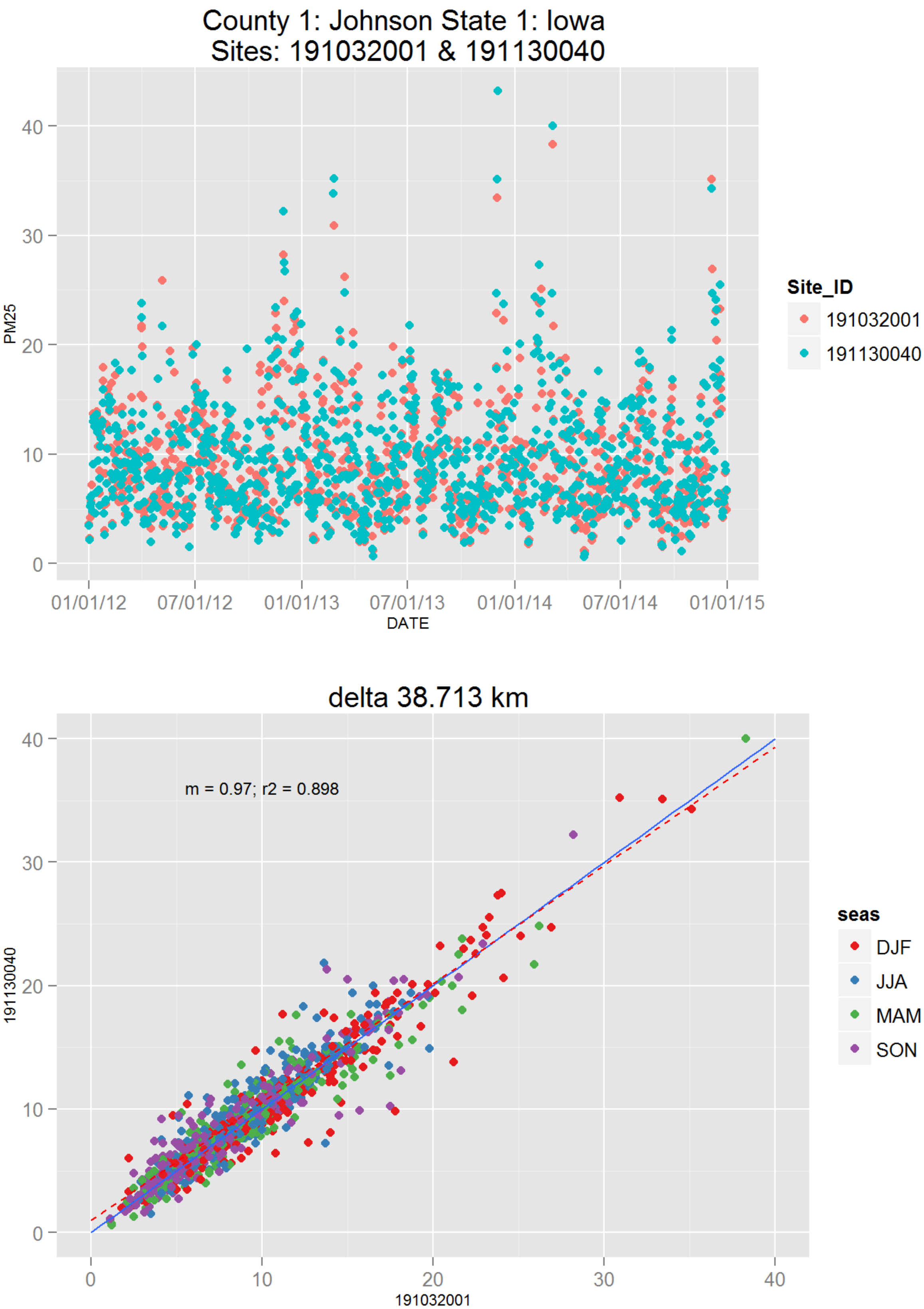
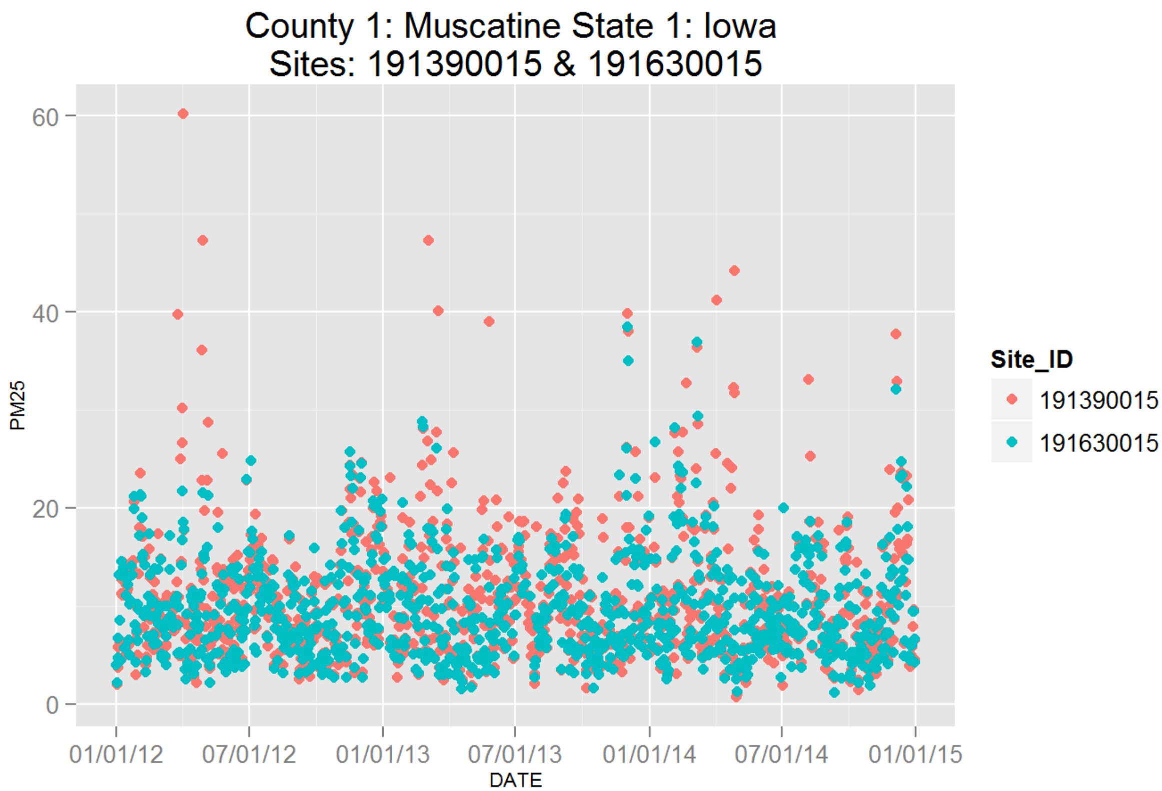


Figure 114: Comparison of $PM_{2.5}$ data for sites 191032001 and 191032001. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.



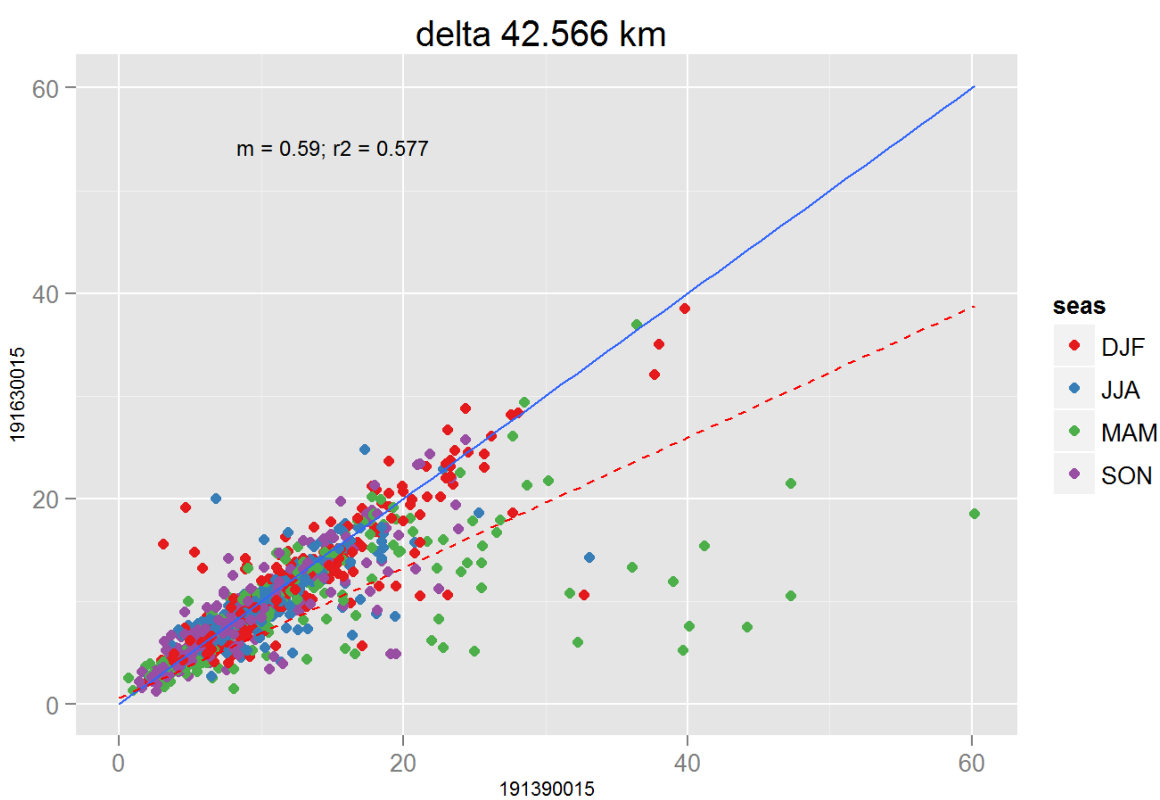


Figure 115: Comparison of $PM_{2.5}$ data for sites 191390015 and 191390015. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

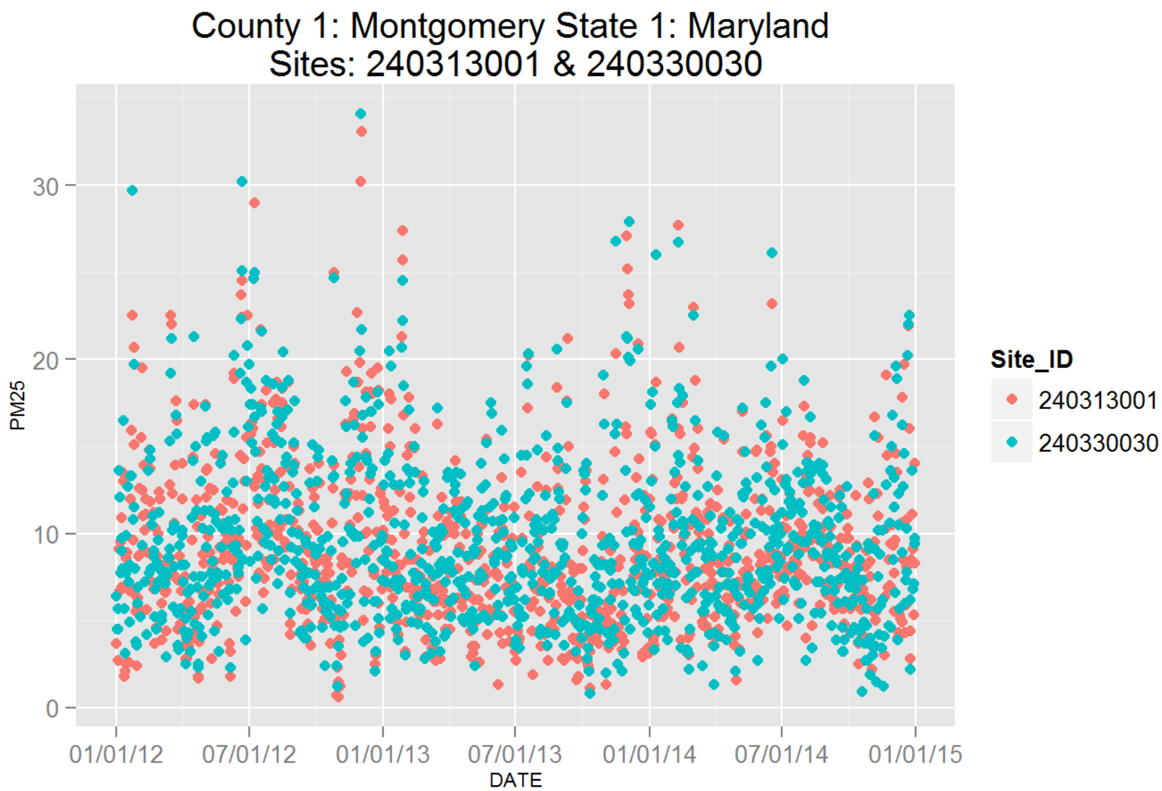
Figure 116: Comparison of $PM_{2.5}$ data for sites 211110051 and 211110051. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

Figure 117: Comparison of $PM_{2.5}$ data for sites 220330009 and 220330009. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

Figure 118: Comparison of $PM_{2.5}$ data for sites 240150003 and 240150003. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

Figure 119: Comparison of $PM_{2.5}$ data for sites 240251001 and 240251001. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

Figure 120: Comparison of $PM_{2.5}$ data for sites 240290002 and 240290002. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.



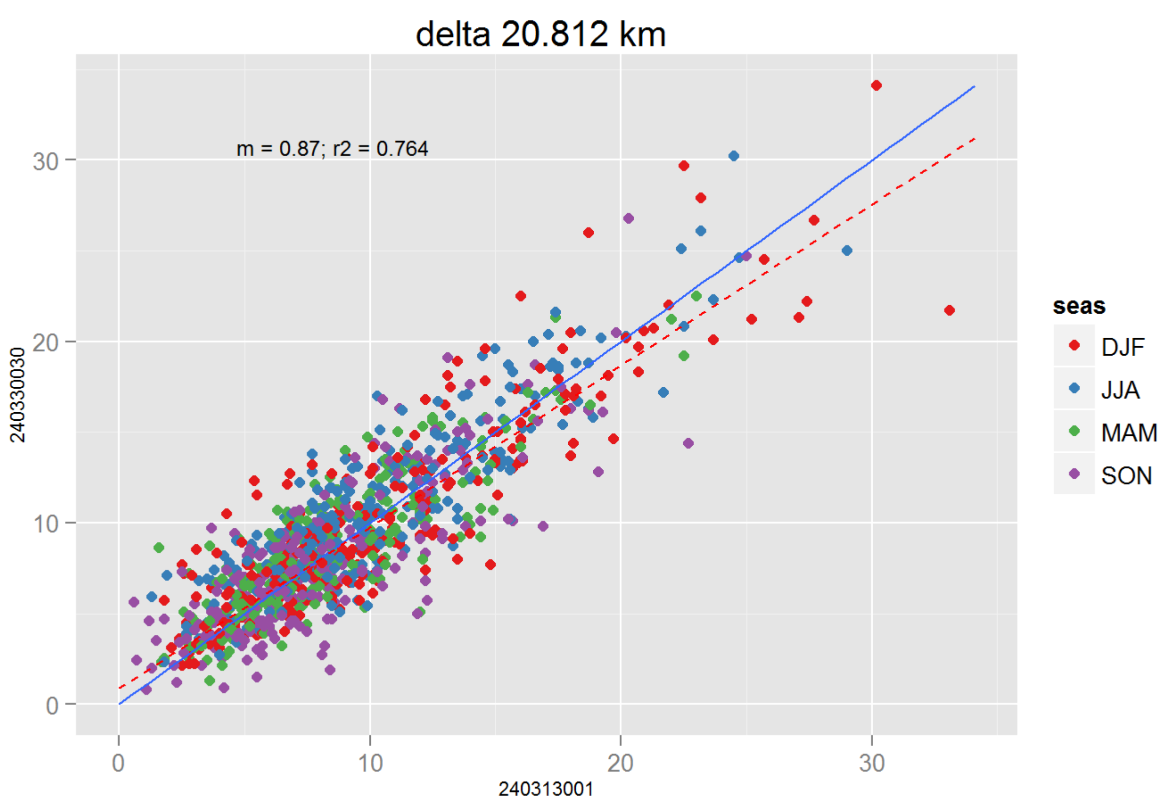
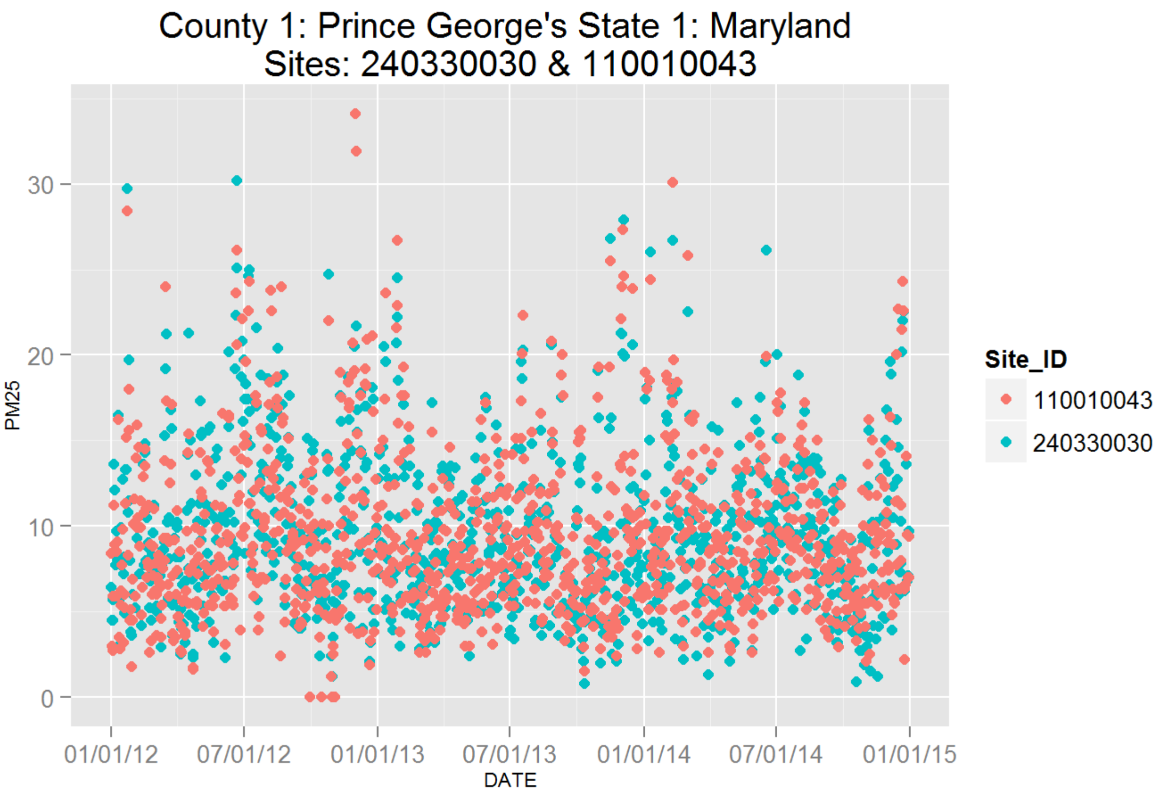


Figure 121: Comparison of $PM_{2.5}$ data for sites 240313001 and 240313001. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.



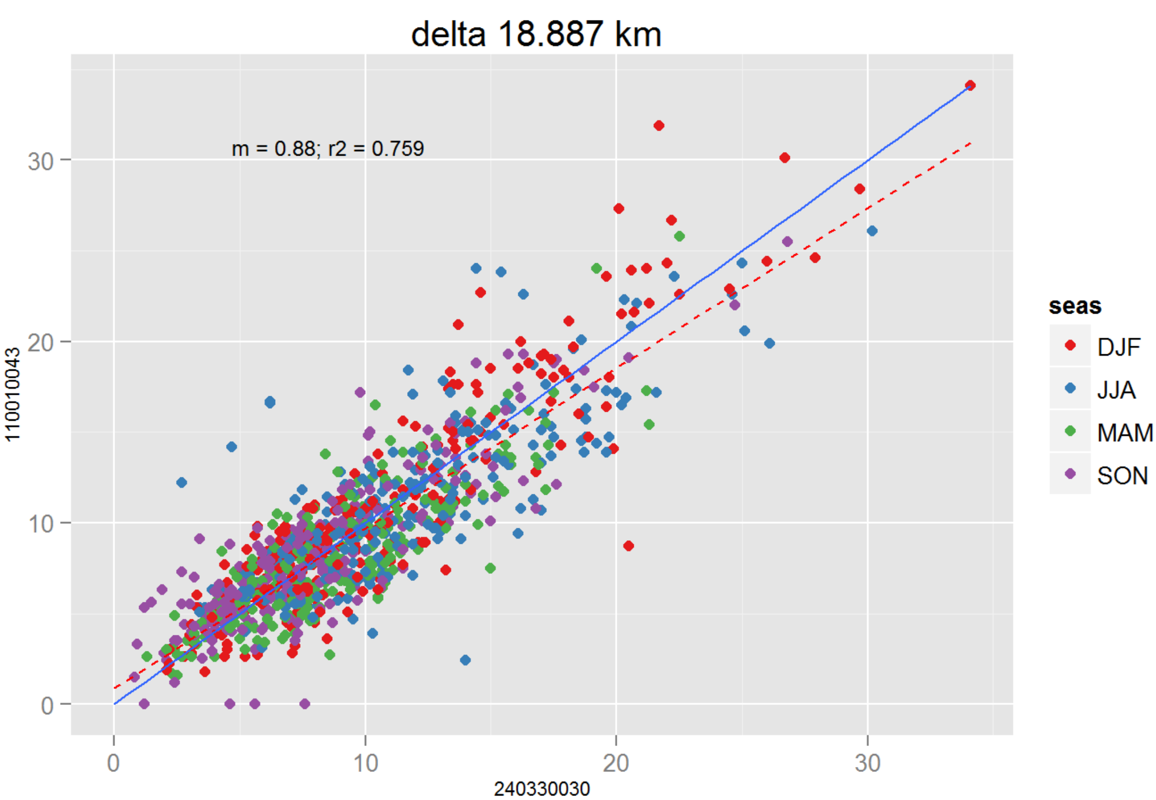


Figure 122: Comparison of $PM_{2.5}$ data for sites 240330030 and 240330030. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

Figure 123: Comparison of $PM_{2.5}$ data for sites 261630001 and 261630001. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

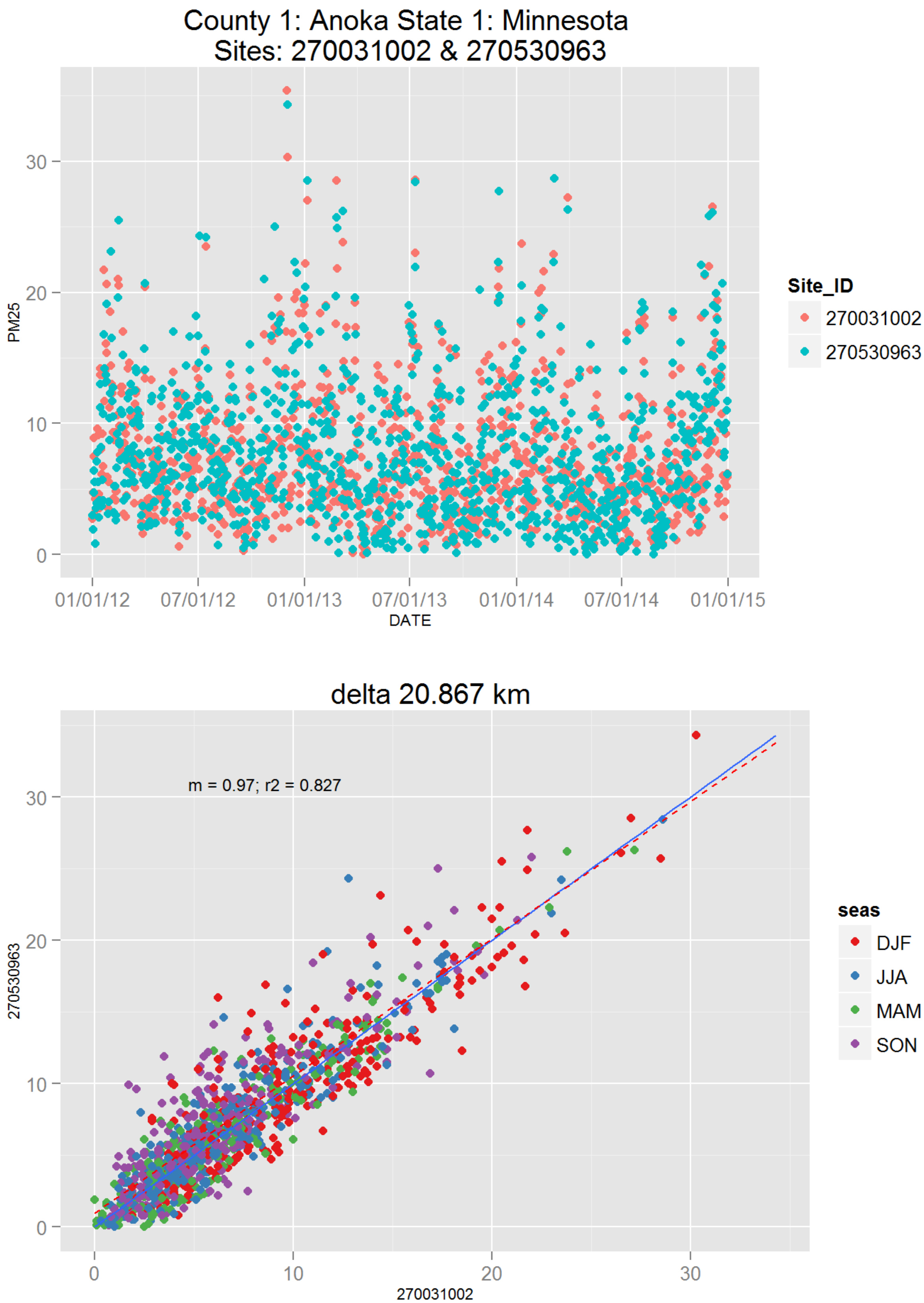


Figure 124: Comparison of $PM_{2.5}$ data for sites 270031002 and 270031002. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

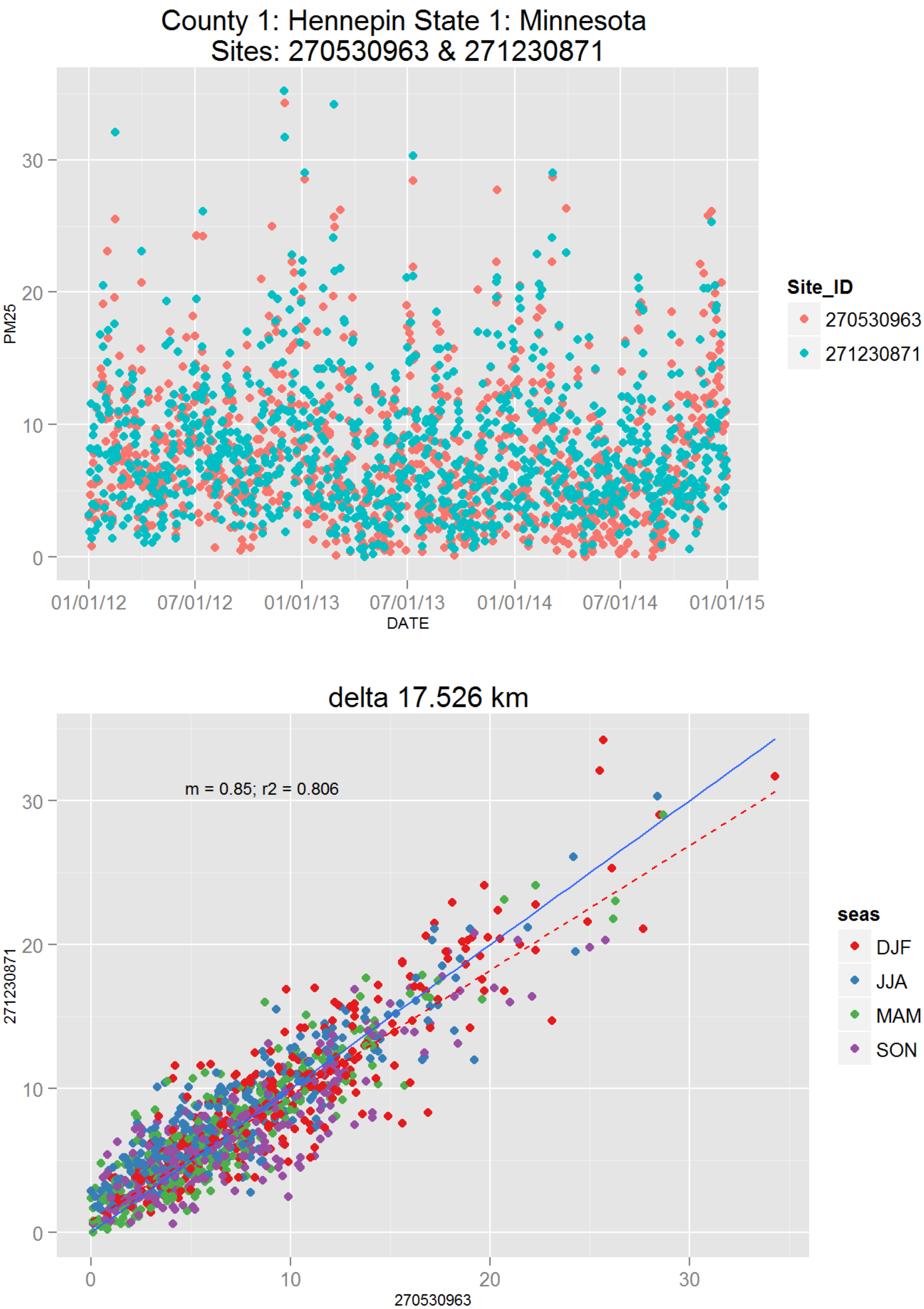


Figure 125: Comparison of $PM_{2.5}$ data for sites 270530963 and 270530963. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

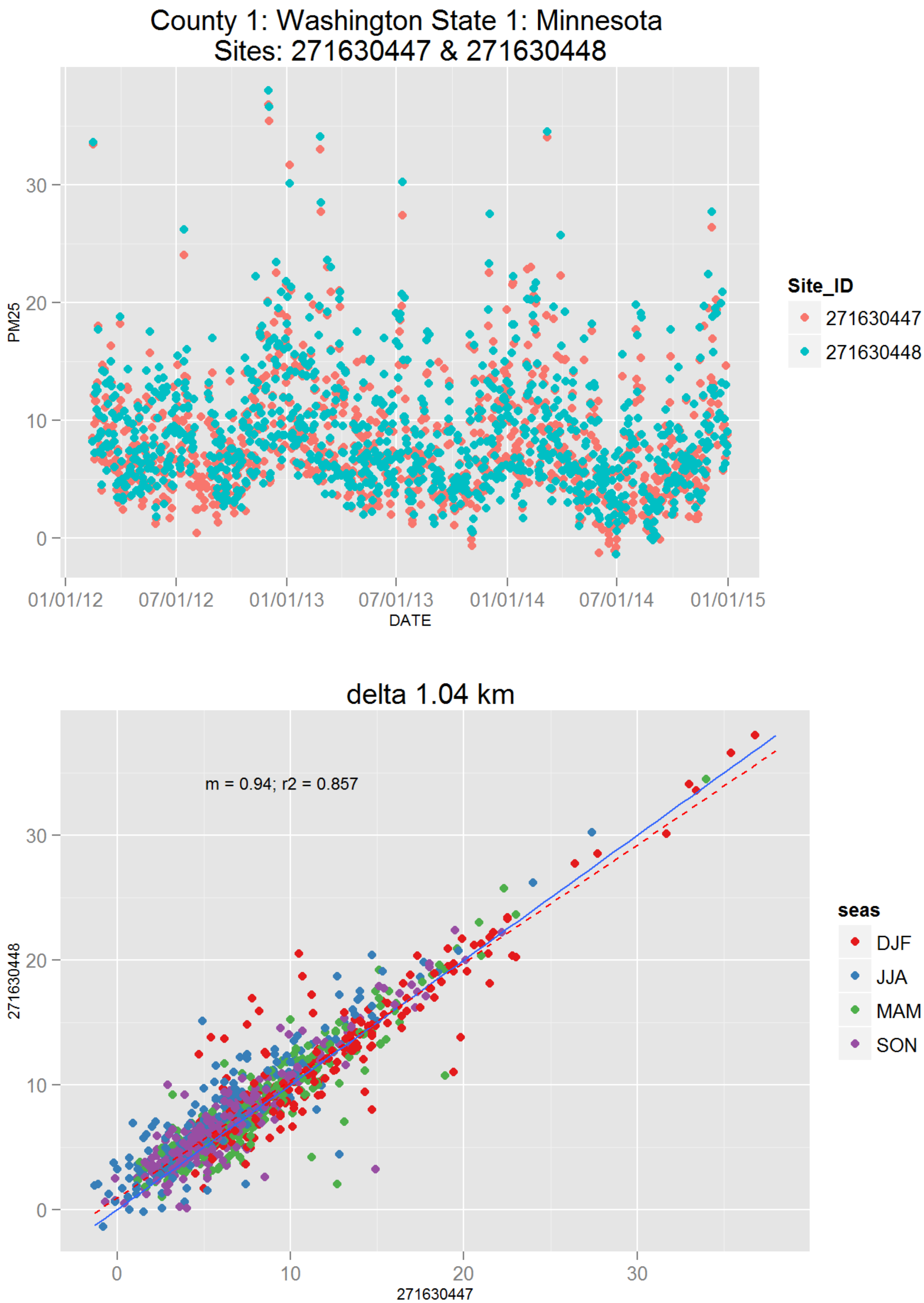
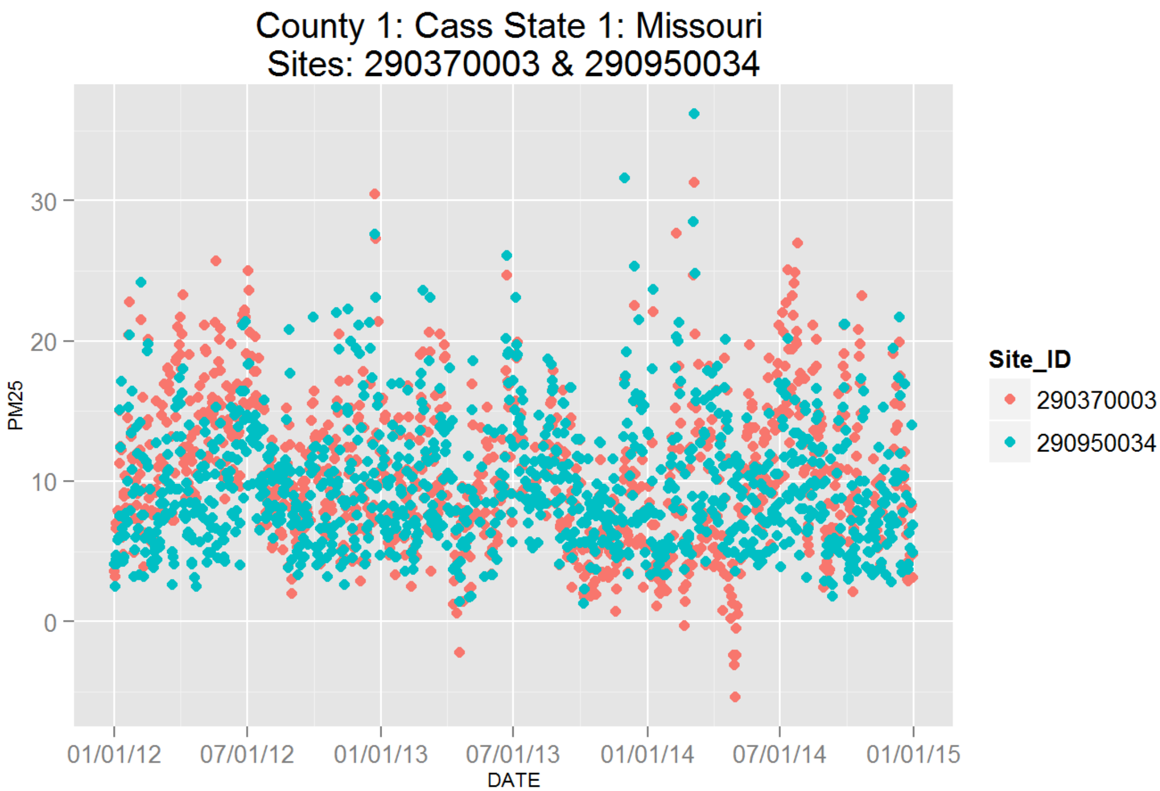


Figure 126: Comparison of $PM_{2.5}$ data for sites 271630447 and 271630447. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.



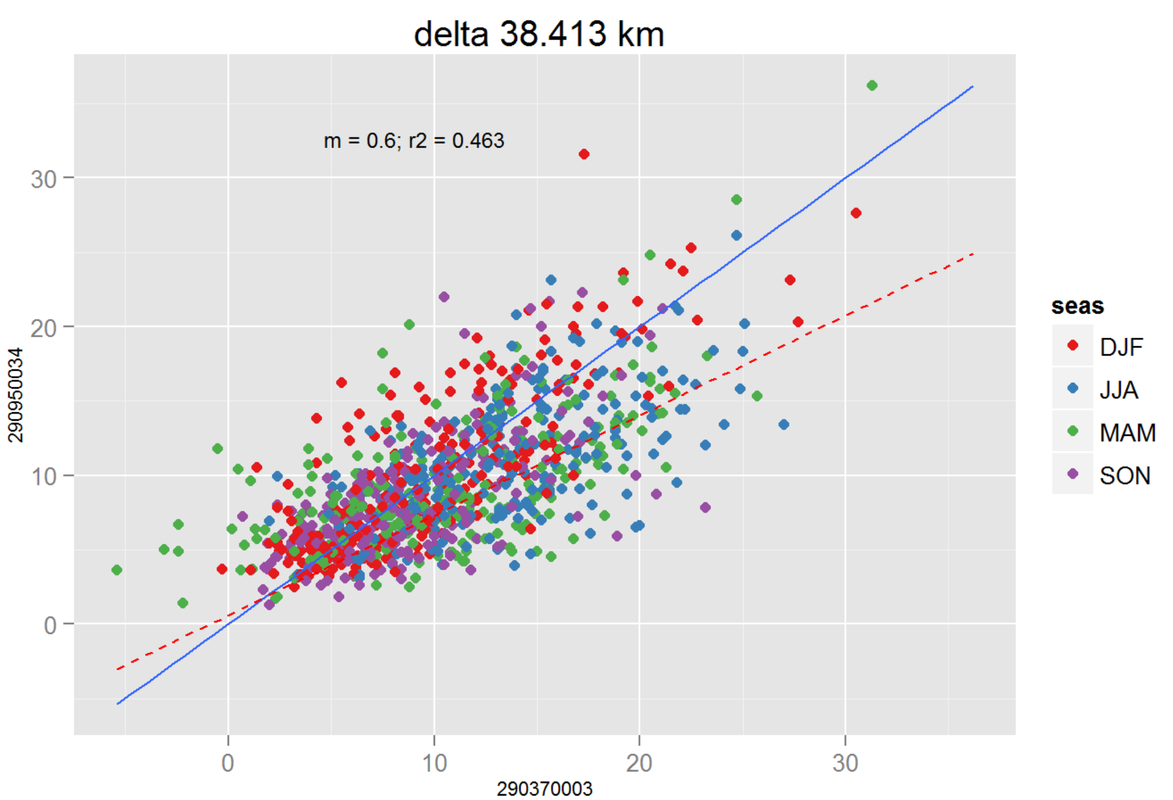
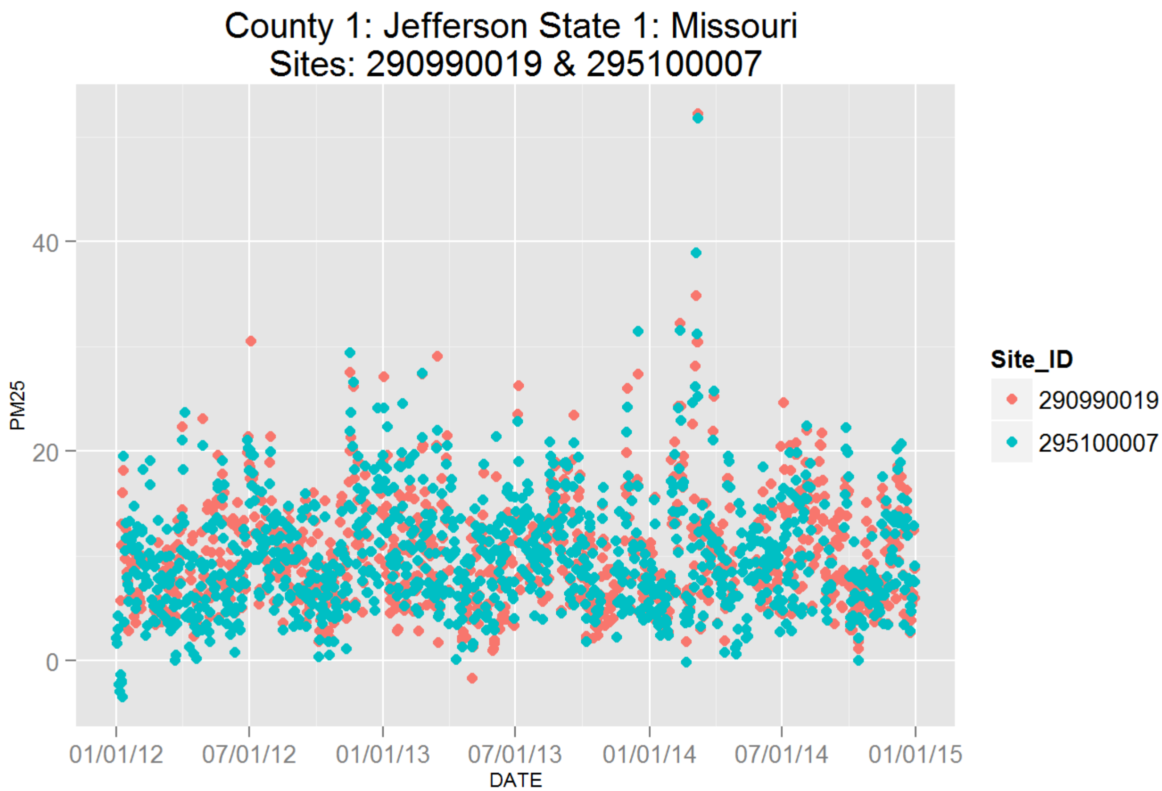


Figure 127: Comparison of $PM_{2.5}$ data for sites 290370003 and 290370003. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

Figure 128: Comparison of $PM_{2.5}$ data for sites 290470005 and 290470005. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.



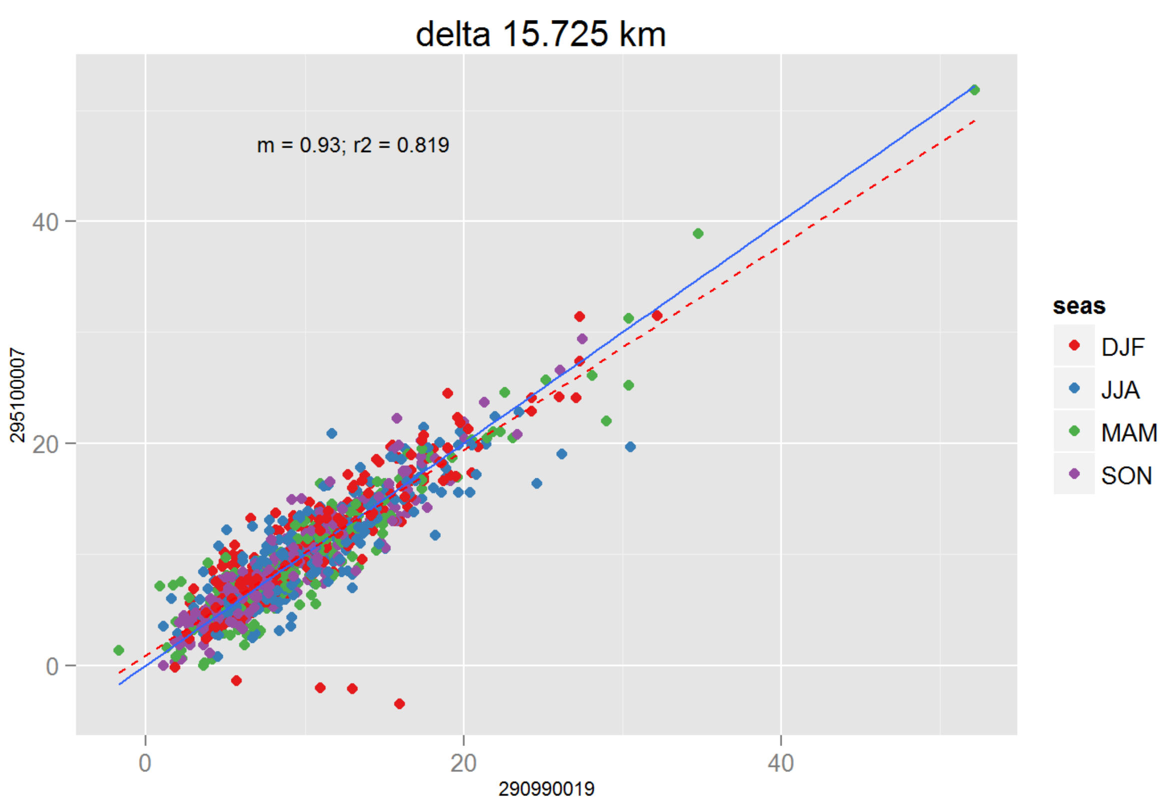
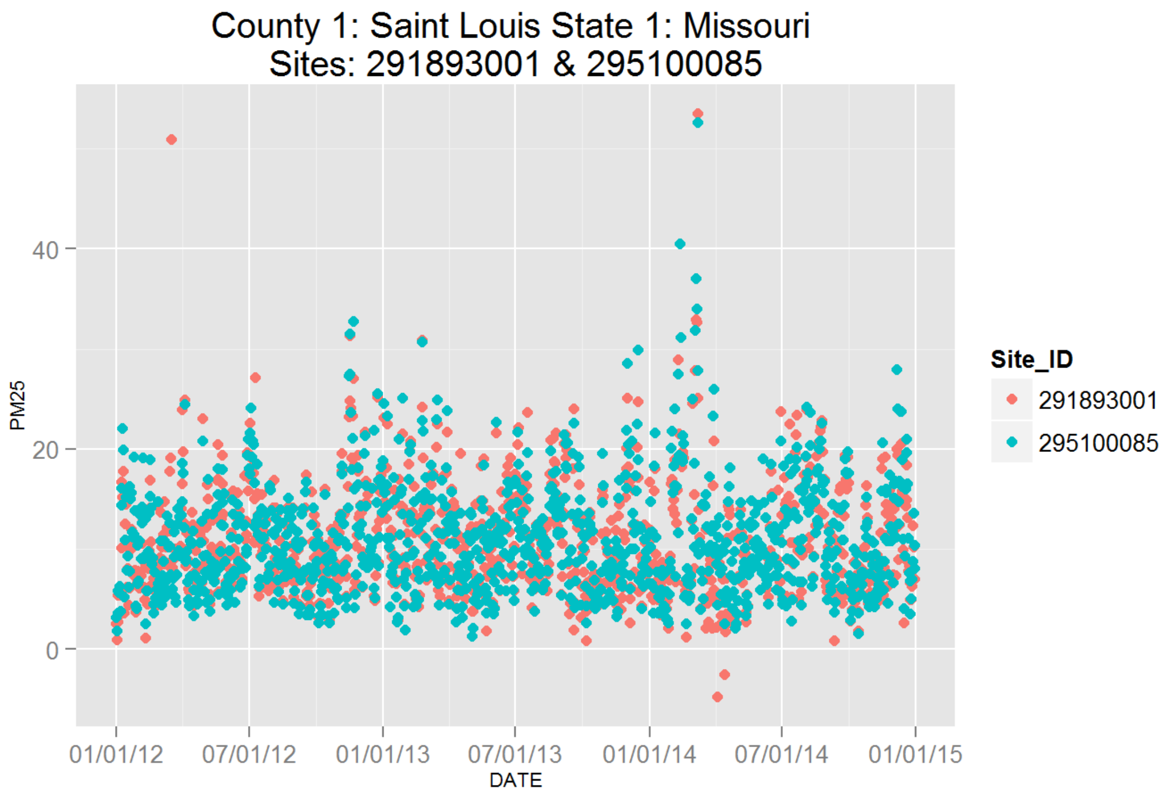


Figure 129: Comparison of $PM_{2.5}$ data for sites 290990019 and 290990019. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.



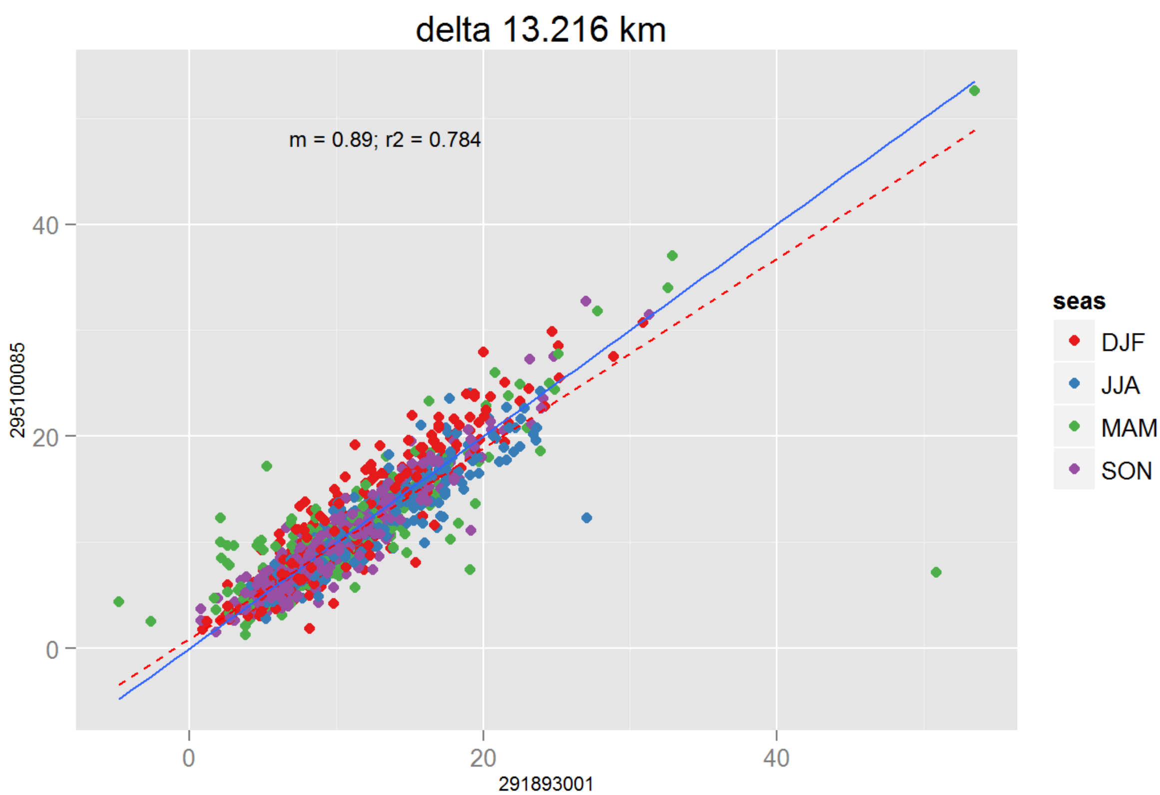
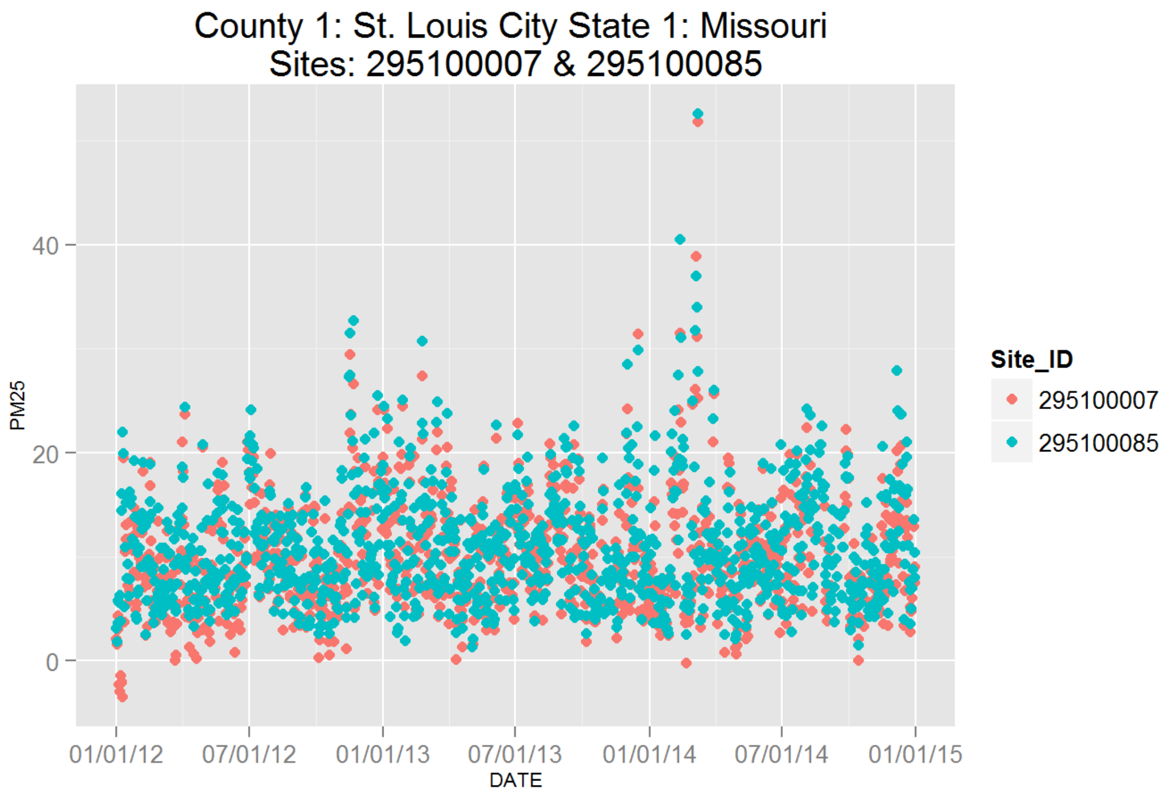


Figure 130: Comparison of $PM_{2.5}$ data for sites 291893001 and 291893001. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.



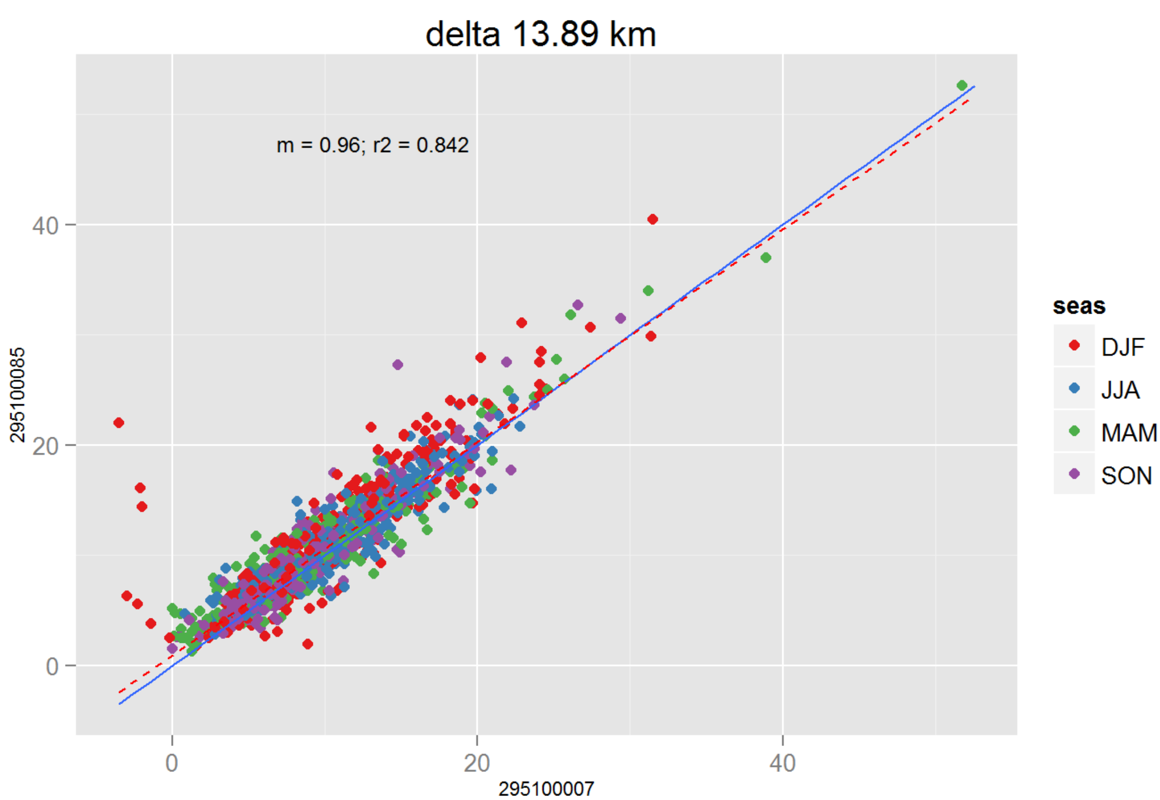


Figure 131: Comparison of $PM_{2.5}$ data for sites 295100007 and 295100007. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

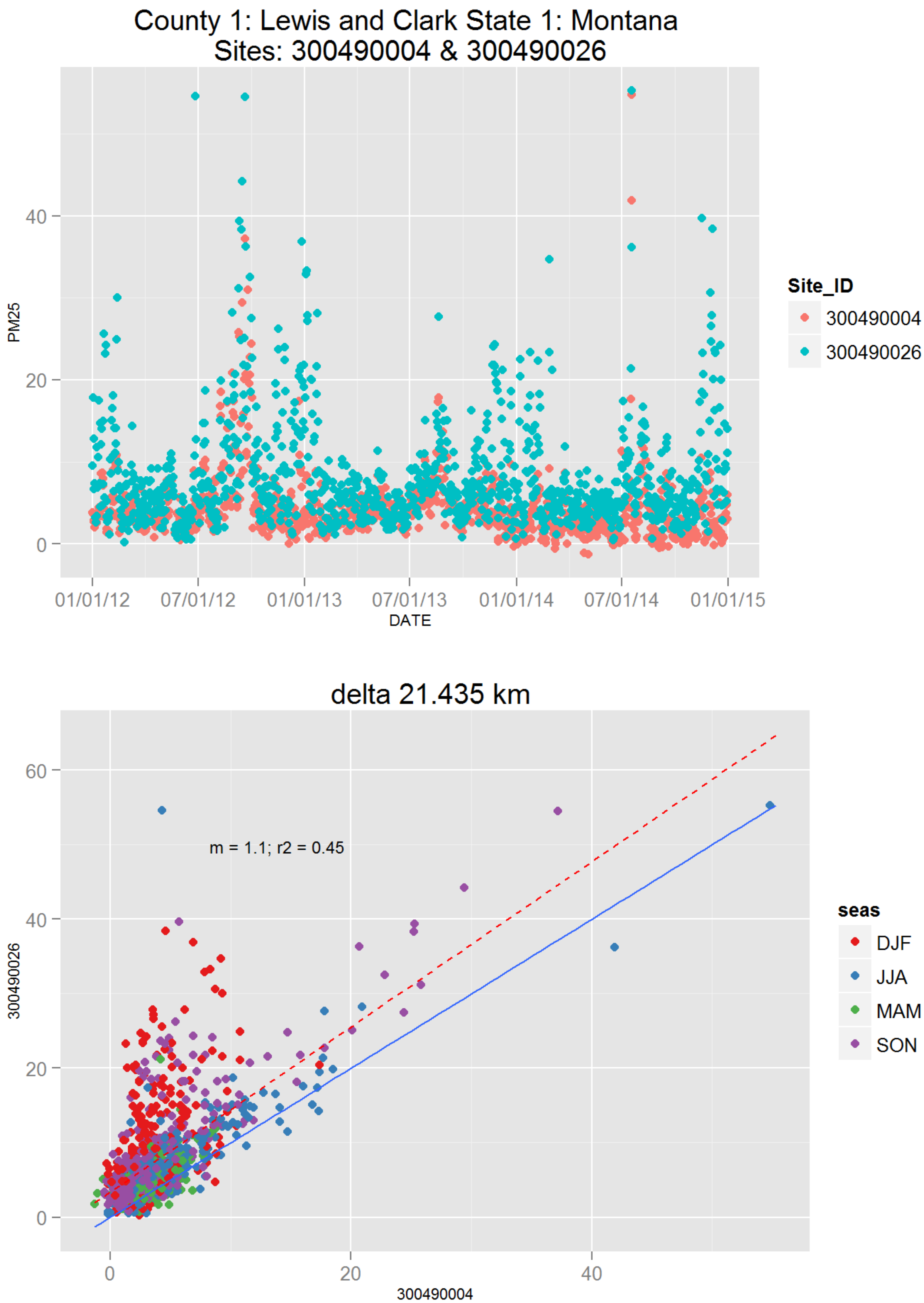
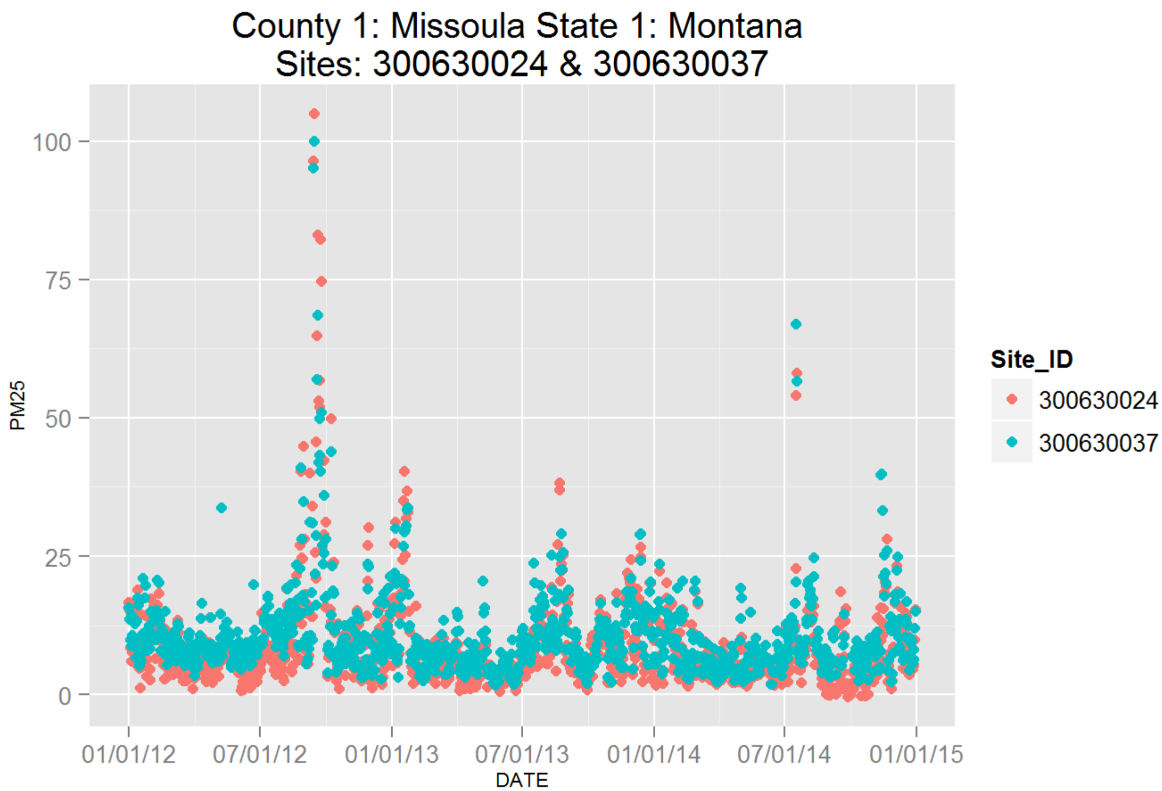


Figure 132: Comparison of $PM_{2.5}$ data for sites 300490004 and 300490004. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.



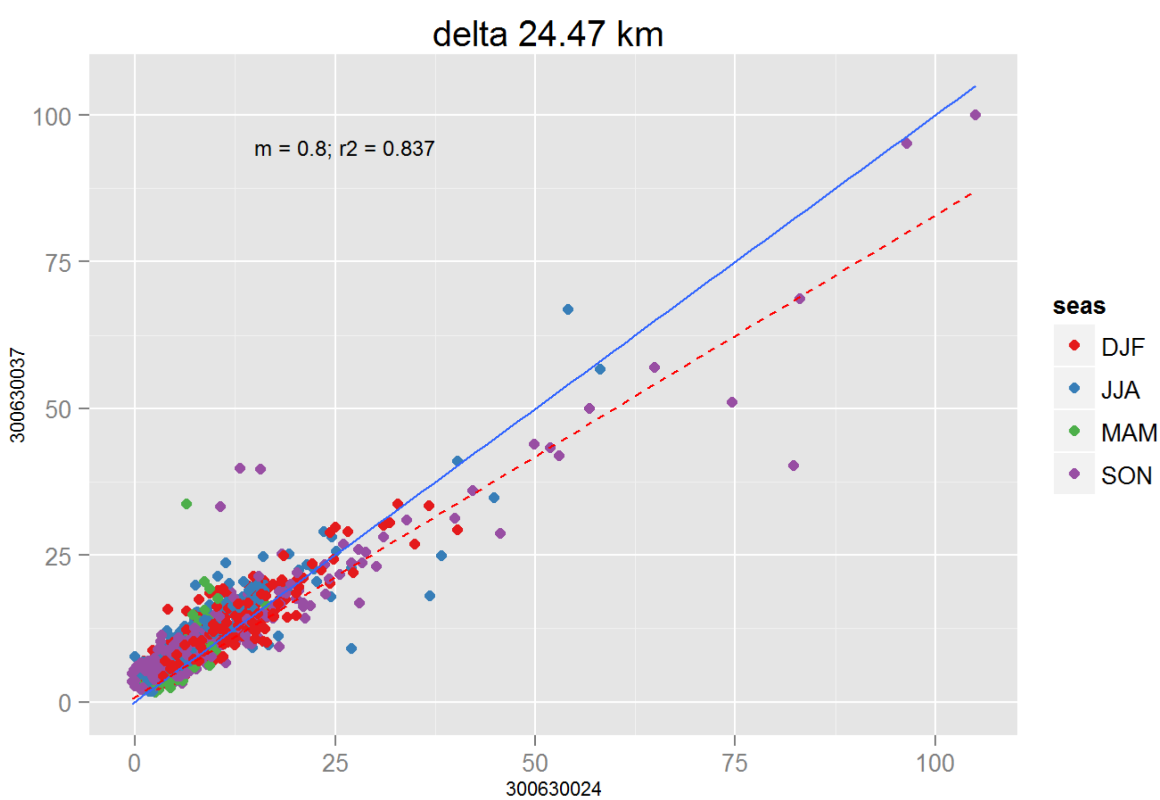
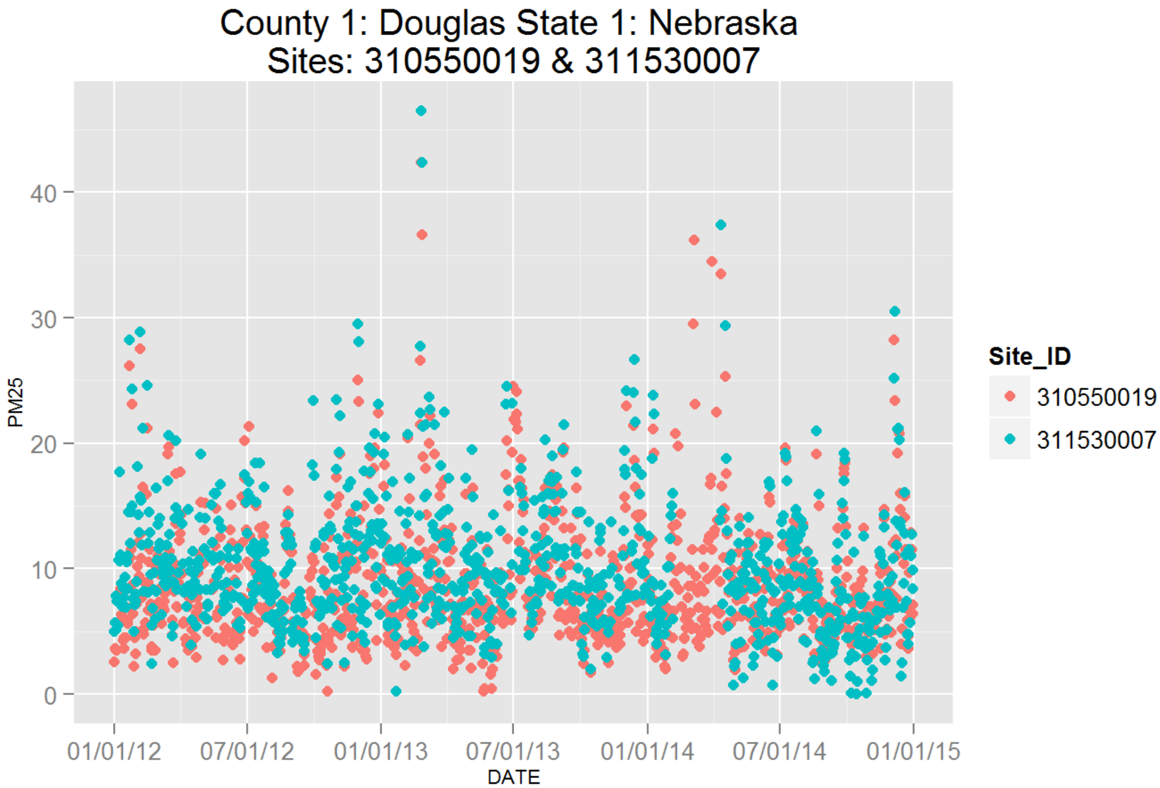


Figure 133: Comparison of $PM_{2.5}$ data for sites 300630024 and 300630024. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.



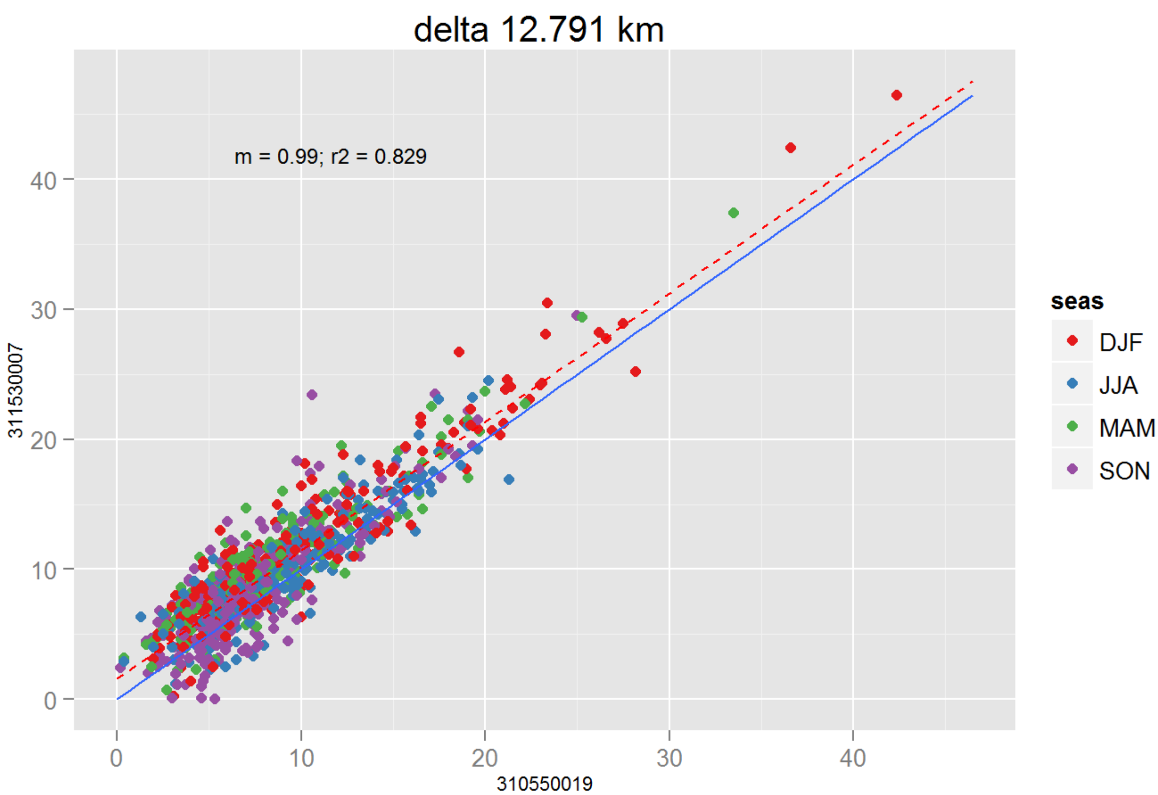
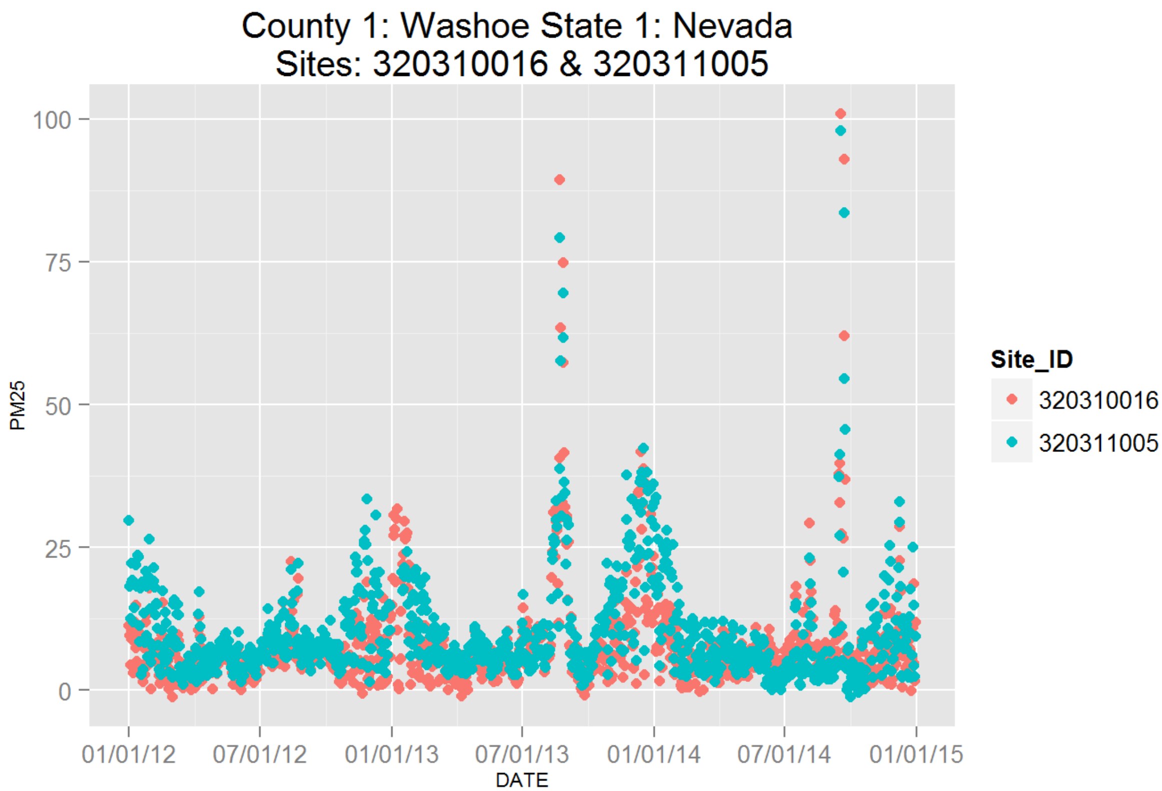


Figure 134: Comparison of $PM_{2.5}$ data for sites 310550019 and 310550019. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.



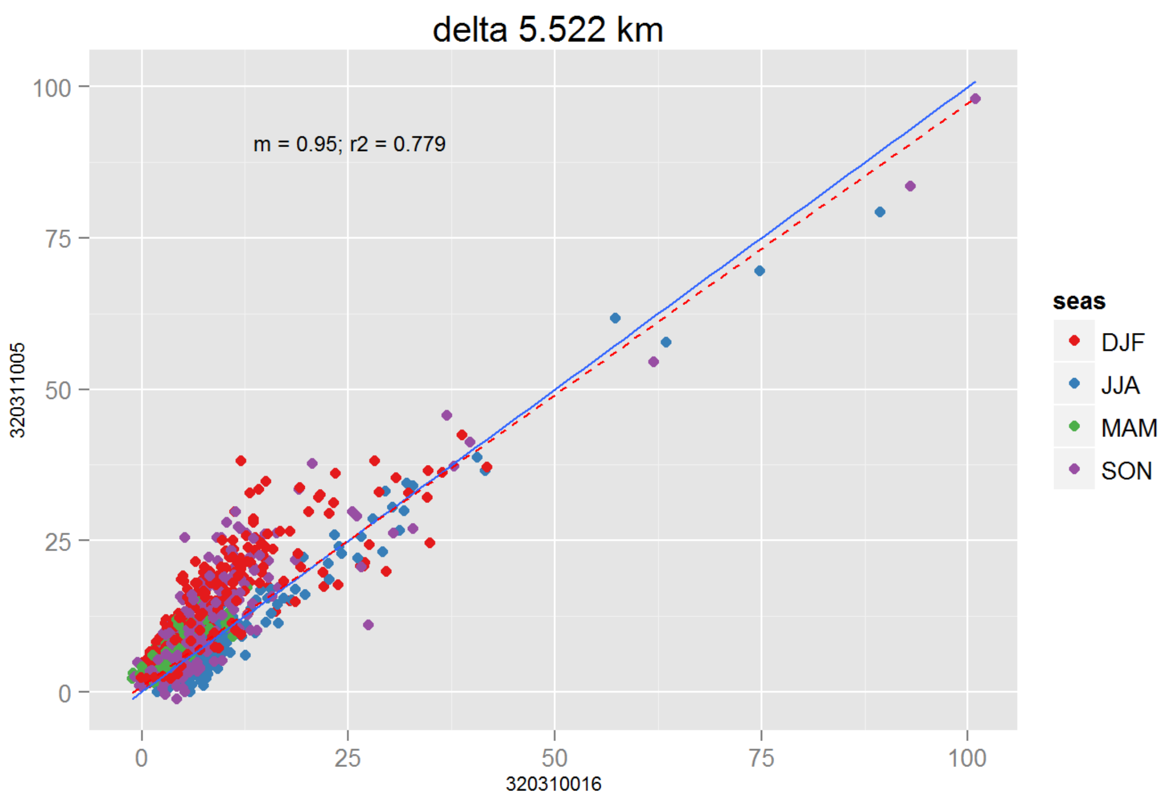


Figure 135: Comparison of $PM_{2.5}$ data for sites 320310016 and 320310016. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

Figure 136: Comparison of $PM_{2.5}$ data for sites 330050007 and 330050007. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

Figure 137: Comparison of $PM_{2.5}$ data for sites 330150018 and 330150018. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

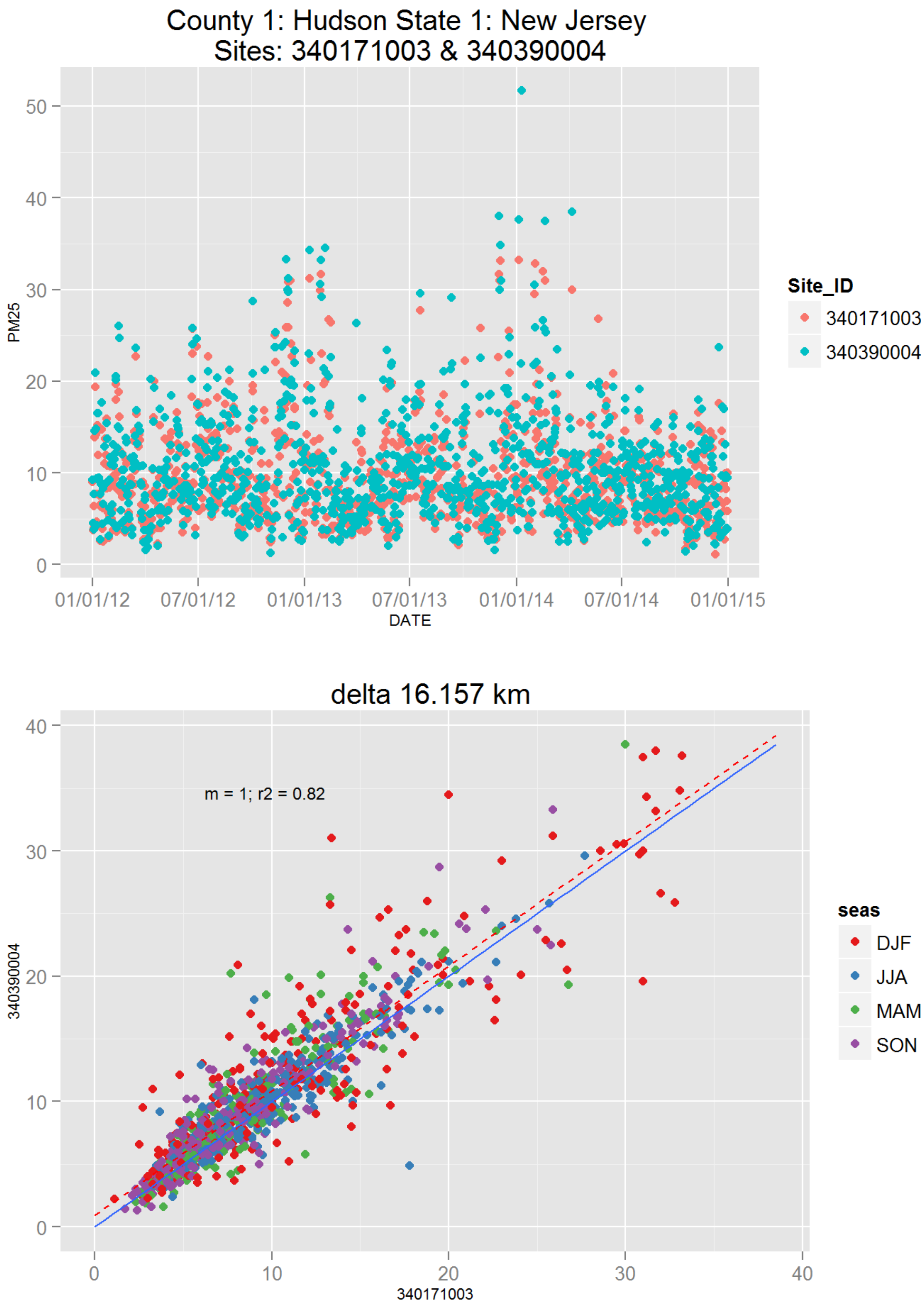


Figure 138: Comparison of $PM_{2.5}$ data for sites 340171003 and 340171003. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

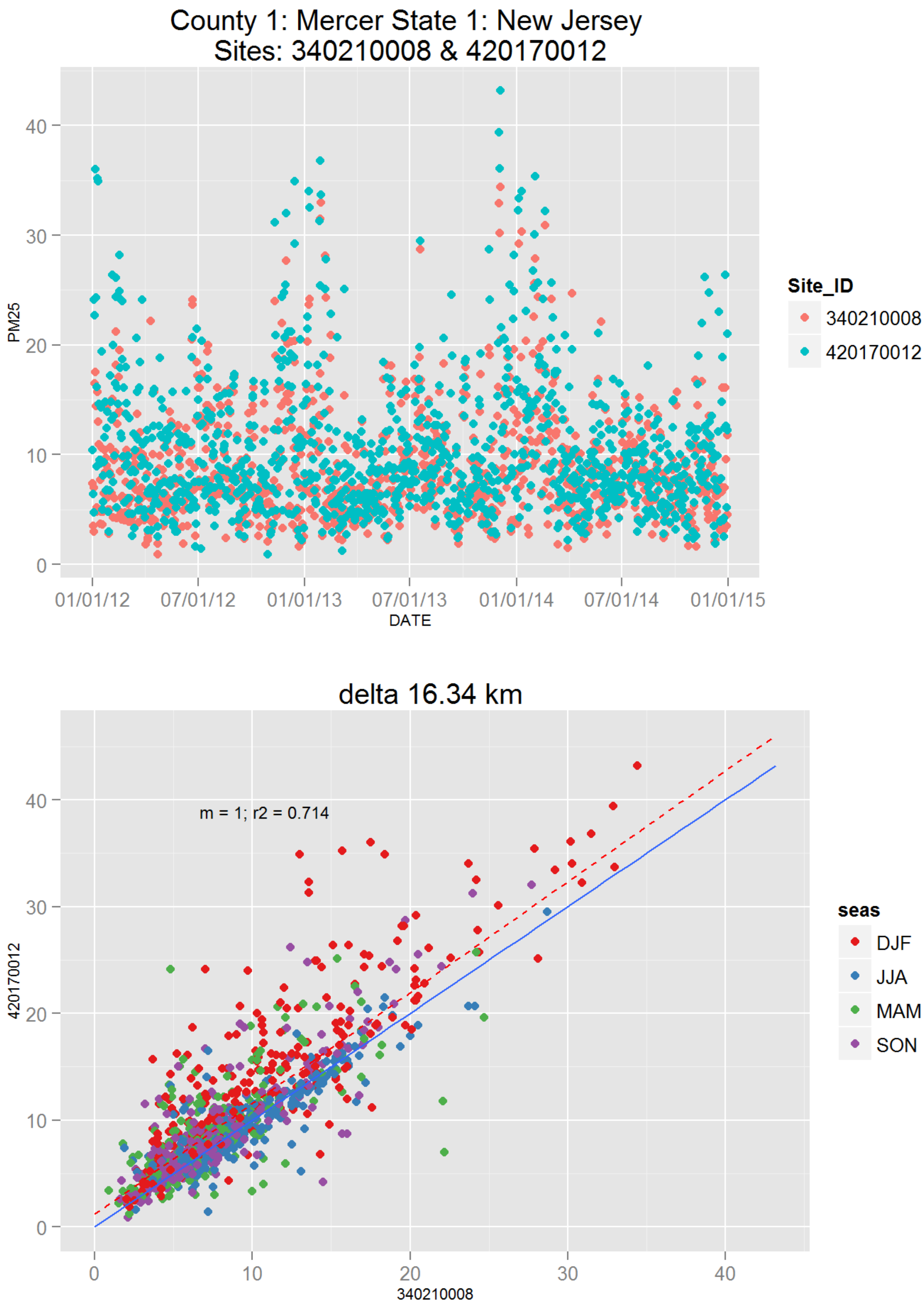


Figure 139: Comparison of $PM_{2.5}$ data for sites 340210008 and 340210008. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

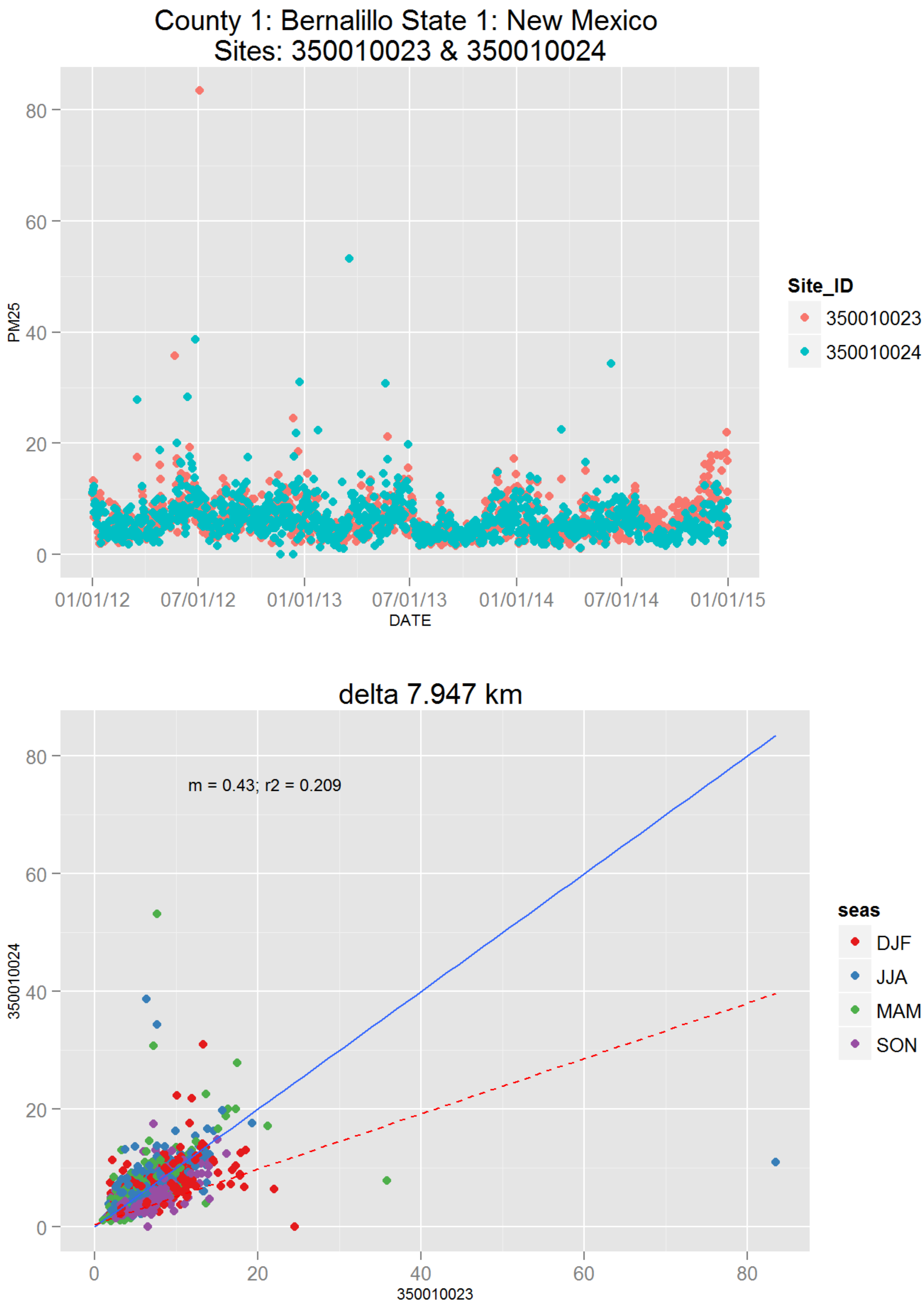
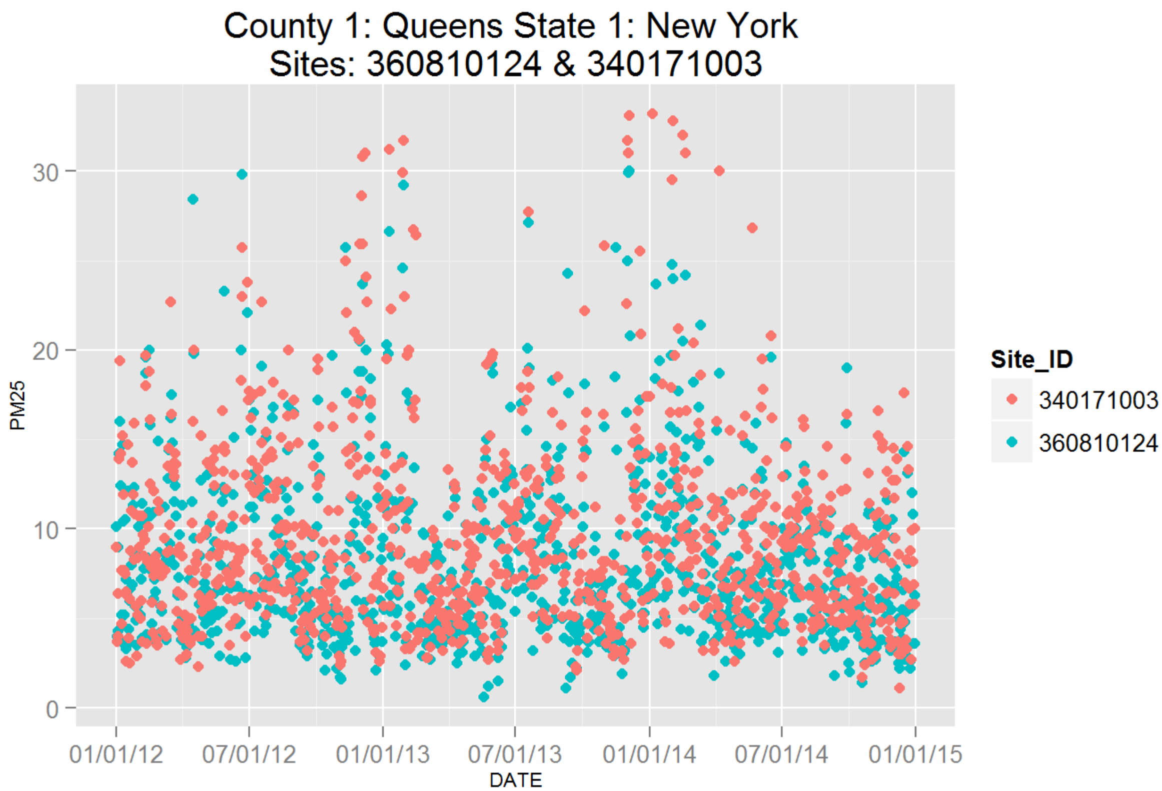


Figure 140: Comparison of $PM_{2.5}$ data for sites 350010023 and 350010023. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.



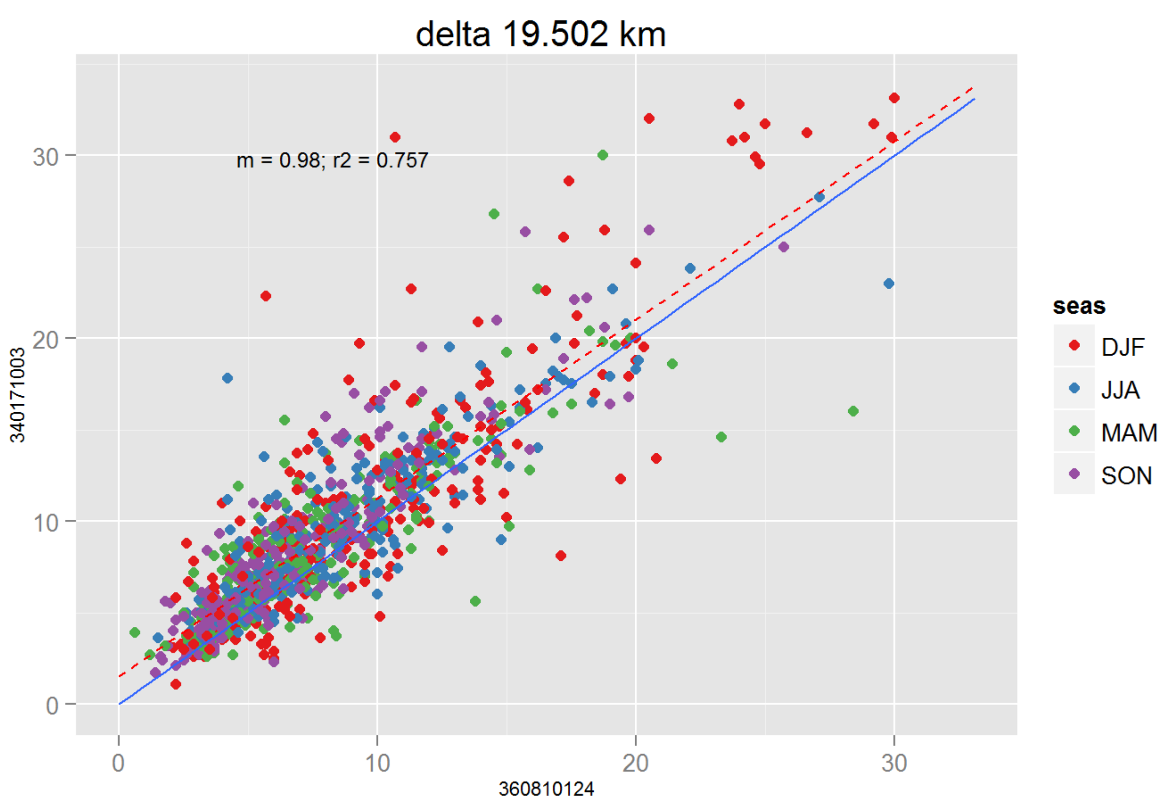


Figure 141: Comparison of $PM_{2.5}$ data for sites 360810124 and 360810124. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

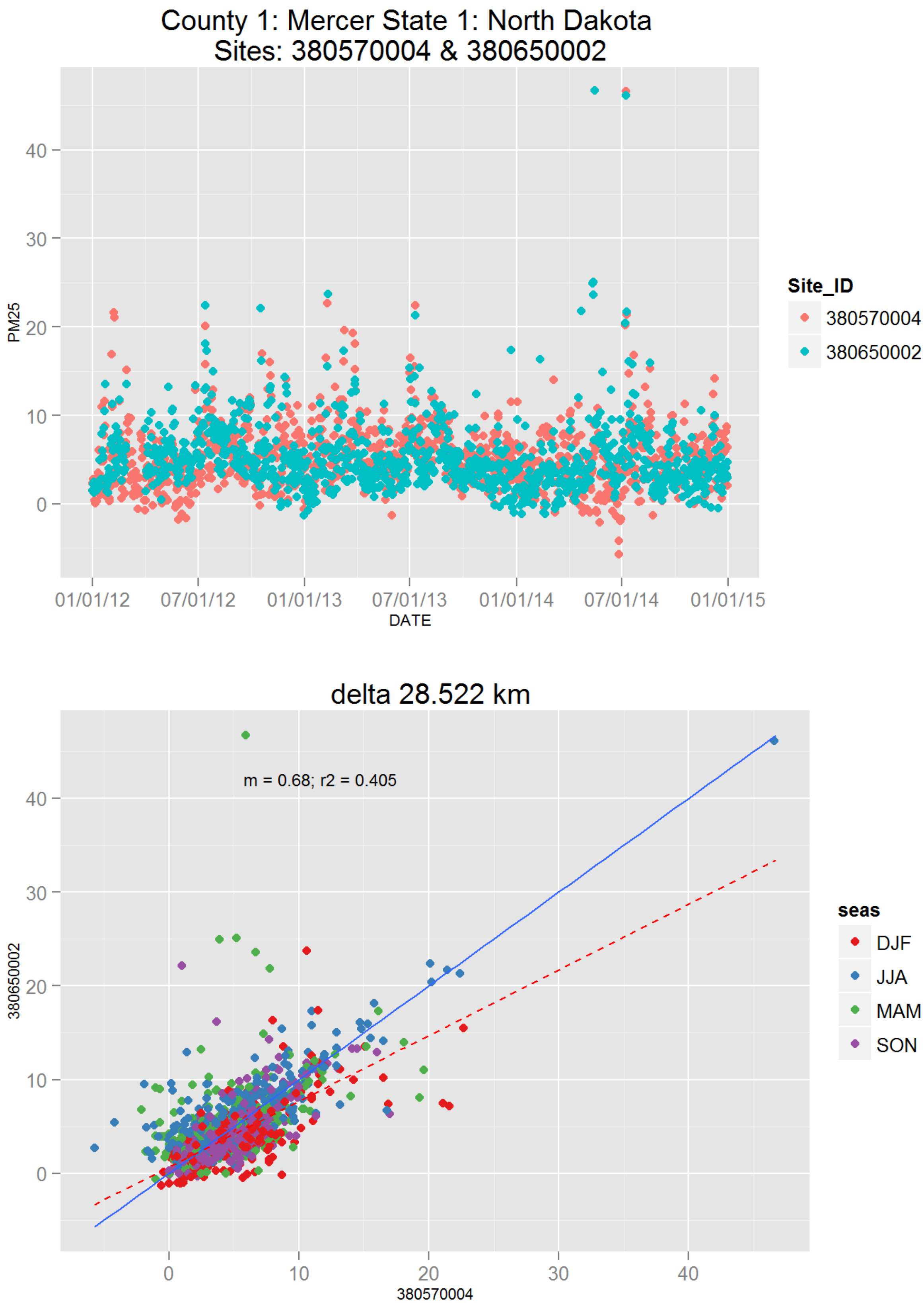
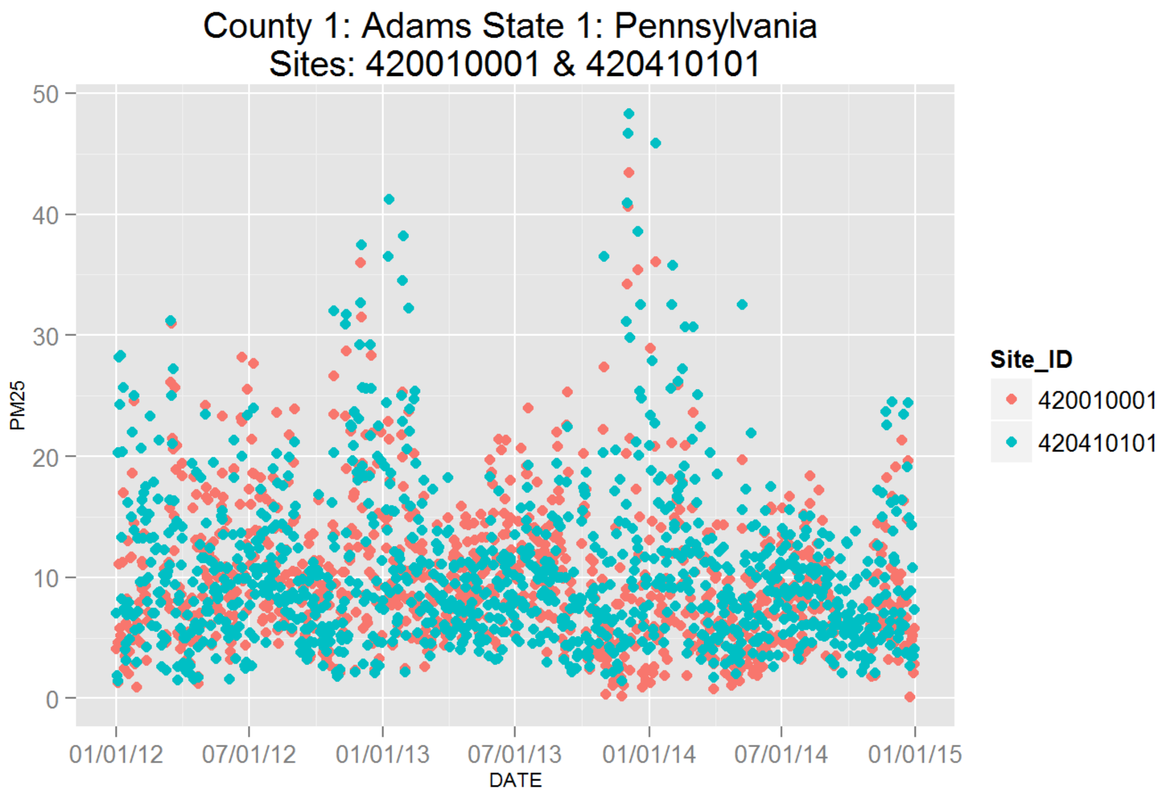


Figure 142: Comparison of $PM_{2.5}$ data for sites 380570004 and 380570004. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.



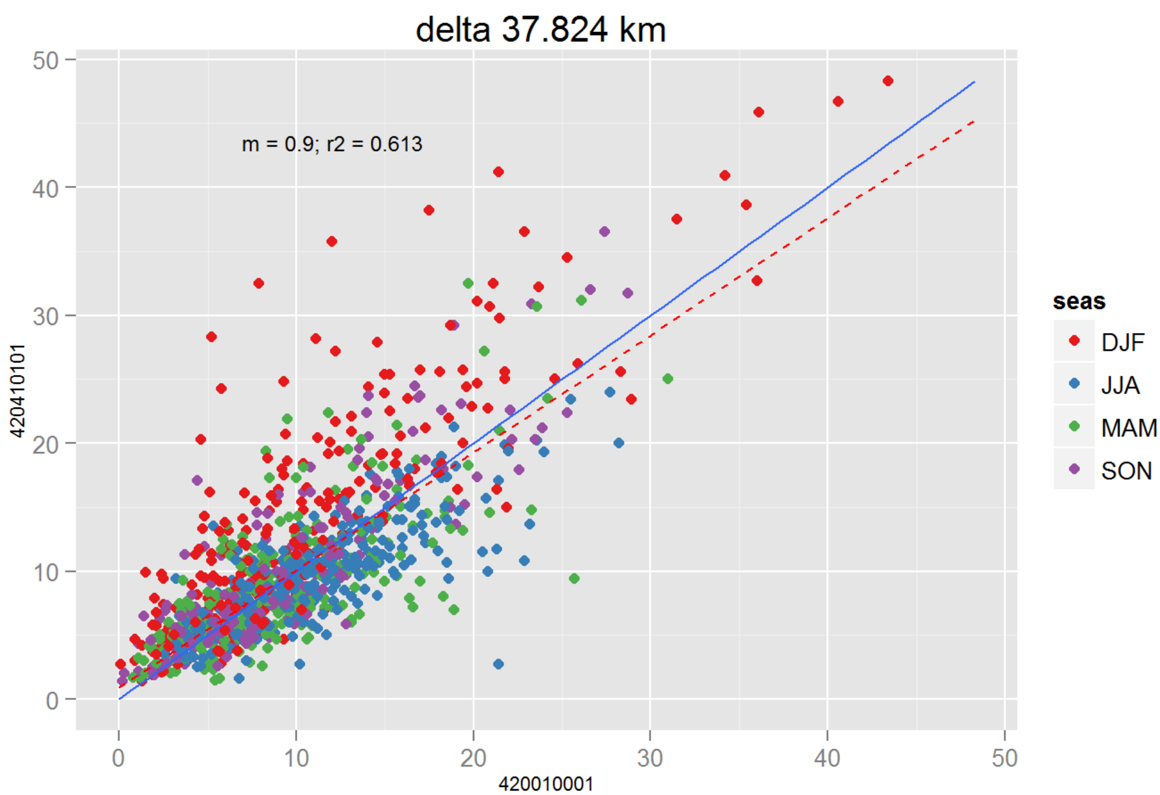
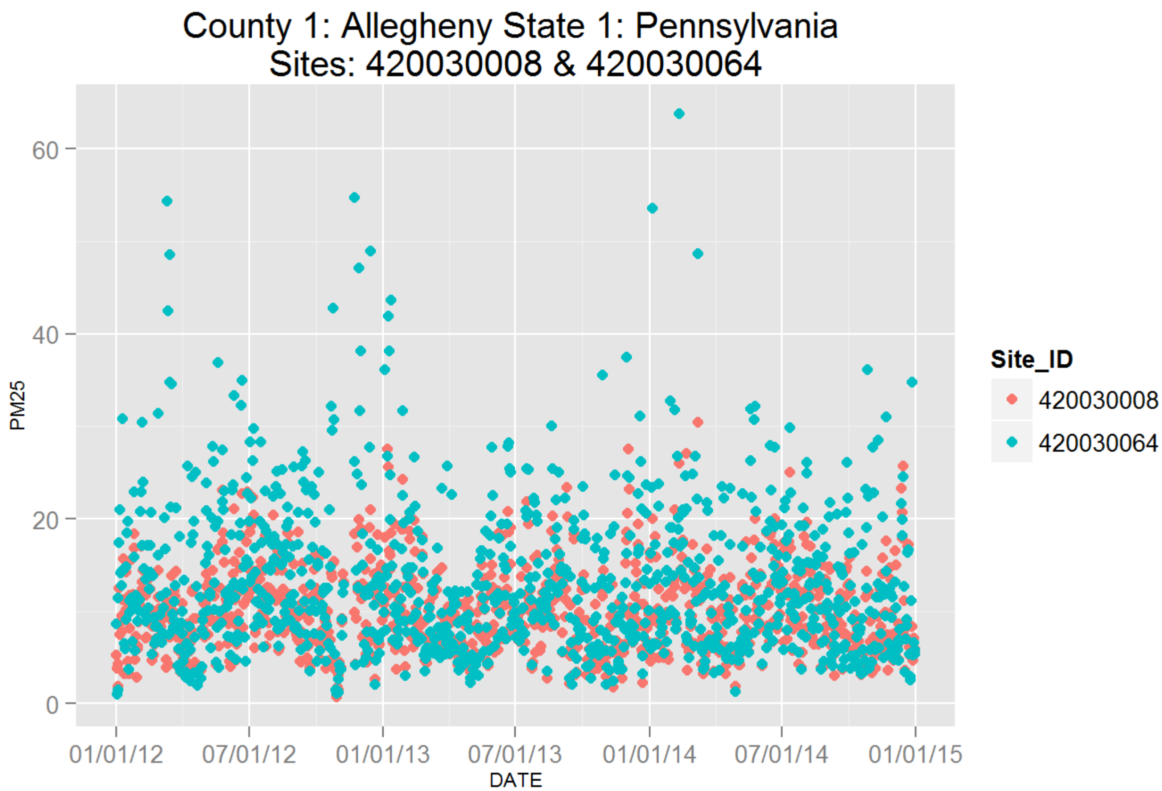


Figure 143: Comparison of $PM_{2.5}$ data for sites 420010001 and 420010001. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.



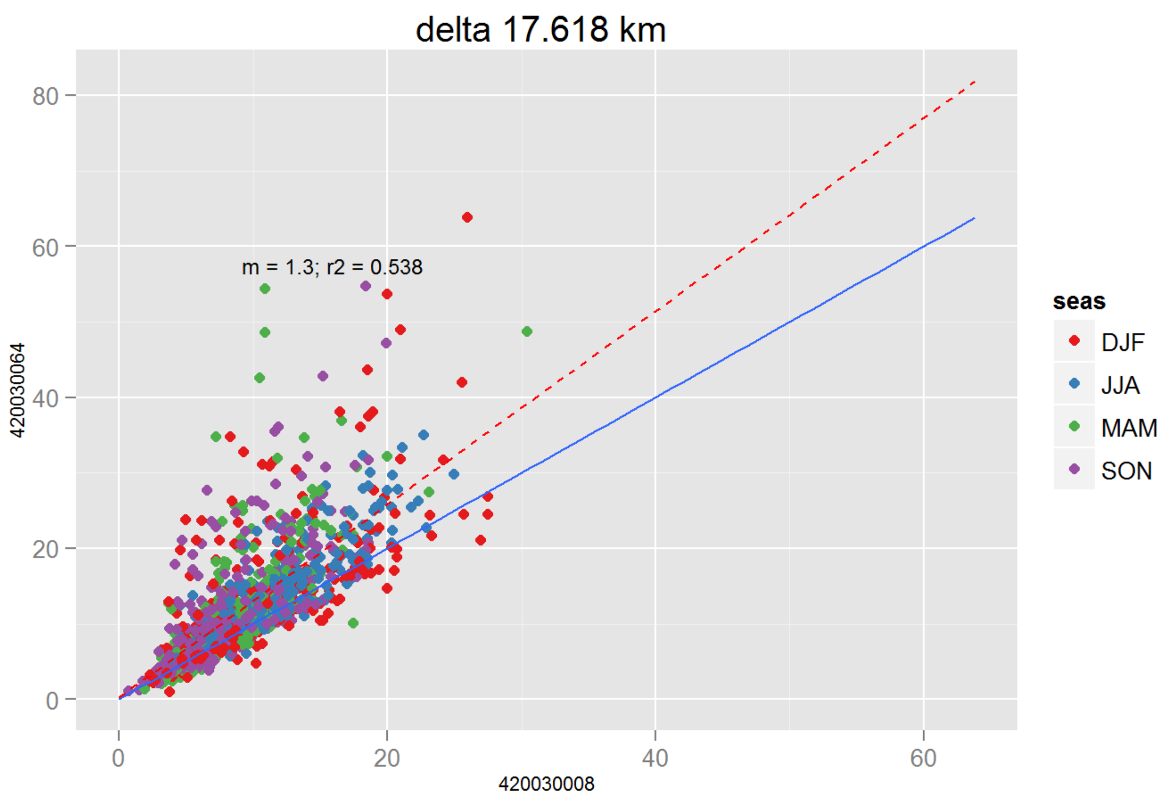


Figure 144: Comparison of $PM_{2.5}$ data for sites 420030008 and 420030008. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

Figure 145: Comparison of $PM_{2.5}$ data for sites 420070014 and 420070014. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

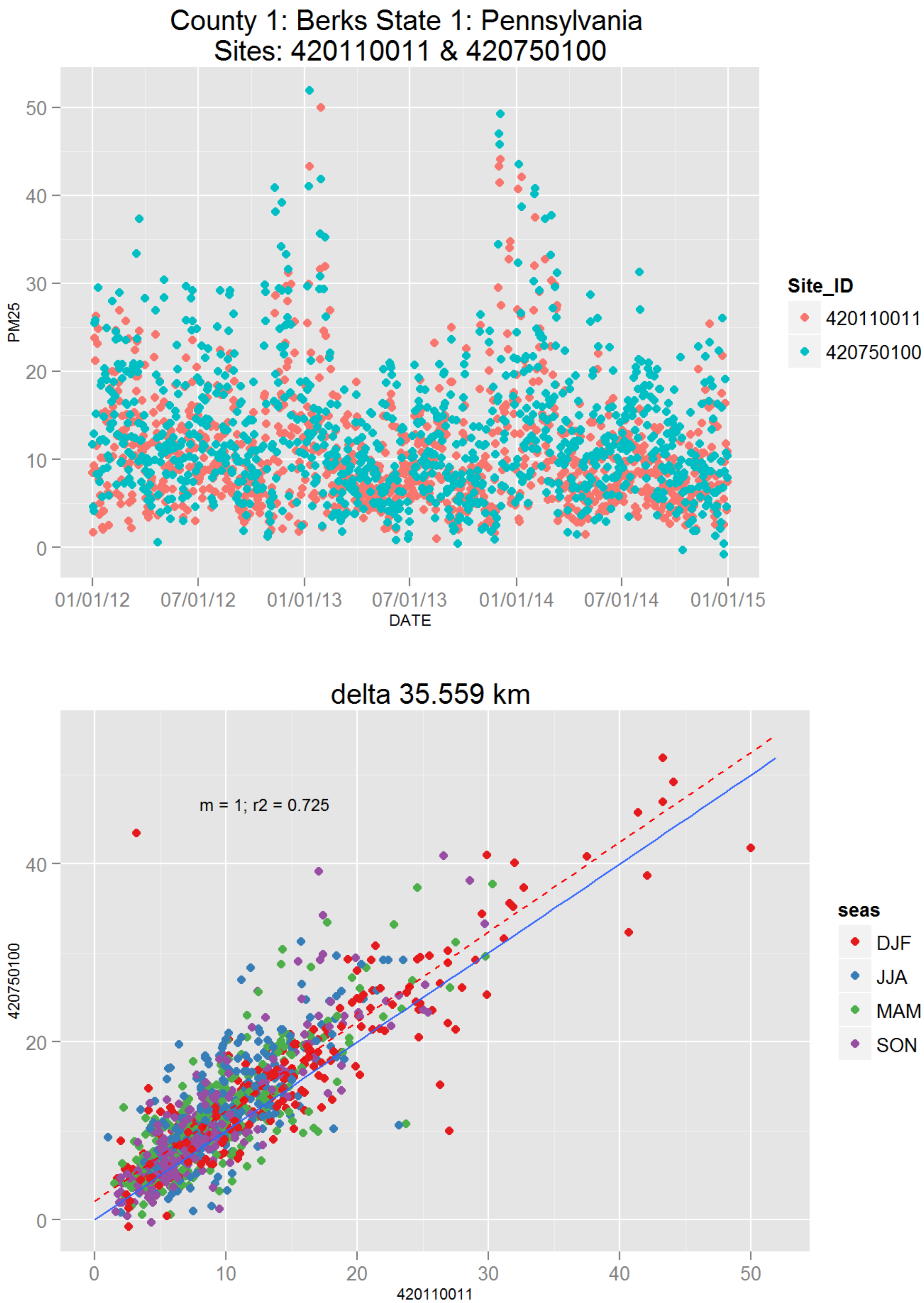
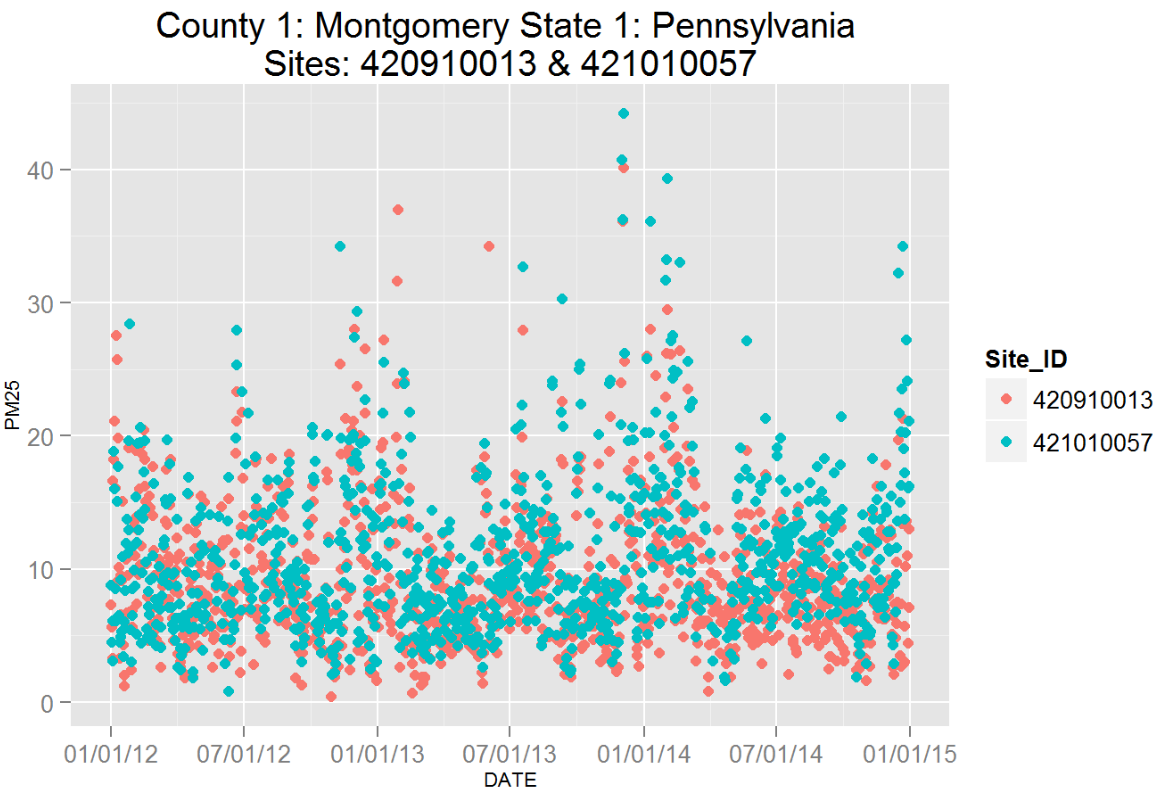


Figure 146: Comparison of $PM_{2.5}$ data for sites 420110011 and 420110011. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

Figure 147: Comparison of $PM_{2.5}$ data for sites 420410101 and 420410101. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

Figure 148: Comparison of $PM_{2.5}$ data for sites 420450002 and 420450002. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

Figure 149: Comparison of $PM_{2.5}$ data for sites 420710007 and 420710007. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.



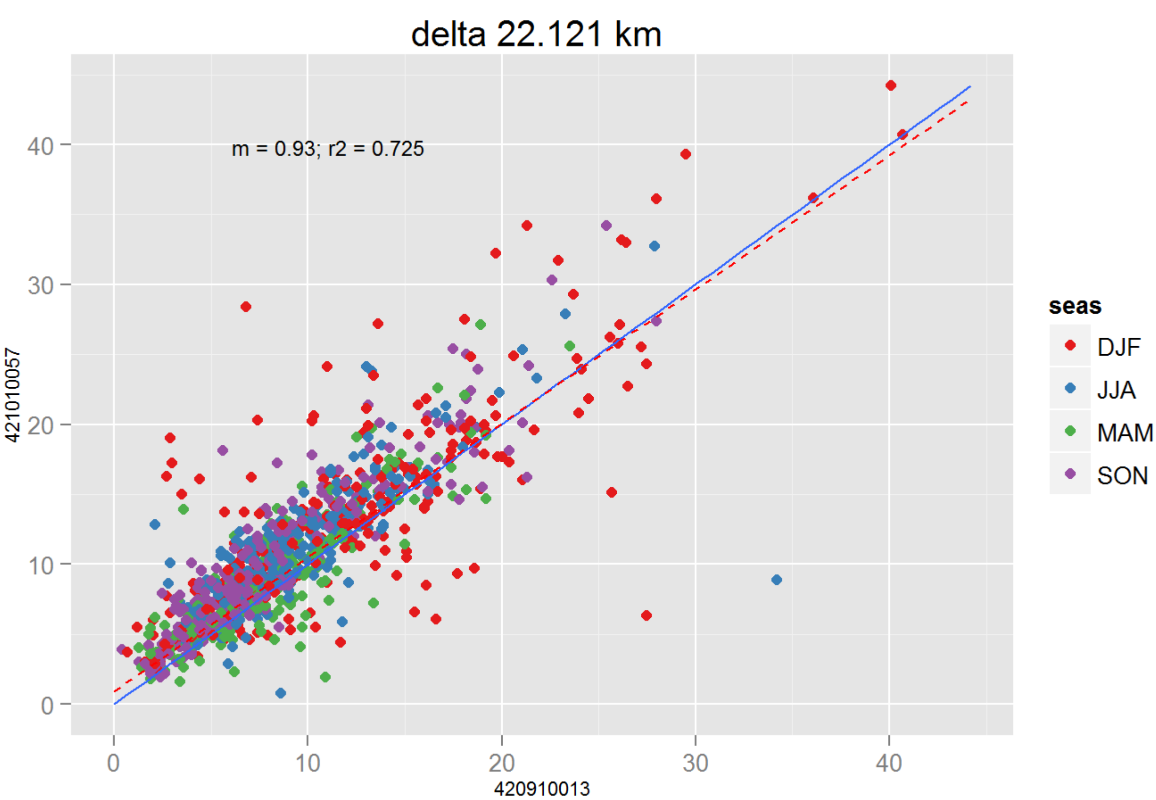


Figure 150: Comparison of $PM_{2.5}$ data for sites 420910013 and 420910013. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

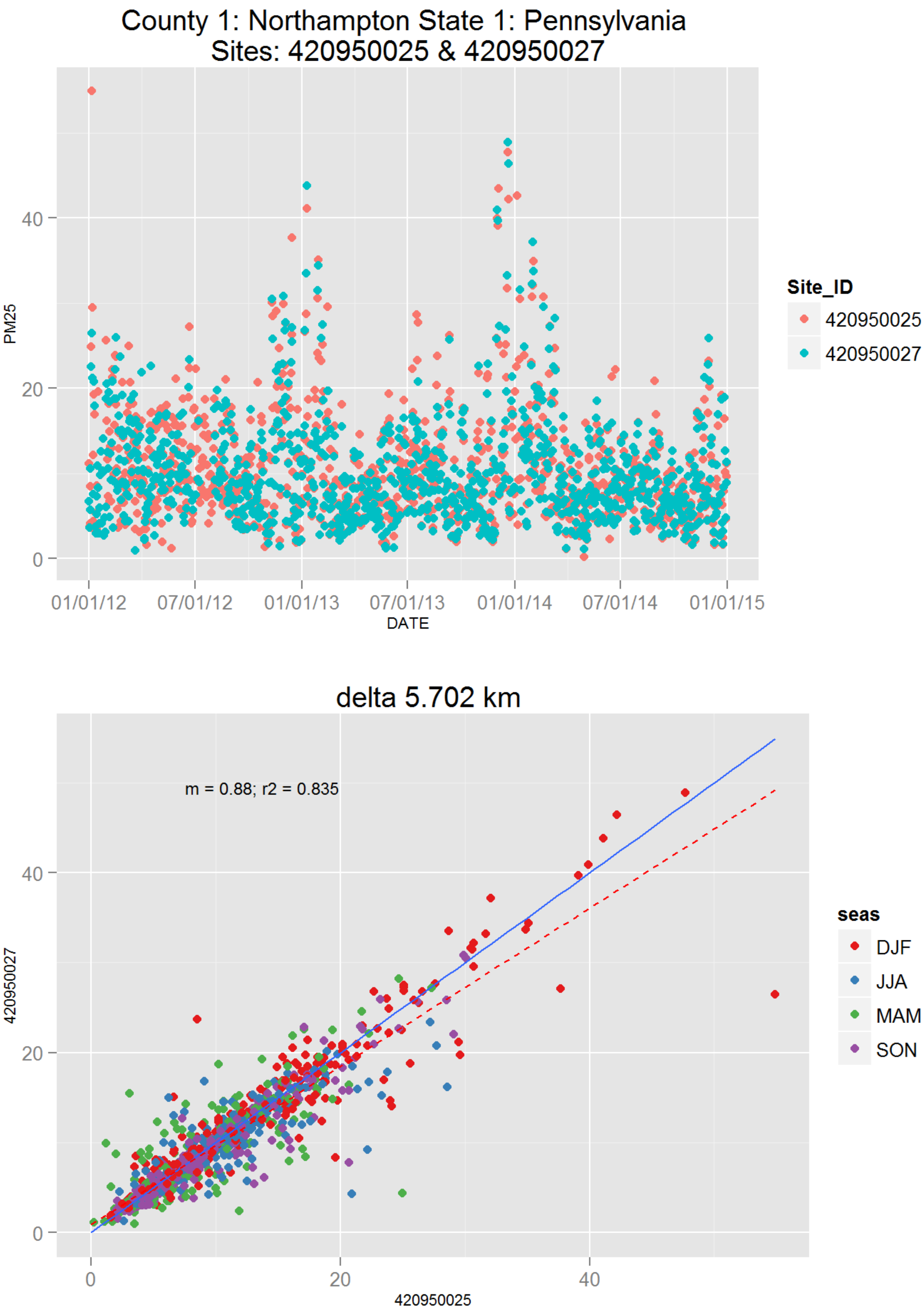
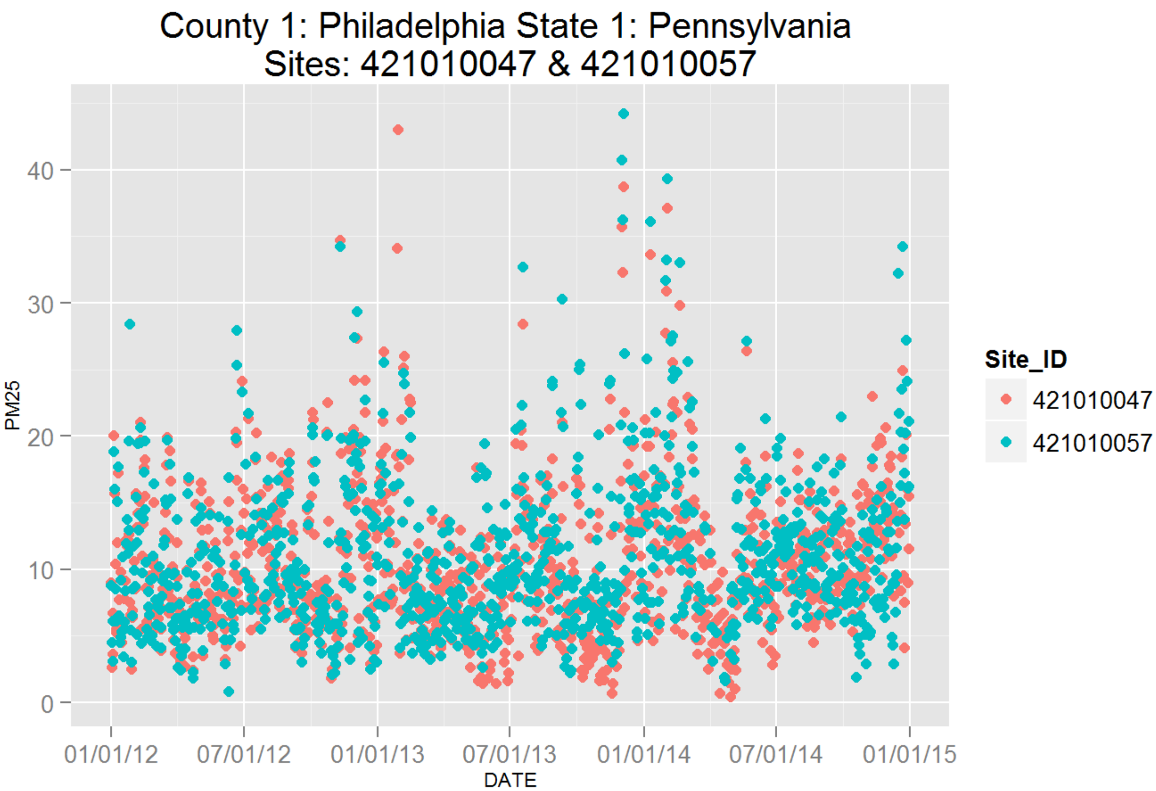


Figure 151: Comparison of $PM_{2.5}$ data for sites 420950025 and 420950025. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.



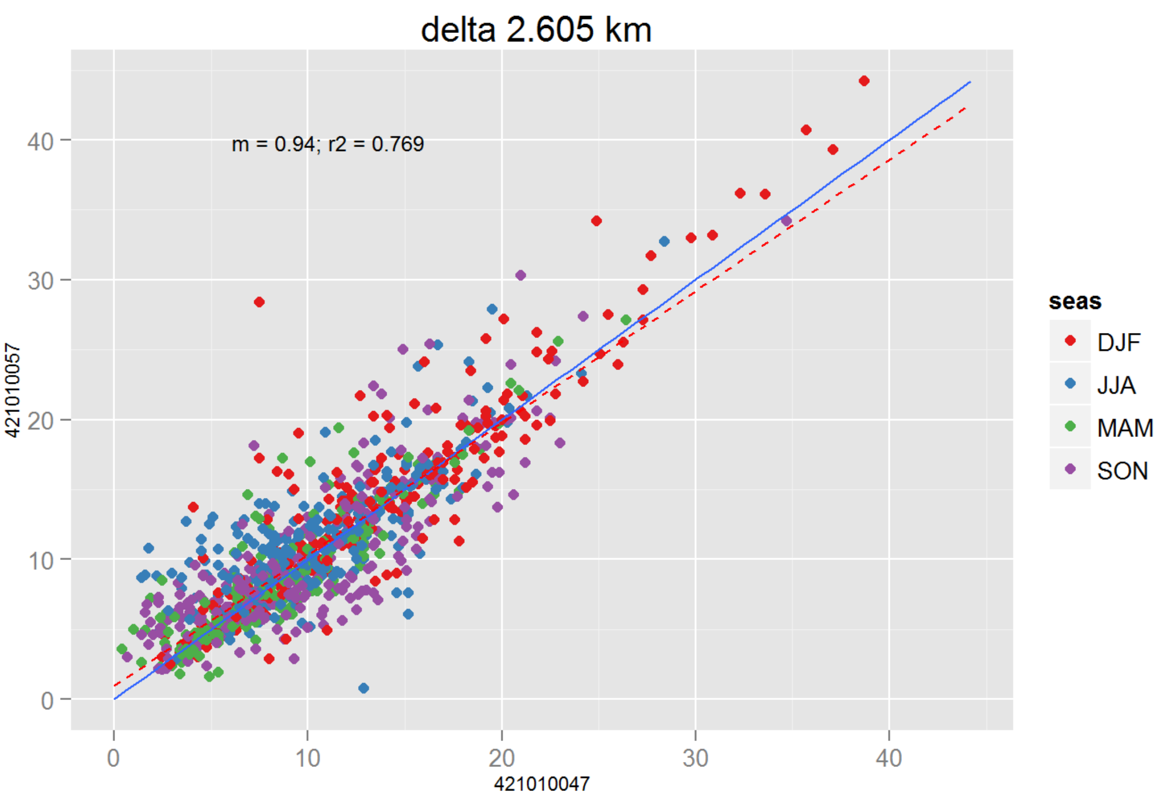


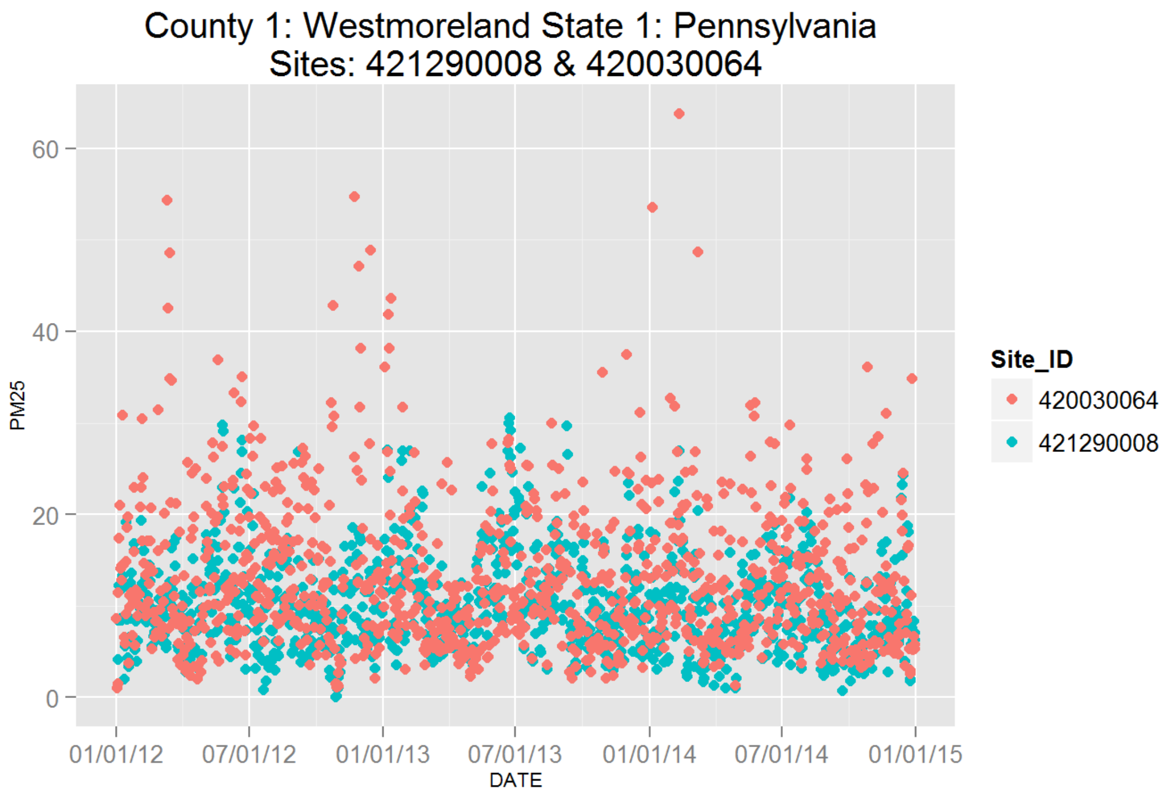
Figure 152: Comparison of $PM_{2.5}$ data for sites 421010047 and 421010047. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

Figure 153: Comparison of $PM_{2.5}$ data for sites 421010055 and 421010055. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

Figure 154: Comparison of $PM_{2.5}$ data for sites 421250005 and 421250005. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

Figure 155: Comparison of $PM_{2.5}$ data for sites 421250200 and 421250200. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

Figure 156: Comparison of $PM_{2.5}$ data for sites 421255001 and 421255001. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.



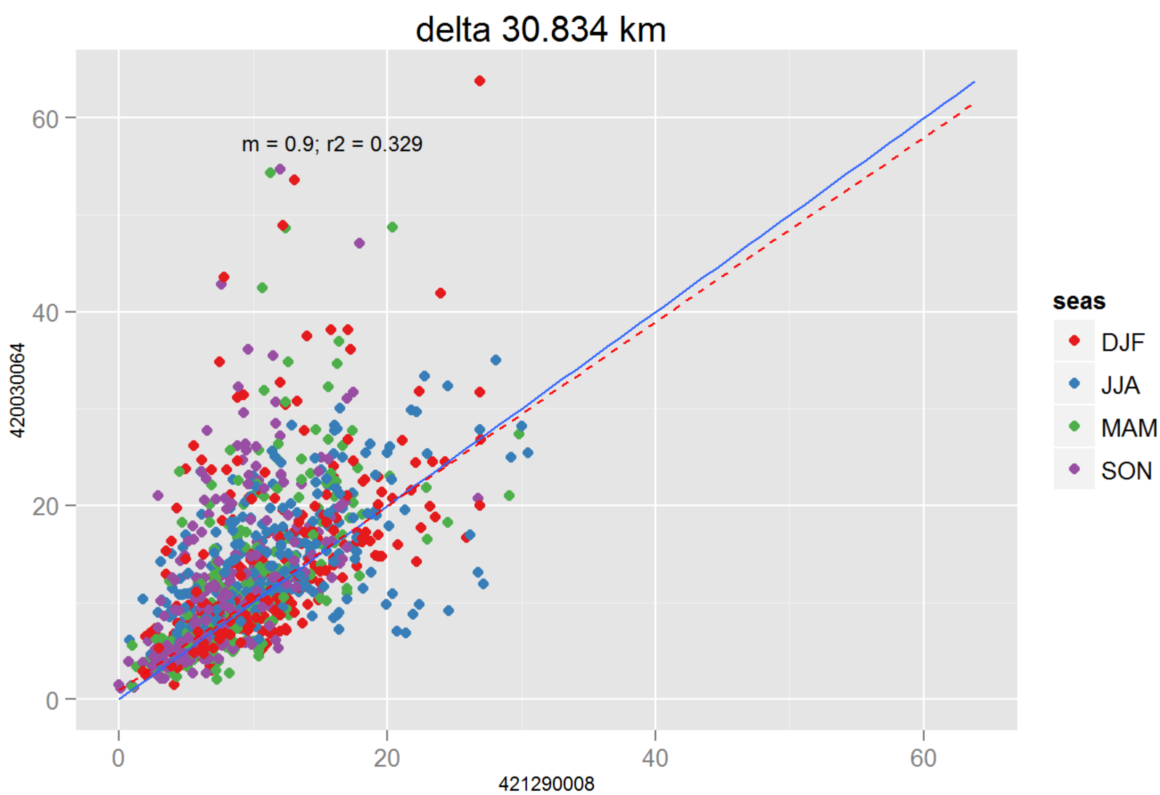
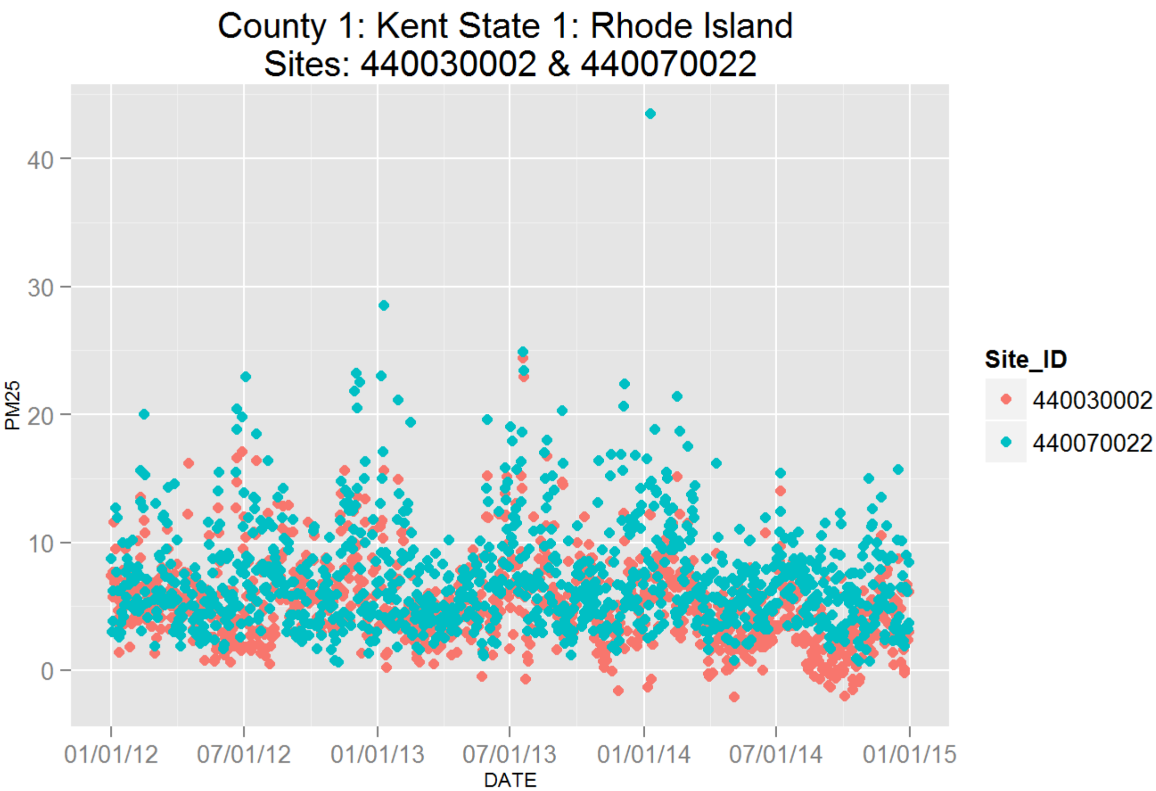


Figure 157: Comparison of $PM_{2.5}$ data for sites 421290008 and 421290008. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

Figure 158: Comparison of $PM_{2.5}$ data for sites 421330008 and 421330008. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.



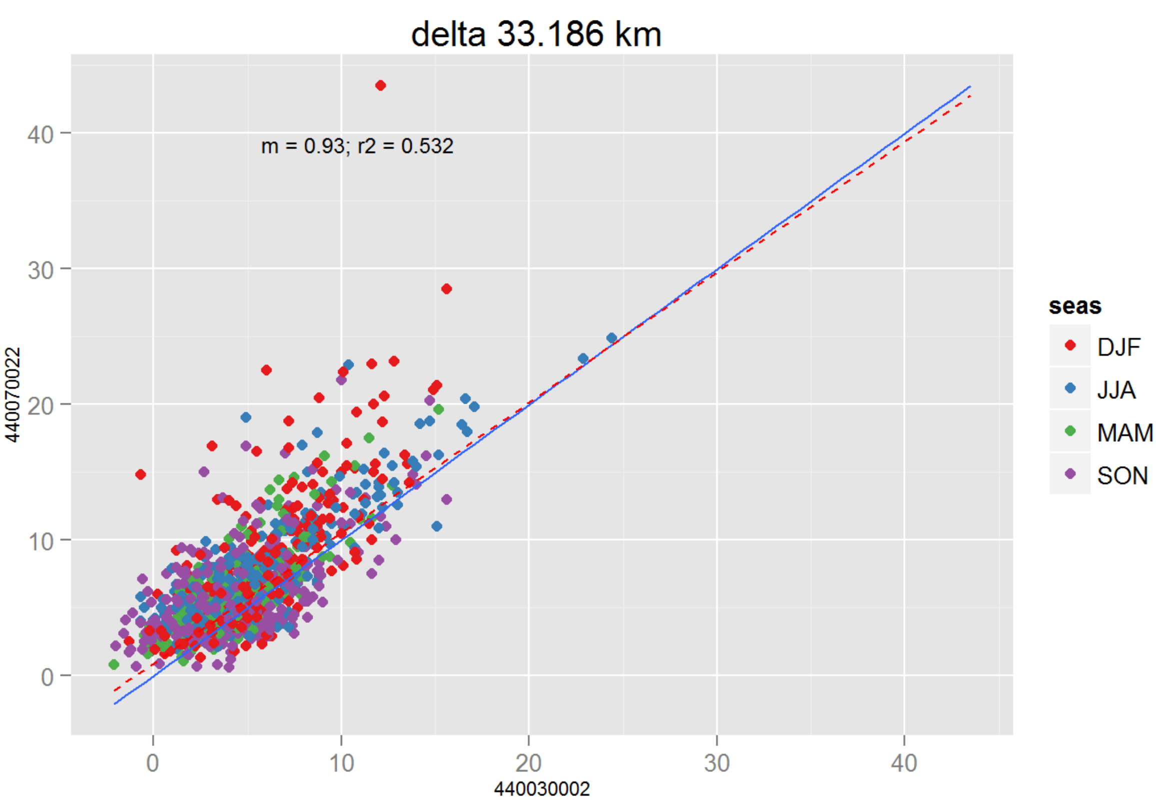
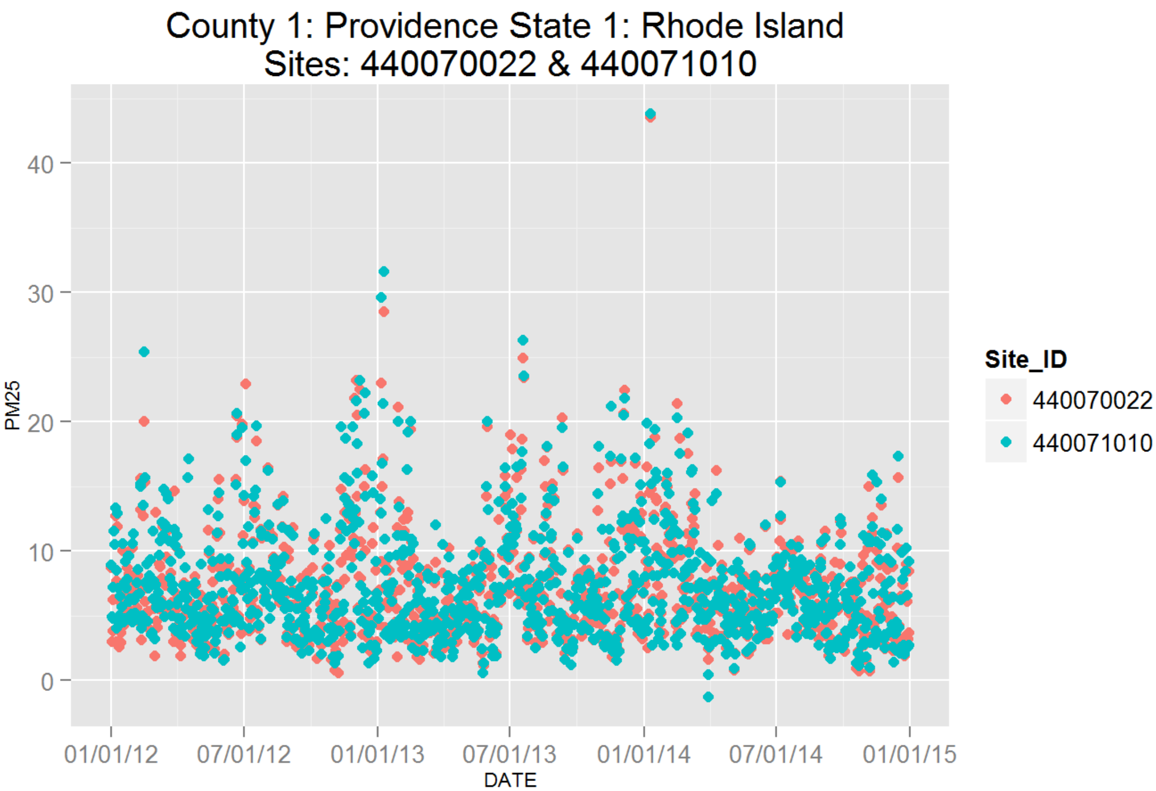


Figure 159: Comparison of $PM_{2.5}$ data for sites 440030002 and 440030002. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.



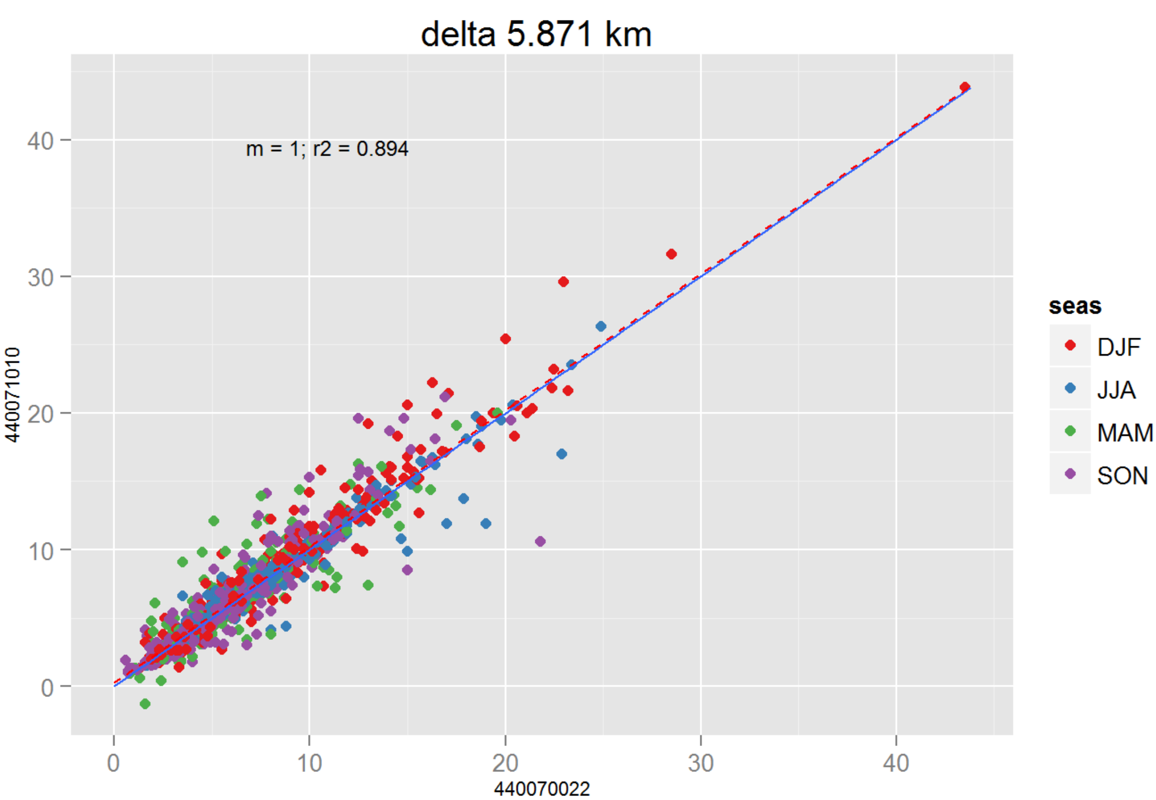


Figure 160: Comparison of $PM_{2.5}$ data for sites 440070022 and 440070022. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

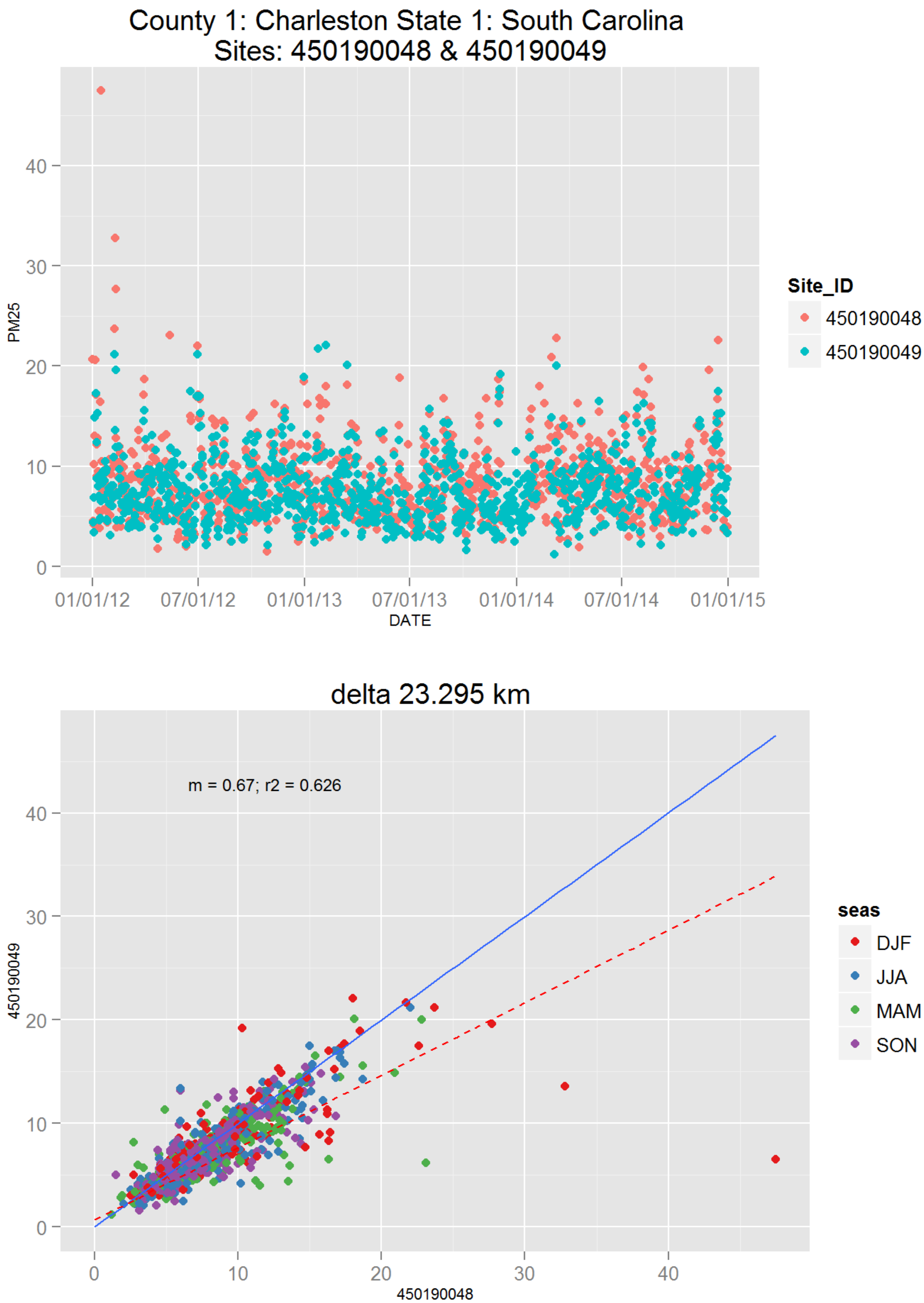
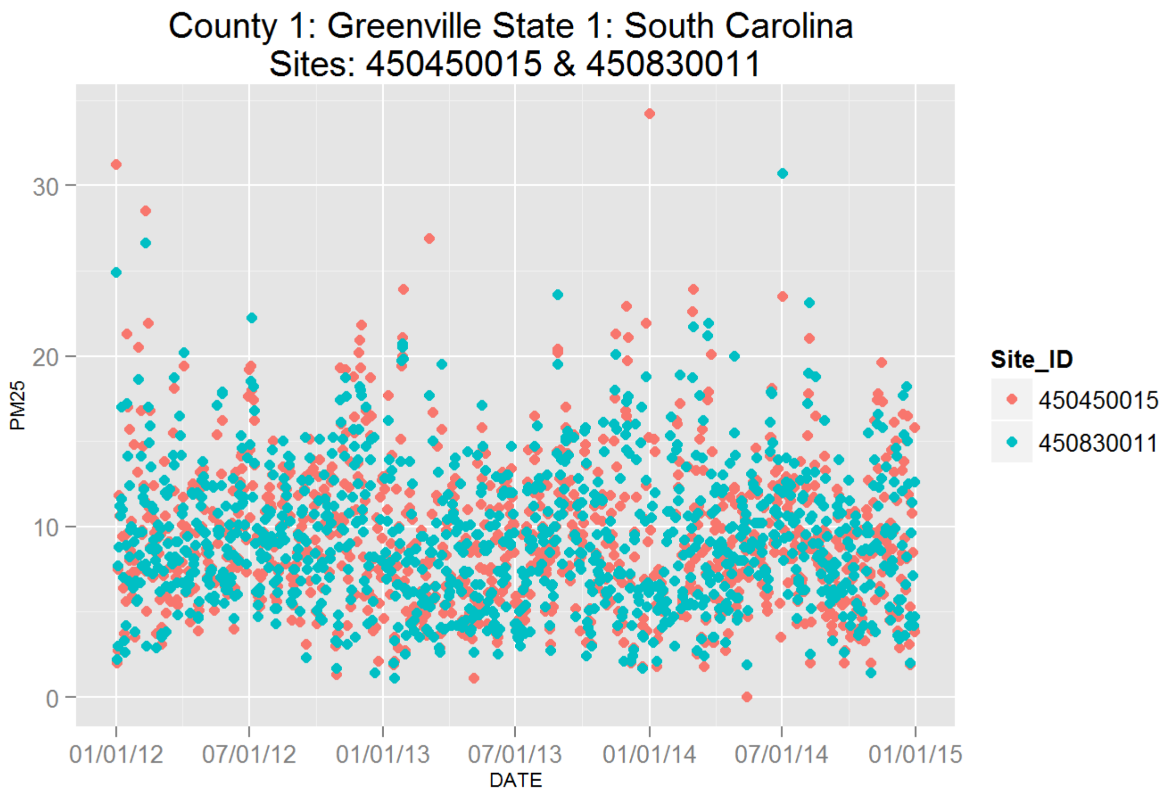


Figure 161: Comparison of $PM_{2.5}$ data for sites 450190048 and 450190048. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.



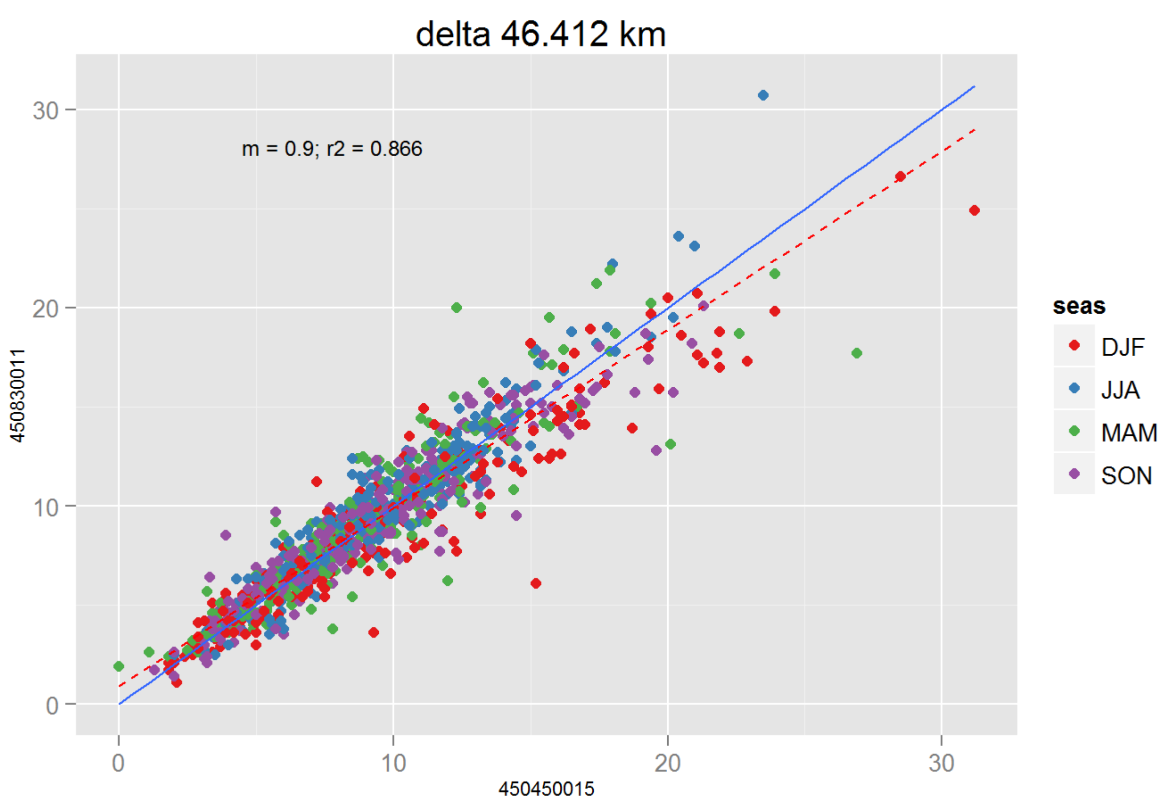
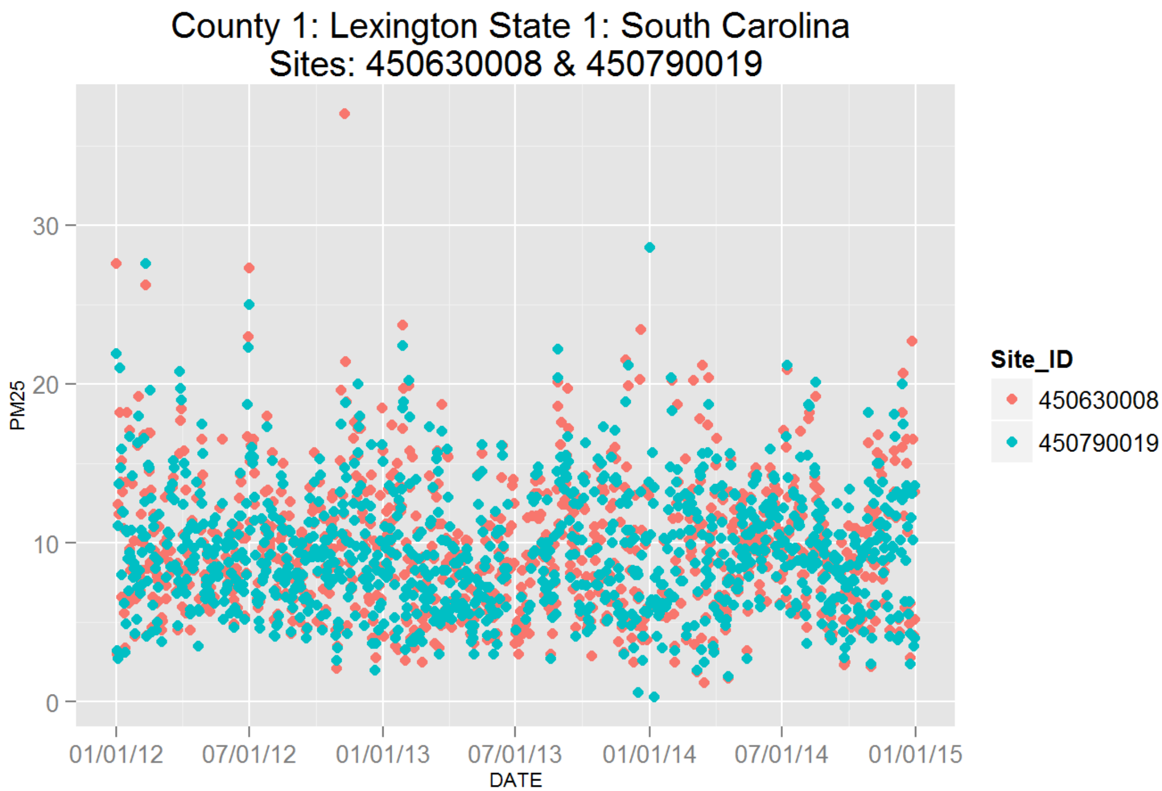


Figure 162: Comparison of $PM_{2.5}$ data for sites 450450015 and 450450015. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.



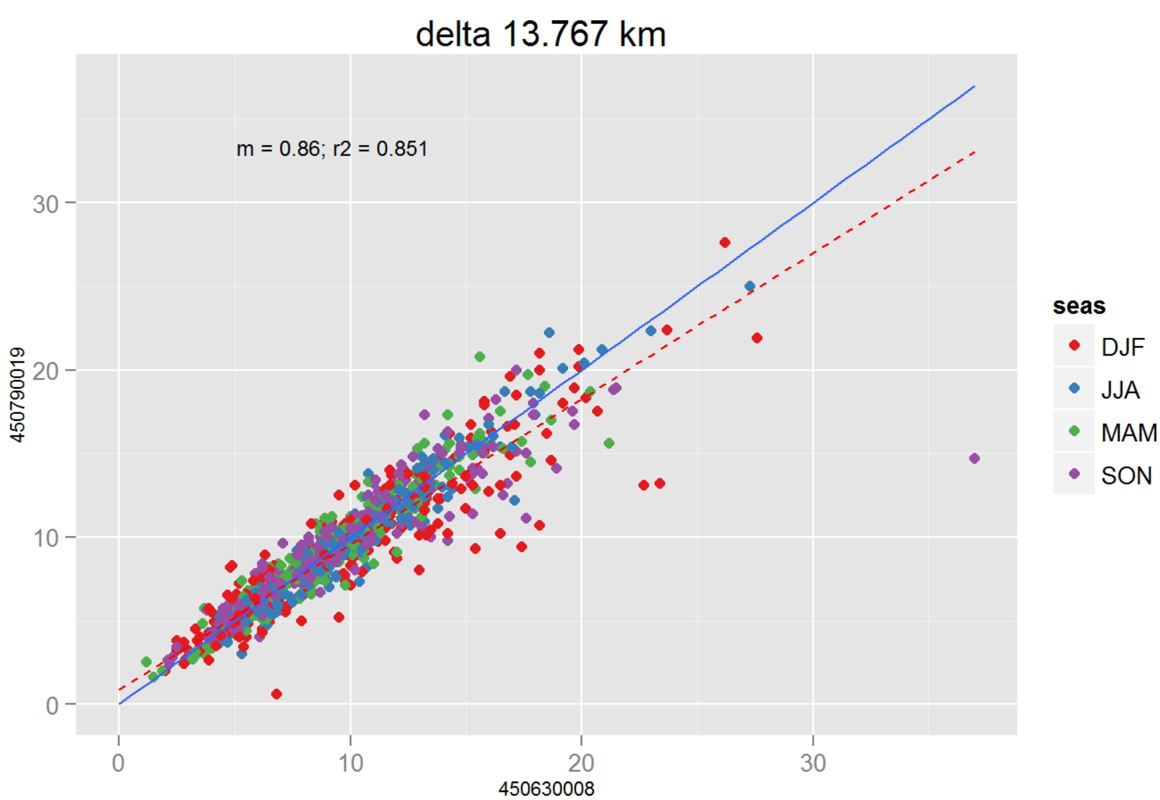


Figure 163: Comparison of $PM_{2.5}$ data for sites 450630008 and 450630008. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

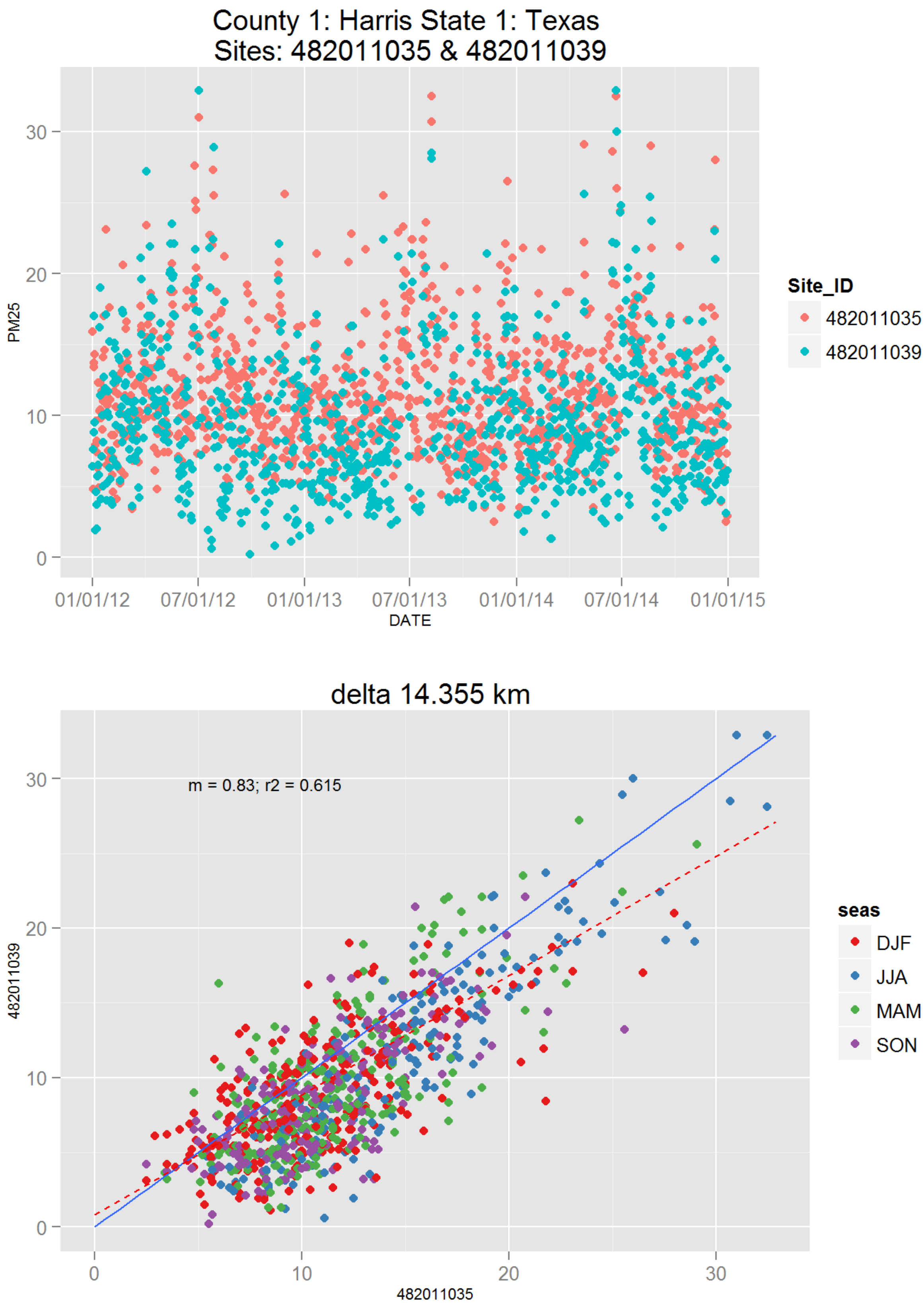


Figure 164: Comparison of $PM_{2.5}$ data for sites 482011035 and 482011035. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with the data points are colored by month.

Figure 165: Comparison of $PM_{2.5}$ data for sites 490353006 and 490353006. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

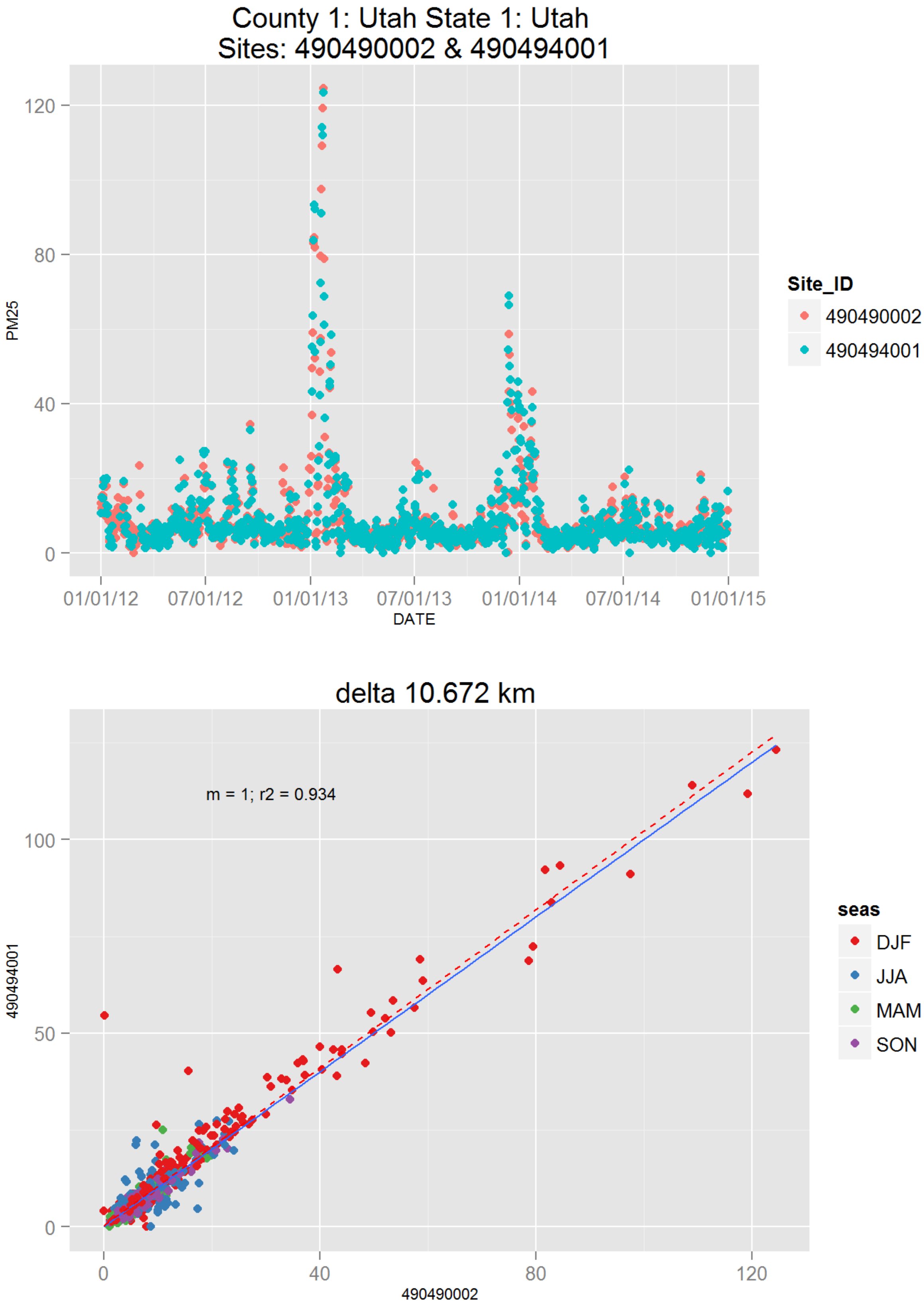


Figure 166: Comparison of $PM_{2.5}$ data for sites 490490002 and 490490002. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

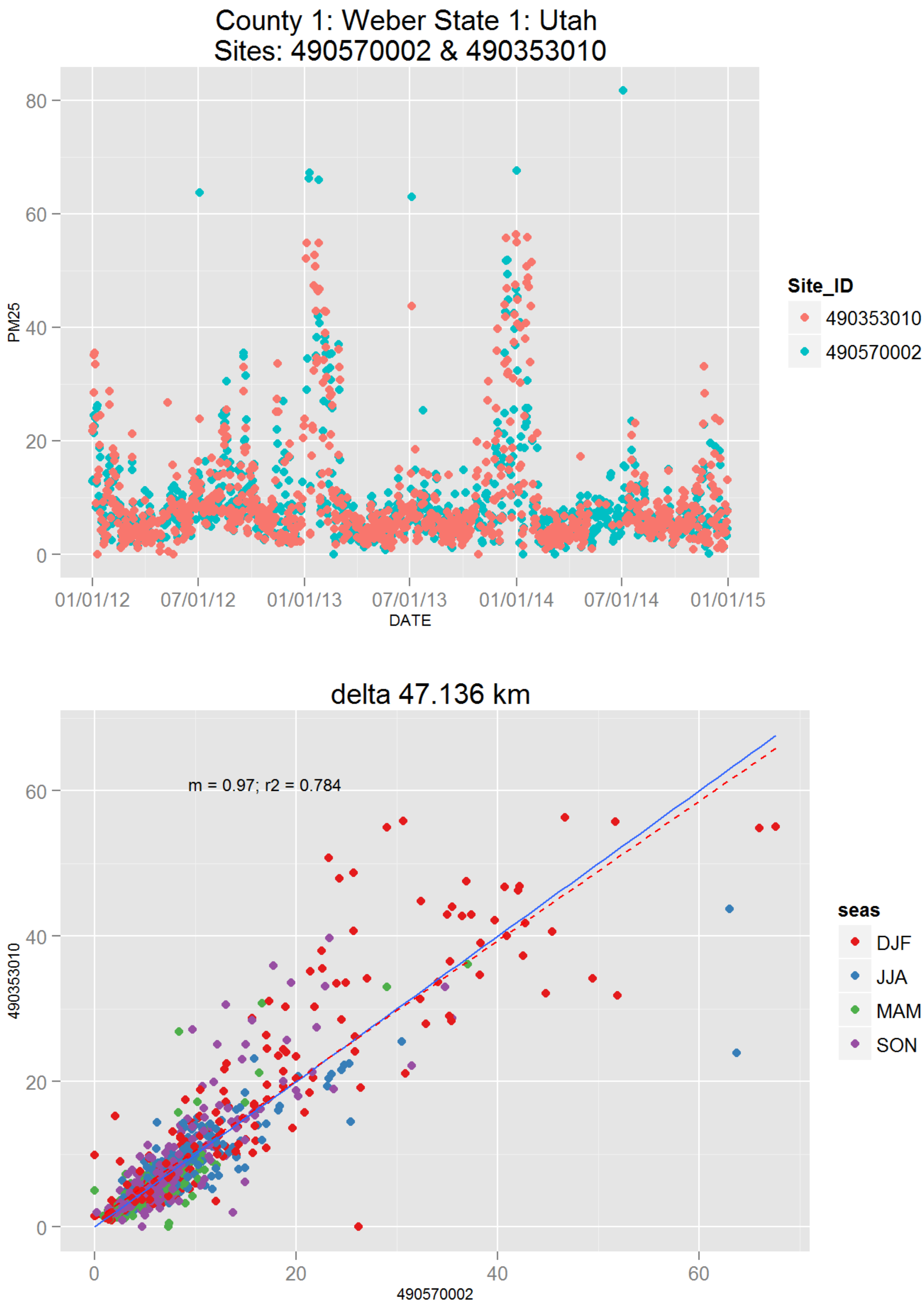
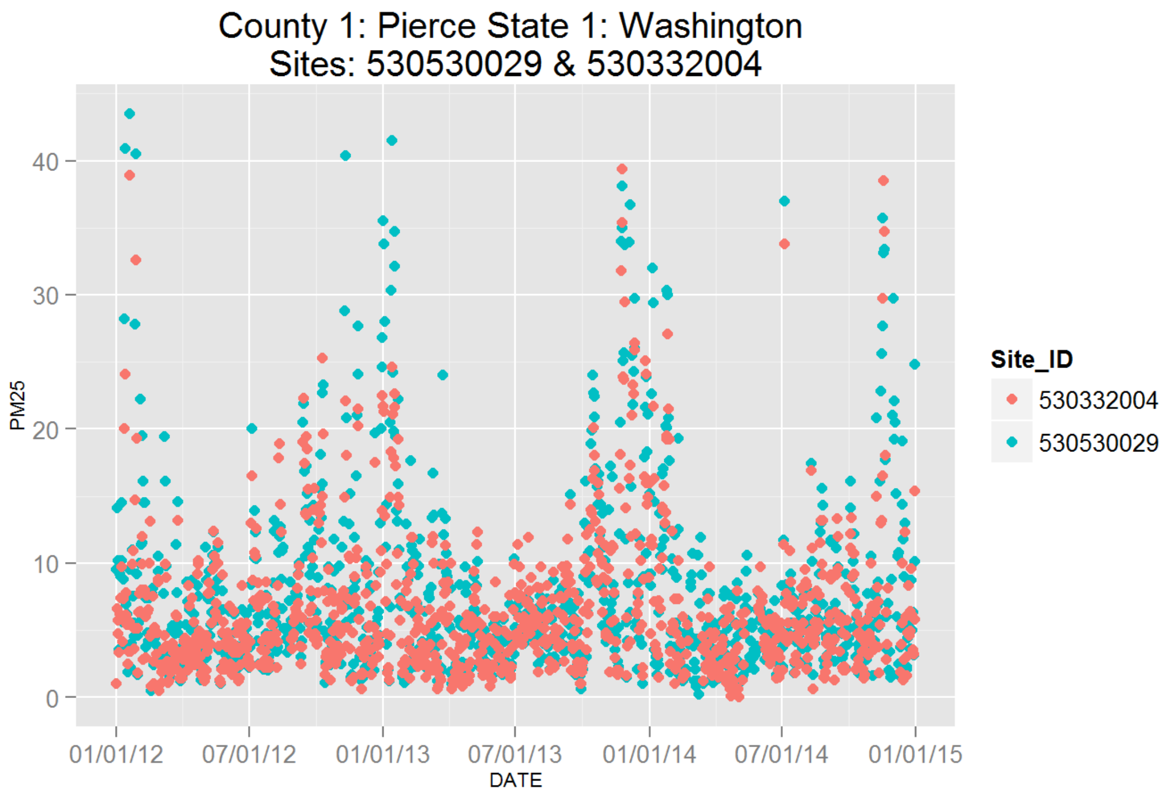


Figure 167: Comparison of $PM_{2.5}$ data for sites 490570002 and 490570002. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.



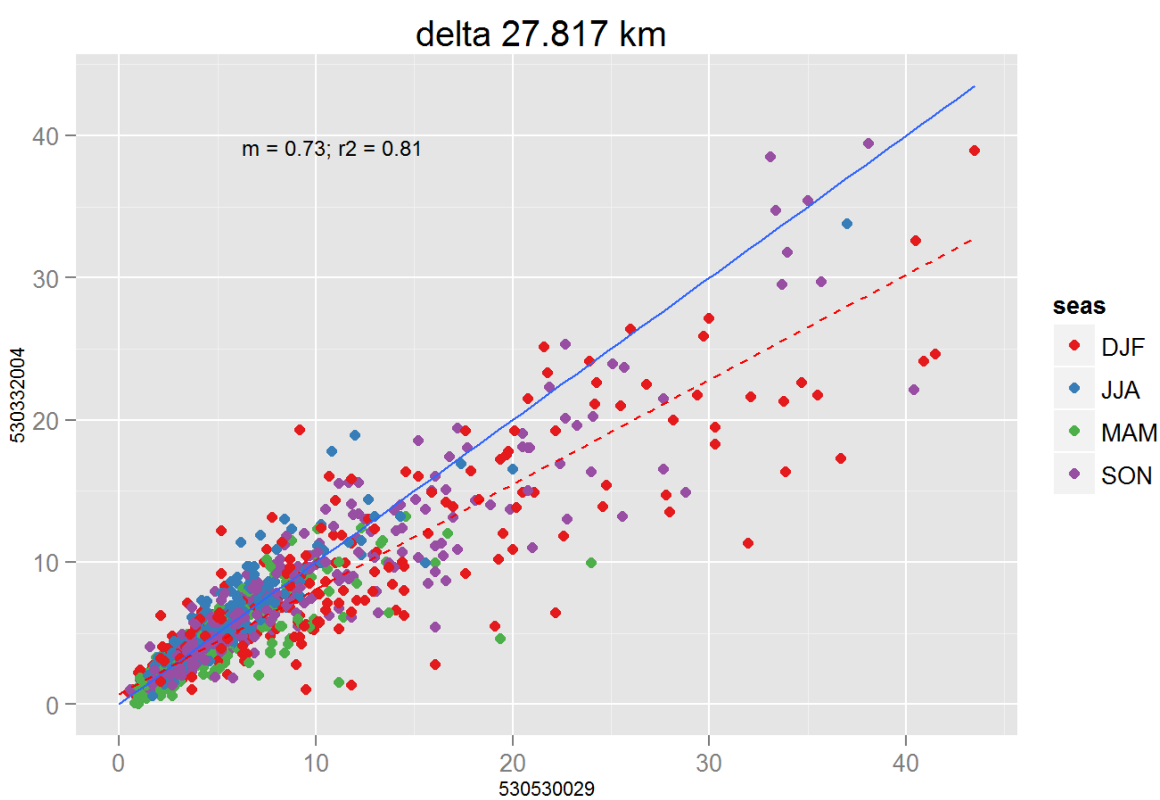
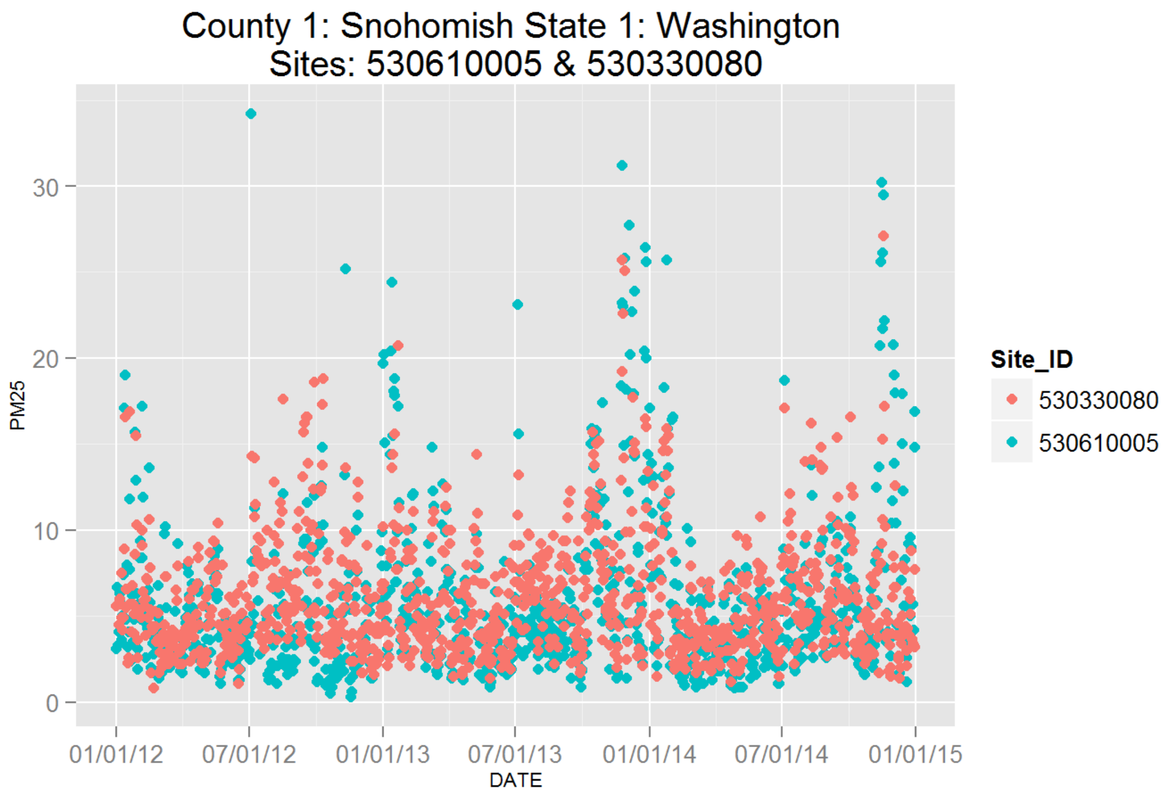


Figure 168: Comparison of $PM_{2.5}$ data for sites 530530029 and 530530029. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.



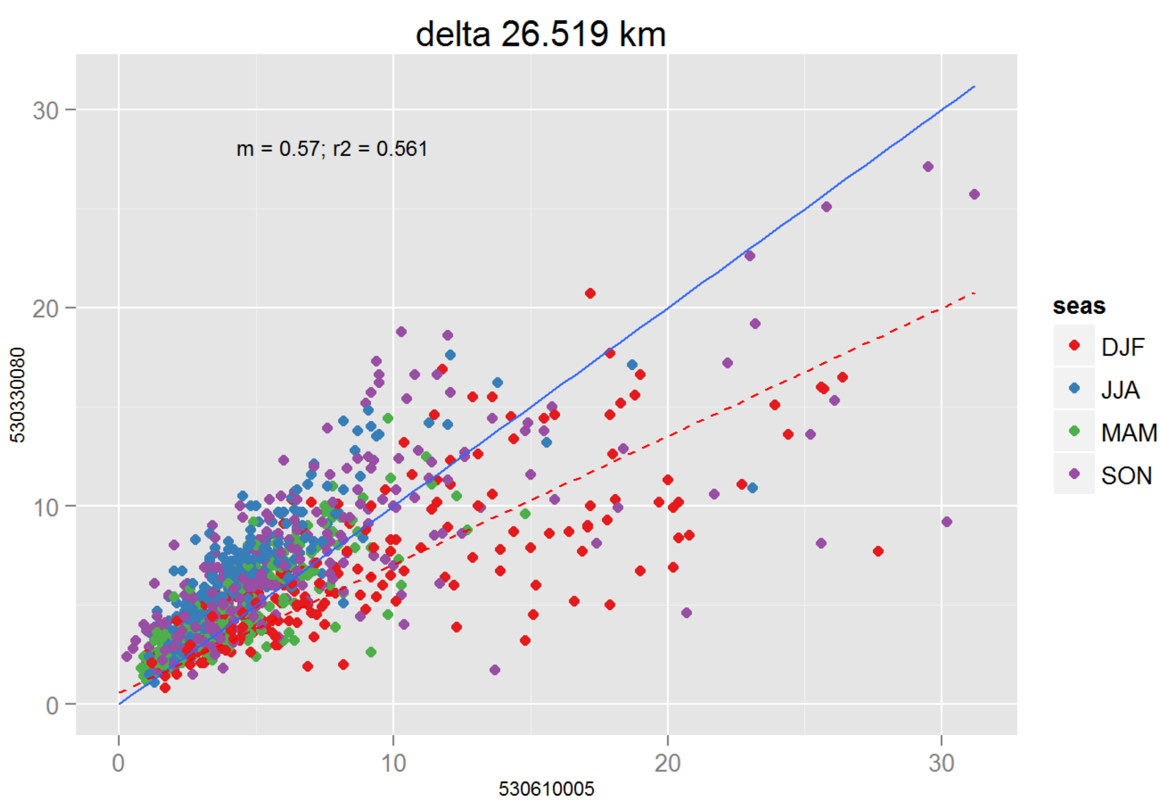


Figure 169: Comparison of $PM_{2.5}$ data for sites 530610005 and 530610005. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

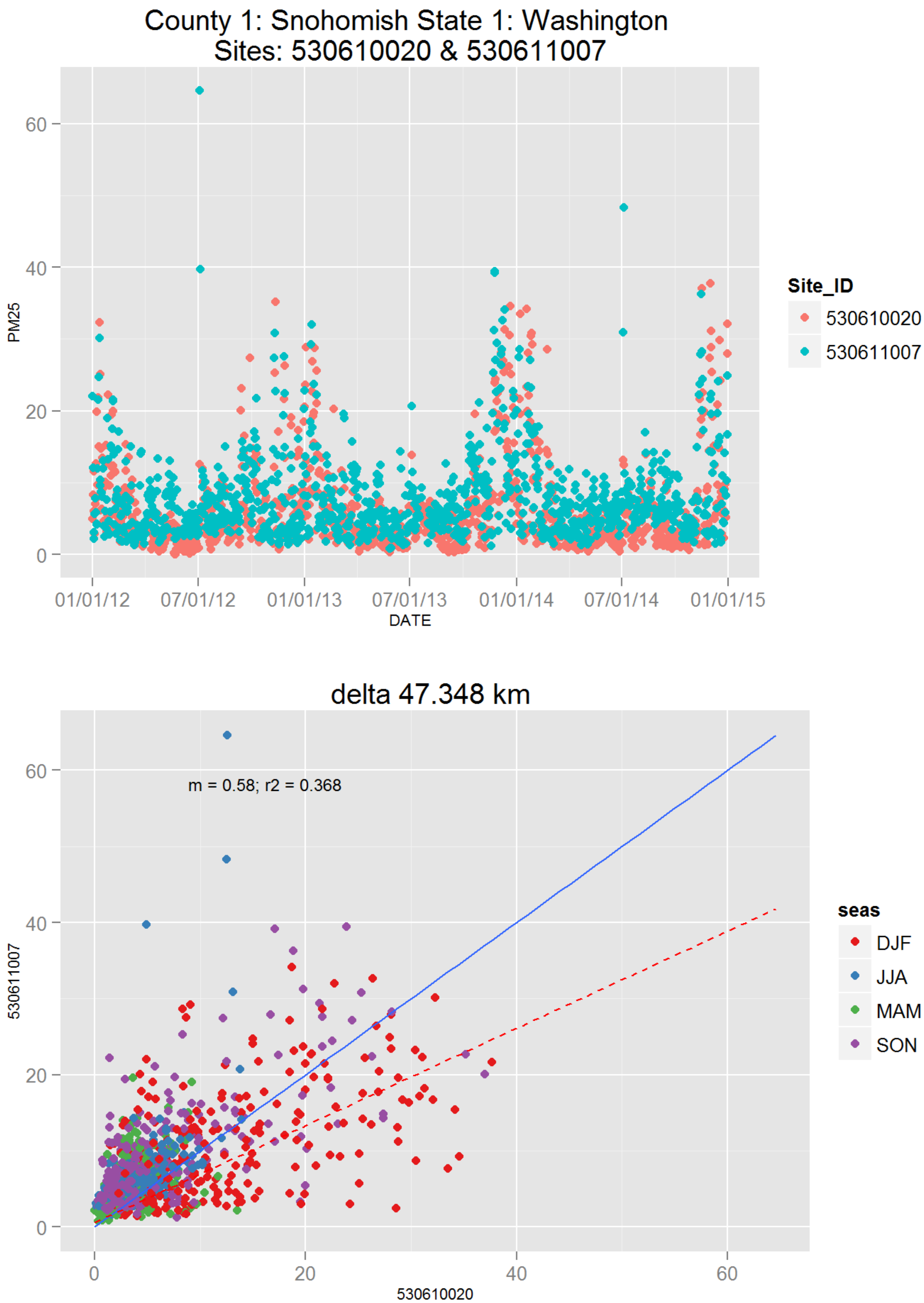
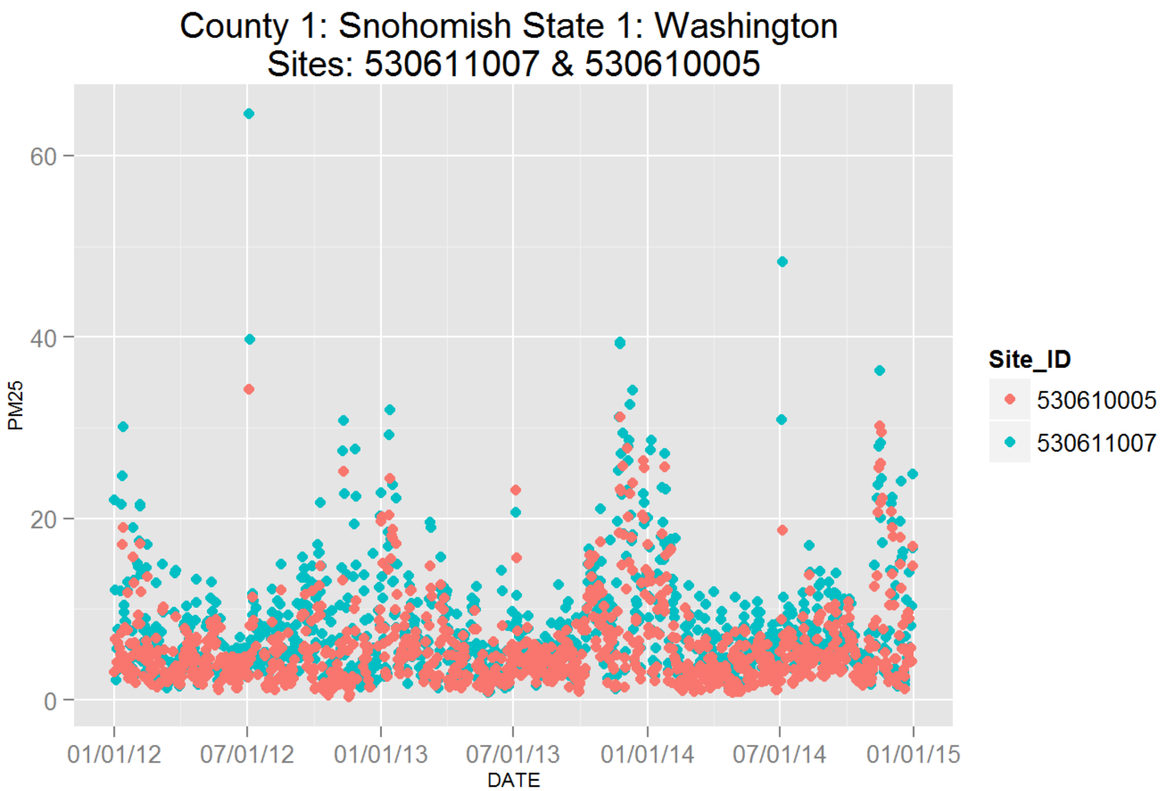


Figure 170: Comparison of $PM_{2.5}$ data for sites 530610020 and 530610020. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.



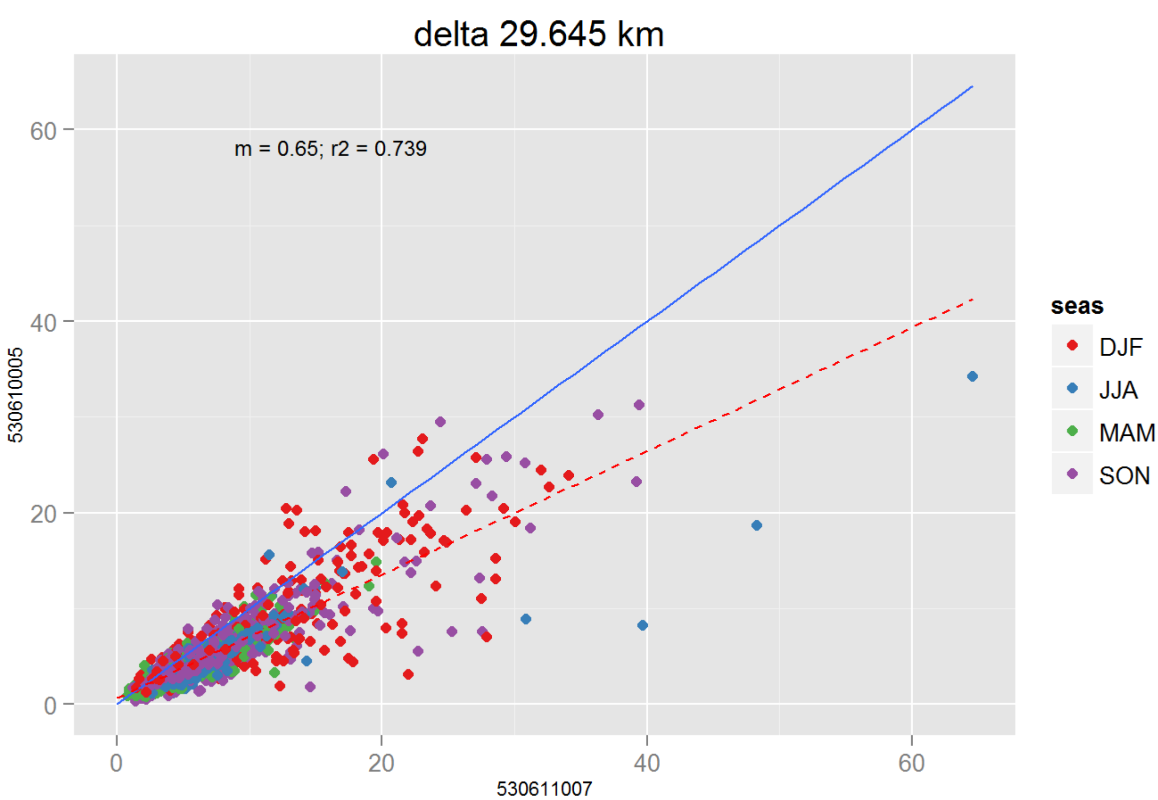


Figure 171: Comparison of $PM_{2.5}$ data for sites 530611007 and 530611007. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

Figure 172: Comparison of $PM_{2.5}$ data for sites 550090005 and 550090005. Top panel shows time series for both sites for years 2012-2014. Bottom panel shows scatter plot of paired data, along with slope for the linear regression and correlation coefficient (r2), with data points are colored by month.

5 Comparison of air quality variability for ozone sensitivity tests

Results from the ozone sensitivity analysis discussed in Section 2.2.3.

5.1 All available data, no quarterly subsets

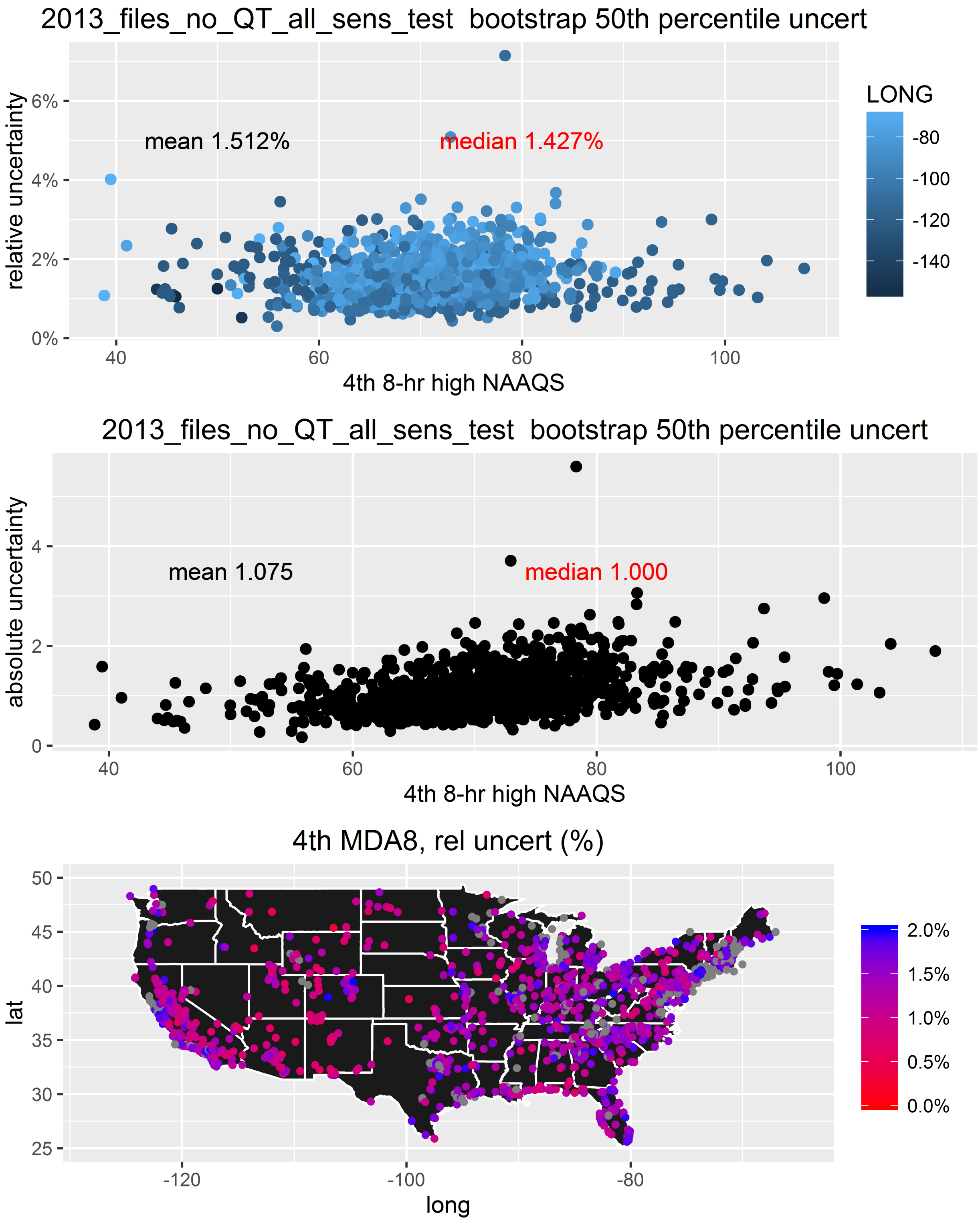


Figure 173: Bootstrap results for the ozone 2013 DVs, showing the 50%, 68%, 75%, and 95% CIs, along with the mean and median bootstrap DVs. The top panel shows the DVs at the various CIs, the middle panel shows the relative difference between the CI and the actual DV, and the bottom panel shows the distribution of the relative differences between the CI and the actual DV.

5.2 All available data, with quarterly subsets

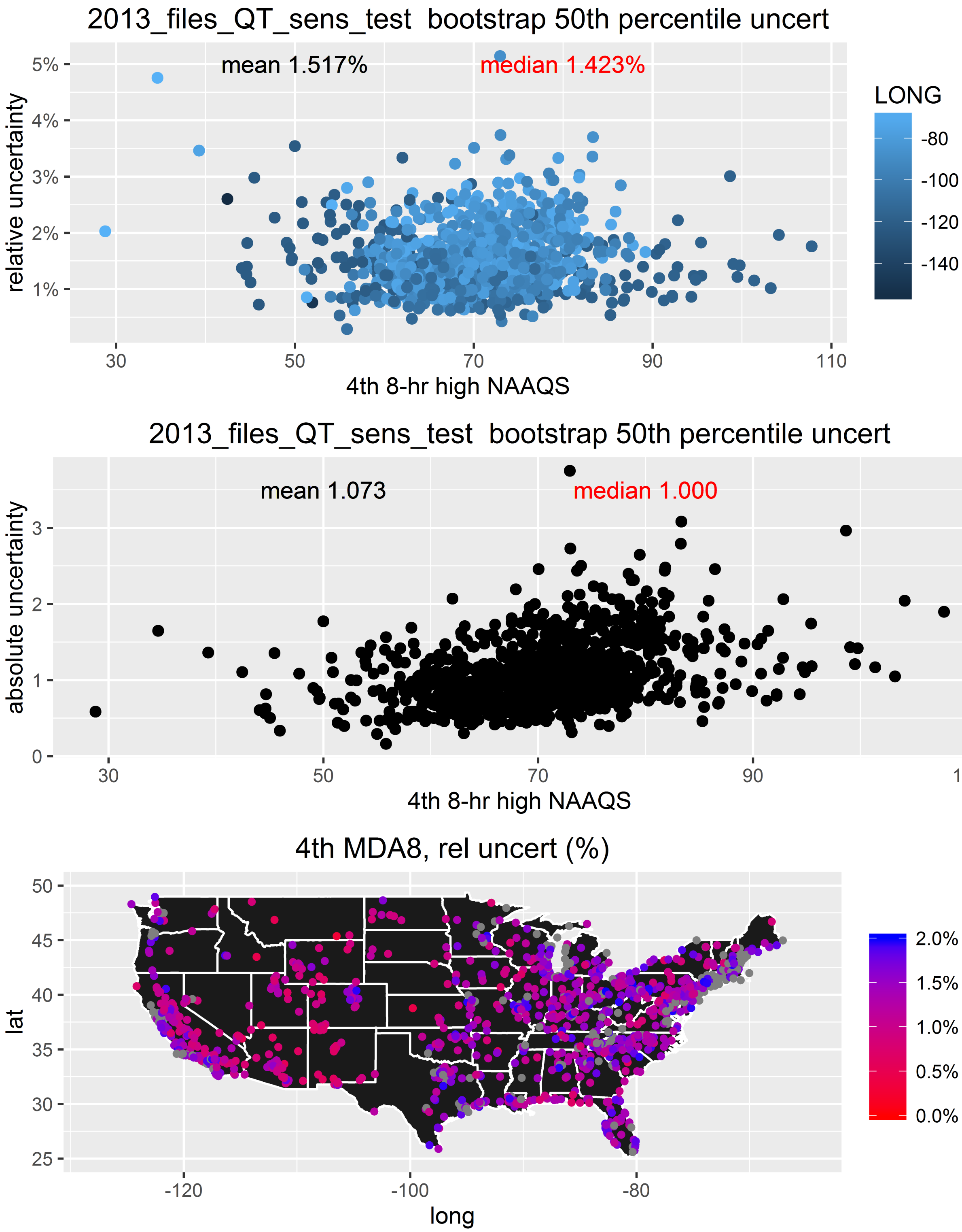


Figure 174: Bootstrap results for the ozone 2013 DVs, showing the 50%, 65%, 75%, and 95% CIs, along with the mean and median bootstrap DVs. The top panel shows the DVs at the various CIs, the middle panel shows the relative difference between the CI and the actual DV, and the bottom panel shows the distribution of the relative differences between the CI and the actual DV.

6 Analysis of temporal lag on ozone data and results from a blocked bootstrap sensitivity analysis

This section presents results from an analysis to examine the temporal correlation of air quality levels, i.e., the tendency of high concentration days to occur after other days with high concentrations. Such behavior, if present, would be a function of both emission trends (e.g., weekday traffic versus weekend traffic) and meteorology (e.g., high pressure systems often hinder the transport of pollutants and also accompany higher temperatures, which tend to increase the formation of ozone). The primary motivation for this assessment is to determine whether the implementation of a block bootstrap procedure is needed for the bootstrapping analysis described in Section 2.2.3 in order to account for possible temporal correlation, and if so, what is the appropriate block size. If not properly accounted for, correlation can affect the assessment of uncertainty (i.e., the standard errors used to calculate confidence intervals). While in this analysis confidence intervals were constructed using empirical percentiles, it is important to consider whether autocorrelation may be affecting the distributional characteristics of the bootstrapped data. Thus a sensitivity analysis is considered.

A block bootstrap method can be used in the presence of autocorrelation to replicate the correlation structure in the data. Blocks are designed such that the dependence between adjacent or closely spaced measurements is contained within a block, and there is induced independence between measurements in adjacent blocks. Block size selection can be tricky, as the blocks should be large enough to induce independence but small enough to retain important characteristics of the data, including natural variation and overall trends (i.e., the variance-bias trade-off for avoiding over-smoothing). There is no one agreed upon method for selection of block size for bootstrapping procedures. Many considerations can come into play, including practical issues and subject-matter scientific expertise. The analysis presented here first attempts to determine the "length of lag" in the ambient ozone data (i.e., how long do correlations of concentrations between MDA8 values persist). Based on the lag analysis, a secondary "blocked" bootstrap analysis was completed which sampled blocks of days corresponding to the lag found in the initial analysis. Ultimately, a 7-day lag was selected from the lag analysis. The resultant bootstrap results were similar to the original non-parametric bootstrap, which sampled individual days rather than blocks of 7-days.

6.1 Analysis procedure and results

The R software package [R Core Team, 2017] was used to conduct the lag analysis. The acf (autocorrelation function) and pacf (partial autocorrelation function) were used to determine the autocorrelation of the time series of MDA8 values at each measurement site for all data available from 2016. The results from the network-wide correlations were summarized in Figures 175 and 176.

The results from the acf analysis suggest that autocorrelations drop off after lag 3, as the mean and median correlation coefficients at 4, 5 and 6 days lag are equivalent (top panel of Figure 175). While the correlations are still within the 95% confidence interval returned from acf out to the 6-day lag, the fact that the distribution of the differences between the individual correlation and the confidence intervals (middle and bottom panel of Figure 175) are virtually identical starting at 3 days of lag suggest that correlations at this level would be found at any lag period. The pacf analysis accounts for the autocorrelation found in the previous lag periods (i.e., the correlation found for the 2-day lag removes the correlation found from the 1-day lag). The results from this analysis suggest that the autocorrelation is only significant to one day. Taking these results into account, a 3-day lag should appropriately account for any autocorrelation in the ozone data. This is implemented in the bootstrapping analysis via a 7-day block size to account for +- 3 days surrounding the sampled daily value. Thus, prior to bootstrapping, the data is grouped into fixed blocks of size n $\bar{7}$ and the sampline with replacement is performed on the blocks. This is also consistent with block sizes used in Inoue and Shintani [2006] and Hall and Horowitz [1996]. A 7-day block size also addresses the consideration of weekly (7-day) pollution patterns across weekday to weekend that may exist.

A second bootstrap analysis was completed for the 2016 ozone data using a block sampling method, with the 7-day block sample size, in order to determine the effect of possible temporal autocorrelation on the bootstrap confidence intervals. The analysis was conducted with the R "boot" package with the tsboot (time seties bootstrap) package with block resampling with fixed block lengths. Simple blocking, rather than overlapping blocks of randomly varying widths, should suffice for initial consideration of possible effect of dependence on bootstrapped confidence intervals (Lahiri [1999] and Andrews [2002]). The results from this bootstrap approach are shown in Figures 177 and 178 (which can be compared to the results from the non-paremetric bootstrap in Sections 3.1 and 4.2 in the main document). The results detailed in Figure 178 indicate slightly greater variability in the blocked bootstrap result, with the mean variability from the blocked bootstrap was 1.62%, versus 1.42% from the non-parametric bootstrap, and the median was 1.55%, versus 1.47% from the non-parametric bootstrap. while there are a few sites with notable larger variability, as with the non-parametric bootstrap, there is no large-scale trend in the variability. The only location of note is perhaps the Uinta Basin in Utah, where a cluster of sites are grayed out in the map, indicating variability greater than the color scale. These sites have the highest variability from the blocked bootstrap. The Uinta Basin is known to have a unique patter in high-ozone days, with the maximum concentrations occurring in the winter during unique meteorological events, such that the high days are always clustered together. As a result, this highly unique ozone pattern has distinctly different results in the blocked bootstrap as compared to the non-parametric bootstrap.

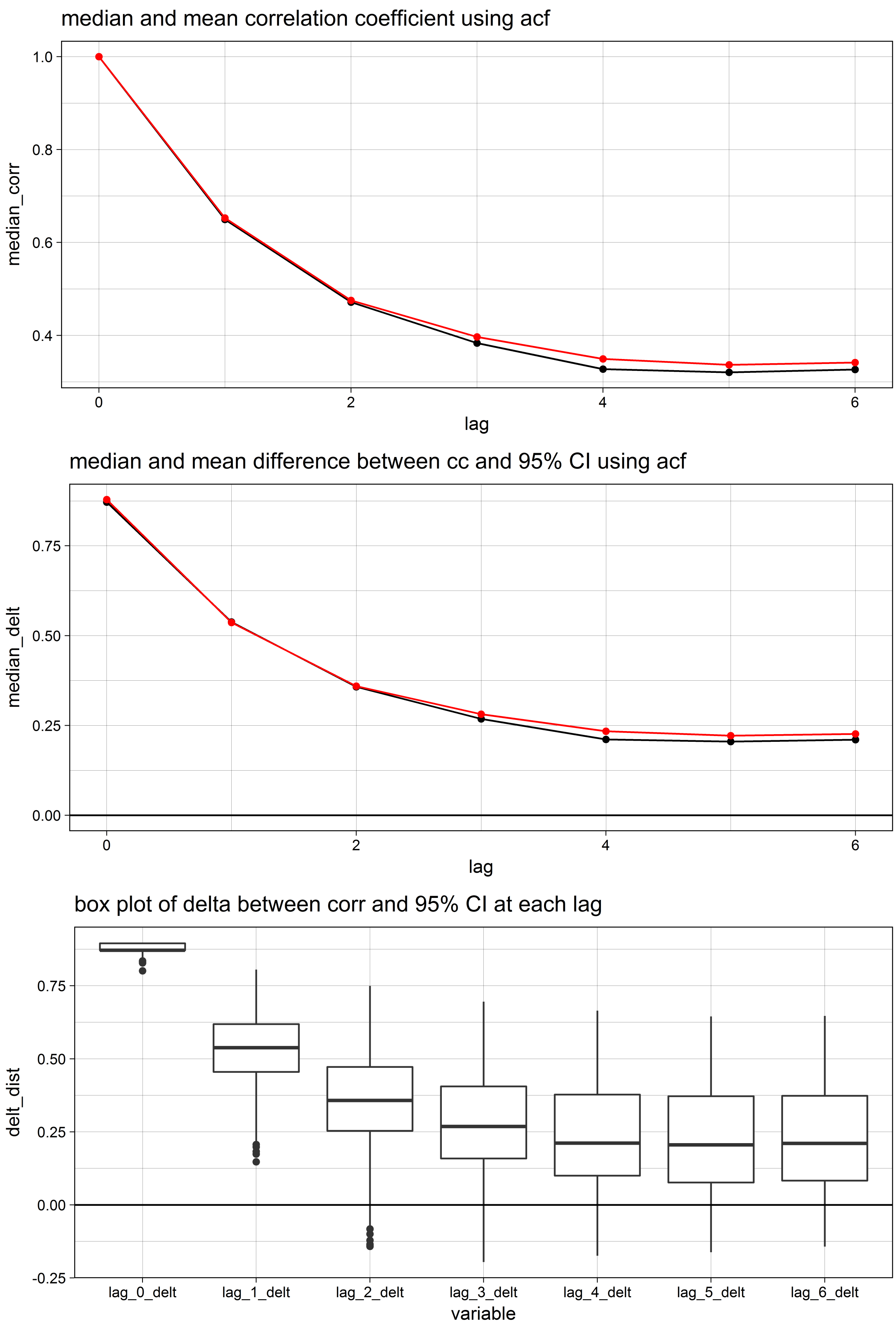


Figure 175: Mean (red lines) and median (black lines) correlations from the acf analysis for ozone data from 2016.

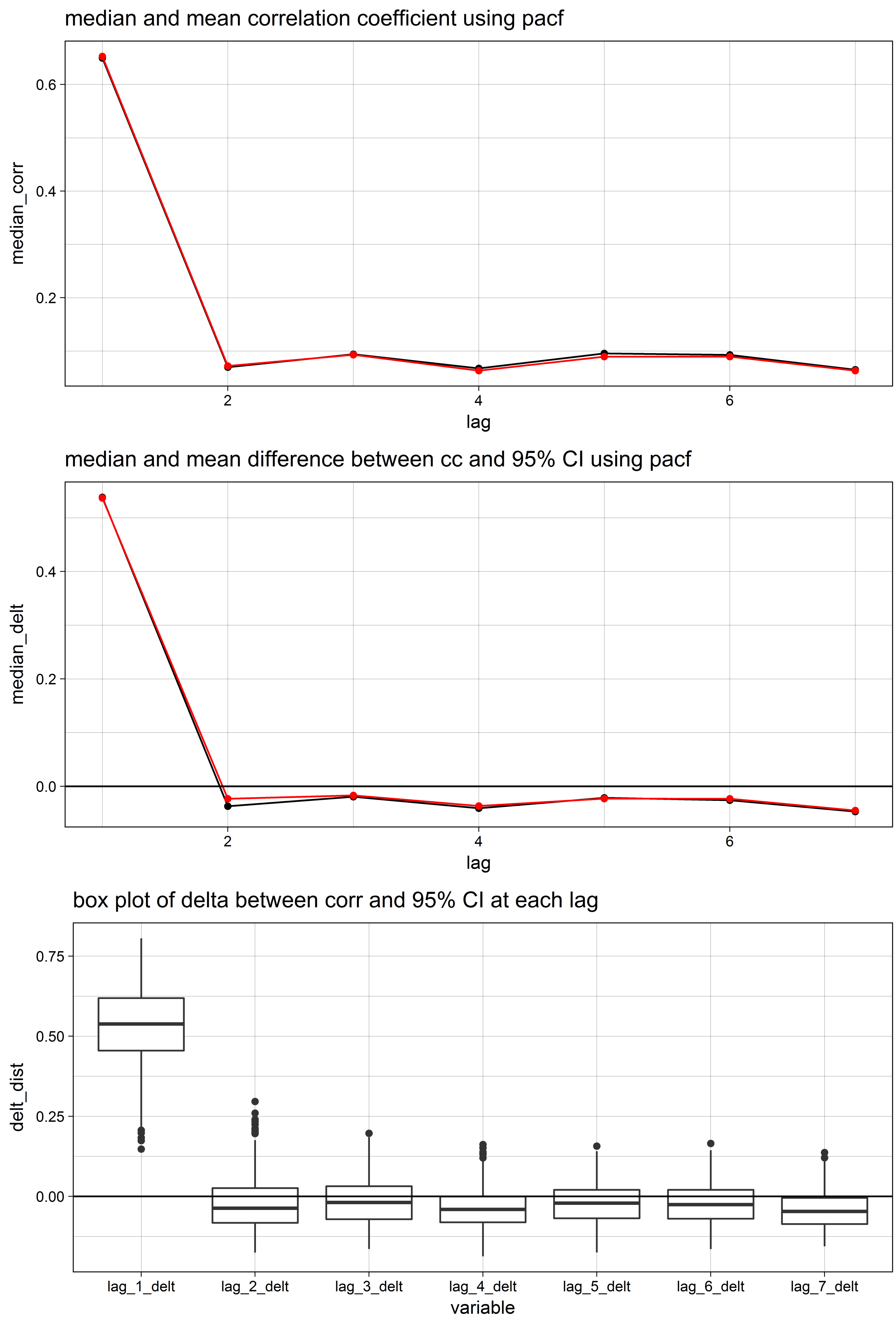


Figure 176: Mean (red lines) and median (black lines) correlations from the pacf analysis for ozone data from 2016.

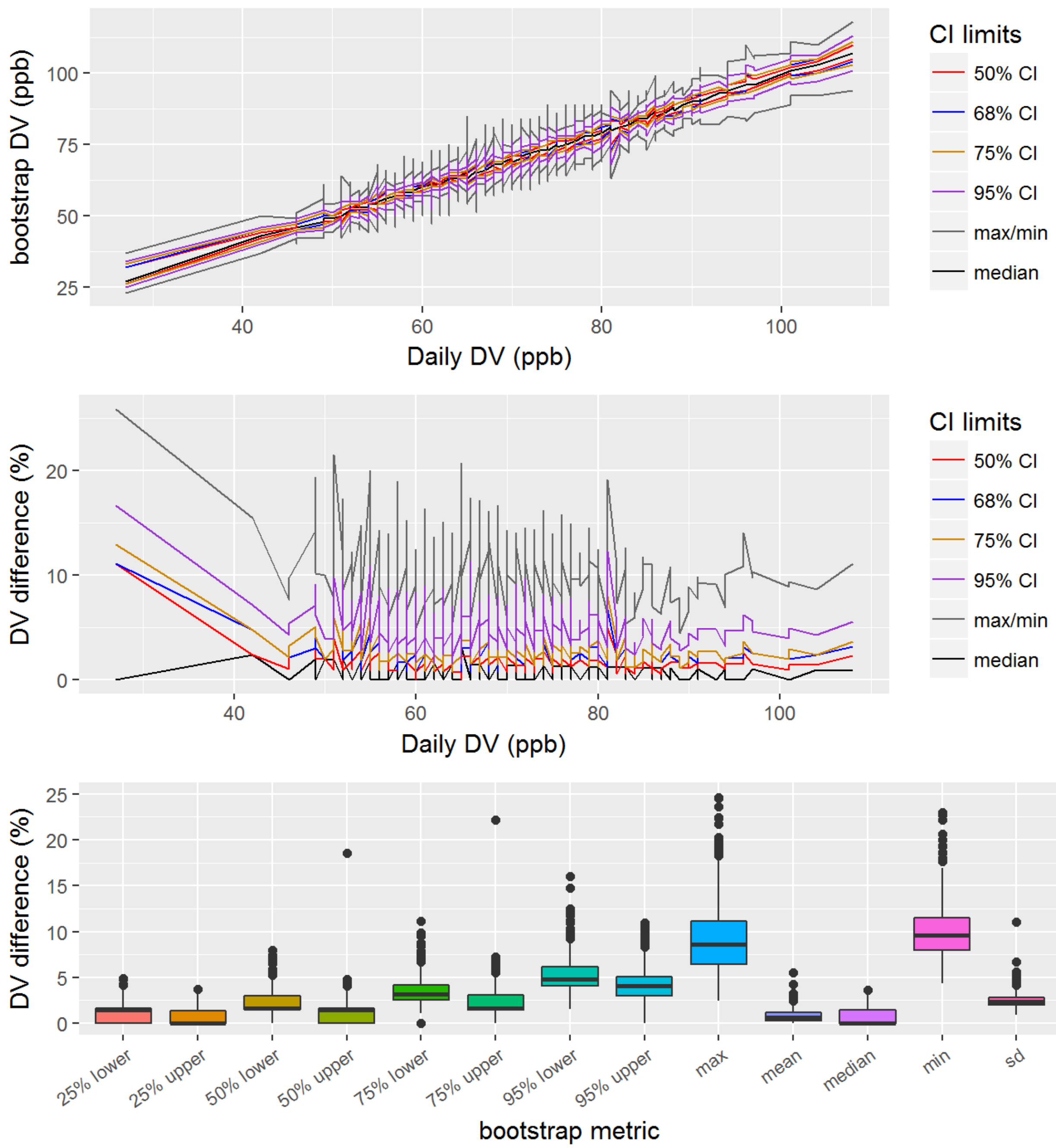


Figure 177: Blocked-bootstrap results for the ozone 2016 DVs, showing the 50%, 68%, 75%, and 95% CIs, along with the mean and median bootstrap DVs. The top panel shows the DVs at the various CIs, the middle panel shows the relative difference between the CI and the actual DV, and the bottom panel shows the distribution of the relative differences between the CI and the actual DV.

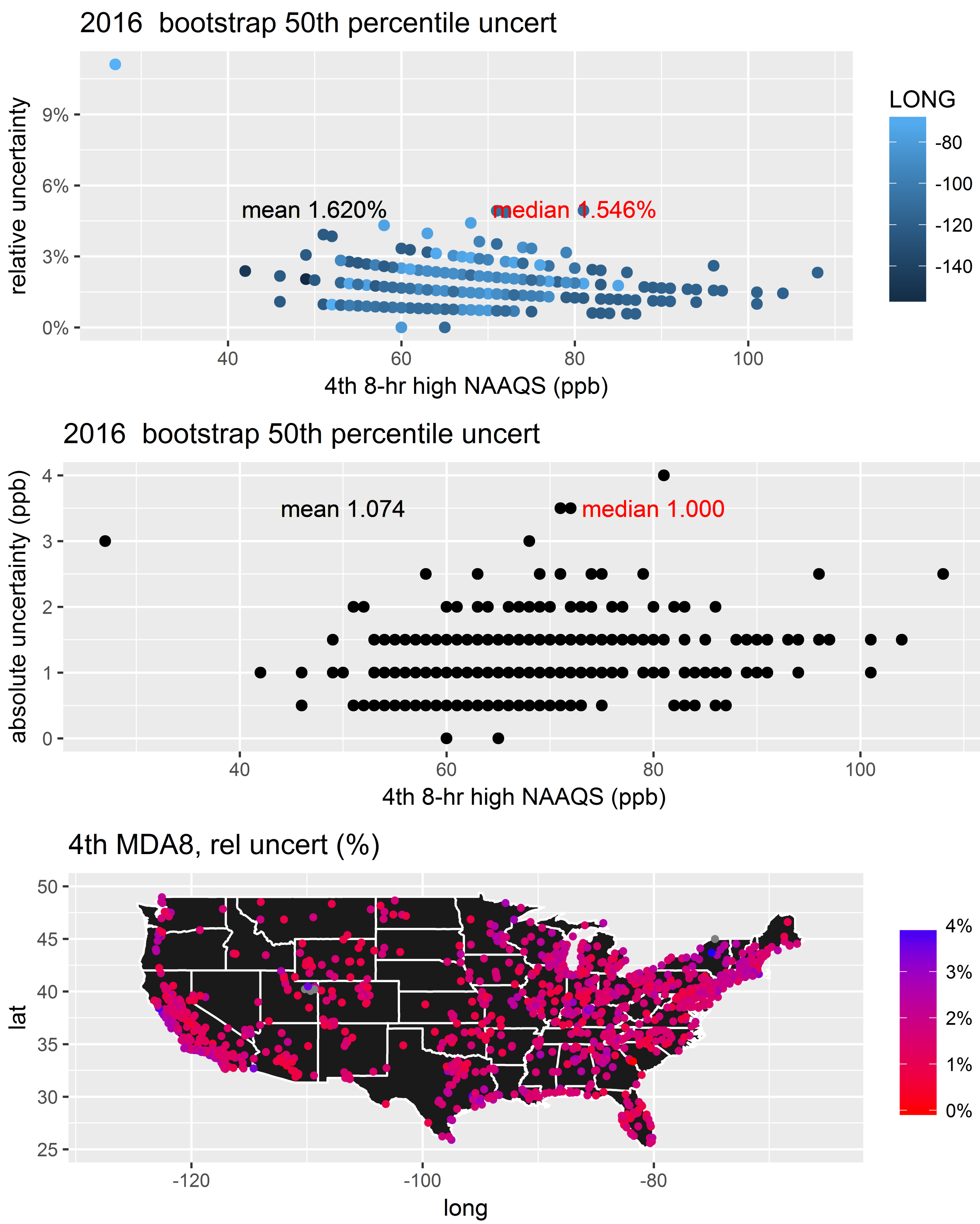


Figure 178: Blocked-bootstrap results from the 50% CIs for the 2016 ozone DVs. The top panel shows the relative difference between the CI and the actual DV, the middle panel shows the absolute difference between the values for the DVs at each site and the CI, and the bottom panel shows the spatial distribution of the relative difference between the 50% CIs for the 2015 ozone DV at each site.

7 Results from cluster analyses and other spatial groupings

This section presents results from several cluster analyses and other analyses conducted to examine the presence of spatial groupings or trends. If strong correlation in the variability can be found in natural spatial groupings, there may be reason to consider the variability at a regional, rather than national, level. The primary purpose of this analysis is to attempt to identify natural spatial groupings and determine if there strong correlations in the variability within these spatial groupings and if the variability between spatial groupings are significantly different. Since there is no clear pathway to determine the spatial correlations, the analysis presented here consists of several iterations of cluster analysis as well as an analysis of variability based on well-established climate regions to explore this issue from various perspectives.

7.1 Cluster analyses

Cluster analysis is an analysis technique that attempts to group data by similar characteristics of the data in question. This is generally done by assigning quantitative values to each characteristic and measuring and minimizing the "distance" between the existing clusters. The "distance" parameter can be calculated in a variety of ways, but the most common (and the one used here) is simply the Euclidian distance between the input variables. Two types of clustering algorithms are applied, a K-means algorithm and a hierarchical algorithm. The K-mean algorithm uses a pre-determined number of clusters and initially randomly assigns all items to clusters. The distance between cluster centers and all individuals are calculated, then individuals are reassigned to their closest cluster. The algorithm repeats a set number of times or until a minimum convergence threshold is reached. Hierarchical algorithms do not use a predetermined a number of clusters, but instead start with each individual as part of their own cluster. The first step in a hierarchical analysis combines the two closest clusters (which are just the two closest members at the first step). Each subsequent step combines the next closest clusters, until only 2 clusters are left. The R software package [R Core Team, 2017] was used to conduct the cluster analysis, using the kmeans and hclust functions. The analysis was performed on the results from the 2014-2016 PM variability results, as described in the following sections.

7.1.1 Cluster analysis with latitude, longitude, and variability values

This cluster analysis used the latitude, longitude (both in degrees eastwest and northsouth), and the relative variability (as a percentage of the site's DV). Thus, the distance between individuals and clusters is defined as the difference between the latitude, longitude, and relative variability. Since the longitudes and latitude varies on a much larger scale between sites (longitude ranges from -64 to -160 degrees, latitude ranges from 17 to 64 degrees) than the relative variability (0-5 percent for the annual and 0-75 percent for the daily DVs), the spatial input component will have a greater impact on the resulting than the site-level variability (clusters for the annual and daily DVs were computed separately). That is, the spatial closeness will be the primary factor in forming these clusters, but the analysis will then try to group nearby sites with similar levels of variability. Hierarchical and K-mean clustering were applied independently.

The clusters formed from this analysis is shown in Figures 179 and 180 and statistics are summarized in Tables 1- 4. The K-means analysis used 10 clusters, which was picked based on the number of EPA Regions. The figure also shows the hierarchical cluster results at 10 clusters for comparison. The clusters from the hierarchical analysis have relatively little recognizable geographic correlation. For example, cluster 1 (orange circles in both the annual and 24-hr figures) consists of a group of sites over California and Arizona and a group over the south eastern US (Florida, Georgia, Alabama), with a major discontinuity in this grouping, with no data points in New Mexico, Texas, Louisiana, and Mississippi. Table 1 and 2 show the statistics from the hierarchical clusters for the annual and 24-hr standards. The table includes a comparison of the mean variability from each cluster to the mean from the entire dataset using a Welch Modified Two-Sample t-Test (determined from the tsum.test function from the BSDA package in R) to determine if the means are significantly different. For the annual standard, the p values are all fairly high, with the smallest value just over 0.1, which is well above the nominal p value of 0.5 typically identified as an indicator that the means may be different. For the 24-hr standard, there are 2 clusters with p values less than 0.05. Cluster 7 has a p value of 0.02, which may be different than the annual mean, but the sites in this cluster (light blue squares with an "x") are spread across the country, i.e., they are not spatially distinct. Cluster 8 has the smallest p value (0.006) and has the smallest mean variability. For the most part, this cluster is in the same region (purple asterisk), in the eastern US, from North Carolina up to New York. However, this cluster is interspersed with several other clusters. Thus, while it has distinct variability values and is spatially correlated, it is not spatially distinct.

The results from the K-mean cluster analysis are starkly different from the hierarchical analysis. The clusters are all geographically distinct and the results of the t-test indicate that several of the clusters are distinctly different from the mean dataset. For the annual standard, half of the clusters have p-values less than 0.05 (2, 3, 4, 8, and 10) while 7 clusters have p-values less than 0.05 for the daily standard (1, 2, 3, 4, 5, 6, and 9, though cluster 1 and 9 only have a few members and cluster 2 is close enough to 0.05 to discount as significantly different, leaving only 4 clusters of note). On the surface, this suggests there are regional differences in the variability. However, the differences between the results from the annual and daily standards suggest the result is less certain. For example, cluster 4 in the annual analysis stands out as having the largest mean variability and a very small p-value, suggesting the variability in this subset is significantly different from the mean dataset. However, these sites are part of a larger cluster in the daily results (cluster 8), which include California sites, and has lower mean variability than the mean from the dataset (though not significantly different, with a p-value of 0.31). Another example of inconsistency between the annual and daily results is cluster 3 in the annual results, which roughly correlates to cluster 6 in the daily results. In this case, the clusters represent approximately the same geographic region. However, for the annual result, cluster 3 has mean variability that appears to be significantly higher than the mean dataset's (p-value of 0.016), but significantly lower mean variability for the daily standard than the mean dataset's (p-value of 0.018). Thus, these particular geographic areas have higher than average variability in the long-term, but lower than average variability in the short-term. The inconsistent results from the K-means analysis make it difficult draw specific conclusions about the geographic nature of the variability as estimated by this analysis.

Table 1: Comparison of hierarchical clusters for lat-long-annual variability

n.sites	grp	mean	median	sd	ann.pval
107	1	1.628	1.538	0.7296	0.1969
44	2	1.917	1.786	0.7377	0.1052
43	3	1.759	1.724	0.7835	0.7972
57	4	1.759	1.744	0.5889	0.7024
43	5	1.659	1.630	0.5025	0.4108
44	6	1.745	1.765	0.4914	0.8207
73	7	1.736	1.754	0.5604	0.8929
26	8	1.674	1.714	0.6251	0.6762
49	9	1.816	1.744	0.7380	0.4171
38	10	1.660	1.453	0.8483	0.6384
524	All	1.727	1.705	0.6700	1.0000

Table 2: Comparison of hierarchical clusters for lat-long-24-hr variability

n.sites	grp	mean	median	sd	TF.pval
107	1	5.985	3.846	8.882	0.718242
44	2	6.396	5.013	5.046	0.365465
43	3	6.339	5.556	6.332	0.501141
57	4	5.150	4.054	3.758	0.370173
43	5	6.192	4.762	6.745	0.618803
44	6	5.098	4.692	3.403	0.336031
73	7	4.720	4.167	2.598	0.021686
26	8	4.338	4.006	1.917	0.005972
49	9	5.989	4.545	8.297	0.787386
38	10	6.217	4.583	6.185	0.594067
524	all	5.659	4.447	6.163	1.000000

Table 3: Comparison of K-means clusters for lat-long-annual variability

n.sites	grp	mean	median	sd	ann.pval
61	1	1.634	1.613	0.5826	2.508e-01
26	2	1.425	1.291	0.5835	1.618e-02
96	3	1.863	1.796	0.4657	1.599e-02
26	4	2.487	2.395	0.7843	4.550e-05
58	5	1.661	1.658	0.5383	3.941e-01
38	6	1.619	1.695	0.6162	3.044e-01
57	7	1.573	1.471	0.5912	7.068e-02
66	8	1.423	1.373	0.4362	2.550e-06
57	9	1.663	1.471	0.8888	6.006e-01
39	10	2.266	2.174	0.7903	1.556e-04
524	all	1.727	1.705	0.6700	1.000e+00

Table 4: Comparison of K-means clusters for lat-long-24-hr variability

n.sites	grp	mean	median	sd	TF.pval
3	1	61.111	58.333	12.729	1.701e-02
59	2	4.794	4.762	2.367	3.604e-02
44	3	9.083	8.477	3.979	2.320e-06
104	4	3.807	3.333	2.131	8.750e-08
33	5	4.041	2.941	3.060	9.146e-03
109	6	4.751	4.545	2.849	1.831e-02
38	7	4.994	4.762	2.825	2.154e-01
68	8	5.165	4.202	3.342	3.111e-01
7	9	30.768	30.952	9.002	3.049e-04
59	10	5.020	3.846	3.136	1.938e-01
524	all	5.659	4.447	6.163	1.000e+00

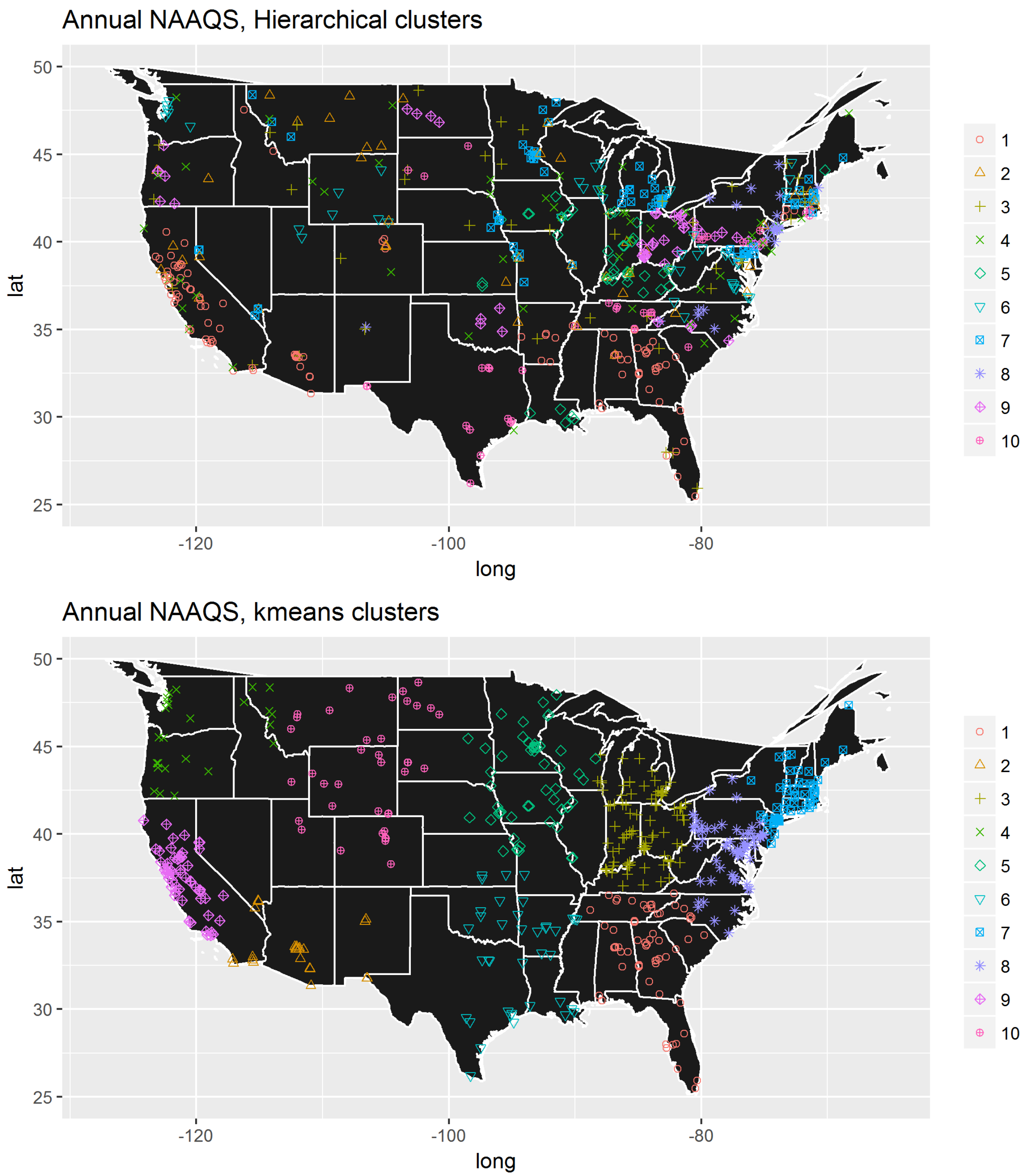
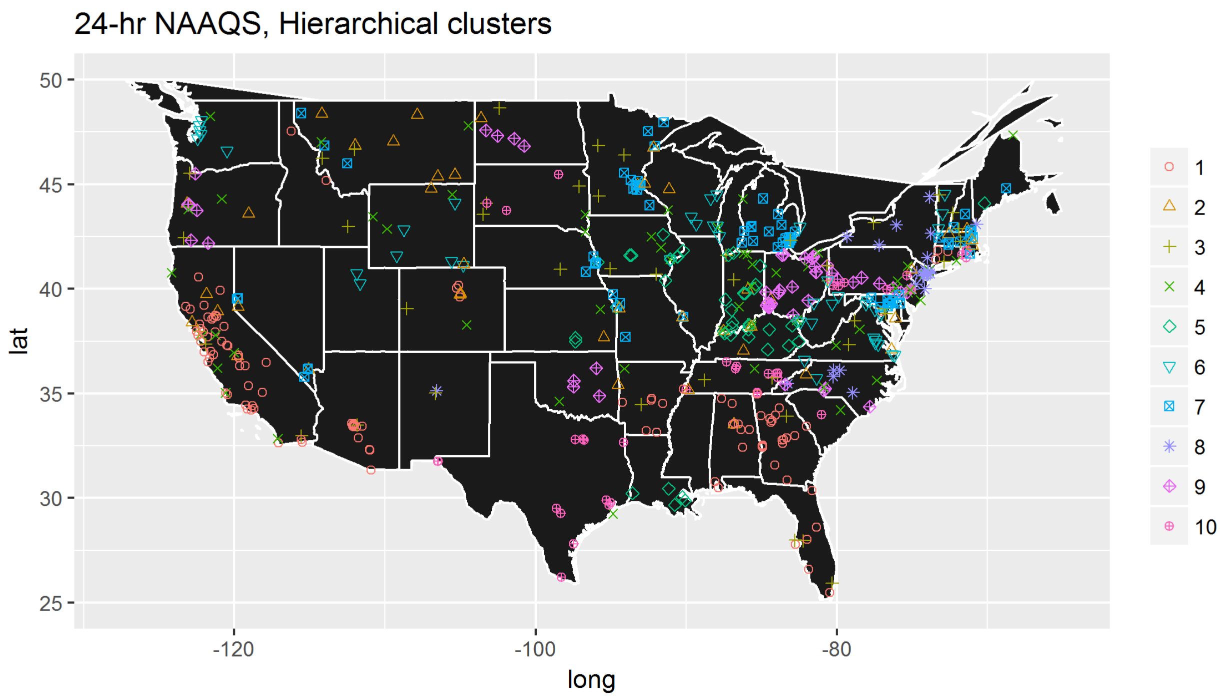


Figure 179: Hierarchical and K-means clusters for the annual variability for 2016.



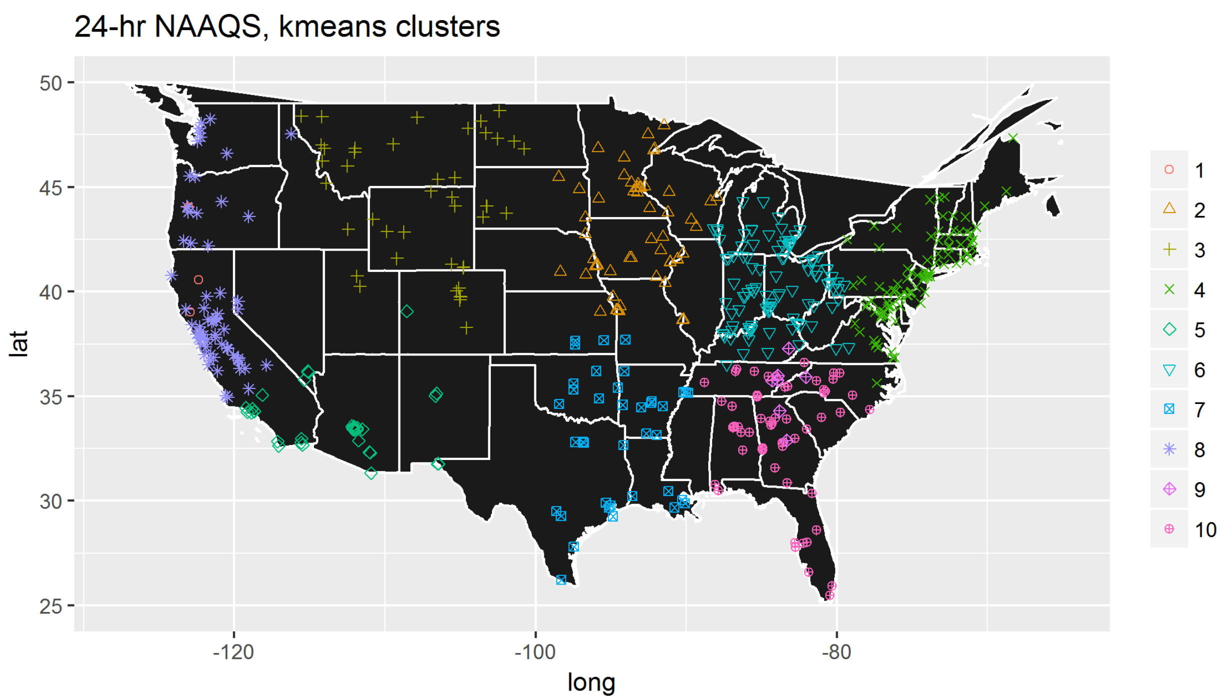


Figure 180: Hierarchical and K-means clusters for the 24-hr variability for 2016.

7.1.2 Cluster analysis with time series of variability values (2014-2016)

This cluster analysis used the relative variability (as a percentage of the site's DV) from each site over 3 DV periods (2014-2016). Thus, the distance between individuals and clusters is defined as the difference between each year's variability values (i.e., the variability from 2014 data, the variability from 2015 data, the variability from 2016 data) for a particular standard. Unlike the previous analysis, all input variables are on the same scale, such that no one parameter is driving the cluster formation. Therefore, this analysis attempts to group sites with similar levels of variability over time in order to see if those variability trends have spatial correlation. Since this approach incorporates the variability over time, it reflects the final composite variability value determined in the main analysis, which is the average over 3 DV periods. Hierarchical and K-mean clustering were applied independently.

The clusters formed from this analysis is shown in Figures 181 and 182 and statistics are summarized in Tables 5- 8. The K-means analysis used 10 clusters, which was picked based on the number of EPA Regions. The figure also shows the hierarchical cluster results at 10 clusters for comparison. As with the latlongvariability analysis presented in the previous section, the clusters from the hierarchical analysis have relatively little recognizable geographic correlation. However, most of the clusters have mean variability levels that are distinctly different from the mean dataset (note that the mean values presented here represent the mean from all years), though this approach also resulted in more clusters with very few members, such that the annual results only had 5 clusters with 20 or more members and the 24-hr results only had 3 clusters with 20 or more members. The spatial distribution of the results from the K-means analysis was similar to the hierarchical results, in that relatively little recognizable geographic correlation. However, the K-mean algorithm resulted in more meaningful clusters in terms of number of members and the statistical significance. Thus, while the cluster analysis conducted with the 3-year variability trends resulted in groups that were distinct with respect to their variability levels, it showed essentially no spatial correlation, suggesting that geographic differences in variability do not need to be taken into account.

Table 5: Comparison of hierarchical clusters for 2014-2016 variability, annual

n˙sites	grp	mean	median	sd	ann˙pval
86	1	1.027	1.053	0.2387	1.418e-49
84	2	1.475	1.449	0.2964	9.408e-08
30	3	2.422	2.381	0.2443	2.756e-19
147	4	1.877	1.852	0.2730	1.498e-05
20	5	1.488	1.559	0.4272	3.258e-02
7	6	2.185	2.439	0.8897	2.129e-01
12	7	3.017	3.008	0.3734	3.841e-08
9	8	3.099	3.125	0.4706	1.585e-05
1	9	4.843	5.128	0.8896	NA
2	10	3.987	4.139	0.5126	9.888e-02
398	11	1.716	1.666	0.5855	1.000e+00

Table 6: Comparison of hierarchical clusters for 2014-2016 variability, 24-hr

n˙sites	grp	mean	median	sd	TF˙pval
272	1	3.630	3.333	1.662	1.499e-08
49	2	8.034	7.895	2.365	1.311e-11
58	3	4.848	4.545	2.134	7.556e-01
9	4	11.821	11.765	2.320	1.106e-05
4	5	9.097	5.409	7.870	3.701e-01
1	6	18.480	20.312	4.399	NA
1	7	55.556	75.000	41.107	NA
1	8	25.966	29.630	7.354	NA
2	9	13.220	14.460	6.831	3.363e-01
1	10	41.250	46.667	20.340	NA

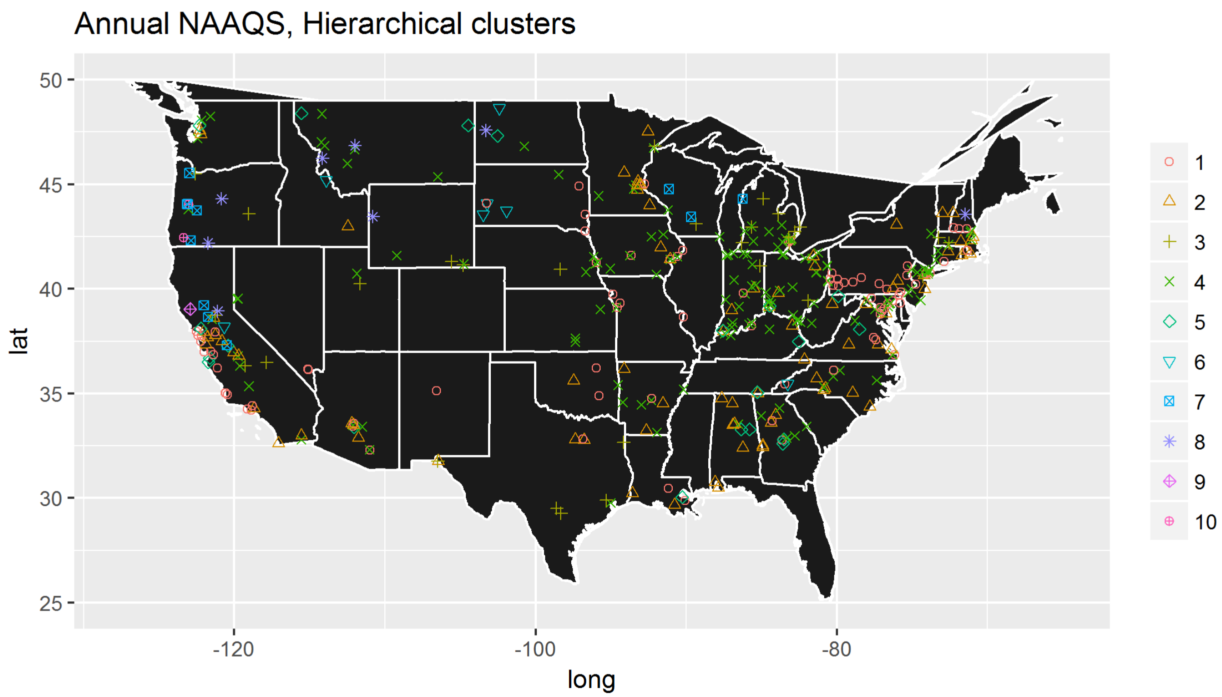
n˙sites	grp	mean	median	sd	TF˙pval
398	11	4.957	4.129	4.143	1.000e+00

Table 7: Comparison of K-means clusters for 2014-2016 variability, annual

n˙sites	grp	mean	median	sd	ann˙pval
23	1	1.027	1.053	0.2387	1.068e-14
3	2	1.475	1.449	0.2964	2.939e-01
61	3	2.422	2.381	0.2443	1.781e-38
66	4	1.877	1.852	0.2730	3.753e-04
48	5	1.488	1.559	0.4272	1.382e-03
5	6	2.185	2.439	0.8897	3.038e-01
68	7	3.017	3.008	0.3734	4.822e-50
28	8	3.099	3.125	0.4706	3.779e-16
27	9	4.843	5.128	0.8896	9.066e-17
69	10	3.987	4.139	0.5126	2.528e-56
398	11	1.716	1.666	0.5855	1.000e+00

Table 8: Comparison of K-means clusters for 2014-2016 variability, 24-hr

n˙sites	grp	mean	median	sd	TF˙pval
2	1	3.630	3.333	1.662	4.577e-01
59	2	8.034	7.895	2.365	2.016e-13
44	3	4.848	4.545	2.134	7.770e-01
104	4	11.821	11.765	2.320	4.609e-65
37	5	9.097	5.409	7.870	3.098e-03
46	6	18.480	20.312	4.399	1.172e-26
2	7	55.556	75.000	41.107	3.319e-01
22	8	25.966	29.630	7.354	6.561e-12
73	9	13.220	14.460	6.831	7.360e-16
9	10	41.250	46.667	20.340	6.812e-04
398	11	4.957	4.129	4.143	1.000e+00



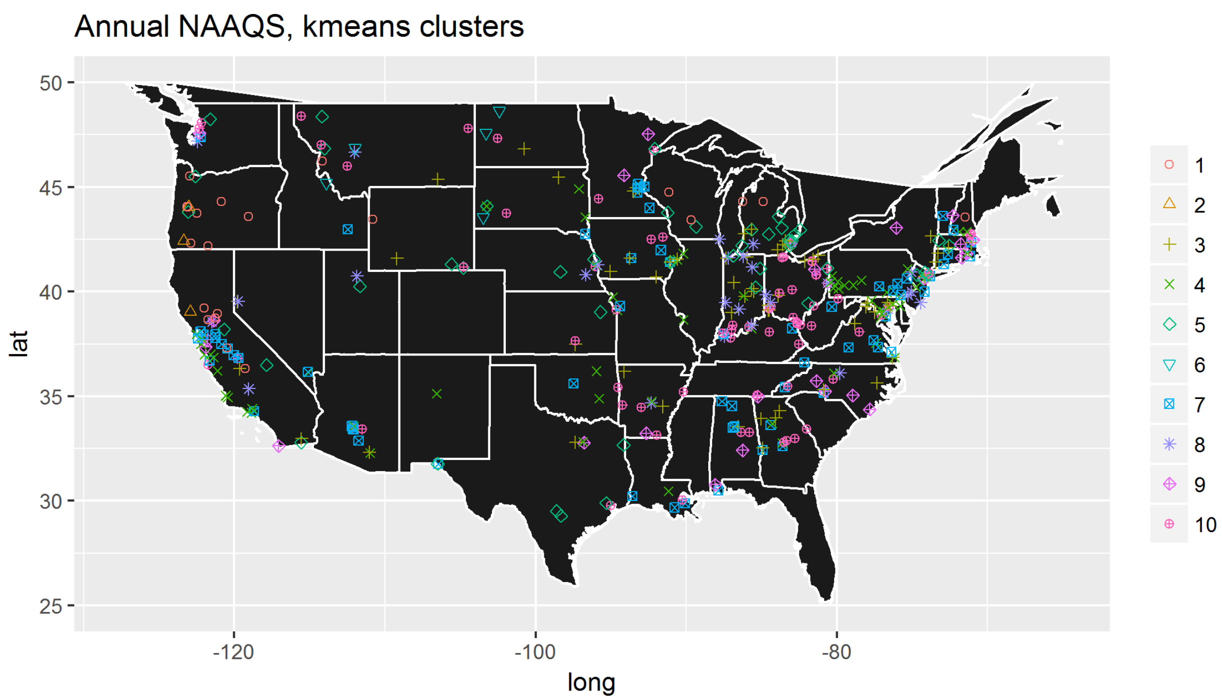


Figure 181: Hierarchical and K-means clusters for the annual variability for 2014-2016.

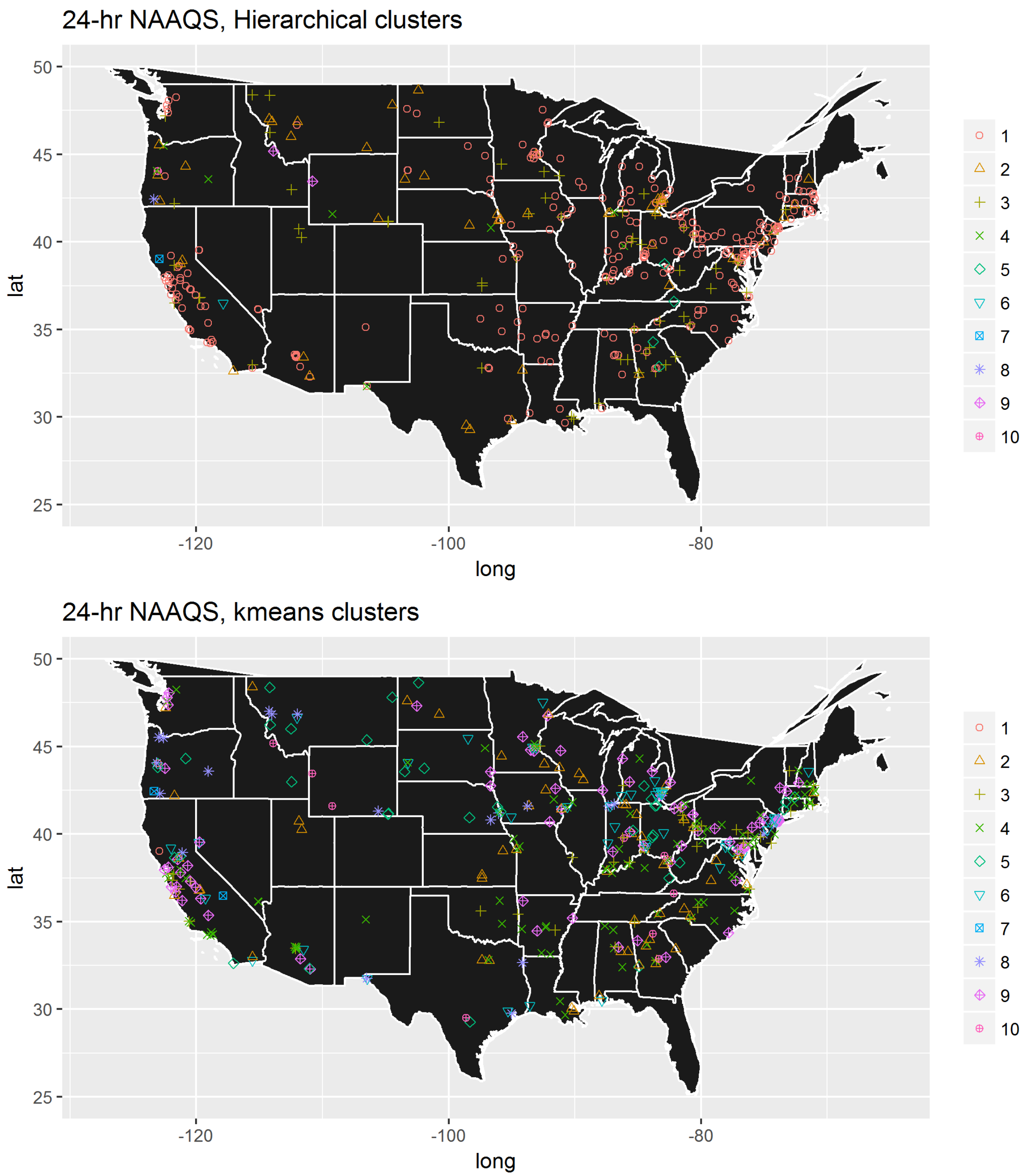


Figure 182: Hierarchical and K-means clusters for the 24-hr variability for 2014-2016.

7.1.3 Cluster analysis with time series of variability values (2012-2016)

This cluster analysis used the relative variability (as a percentage of the site's DV) from each site over 5 DV periods (2012-2016). Thus, the distance between individuals and clusters is defined as the difference between each year's variability value (i.e., the variability from 2012 data, ..., the variability from 2016 data) for a particular standard. Unlike the previous analysis, all input variables are on the same scale, such that no one parameter is driving the cluster formation. Therefore, this analysis attempts to group sites with similar levels of variability over time in order to see if those variability trends have spatial correlation. Since this approach incorporates the variability over time, it partly reflects the final composite variability value determined in the main analysis, which is the average over 3 DV periods. The extended period is evaluated in addition to the 3 DV periods presented above in order to improve correlations that may exist with a longer data record. Hierarchical and K-mean clustering were applied independently.

The clusters formed from this analysis is shown in Figures 183 and 184 and statistics are summarized in Tables 9- 12. The K-means analysis used 10 clusters, which was picked based on the number of EPA Regions. The figure also shows the hierarchical cluster results at 10 clusters for comparison. The results from this analysis are fairly similar to the results using the 3 DV periods. The clusters are not spatially distinct; many clusters have few members, though most have distinct variability levels. Thus, while the cluster analysis conducted with the 5-year variability trends resulted in groups that were distinct with respect to their variability levels, it showed essentially no spatial correlation, suggesting that geographic differences in variability do not need to be taken into account.

Table 9: Comparison of hierarchical clusters for 2012-2016 variability, annual

n˙sites	grp	mean	median	sd	ann˙pval
76	1	1.138	1.099	0.2967	1.272e-27
30	2	2.292	2.273	0.3028	6.841e-12
55	3	1.533	1.531	0.3474	9.338e-04
105	4	1.809	1.802	0.3136	6.702e-02
1	5	3.295	2.632	1.2706	NA
16	6	2.929	2.871	0.3333	1.926e-11
2	7	3.691	3.562	1.0601	2.315e-01
1	8	3.117	3.670	0.8943	NA
1	9	4.096	4.348	0.5319	NA
3	10	2.185	2.232	0.7266	3.884e-01
290	11	1.727	1.678	0.5550	1.000e+00

Table 10: Comparison of hierarchical clusters for 2012-2016 variability, 24-hr

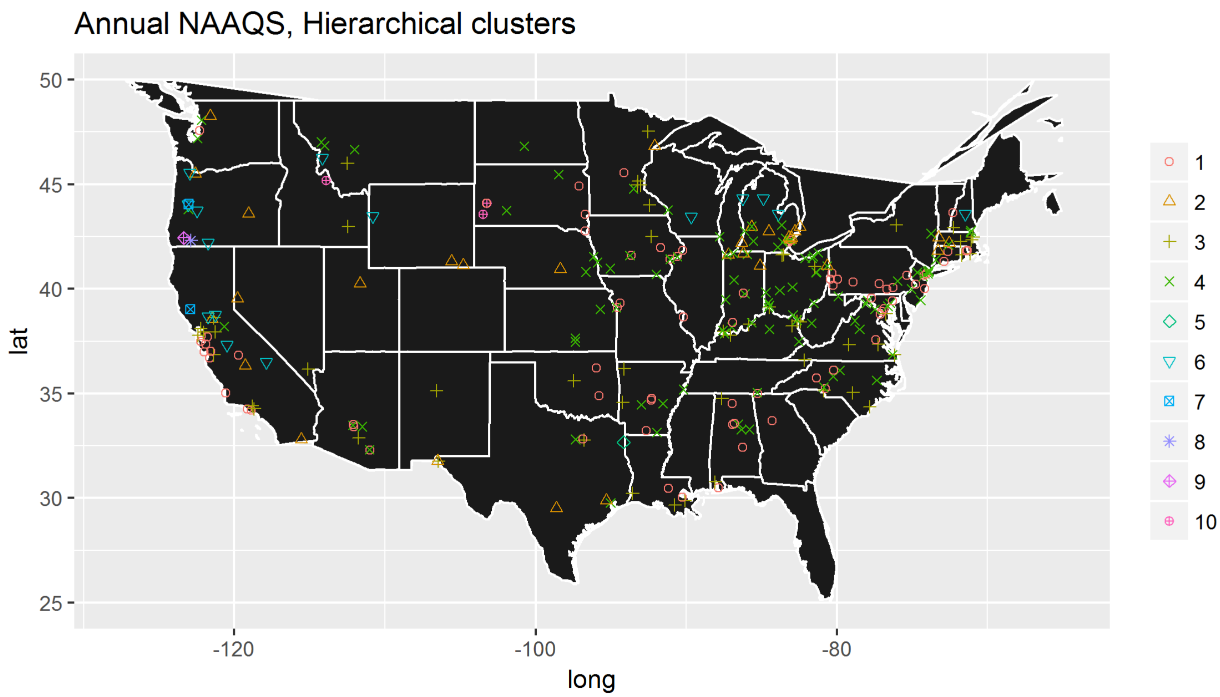
n˙sites	grp	mean	median	sd	TF˙pval
193	1	3.460	3.226	1.469	6.142e-09
25	2	7.603	7.143	2.917	1.610e-04
32	3	5.251	4.762	2.776	5.508e-01
33	4	7.187	7.143	2.787	1.074e-04
1	5	36.947	10.000	37.861	NA
1	6	13.357	7.692	10.734	NA
2	7	14.391	12.812	5.456	2.455e-01
1	8	36.444	10.000	39.143	NA
1	9	23.033	28.571	9.620	NA
1	10	27.333	18.750	23.875	NA
290	11	4.927	4.132	3.802	1.000e+00

Table 11: Comparison of K-means clusters for 2012-2016 variability, annual

n˙sites	grp	mean	median	sd	ann˙pval
46	1	1.138	1.099	0.2967	1.113e-18
39	2	2.292	2.273	0.3028	5.306e-15
56	3	1.533	1.531	0.3474	8.645e-04
51	4	1.809	1.802	0.3136	1.358e-01
34	5	3.295	2.632	1.2706	2.922e-08
3	6	2.929	2.871	0.3333	2.210e-02
3	7	3.691	3.562	1.0601	8.448e-02
21	8	3.117	3.670	0.8943	5.988e-07
21	9	4.096	4.348	0.5319	5.424e-16
16	10	2.185	2.232	0.7266	2.442e-02
290	11	1.727	1.678	0.5550	1.000e+00

Table 12: Comparison of K-means clusters for 2012-2016 variability, 24-hr

n˙sites	grp	mean	median	sd	TF˙pval
1	1	3.460	3.226	1.469	NA
72	2	7.603	7.143	2.917	1.189e-09
67	3	5.251	4.762	2.776	4.262e-01
20	4	7.187	7.143	2.787	2.264e-03
26	5	36.947	10.000	37.861	2.218e-04
75	6	13.357	7.692	10.734	2.885e-09
2	7	14.391	12.812	5.456	2.455e-01
3	8	36.444	10.000	39.143	2.978e-01
2	9	23.033	28.571	9.620	2.285e-01
22	10	27.333	18.750	23.875	2.491e-04
290	11	4.927	4.132	3.802	1.000e+00



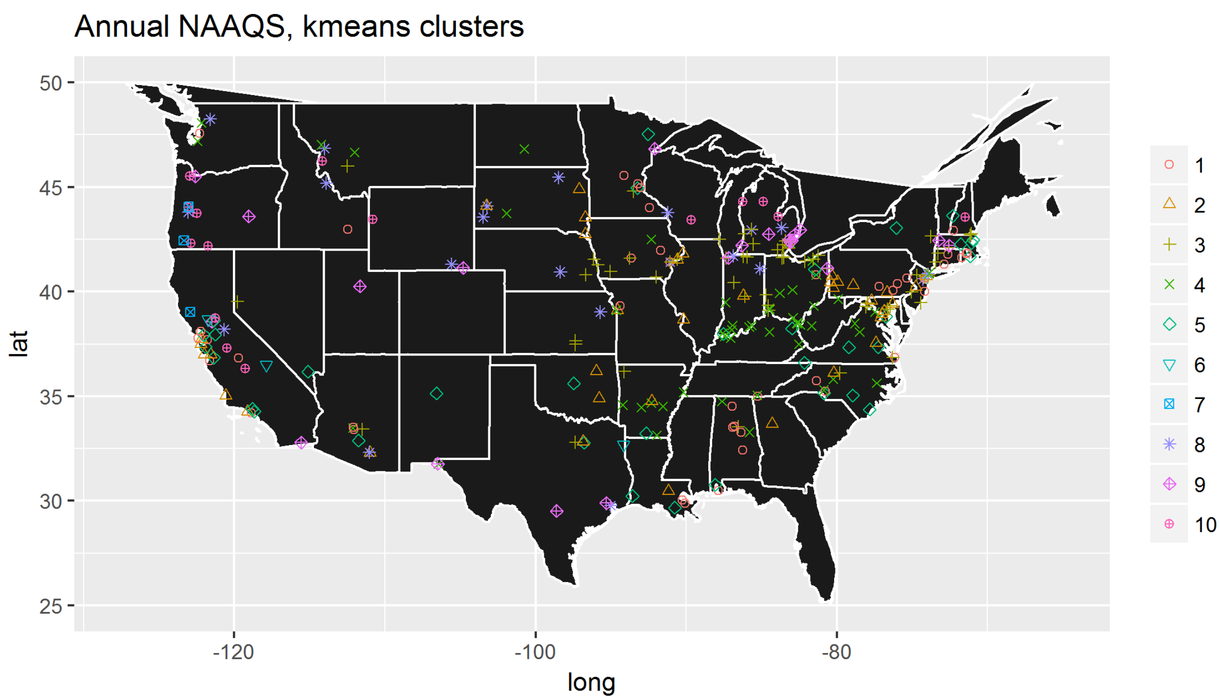
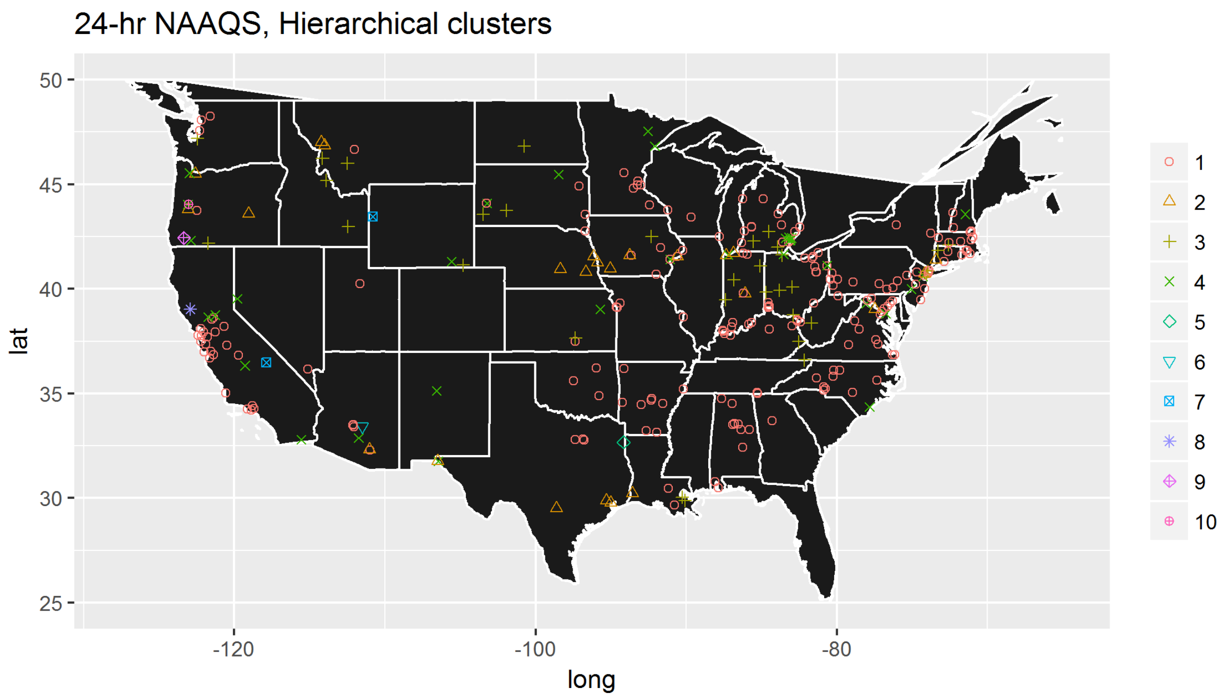


Figure 183: Hierarchical and K-means clusters for the annual variability for 2012-2016.



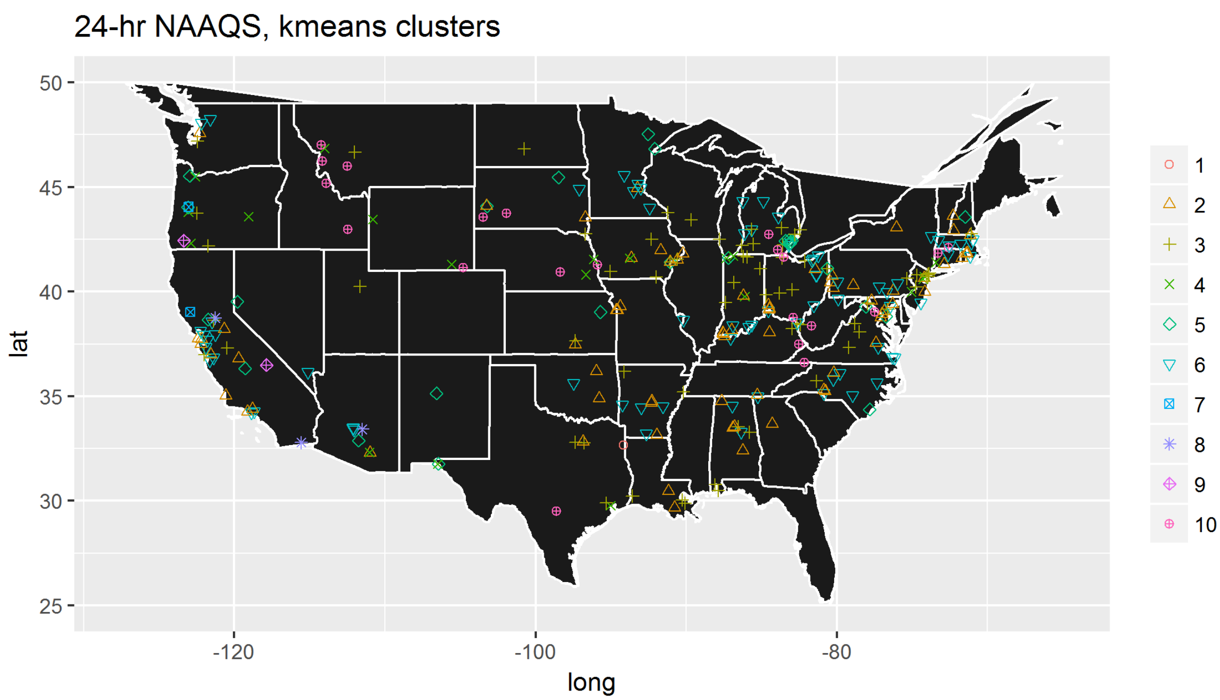


Figure 184: Hierarchical and K-means clusters for the 24-hr variability for 2012-2016.

7.2 Spatial analysis using NOAA Climate Regions

The National Oceanic and Atmospheric Administration (NOAA) has identified 9 "climate regions" [Thomas and Koss, 1984], which have been identified to have distinct climatologically characteristics (more information available at the NOAA website). This spatial grouping thus represents an independent spatial grouping with which to evaluate regional variability characteristics. The mean annual and 24-hr variability values from sites within these regions are compared in Figure 185 and detailed in Tables 13 and 14. The Pacific Northwest (region 4) and the Central Northwest (region 9) stand out as having higher variability, which was seen in the first K-means cluster analysis (using latitude, longitude, and the variability). The p-values for the annual results are less than 0.05 for these two regions (though the p-value for region 9 is just barely less than 0.05 and the p-value for region 4 is still relatively large). The p-values for these two regions from the 24-hr results are well above the nominal value of 0.05 and so are not significantly different from the mean dataset. Thus, the results again make it difficult draw specific conclusions about the geographic nature of the variability as estimated by this analysis, though the overall interpretation of these results is that most regions are not significantly different from the mean dataset.

Table 13: Comparison of variability within NOAA climate regions, annual

Region Number	n sites	ann mean	ann median	ann sd	ann pval
1	96	1.724119413	1.724137931	0.538832598	0.986301656
2	69	1.864963958	1.875	0.501458469	0.547340281
3	94	1.457900874	1.388888889	0.545905166	0.245378188
4	22	2.473011166	2.405978785	0.845285667	0.013781951
5	39	1.719076105	1.704545455	0.622791499	0.970268689
6	70	1.563384611	1.583124478	0.459051349	0.469019357
7	25	1.623411728	1.595744681	0.621819606	0.676247393
8	66	1.610690042	1.405159932	0.84597489	0.627212461
9	42	2.255001105	2.198067633	0.808280411	0.047892639
All	523	1.727970703	1.727970703	0.670071968	1

Table 14: Comparison of variability within NOAA climate regions, 24-hr

Region Number	n sites	TF mean	TF median	TF sd	TF pval
1	96	5.861069073	4.545454545	6.966363904	0.924845491
2	69	4.690627066	4.545454545	1.965280928	0.632945983
3	94	3.628720758	3.125	2.092141946	0.326832786
4	22	10.13095985	7.417582418	11.55838373	0.165608839
5	39	5.280129178	5	3.051945505	0.853550834
6	70	5.893319944	4.653679654	5.146873277	0.911597545
7	25	5.383605813	4	3.836720273	0.896970478
8	66	5.93939837	3.923076923	10.5914199	0.90680014
9	42	8.69767221	7.692307692	3.979459168	0.166016078
All	523	5.660892679	4.347826087	6.168440212	1

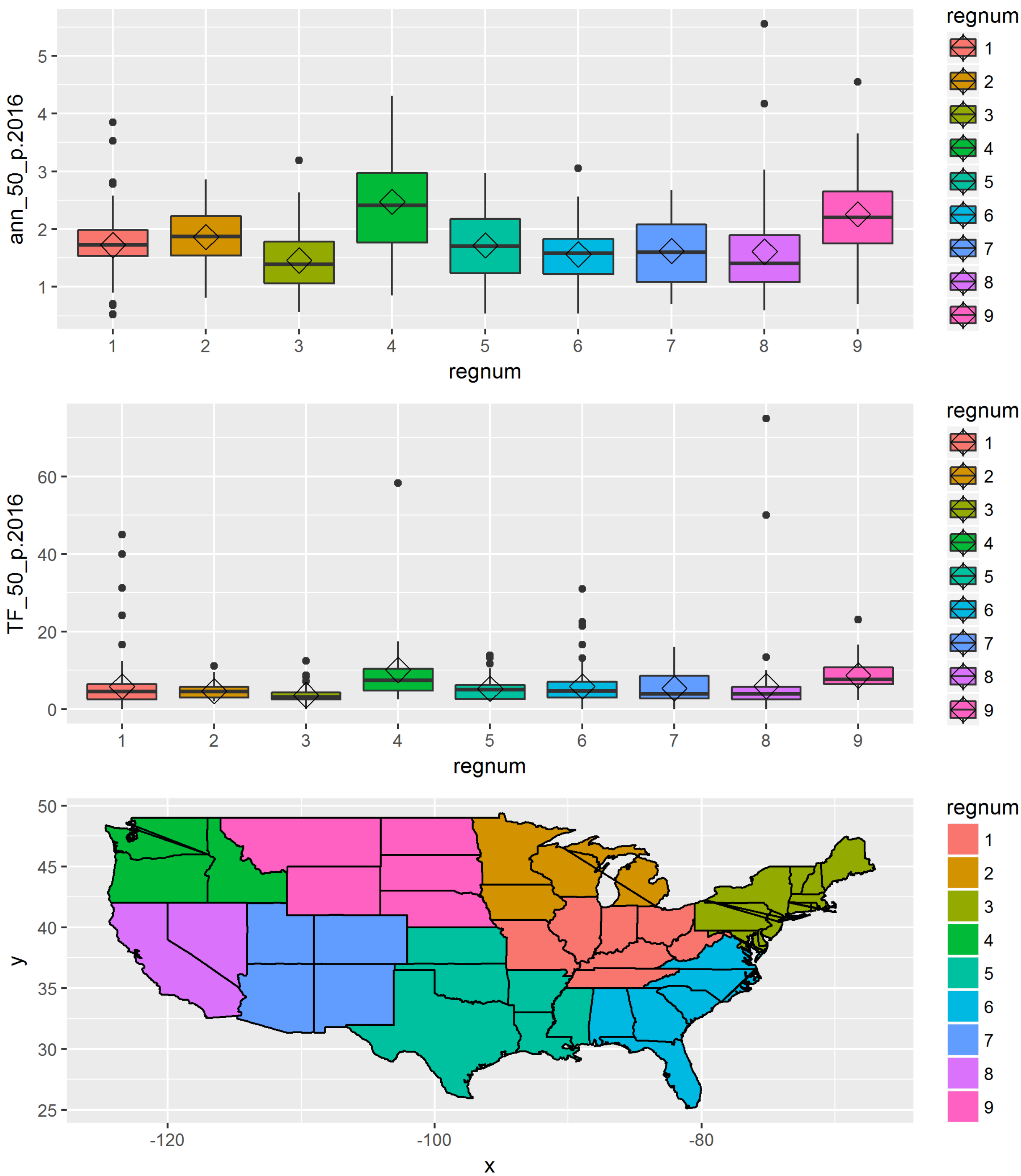


Figure 185: Comparison of variability within NOAA climate regions for 2016.

References

D. W. Andrews. The Block-block Bootstrap: Improved Asymptotic Refinements. Cowles Foundation Discussion Papers 1370, Cowles Foundation for Research in Economics, Yale University, May 2002. URL `https://ideas.repec.org/p/cwl/cwldpp/1370.html`.

P. Hall and J. L. Horowitz. Bootstrap critical values for tests based on generalized-method-of-moments estimators. *Econometrica*, 64(4):891–916, 1996. ISSN 0165-0270. doi: 10.2307/2171849. URL `http://www.jstor.org/stable/2171849`.

A. Inoue and M. Shintani. Bootstrapping gmm estimators for time series. *Journal of Econometrics*, 133(2):531–555, 2006. URL `http://EconPapers.repec.org/RePEc:eee:econom:v:133:y:2006:i:2:p:531-555`.

S. N. Lahiri. Theoretical comparisons of block bootstrap methods. *The Annals of Statistics*, 27(1):386–404, 1999.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2017.

R. K. Thomas and W. J. Koss. Regional and national monthly, seasonal, and annual temperature weighted by area, 1895-1983. *Historical Climatology Series 4-3*, 1984.

United States Environmental Protection Agency	Office of Air Quality Planning and Standards Air Quality Assessment Division Research Triangle Park, NC	Publication No. EPA-454/R-18-001 April, 2018