Air pollution and daily deaths in California

S. Stanley Young, National Institute of Statistical Sciences, young@niss.org
Paul Fogel, Consultant, Paris, France, paul_fogel@hotmail.com

Abstract
The current environmental epidemiology paradigm is that air pollution is an acute killer. The EPA is relying on environmental observational data. There is a need to examine this claim. Daily data for PM2.5, ozone and deaths was obtained for eight California air basins and for years 2000-2012. Here we focus on Los Angeles County and the years 2007-2010. The data consists of three time series, deaths, ozone and PM2.5. Two JMP Addins were written. A JMP Addin was written to compute four types of time series filters/smoothers. A second JMP Addin was written to provide a new p-value plot useful for evaluation of multiple comparisons. Deviations from the three time series, PM2.5, ozone, and daily deaths were analyzed to see if increases in air pollution are associated with increases in death. No associations of air pollution with daily deaths in California were found in this data set.

1.0 Introduction
There are two reasons you should consider reading this paper. The technical reasons are in dealing with time series data and multiple testing. We present a JMP Addin useful for smoothing/filtering a time series to track the central nature of the data and examination of deviations from the central smoother. The methods are non-parametric, easy to understand and work for non-stationary time series as well as stationary. Multiple testing comes up and we present a new p-value plot JMP Addin that allows a visual assessment of possible effects. The non-technical aspect of the paper concerns the reliability of claims made about air pollution associated with acute deaths. This claim is central to the EPA's call to increase regulations on electric power plants. These regulations are expected to cost $270B/yr, or about $900/yr for every person in the US. Statisticians and data analysts are busy, but you might want to lift your head and take a look.

Let's go over some air pollution history. The Great Smog of London, 1952, is estimated to have killed four thousands or more people. The pollution disaster alerted all to the hazards of high levels of air pollution. A key paper published in the New England Journal of Medicine in 1993 asserted that much lower levels of air pollution caused deaths. Data used in that paper is not publicly available and we will comment later on that. In the US, there has been great progress in reducing air pollution over the last 30 years. See Figures 1 and 2. (Figures are given in Appendix 4.) However the EPA takes an aggressive position on the effects of air pollution on acute deaths; the then head of the EPA said to a congressional committee, “Air pollution can kill you right now.” Current literature is mixed on an association of air pollution with acute deaths. See Milojevic et al. (2014) and the references therein. This paper uses a very large data set and essentially they find no evidence for air pollution leading to acute heart attacks, one of the alleged causes on acute air pollution deaths. So we have something of a mystery. Does the current level of air pollution lead to acute deaths? Some say yes and others no. As researchers in this area typically don't make their data available it is difficult to decide. To address the questions we assembled from public sources a time series data set to look at the question.

Daily deaths and air pollution levels, as measured by PM2.5 and ozone, were obtained for the years 2000-2012 for the eight most populated California air basins. PM2.5 is composed of small particles, so small they will pass through a 2.5 micron grid. These particles are not chemically defined and consist of combustion products, dust and other things. When inhaled they can go deep into the lungs. Ozone is a reactive oxygen species, O3 rather than the normal O2 that supports our life. Here we report on our preliminary analysis of data from the South Coast air basin, which includes Los Angeles, for the years 2007-2010.

Of interest is the fact that California has a fire season where massive fire can greatly increase air pollution levels. See Figure 3 for a dramatic satellite picture of wild fires in Southern California. There can be dramatic spikes in the levels of PM2.5. These so called natural experiments can be used to test for the presumed acute death effects of air pollution. The term "natural experiment" is used when exposure to the event or intervention of interest occurs in the normal course of events and has not been manipulated by the researcher. As this exposure is not random care is needed in reporting and interpretation of evidence from natural experimental studies. Smaller spikes in air pollution can be used to test for low level effects of air pollution. We will focus on heart and lung deaths of people 65 and older as literature points to these people as being the most sensitive to air pollution effects.

This paper is divided into sections including Data, Methods, Analysis Strategy, and Discussion.

Data

Air pollution data for the eight California air basins was obtained from the California Air Resources Board web site, www.arb.ca.gov. Air pollution data is public and can be downloaded. Mortality data was obtained from California Department of Public Health, www.cdph.ca.gov. These public use files require payment of a fee and there are restrictions on disclosure.

Methods: **Base Methods**

A moving median can be used to track the central tendency of the data over time. The median is a measure of central tendency that should be minimally affected by extreme values. The JMP addin we use allows control over the number of time points used in the moving window and also has an option to exclude points from the center of the moving median. For example, 21-5 denotes a moving median over 21 points with the central five points excluded.

A moving average takes the average of k consecutive points, drops the first point and adds the next point in the time series and the average is recomputed.

A smoothing spline, Eubank (1999), is a function that varies in smoothness (or flexibility) according to a tuning parameter lambda (λ). As the value of λ decreases the model has more weight and the fit becomes more flexible and curved. As the value of λ increases, the fit becomes stiffer, finally approaching a straight line.

A partial correlation coefficient gives the correlation between two variables (e.g. air pollutants) when conditioning on one or more other variables, e.g. time. For example, what exactly is the correlation *Rxy.z* between variables *x* and *y* when accounting on variation in *z*? It is the correlation between the parts of *x* and *y* that are linearly uncorrelated with *z*. To obtain these parts of *x* and *y*, they are both regressed on *z*. The residuals of the regression are then the parts of *x* and *y* that are uncorrelated with *z*. The correlation between these residuals of *x* and *y* is the partial correlation between *x* and *y* when conditioning on *z*. The order of the partial correlation coefficient is determined by the number of variables it is conditioned on. For example, *Rxy.z* is a first-order partial correlation coefficient, because it is conditioned solely on one variable (*z*).

A p-value plot can be used to evaluate multiple statistical tests, Schweder and Spjøtvoll (1982). The standard p-value plot rank orders the p-values from smallest to largest and plots them against the integers 1, 2, 3, …. A different p-value plot was constructed by Dmitri Zaykin (personal communication) and coded into a JMP Script by Paul Fogel. In this plot the negative log10 is plotted against the expected values of p-values from a uniform distribution.

A JMP Addin Moving Median is described in Appendix 1.

Methods: Analysis Strategy
We use three time series, HeartLung daily deaths, PM2.5 daily values, and ozone daily values for the four years 2007-2010, 1,461 days. We would like to know if air pollution is associated with deaths. Death can lag exposure, so we examine lags of 0, 1, and 2 days. Partial correlations were computed for deaths, ozone, PM2.5, min and max temp. They showed no positive association of air pollution and deaths, Figure 4. A simple analysis of deaths/ozone is usually considered problematic as deaths increase in the winter while ozone goes down in the winter; both variables have seasonal effects, Figure 5 (a). So deaths and ozone are inversely related, Figure 5 (b). Again, see the partial correlations, Figure 4.

The usual practice is to consider the large time relationships as seasonal and ignore this inverse relationship. Deviations from the time trends for the three time series are used to examine possible relationships, Figure 6. For each time series we examine splines, moving averages and moving medians. As we will compute a deviation from this moving central value, we wanted that value to not be unduly influenced by the daily value. So we did a small study of methods that can be used to measure the central tendency the time trends. We discovered that there is a high correlation between each pair of methods, typically over 0.95. We wanted the correlation between the estimated value and the observed value to be low, implying that the fit was not being pulled to the observed value. Using this reasoning, a moving median of 21 days with a gap of five observations in the middle appeared marginally better than the other fitting methods, Table 1. For deaths and ozone all of the methods could be tuned to give satisfying results. A moving median of 21 days with five central days removed was used for deaths and ozone. PM2.5 has a non-stationary time series so we simply used the median value of PM2.5. Partial correlations were computed for min and max temp with deviations for deaths, ozone, PM2.5. They showed no positive association of air pollution and deaths, Figure 7.

Results

Our analysis is based on multiple 2-way plots, deaths (lags of 0, 1, 2 days) versus air pollution levels (ozone and PM2.5) for four years, 3x2x4=24 figures. All 24 figures are given in Appendix 2. Two of those two-way plots are given in Figure 8.

In Figure 8 (a), we see deaths (heart-lung deaths deviation from the moving median trend line) versus ozone levels and ozone (deviations trend lines). If ozone was causing deaths then we should see the data points flow from lower left to upper right. Non-parametric density contours are added to the figure as there is overprinting of the data points. A similar figure for PM2.5 is given in Figure 8 (b). Again, if there were ozone or PM2.5 related increase in deaths, then these contours would flow from lower left to upper right. In fact, data points appear to without any particular trend. From these two figures there is no apparent air pollution association with deaths. Appendix 2 gives all 24 figures.

Within each 2-way plot a regression was computed testing for a relationship between deaths and air pollution; 24 p-values resulted. These 24 p-values are displayed in two p-value plots, Figure 9 (a) and (b), standard and new, respectively.

Three p-values are nominally statistically significant, Appendix 3, but they do not support cause and effect relationships for the following reasons. First, the smallest p-value of 0.00346 does not replicate over the other three years. The p-values are 0.6514, 0.6418, 0.0035 and 0.2406 for the years 2007-2010 respectively. The other two nominally significant p-values do not replicate either. Using the False Discovery Addin by John Sall gives an adjusted p-value of 0.084 for the smallest p-value.

It is instructive to examine two of the “natural experiments” in the data set. See Figure 10 a,b,c. For example there is a dramatic increase in PM2.5 between days 250 and 350, Figure 10a, but daily deaths show no complimentary pattern, Figure 10b. A plot of deaths versus PM2.5 shows no pattern, Figure 10c.

We give a second natural experiment Figure 11a,b,c, days 450 to 600. We see a pronounced increase in PM2.5 from day ~520 to ~580, Figure 11a, with no apparent effect on deaths, Figure 11b. Again directly plot of deaths versus PM2.5 (deviations for both), Figure 11c, indicates no relationship.

Discussion

Access to data used in air pollution epidemiology papers is critical for effective examination/ scientific oversight of claims made in papers. Peng et al. (2006) made a call that environmental epidemiology data sets be made public. Still, access to data remains problematic. Cecil and Griffin (1985) commented, “As an abstract principle, the sharing of research data is a noble goal and meets with little opposition. However, when data sharing is attempted in a particular circumstance, the conflicting interests of the parties can thwart the exchange. A glance at the benefits and obstacles to data sharing ….. reveals the reason: few of the benefits and most of the burdens fall to the possessor of a data set.” I have personally requested over 50 air pollution data sets and received only one. The EPA has not provided data sets requested by Congress, Smith (2014). Here we resort to assembling our data set from public sources.

When an association is observed in an observational data set, cause and effect is possible. For this data set, no associations were found between air pollution and deaths so speculations on cause and effect are not warranted. Where associations have been found previously, they occurred in older individuals and appeared more likely due to lung or cardiovascular affects. Our analysis is focused on the most susceptible individuals and the most likely cause of deaths.

Multiple papers find no effect of air pollution on mortality in California. See papers cited in Young and Xia (2013).

Note that within a narrow timeframe, in our case the 21 day moving median for deaths, the effect of many unmeasured covariates is expected to be minimized, e.g age distribution, gender distribution, income, etc.

Let me remind the reader of several things. Data use for making claims in the area of air pollution is largely unavailable. We apparently have "trust me" science. Where data is available, claims cited by EPA are very often not supported, e.g. Young and Xia (2013). You do have some "skin in the game" as proposed regulations by the EPA are estimated to be very costly, $270B/year or ~$900/person/per year, which projects to a loss of 59 days of life expectancy, Young and Xia (2013).

In summary, seasonal effects were removed using 21-day moving medians to give time-local estimates of deaths and ozone air pollution. Increases in PM2.5 were erratic with large increases most likely related to wild fires. For PM2.5 we used deviations from the overall median level of PM2.5. Death lags of 0, 1, and 2 days were examined. Analyses were computed for two measures of air pollution, four years, and three lags, a total of 24 tests, looking for a consistent, acute effect of air pollution on mortality. A number of data visualization and statistical analyses support the statement that there were no consistent statistical effects of PM2.5 or ozone on acute deaths. We concluded that there is no evidence of an increase in acute deaths due to PM2.5 or ozone in South Coast air basin, Los Angeles for the years 2007-2010.

Data and JMP addins.

The p-value plot and moving median computation addins and the data set used in this paper can be obtained from the 1st author as can the data set used in this paper, young@niss.org. The partial correlation and adjusted p-values addins are available from SAS JMP.

References

Cecil JS, Griffin E. The role of legal polices in data sharing. in Sharing Research Data, Fienberg, Martin and Straff, Editors. National Academic Press. 1985.

Craig P, Cooper C, Gunnell D, Haw S, Lawson K, Macintyre S, Ogilvie D, Petticrew M, Reeves B, Sutton M, Thompson S. (2012) Using natural experiments to evaluate population health interventions: new Medical Research Council guidance *J Epidemiol Community Health* 66, 1182-1186.

Eubank RL. (1999) Nonparametric Regression and Spline Smoothing, Second Edition, Boca Raton, Florida: CRC.

Fogel P. (Personal communication) Moving Median Addin. See www.niss.org/MovingMedian.

Fogel P. (Personal communication) P-value Plot Addin. See www.niss.org/P-valuePlot.

Milojevic A, Wilkinson P, Armstrong B, et al. (2014) Short-term effects of air pollution on a range of cardiovascular events in England and Wales: case-crossover analysis of the MINAP database, hospital admissions and mortality. *Heart* 100: 1093-1098

Peng RD, Dominici F, Zeger SL. (2006) Reproducible epidemiologic research. *American Journal of Epidemiology* 163, 783-789.

Sall J. (accessed Aug 2014) False discovery rate p value. http://support.sas.com/demosdownloads/downarea_t4.jsp?productID=111423&jmpflag=Y

Schweder T, Spjøtvoll E. (1982), Plots of p-values to evaluate many tests simultaneously. *Biometrika* 69, 493–502.

Smith L. (2014) What is the EPA hiding from the public? 23Jun2014. *Wall Street Journal.*

Wikiapedia. (accessed Aug 2014) The great smog. http://en.wikipedia.org/wiki/Great_Smog

Wolfinger R. (2014) Partial Correlation diagram. https://community.jmp.com/docs/DOC-6173

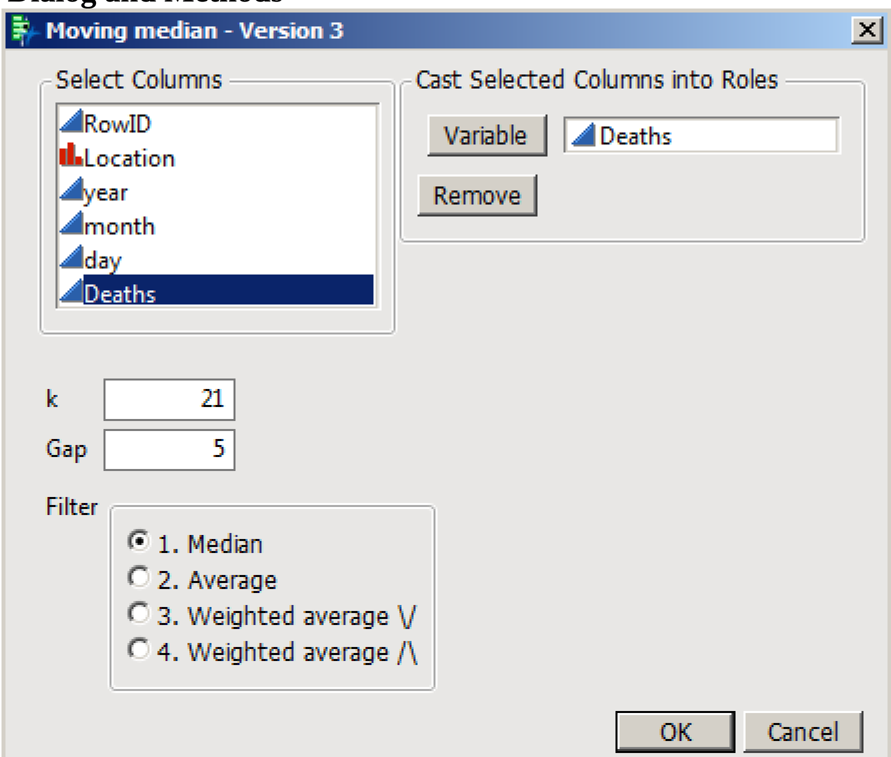
Young SS, Xia JQ. (2013) Assessing geographic heterogeneity and variable importance in an air pollution data set. *Statistical analysis and data mining.* 6, 375-386.

Zaykin D. (personal communication) A new p-value plot method.

Appendix 1. Moving Median Addin

Introduction
With time series data is it often useful to smooth the data to obtain a clearer picture of what is going on. A moving average is a commonly use smoothing function. Data points in the moving window can be weighted in various ways. Near points can be over-weighted or under-weighted. In some situations, the desire is to have a smoothing function that is not unduly weighted by outliers or even the focus point, the center point of the moving window. For example, the difference between the observation and the estimate of the focus point can be of interest.

Dialog and Methods

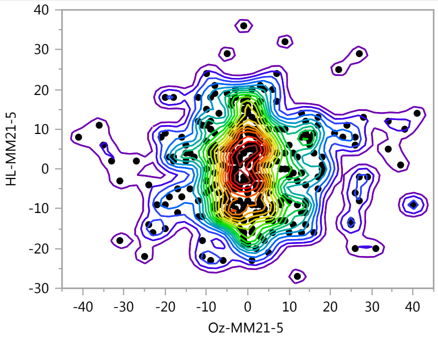


The time series variable is selected into Variable.
Window width is selected as k.
The default gap is 0, A gap width can be entered.

Four Filters are available. The **Median** computes the median within the window width, omitting observations in the gap. On clicking OK, a column of new data is added to the data set. **Average** compute the simple moving average. **Weighted average** \/ down-weights the observation near the center of the window. W**eighted average** /\ down-weights the observations toward the ends of the window.

Fit Group year=2007

Bivariate Fit of HL-MM21-5 By Oz-MM21-5 year=2007



.1 .2 .3 .4 .5 .6 .7 .8 .9 Quantile Density Contours
.1 .2 .3 .4 .5 .6 .7 .8 .9 Quantile Density Contours

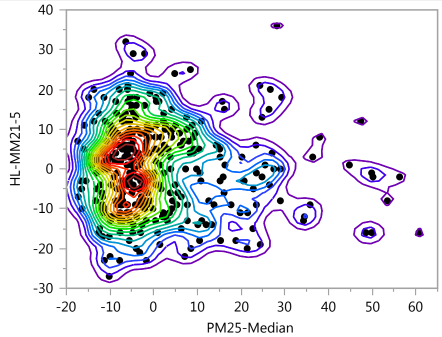
Nonparametric Bivariate Density

Variable	Kernel Std
Oz-MM21-5	2.286564
HL-MM21-5	2.031099

Nonparametric Bivariate Density

Variable	Kernel Std
Oz-MM21-5	2.286564
HL-MM21-5	2.031099

Bivariate Fit of HL-MM21-5 By PM25-Median year=2007



.1 .2 .3 .4 .5 .6 .7 .8 .9 Quantile Density Contours
.1 .2 .3 .4 .5 .6 .7 .8 .9 Quantile Density Contours

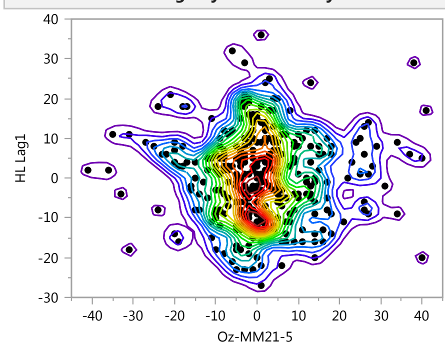
Nonparametric Bivariate Density

Variable	Kernel Std
PM25-Median	2.563506
HL-MM21-5	2.031099

Nonparametric Bivariate Density

Variable	Kernel Std
PM25-Median	2.563506
HL-MM21-5	2.031099

Bivariate Fit of HL Lag1 By Oz-MM21-5 year=2007



.1 .2 .3 .4 .5 .6 .7 .8 .9 Quantile Density Contours
.1 .2 .3 .4 .5 .6 .7 .8 .9 Quantile Density Contours

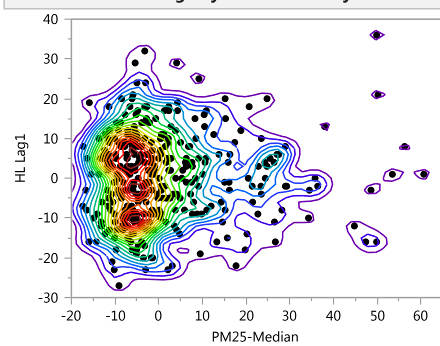
Nonparametric Bivariate Density

Variable	Kernel Std
Oz-MM21-5	2.286564
HL Lag1	2.029763

Nonparametric Bivariate Density

Variable	Kernel Std
Oz-MM21-5	2.286564
HL Lag1	2.029763

Bivariate Fit of HL Lag1 By PM25-Median year=2007

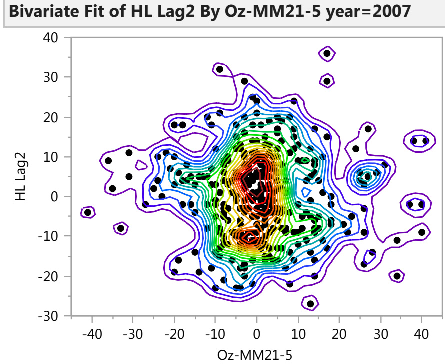


.1 .2 .3 .4 .5 .6 .7 .8 .9 Quantile Density Contours

Nonparametric Bivariate Density

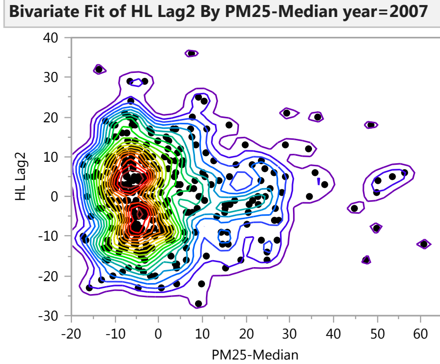
Variable	Kernel Std
PM25-Median	2.563939
HL Lag1	2.027847

Bivariate Fit of HL Lag2 By Oz-MM21-5 year=2007



.1 .2 .3 .4 .5 .6 .7 .8 .9 Quantile Density Contours

Bivariate Fit of HL Lag2 By PM25-Median year=2007



.1 .2 .3 .4 .5 .6 .7 .8 .9 Quantile Density Contours

Fit Group year=2007

Bivariate Fit of HL Lag2 By Oz-MM21-5 year=2007

Nonparametric Bivariate Density

Variable	Kernel Std
Oz-MM21-5	2.286564
HL Lag2	2.030861

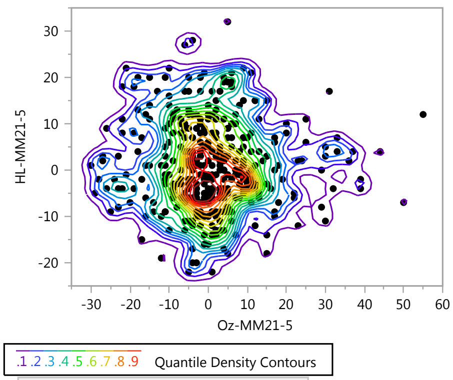
Bivariate Fit of HL Lag2 By PM25-Median year=2007

Nonparametric Bivariate Density

Variable	Kernel Std
PM25-Median	2.560557
HL Lag2	2.025171

Fit Group year=2008

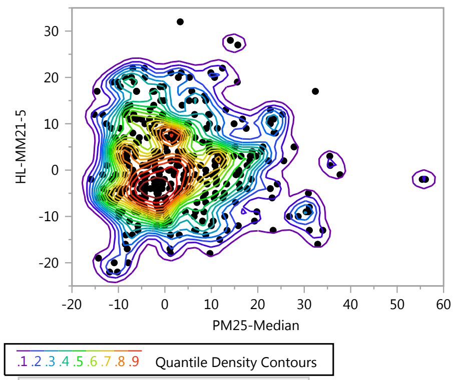
Bivariate Fit of HL-MM21-5 By Oz-MM21-5 year=2008



Nonparametric Bivariate Density

Variable	Kernel Std
Oz-MM21-5	2.362577
HL-MM21-5	1.775002

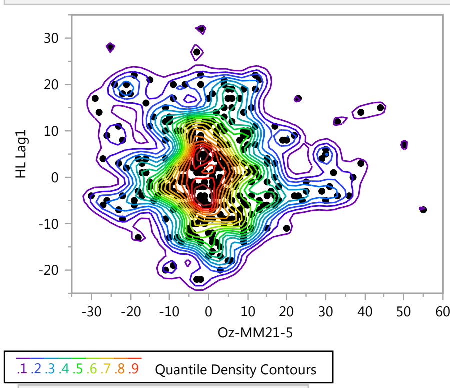
Bivariate Fit of HL-MM21-5 By PM25-Median year=2008



Nonparametric Bivariate Density

Variable	Kernel Std
PM25-Median	2.152363
HL-MM21-5	1.775002

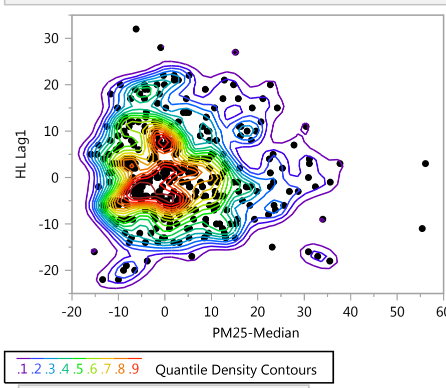
Bivariate Fit of HL Lag1 By Oz-MM21-5 year=2008



Nonparametric Bivariate Density

Variable	Kernel Std
Oz-MM21-5	2.362577
HL Lag1	1.775002

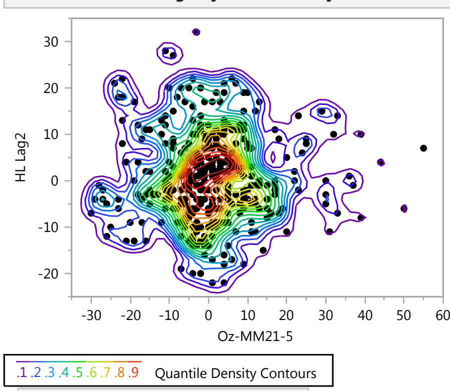
Bivariate Fit of HL Lag1 By PM25-Median year=2008



Nonparametric Bivariate Density

Variable	Kernel Std
PM25-Median	2.152363
HL Lag1	1.775002

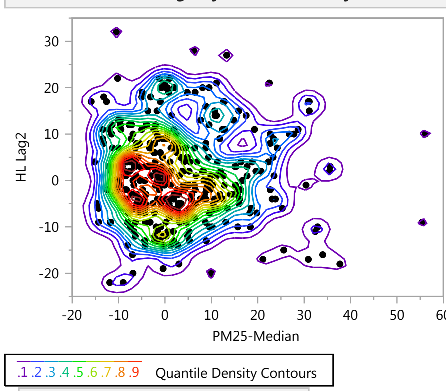
Bivariate Fit of HL Lag2 By Oz-MM21-5 year=2008



Nonparametric Bivariate Density

Variable	Kernel Std
Oz-MM21-5	2.362577
HL Lag2	1.774472

Bivariate Fit of HL Lag2 By PM25-Median year=2008



Nonparametric Bivariate Density

Variable	Kernel Std
PM25-Median	2.152363
HL Lag2	1.774472

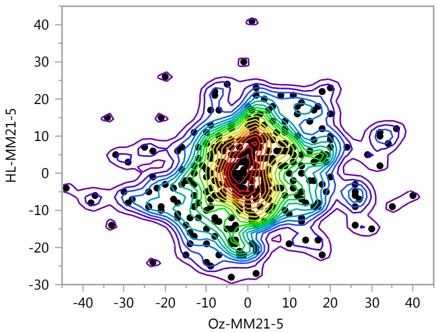
Fit Group year=2008

Bivariate Fit of HL Lag2 By Oz-MM21-5 year=2008

Bivariate Fit of HL Lag2 By PM25-Median year=2008

Fit Group year=2009

Bivariate Fit of HL-MM21-5 By Oz-MM21-5 year=2009

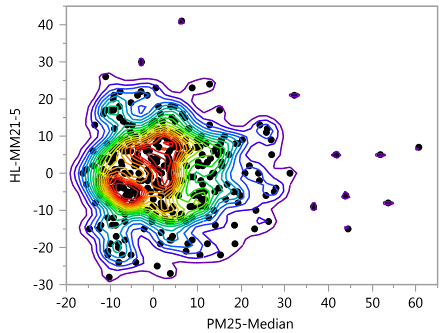


.1 .2 .3 .4 .5 .6 .7 .8 .9 Quantile Density Contours

Nonparametric Bivariate Density

Variable	Kernel Std
Oz-MM21-5	2.343152
HL-MM21-5	1.994588

Bivariate Fit of HL-MM21-5 By PM25-Median year=2009

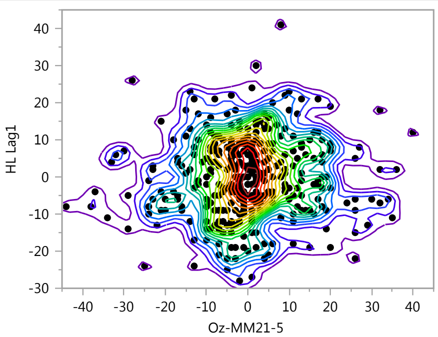


.1 .2 .3 .4 .5 .6 .7 .8 .9 Quantile Density Contours

Nonparametric Bivariate Density

Variable	Kernel Std
PM25-Median	2.182453
HL-MM21-5	1.994588

Bivariate Fit of HL Lag1 By Oz-MM21-5 year=2009

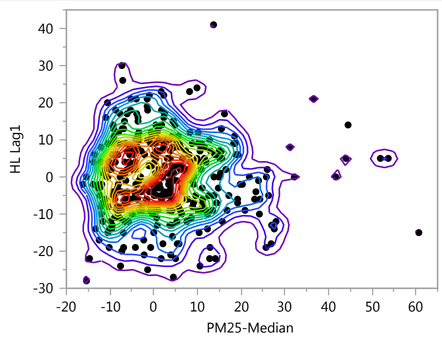


.1 .2 .3 .4 .5 .6 .7 .8 .9 Quantile Density Contours

Nonparametric Bivariate Density

Variable	Kernel Std
Oz-MM21-5	2.343152
HL Lag1	1.994086

Bivariate Fit of HL Lag1 By PM25-Median year=2009

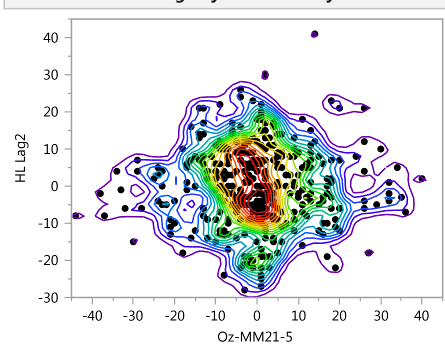


.1 .2 .3 .4 .5 .6 .7 .8 .9 Quantile Density Contours

Nonparametric Bivariate Density

Variable	Kernel Std
PM25-Median	2.182453
HL Lag1	1.994086

Bivariate Fit of HL Lag2 By Oz-MM21-5 year=2009

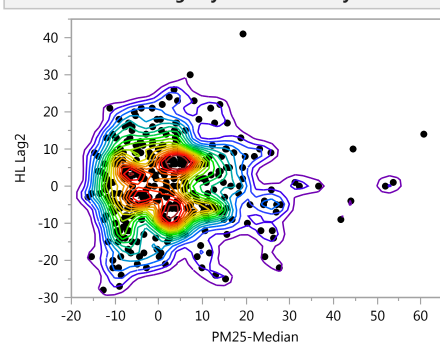


.1 .2 .3 .4 .5 .6 .7 .8 .9 Quantile Density Contours

Nonparametric Bivariate Density

Variable	Kernel Std
Oz-MM21-5	2.343152
HL Lag2	1.996389

Bivariate Fit of HL Lag2 By PM25-Median year=2009



.1 .2 .3 .4 .5 .6 .7 .8 .9 Quantile Density Contours

Nonparametric Bivariate Density

Variable	Kernel Std
PM25-Median	2.182453
HL Lag2	1.996389

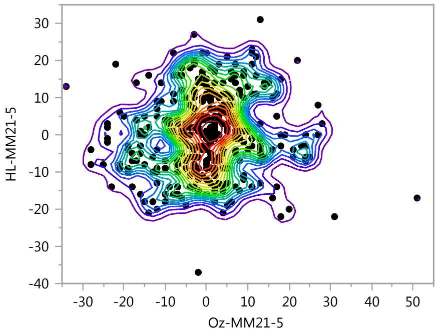
Fit Group year=2009

Bivariate Fit of HL Lag2 By Oz-MM21-5 year=2009

Bivariate Fit of HL Lag2 By PM25-Median year=2009

Fit Group year=2010

Bivariate Fit of HL-MM21-5 By Oz-MM21-5 year=2010

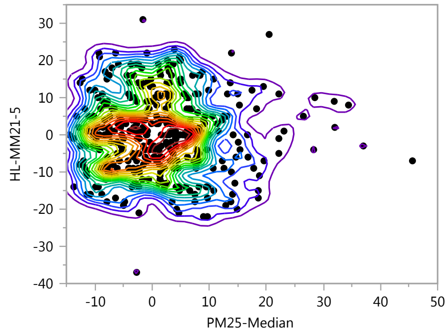


.1 .2 .3 .4 .5 .6 .7 .8 .9 Quantile Density Contours

Nonparametric Bivariate Density

Variable	Kernel Std
Oz-MM21-5	2.047579
HL-MM21-5	1.996005

Bivariate Fit of HL-MM21-5 By PM25-Median year=2010

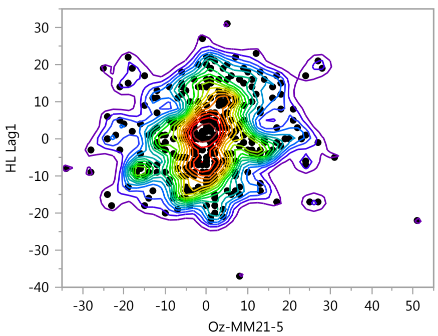


.1 .2 .3 .4 .5 .6 .7 .8 .9 Quantile Density Contours

Nonparametric Bivariate Density

Variable	Kernel Std
PM25-Median	1.748304
HL-MM21-5	1.996005

Bivariate Fit of HL Lag1 By Oz-MM21-5 year=2010

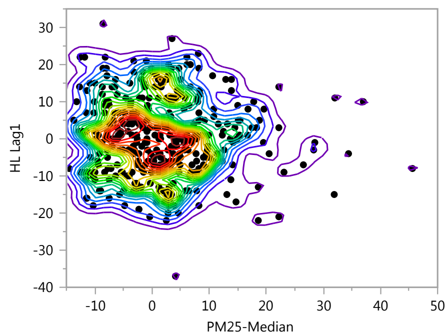


.1 .2 .3 .4 .5 .6 .7 .8 .9 Quantile Density Contours

Nonparametric Bivariate Density

Variable	Kernel Std
Oz-MM21-5	2.0509
HL Lag1	1.999699

Bivariate Fit of HL Lag1 By PM25-Median year=2010

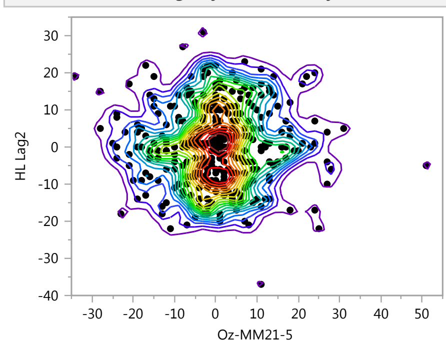


.1 .2 .3 .4 .5 .6 .7 .8 .9 Quantile Density Contours

Nonparametric Bivariate Density

Variable	Kernel Std
PM25-Median	1.748386
HL Lag1	1.999699

Bivariate Fit of HL Lag2 By Oz-MM21-5 year=2010

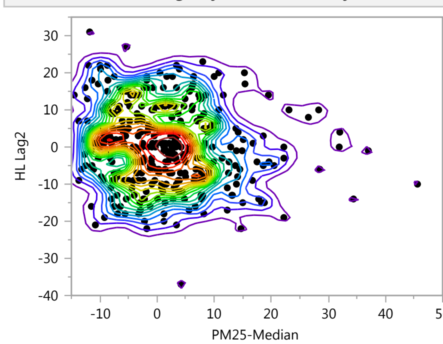


.1 .2 .3 .4 .5 .6 .7 .8 .9 Quantile Density Contours

Nonparametric Bivariate Density

Variable	Kernel Std
Oz-MM21-5	2.054767
HL Lag2	2.000608

Bivariate Fit of HL Lag2 By PM25-Median year=2010



.1 .2 .3 .4 .5 .6 .7 .8 .9 Quantile Density Contours

Nonparametric Bivariate Density

Variable	Kernel Std
PM25-Median	1.746737
HL Lag2	2.000608

False Discovery Rate Adjusted PValues

year	X	Y	PValue	FDR Adj PValue
2009	Oz-MM21-5	HL-MM21-5	0.00346	0.08311
2010	PM25-Median	HL Lag2	0.01561	0.18738
2010	Oz-MM21-5	HL Lag1	0.05165	0.4132
2010	Oz-MM21-5	HL-MM21-5	0.24062	0.92874
2009	Oz-MM21-5	HL Lag1	0.24875	0.92874
2010	PM25-Median	HL Lag1	0.30752	0.92874
2007	PM25-Median	HL-MM21-5	0.32514	0.92874
2008	Oz-MM21-5	HL Lag1	0.33015	0.92874
2007	PM25-Median	HL Lag1	0.36655	0.92874
2008	PM25-Median	HL Lag1	0.38697	0.92874
2009	PM25-Median	HL Lag2	0.53095	0.94132
2008	PM25-Median	HL Lag2	0.63924	0.94132
2008	Oz-MM21-5	HL-MM21-5	0.64178	0.94132
2007	Oz-MM21-5	HL-MM21-5	0.65141	0.94132
2008	PM25-Median	HL-MM21-5	0.69038	0.94132
2007	Oz-MM21-5	HL Lag2	0.79388	0.94132
2010	Oz-MM21-5	HL Lag2	0.8236	0.94132
2009	PM25-Median	HL Lag1	0.857	0.94132
2008	Oz-MM21-5	HL Lag2	0.88249	0.94132
2010	PM25-Median	HL-MM21-5	0.90016	0.94132
2007	Oz-MM21-5	HL Lag1	0.91081	0.94132
2009	Oz-MM21-5	HL Lag2	0.91697	0.94132
2009	PM25-Median	HL-MM21-5	0.9199	0.94132
2007	PM25-Median	HL Lag2	0.94132	0.94132

Appendix 4. Table and Figures

Table 1.

Table 1. Correlations of daily deaths with estimated values from various smoothing methods.

Deaths	1.0000
Deaths Spline 100	0.7670
Deaths Spline 100k	0.7404
Deaths MMed 21-0	0.7464
Deaths MMed 21-5	**0.7094**
Deaths MAv 21-0	0.7535
Deaths MAv 21-5	0.7212
Deaths M\/21-0	0.7197
Deaths M/\21-0	0.7781

The moving median with 21 points and 5 middle points excluded gives the lowest correlation with the central point in the window.

Figure 1. Ozone decrease over time. (National data)

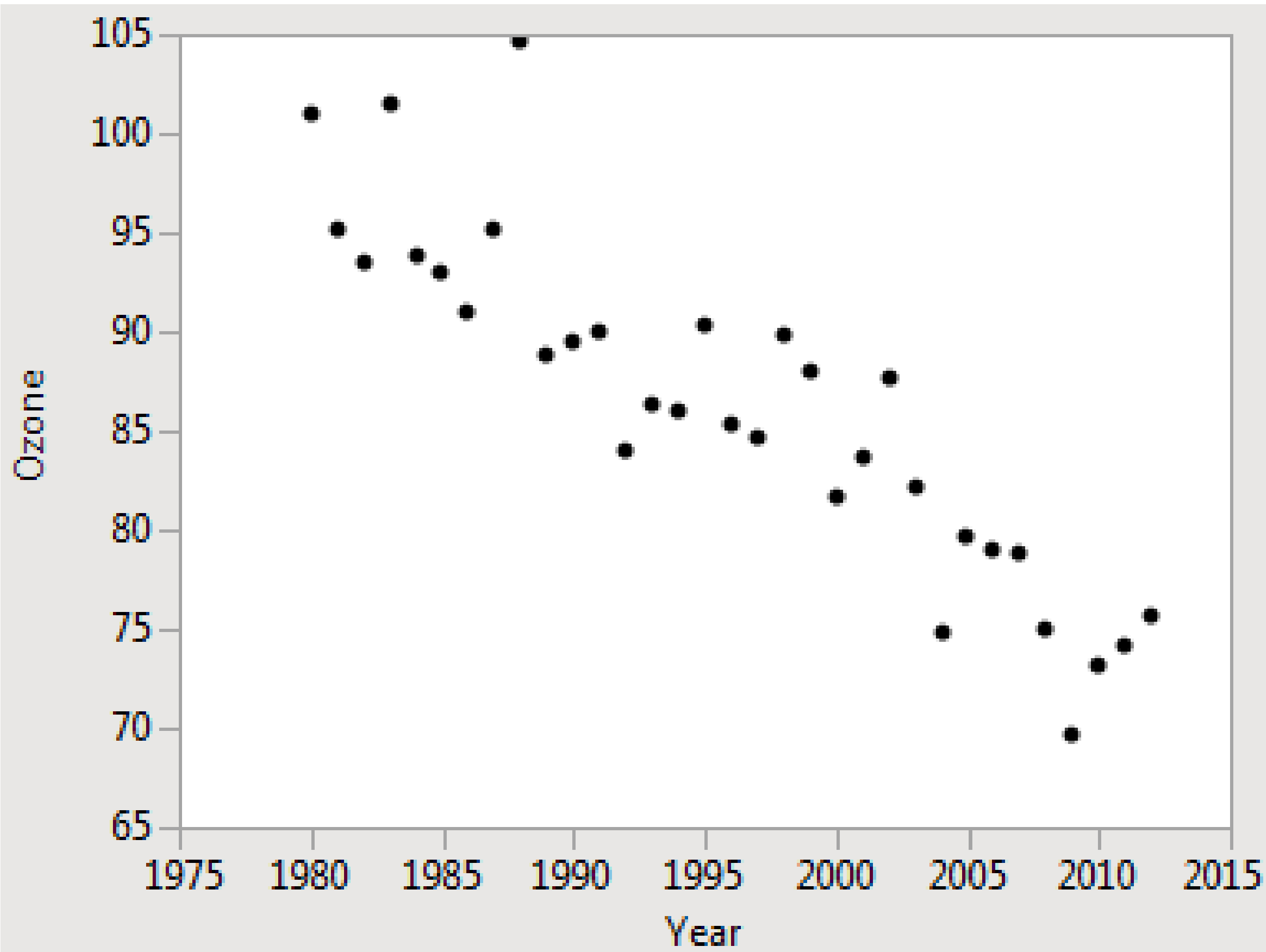


Figure 2. PM2.5 decrease over time. (National data)

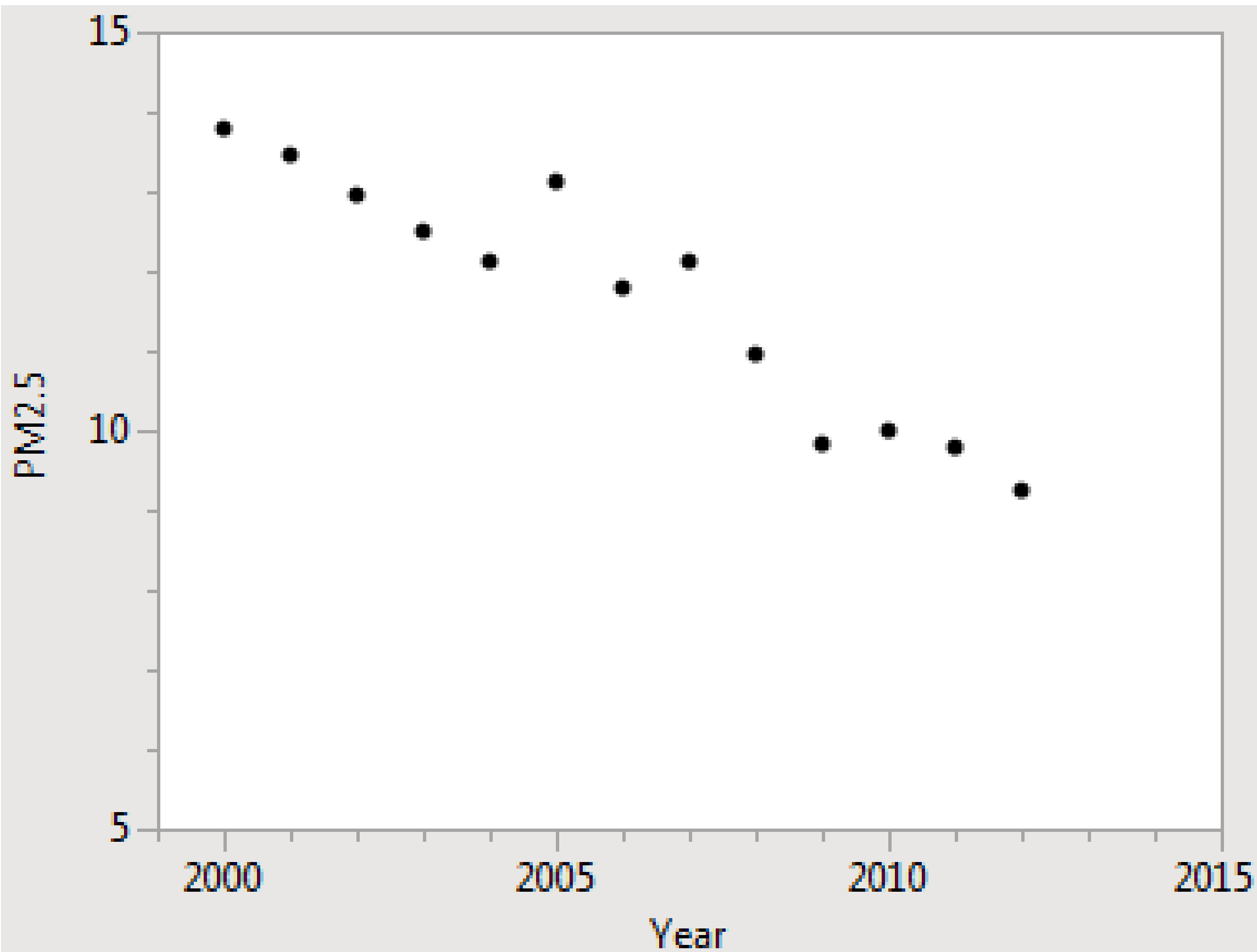
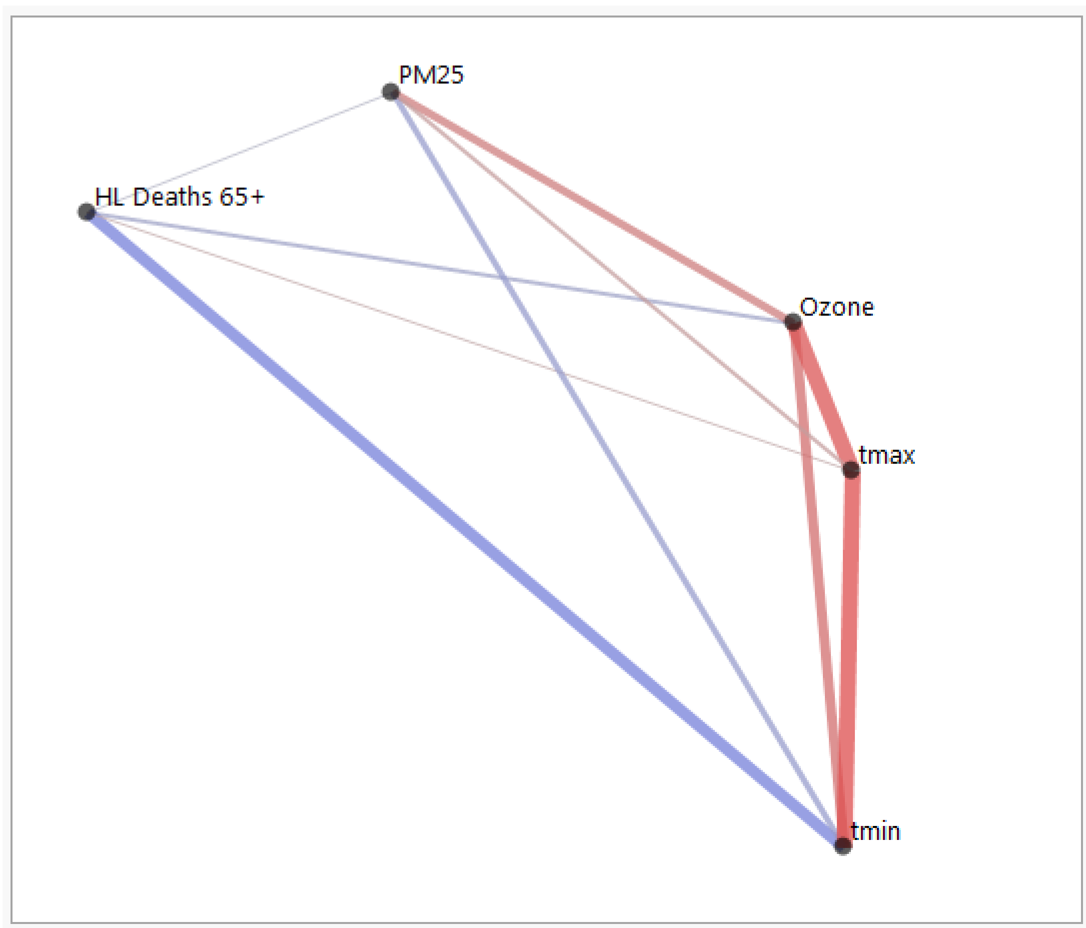


Figure 3. Satellite picture of wild fires in California.

The wildfires as seen from space on July 9, 2008.

Wild fires can dramatically increase PM2.5 levels.

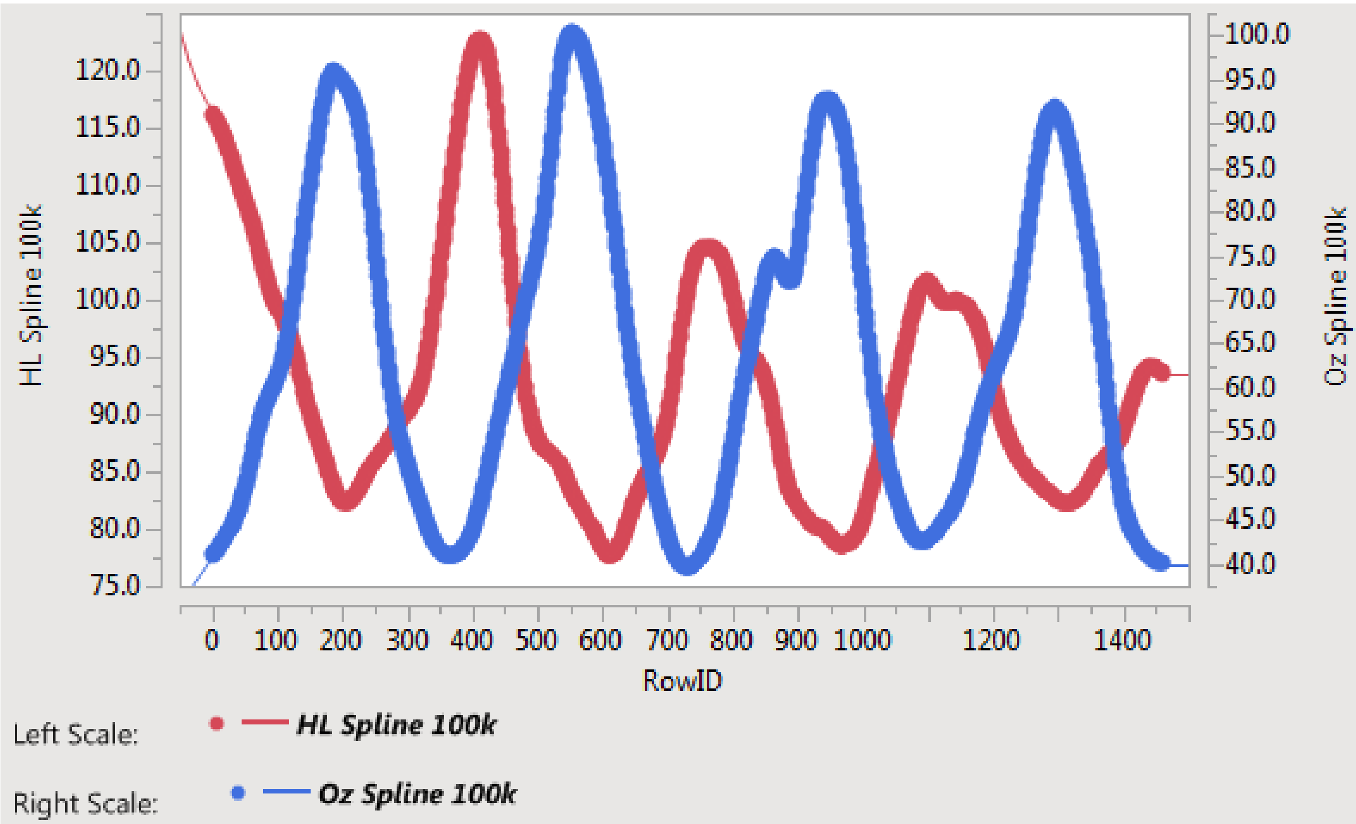
Figure 4. Partial correlations, original data.



Blue: negative partial correlation.
Red : positive partial correlation.

Figure 5. Deaths and ozone are out of phase.

(a)



(b)

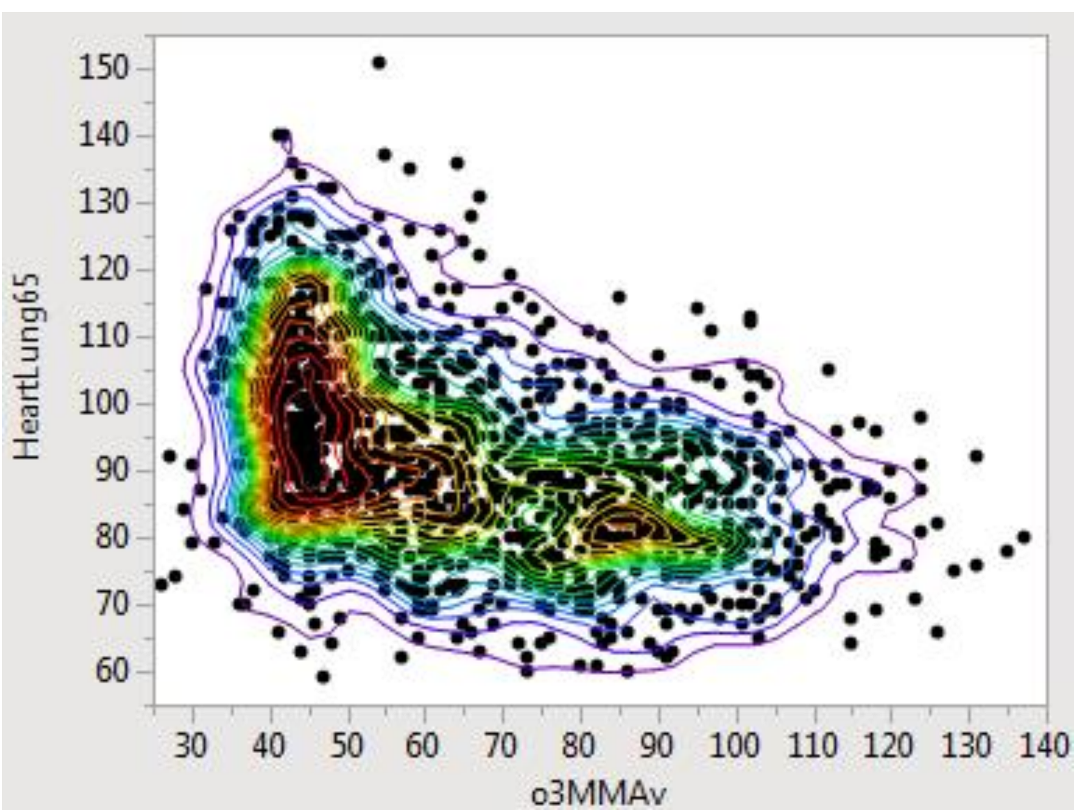
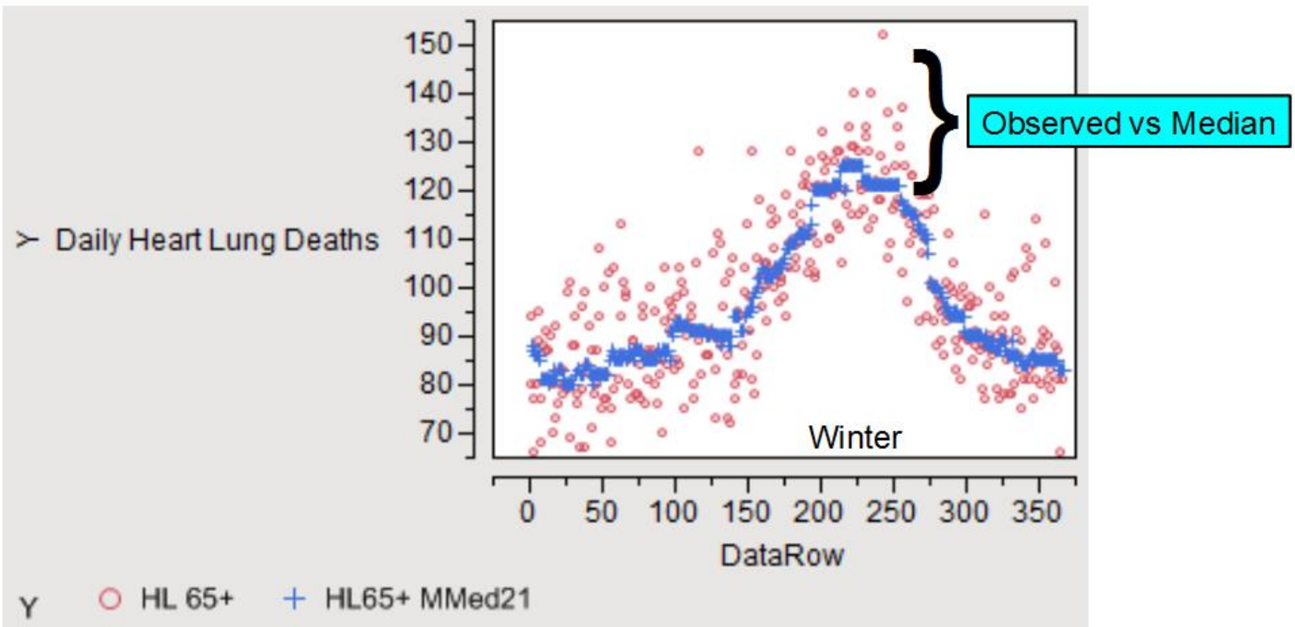
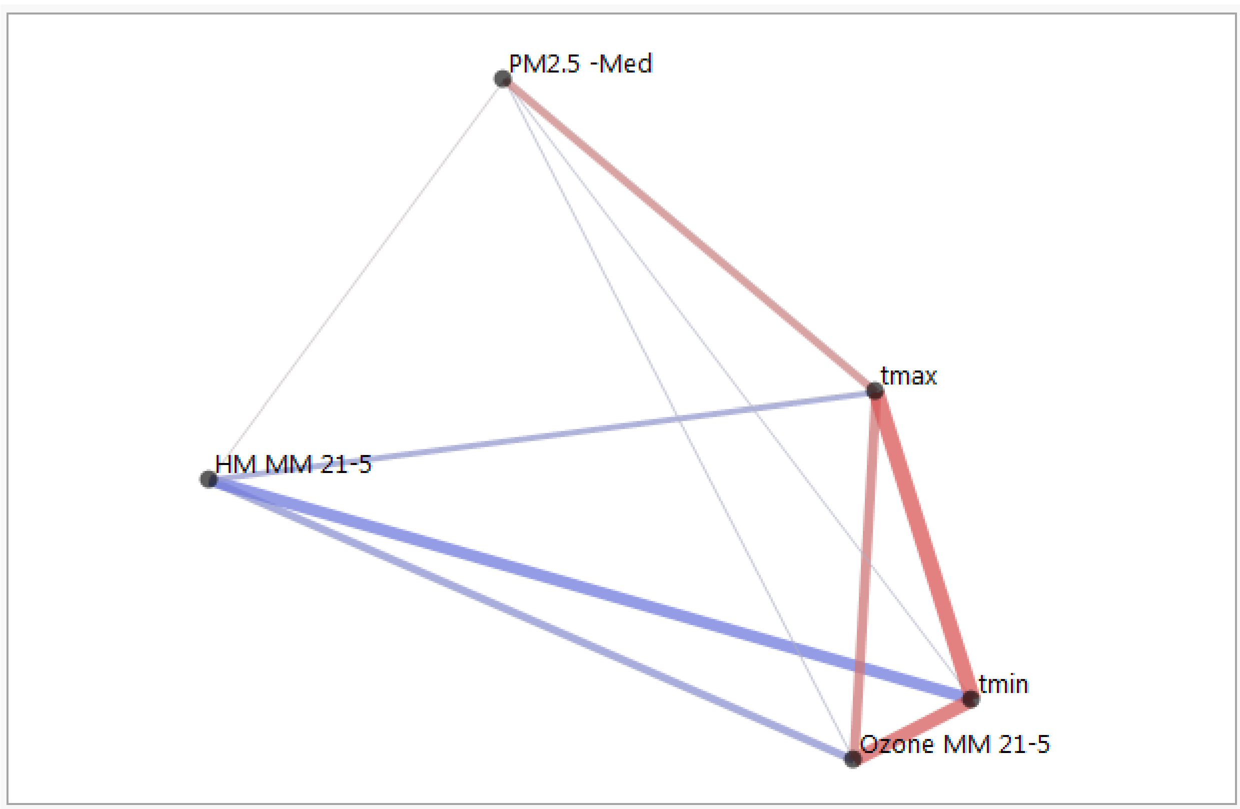


Figure 6. Deviations from moving median.



The moving median in blue tracks the center of the distribution at that day. The red "O" give the number of deaths on each day. The difference between the median and observation is used for analysis. The seasonal time trend is removed.

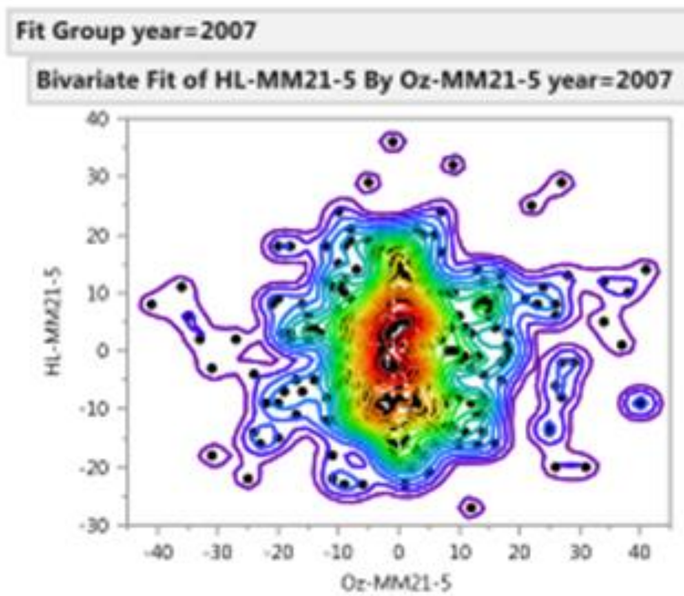
Figure 7. Partial correlations computed on deviations.



The deviations of daily heart/lung deaths from the median trend are negative for ozone and PM2.5. Max and min temperature are positively correlated. Ozone is positively correlated with temperature.

Figure 8.Deaths vs air pollution, ozone and PM2.5.

(a)



(b)

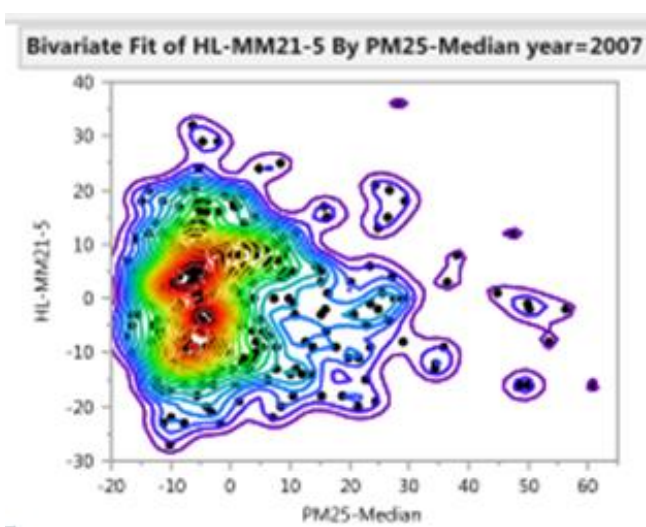
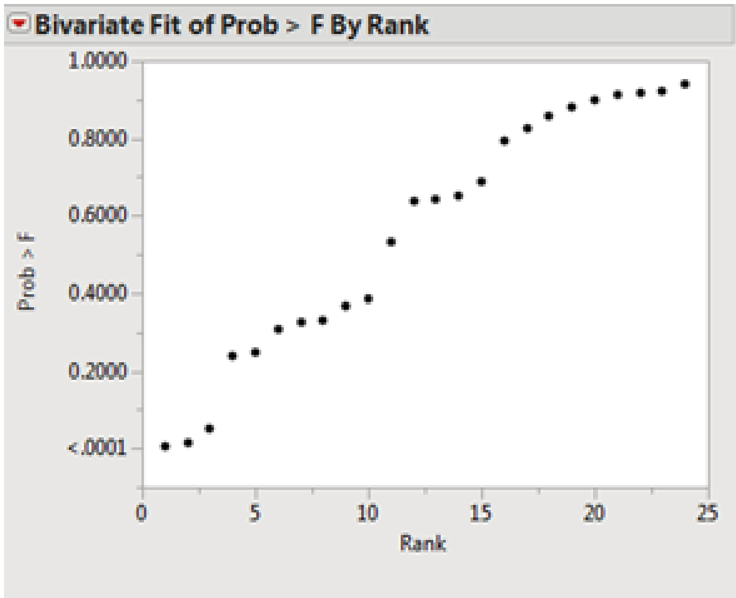
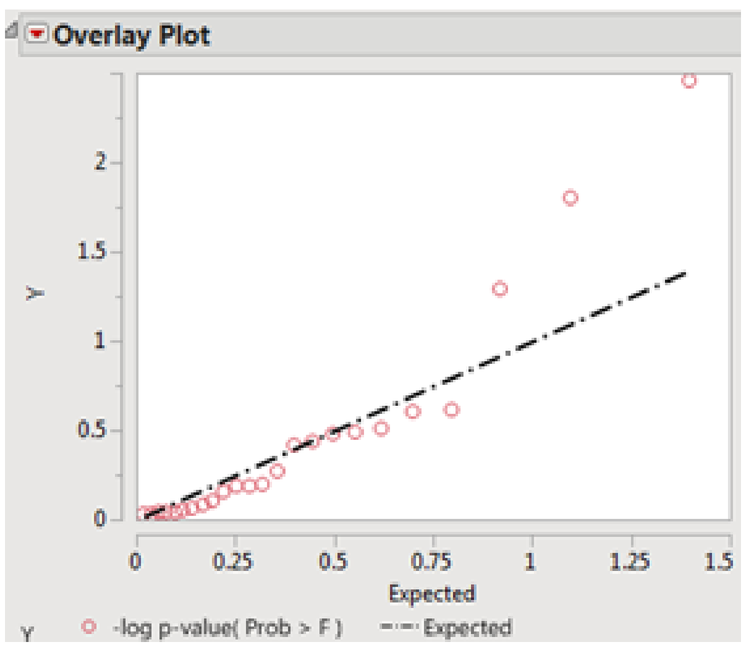


Figure 9. P-value plots.





NB: Using multiplicity adjusted p-values none of the three small p-values is Significant.

Figure 11. PM2.5 and deaths versus time. Deaths versus PM2.5.

(a)

(b)

(c)

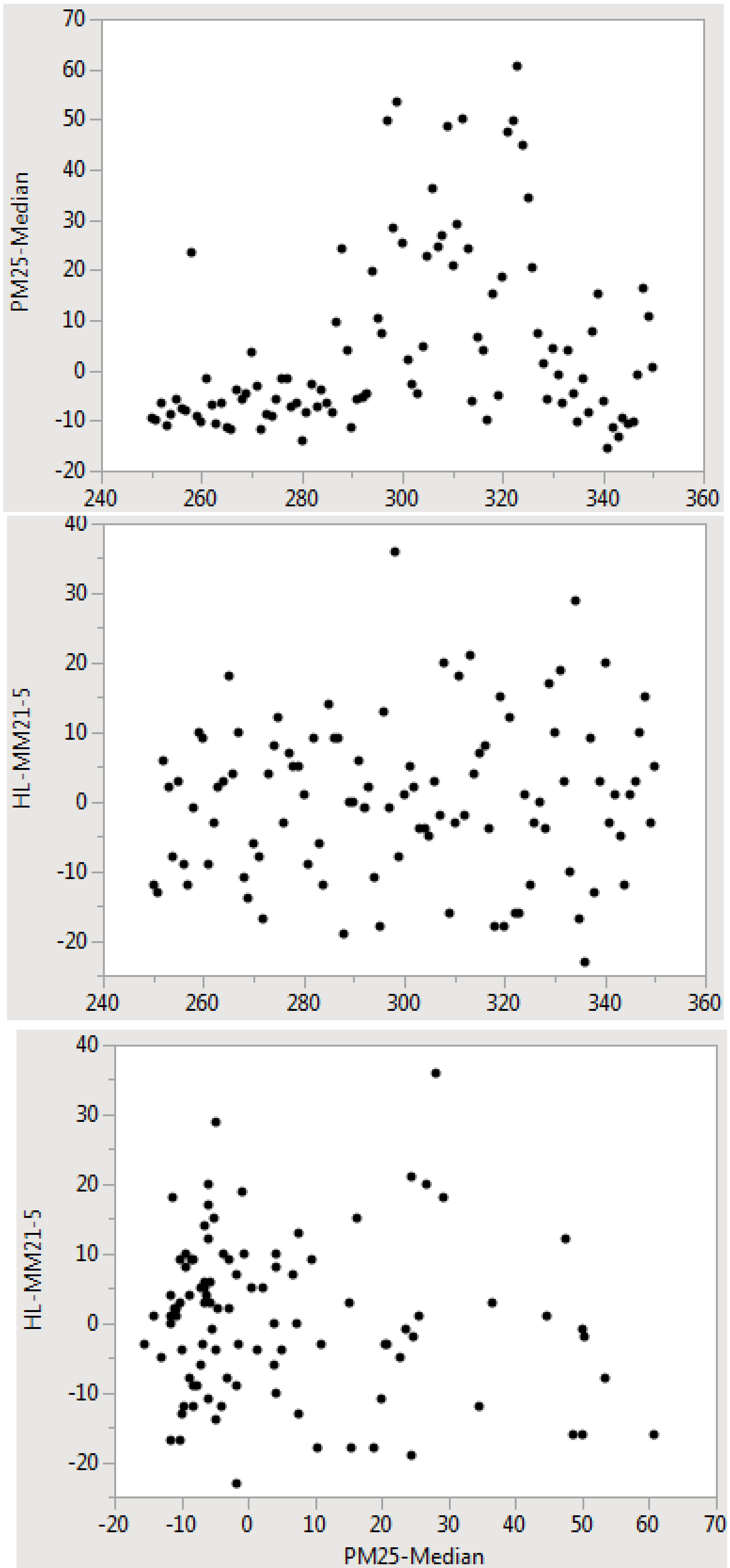


Figure 10. PM2.5 and deaths versus time. Deaths versus PM2.5.

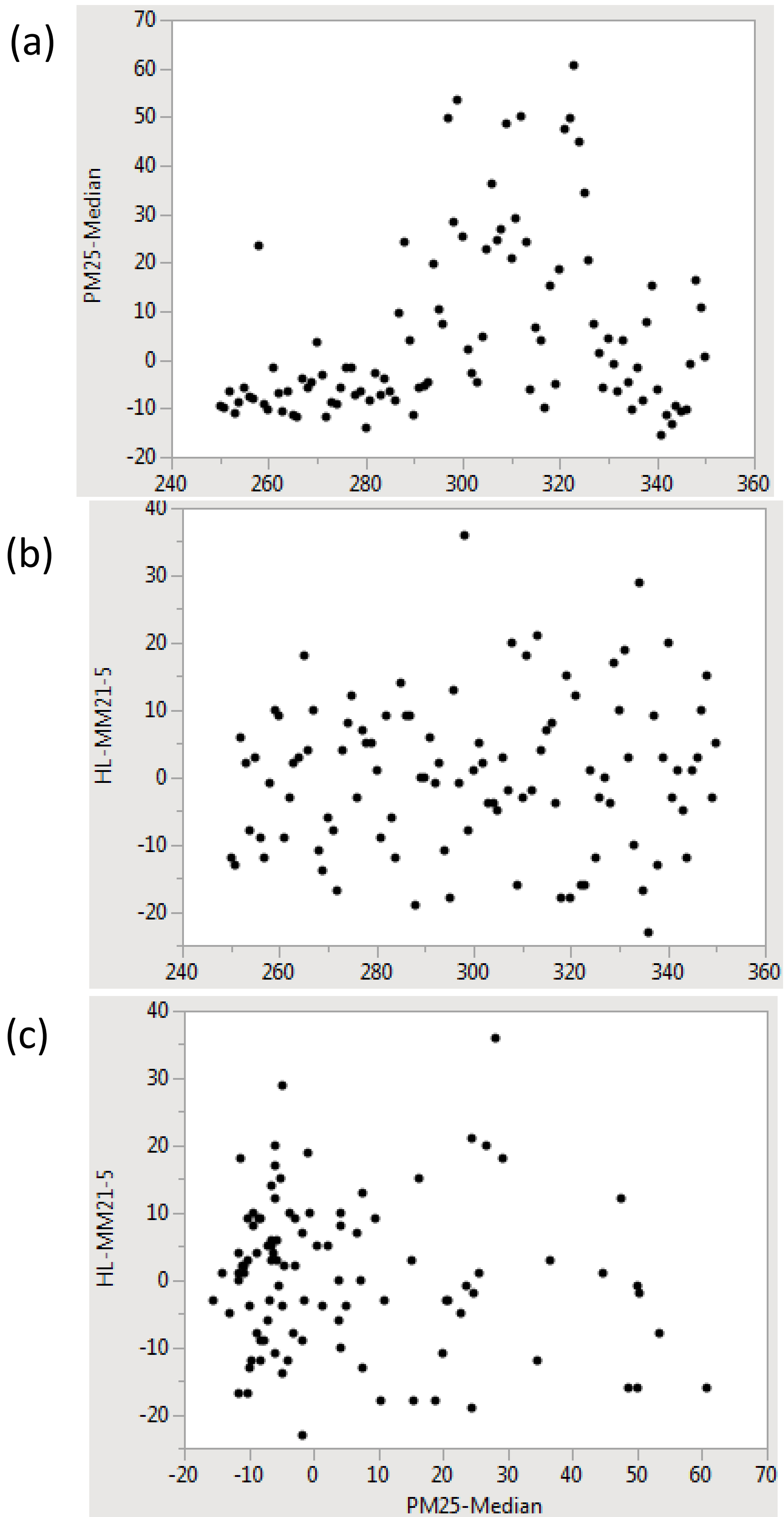


Figure 11. PM2.5 and deaths versus time. Deaths versus PM2.5.