Title
Month, Weekday, and Hour Effects in the Lebec Air Monitor Data

Author
Jan de Leeuw

Publication Date
2011-10-25

MONTH, WEEKDAY, AND HOUR EFFECTS IN THE LEBEC AIR MONITOR DATA

JAN DE LEEUW

Abstract. Data collected between February 2006 and February 2007 with an O3 and PM-2.5 monitor in Lebec, California are analyzed. In this paper we analyze the data using simple least squares imputation, and an additive main effect model to look at the effects of mont-of-the-year, day-of-the-week, and hour-of-the-day.

Date: June 1, 2008.

Contents

List of Figures

1. Introduction

At the request of the *Mountain Communities Town Council* the *California Air Resources Control Board* placed a mobile air pollution monitor at Peace Valley Road in Lebec. The monitor gave hourly measurements of both O3 (ozone) and PM-2.5 (fine particulate matter) between February 2006 and February 2007. Since Lebec already is in the San Joaquin Valley, which is a non-attainment area for ozone and particulate matter, we expect to see fairly high levels of pollution. It is interesting to single out Lebec, because it is at an altitude of over 4000 feet, in a rural area without much agricultural or industrial activity. On the other hand, it is adjacent to I-5, which has an average annual daily traffic count of 70,000 cars, of which 20,000 are trucks, with about 80% of the trucks having 5+ axels. The percentage of cars that are trucks is close to 30%, and it is growing. Lebec is also in the southern part of the San Joaquin Valley, where pollution from the northern part accumulates.

We analyzed the data with descriptive techniques in De Leeuw [2007], emphasizing plots, and violations of the federal and state air quality standards. One problem we ran into is that there is a fair amount of missing data (6% for ozone and 13% for PM). Also, the design is not completely balanced, in the sense that there are more data for some months or days than for others. It seems useful to take both missing data and unbalancedness into account in our presentation of the data.

2. Technique

The data are organized in two data-frames, with days as rows and hours as columns. For ozone the data matrix is 372×24, for PM it is 352×24. Sometimes whole days are missing, sometimes hours are missing within + a day. We approximate the data with a model of the form

$$y_{mdh} \approx \mu + \alpha_m + \beta_d + \gamma_h. \tag{1}$$

Thus the measured ozone or PM at month m and weekday d and hour h is approximated by the sum of a constant μ plus a month-parameter α_m, a day parameter β_d, and an hour parameter γ_h. Parameters are estimated by least squares, i.e. we minimize

$$\sigma(\mu, \alpha, \beta, \gamma) = \sum_{i=i}^{n} \sum_{h=1}^{24} w_{ih}(y_{ih} - (\mu + \alpha_{m(i)} + \beta_{d(i)} + \gamma_h))^2. \tag{2}$$

Here n is the number of days, and $m(i)$ and $d(i)$ are the month and weekday of day i. We use weights w_{ih} to indicate missing data. If hour h of day i is missing then $w_{ih} = 0$, otherwise $w_{ih} = 1$. In the appendix we provide `R` code to compute estimates that minimize the residual sum of squares (2). Note that (1) can also be used to impute missing values.

3. Ozone

In Figure 1 we plot ozone for all time points, as 372 curves on the 24 hours. The hourly trend is clear, but of course generally the plot is a mess.

Insert Figure 1 about here

The predicted values using the parameter estimates obtained by minimizing Equation (2) are plotted in Figure 2. Clearly they consist of 372 parallel curves all showing the same trend over hours.

Insert Figure 2 about here

The effects of month, weekday, and hour are plotted in Figures 3,4, and 5. The effects on month-of-the-year and hour-of-the-day are quite substantial, each of about ± 0.015 ppm, with the summer months and the midday hours being considerably higher in ozone pollution. The effect of day-of-the-week is minimal, only about

$\pm 0.003\ ppm$. There is a small higher effect in the weekend, which is a different conclusion from De Leeuw [2007], where the average maximum ozone level was in the middle of the week.

Insert Figure 3 about here

Insert Figure 4 about here

Insert Figure 5 about here

4. PM-2.5

The situation for PM-2.5 is quite different from that of ozone in many respects. The curves for the 352 days are plotted in Figure 6. We have truncated the vertical axis, because there any many outliers (some of them because of the Day Fire in September). Not many systematic effects can be found in this plot.

Insert Figure 6 about here

The 352 parallel curves with predicted values, using the parameter estimates obtained by minimizing Equation (2), are plotted in Figure 7. They have some small bumps, but otherwise they seem to be rather seriously over-smoothed.

Insert Figure 7 about here

The effects of month, weekday, and hour are plotted in Figures 8,9, and 10. The effect on month-of-the-year is quite substantial. It is about $\pm 5\ \mu g/m^3$, with the winter months being lowest and spring and fall being highest. The effect of day-of-the-week has a smaller effect of about $\pm 2\ \mu g/m^3$, with the weekend being lowest and the midweek highest. Hour-of-the-day is also $\pm 2\ \mu g/m^3$, but quite irregular. It seems there are peaks around 7am and 4pm, indicating perhaps commuter rush hour, and another peak around 1m, indicating perhaps increased truck traffic.

Insert Figure 8 about here

Insert Figure 9 about here

Insert Figure 10 about here

5. Conclusion

The Lebec Air Monitor data show systematic effects of month-of-the-year, day-of-the-week, and hour-of the day. This is true for both ozone and PM-2.5. Although we cannot draw any firm conclusions, it seems reasonable to assume that the effects are caused by a mix of meteorology (temperature) and traffic. The relationships with traffic counts and ambient temperature will be investigated as soon as we have the necessary data.

References

J. De Leeuw. The Lebec Air Monitor. Technical report, UCLA Center for Environmental Statistics, 2007. URL `http://gifi.stat.ucla.edu/CES/airpollution/lebec.pdf`.

Appendix A. R Code

A.1. Imputation. This is the R function that estimates the parameters and does the imputation.

```
imputeLebec<-function(mat,eps=1e-10,itmax=1000,verbose=FALSE) {
dnames<-as.Date(rownames(mat),"%m-%d-%y")
wnames<-weekdays(dnames,abbreviate=TRUE)
mnames<-months(dnames,abbreviate=TRUE)
impdat<-apply(mat,2,function(x) ifelse(is.na(x),0,x))
n<-nrow(mat); oldssq<-Inf; itel<-1
ga<-gd<-gm<-rep(0,n); ah<-rep(0,24)
repeat {
        da<-impdat-outer(gd+gm,ah,"+")
        aa<-mean(da)
        ga<-rep(aa,n)
        dd<-impdat-outer(ga+gm,ah,"+")
        ad<-rowSums(apply(dd,2,function (x) tapply(x,wnames,mean)))/24
        gd<-ad[wnames]
        dm<-impdat-outer(ga+gd,ah,"+")
        am<-rowSums(apply(dm,2,function (x) tapply(x,mnames,mean)))/24
        gm<-am[mnames]
        dh<-impdat-matrix(ga+gd+gm,n,24)
        ah<-apply(dh,2,mean)
        pred<-outer(ga+gd+gm,ah,"+")
        for (i in 1:24) impdat[,i]<-ifelse(is.na(mat[,i]),pred[,i],mat[,i])
        newssq<-sum((impdat-pred)^2)
        if (verbose)
                cat("Iteration: ",formatC(itel,digits=3,width=3),
                        "Previous Loss: ",formatC(oldssq,digits=10,width=20,
                            format="f"),
                        "Current Loss: ",formatC(newssq,digits=10,width=20,
                            format="f"),
                        "\n")
        if (((oldssq-newssq) < eps) || (itel == itmax)) break()
                oldssq<-newssq; itel<-itel+1
        }
return(list(aa=aa,ad=ad,am=am,ah=ah,impdat=impdat,pred=pred,itel=itel,ssq=
    newssq))
}
```

A.2. Plots. And this R does the calculations and makes the plots.

```
dlab<-c("Sun","Mon","Tue","Wed","Thu","Fri","Sat")
mlab<-c("Jan","Feb","Mar","Apr","May","Jun","Jul","Aug","Sep","Oct","Nov","Dec")
pdf("ozfda.pdf")
plot(1:24,leboz[1,],type="l",ylim=c(0,.11),col="BLUE",axes=FALSE,xlab="hours",ylab="ppm")
for (i in 1:372) lines(1:24,leboz[i,],col="BLUE")
axis(1,at=1:24,labels=as.character(0:23)); axis(2)
dev.off()
ozo<-imputeLebec(leboz)
pdf("ozimp.pdf")
plot(ozo$pred[1,],type="l",ylim=c(0,.08),col="BLUE",axes=FALSE,xlab="hours",ylab="ppm")
for (i in 1:372) lines(1:24,ozo$pred[i,],col="BLUE")
axis(1,at=1:24,labels=as.character(0:23)); axis(2)
dev.off()
pdf("ozimph.pdf")
plot(ozo$ah,type="l",col="RED",xlab="hours",ylab="ppm",axes=FALSE,lwd=3)
axis(1,at=1:24,labels=as.character(0:23)); axis(2)
dev.off()
pdf("ozimpd.pdf")
plot(ozo$ad[dlab],type="l",col="RED",xlab="days",ylab="ppm",axes=FALSE,lwd=3)
axis(1,at=1:7,labels=dlab); axis(2)
dev.off()
pdf("ozimpm.pdf")
plot(ozo$am[mlab],type="l",col="RED",xlab="months",ylab="ppm",axes=FALSE,lwd=3)
axis(1,at=1:12,labels=mlab); axis(2)
dev.off()

pdf("pmfda.pdf")
plot(1:24,lebpm[1,],type="l",ylim=c(0,40),col="BLUE",axes=FALSE,xlab="hours",ylab="mg/m3")
for (i in 1:352) lines(1:24,lebpm[i,],col="BLUE")
axis(1,at=1:24,labels=as.character(0:23)); axis(2)
dev.off()
pmo<-imputeLebec(lebpm)
pdf("pmimp.pdf")
plot(pmo$pred[1,],type="l",ylim=c(0,40),col="BLUE",axes=FALSE,xlab="hours",ylab="mg/m3")
for (i in 1:352) lines(1:24,pmo$pred[i,],col="BLUE")
axis(1,at=1:24,labels=as.character(0:23)); axis(2)
dev.off()
```

```
pdf("pmimph.pdf")
plot(pmo$ah, type="l",col="RED",xlab="hours",ylab="mg/m3",axes=FALSE,lwd=3)
axis(1,at=1:24,labels=as.character(0:23)); axis(2)
dev.off()
pdf("pmimpd.pdf")
plot(pmo$ad[dlab], type="l",col="RED",xlab="days",ylab="mg/m3",axes=FALSE,lwd=3)
axis(1,at=1:7,labels=dlab); axis(2)
dev.off()
pdf("pmimpm.pdf")
plot(pmo$am[mlab], type="l",col="RED",xlab="months",ylab="mg/m3",axes=FALSE,lwd=3)
axis(1,at=1:12,labels=mlab); axis(2)
dev.off()
```

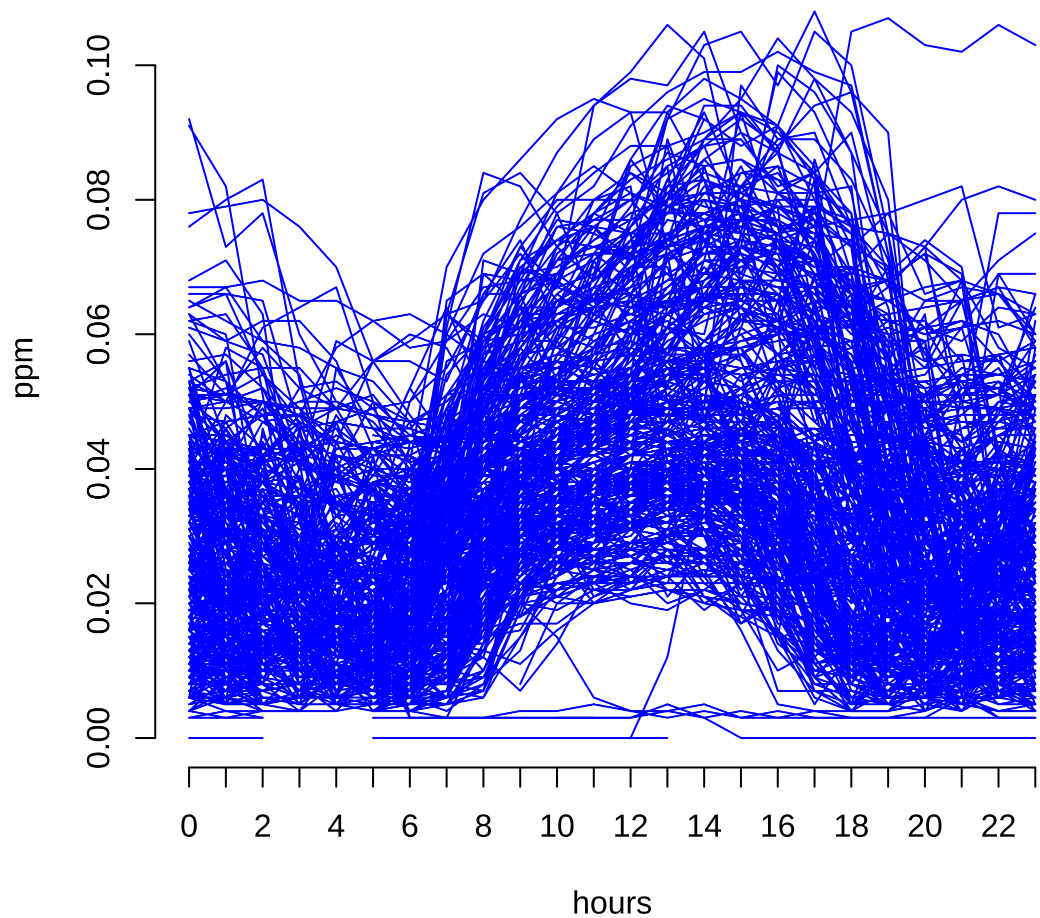

FIGURE 1. Ozone Curves

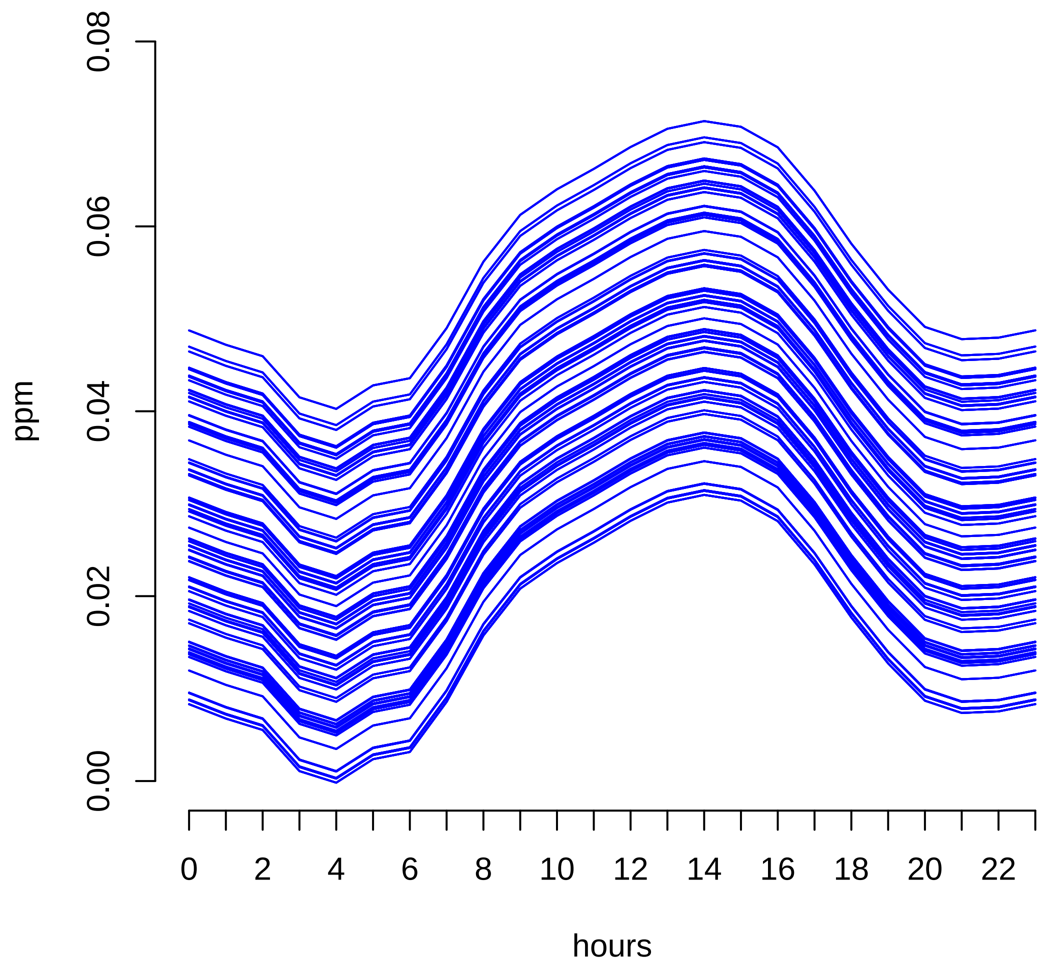

FIGURE 2. Ozone Curves, Smoothed

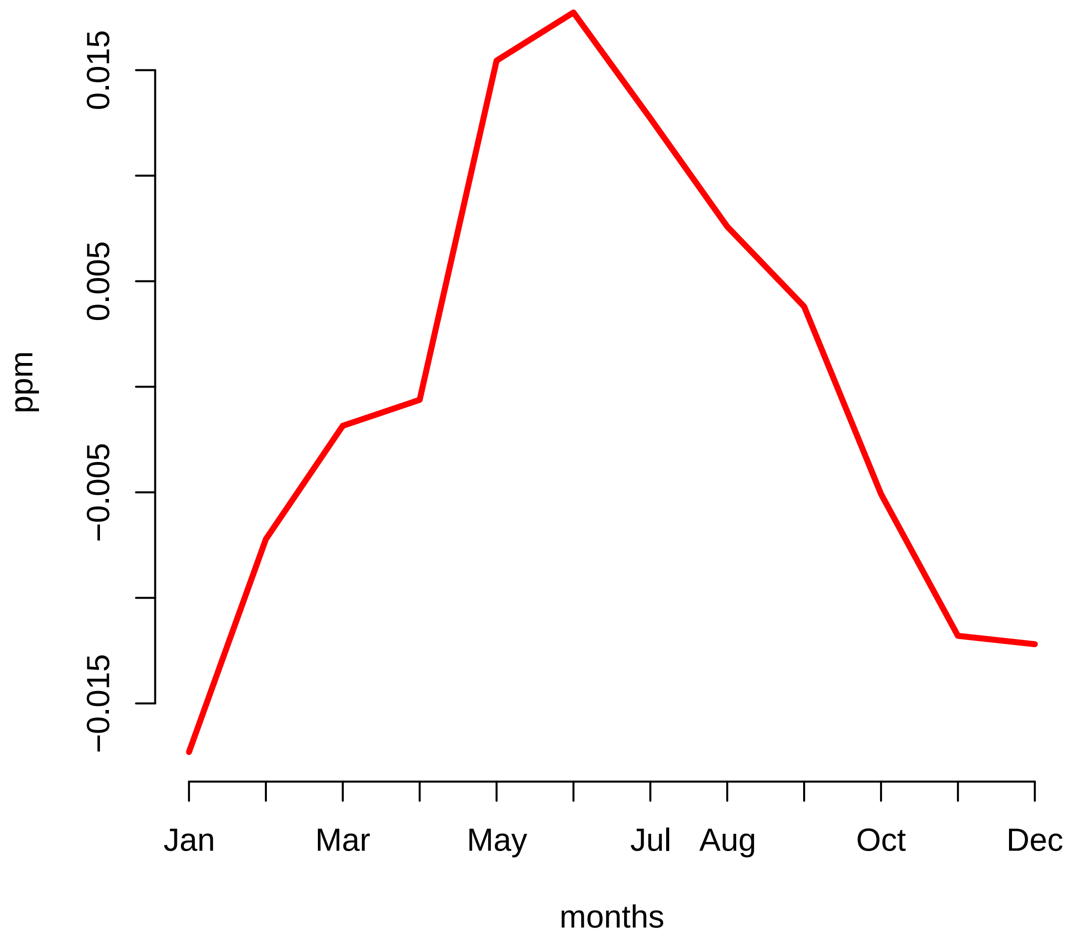

FIGURE 3. Ozone Effect of Month

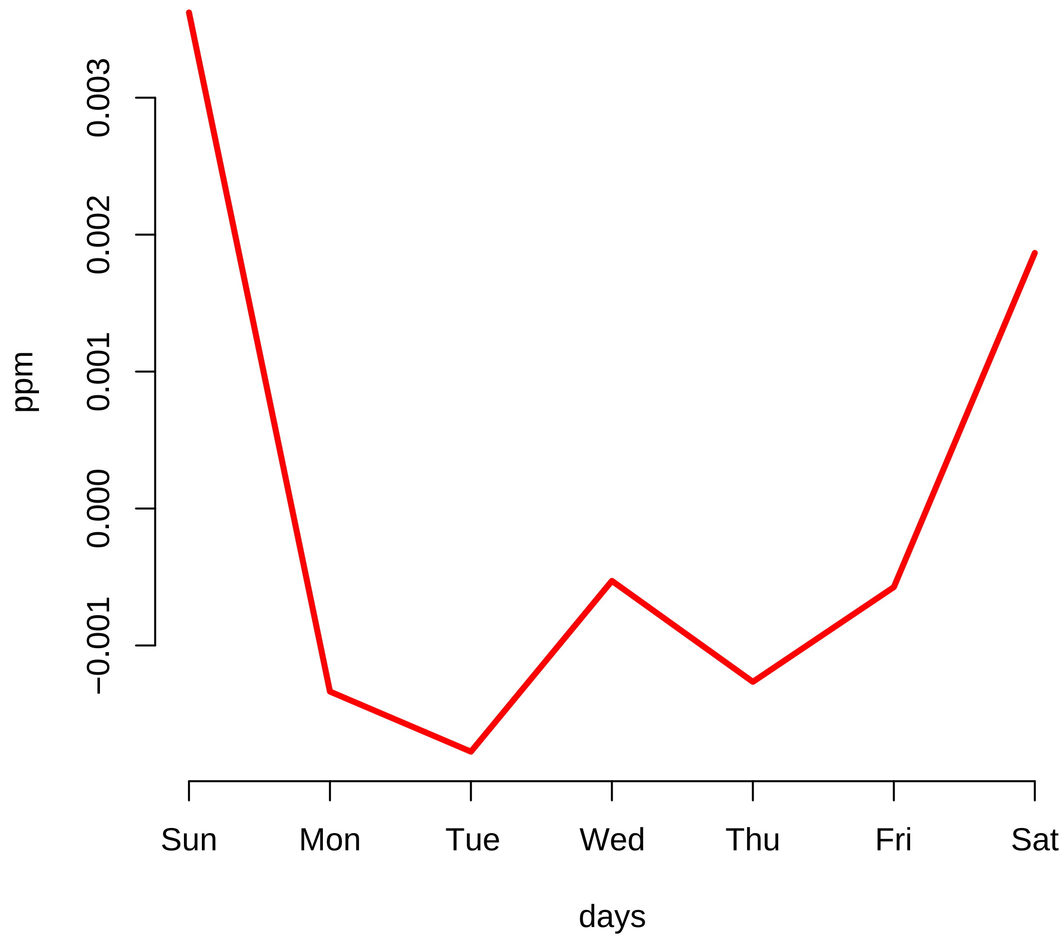

FIGURE 4. Ozone Effect of Weekday

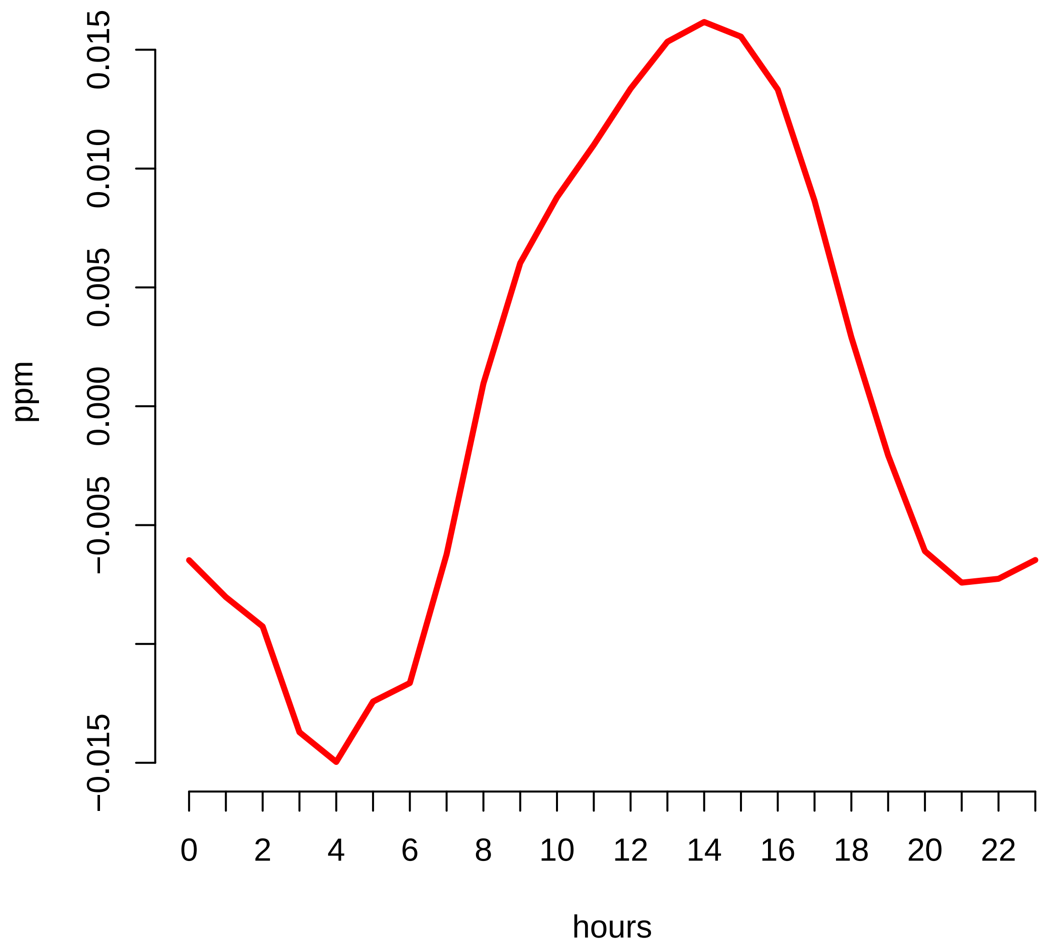

FIGURE 5. Ozone Effect of Hour

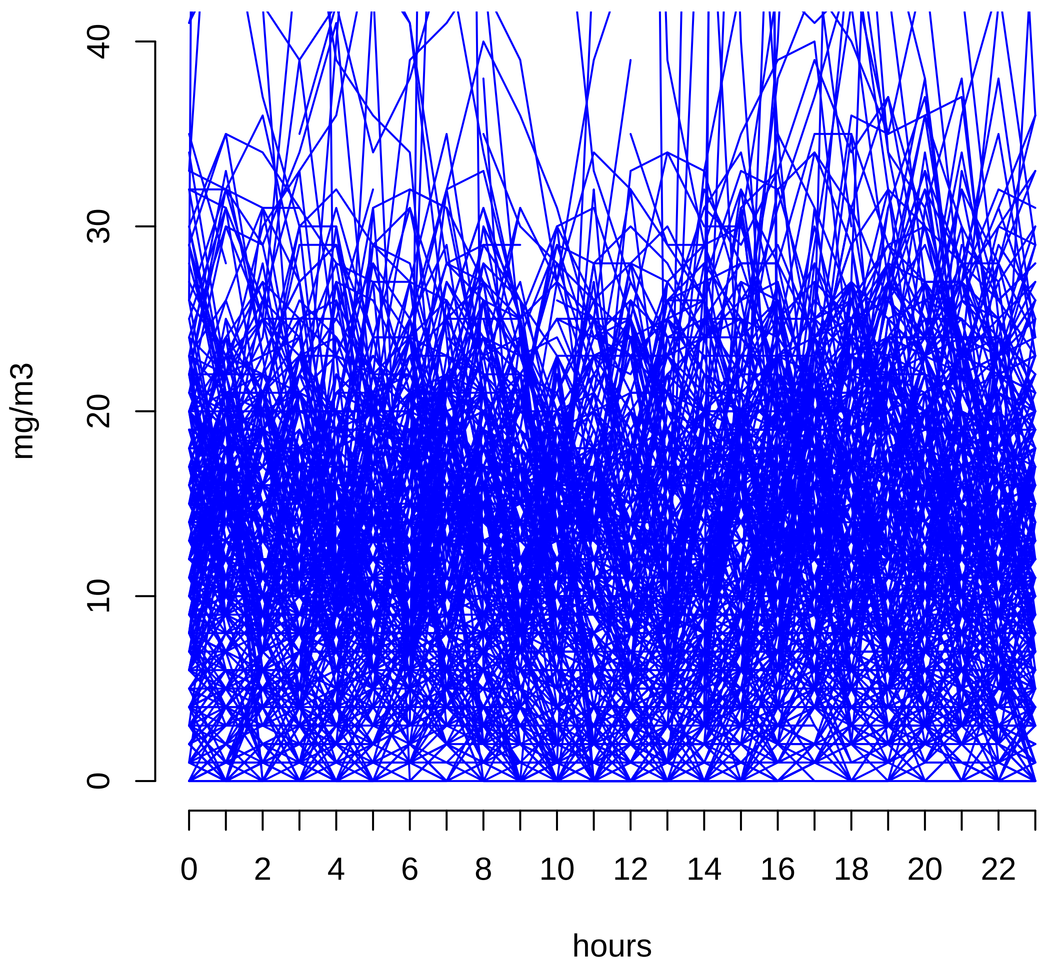

FIGURE 6. PM Curves

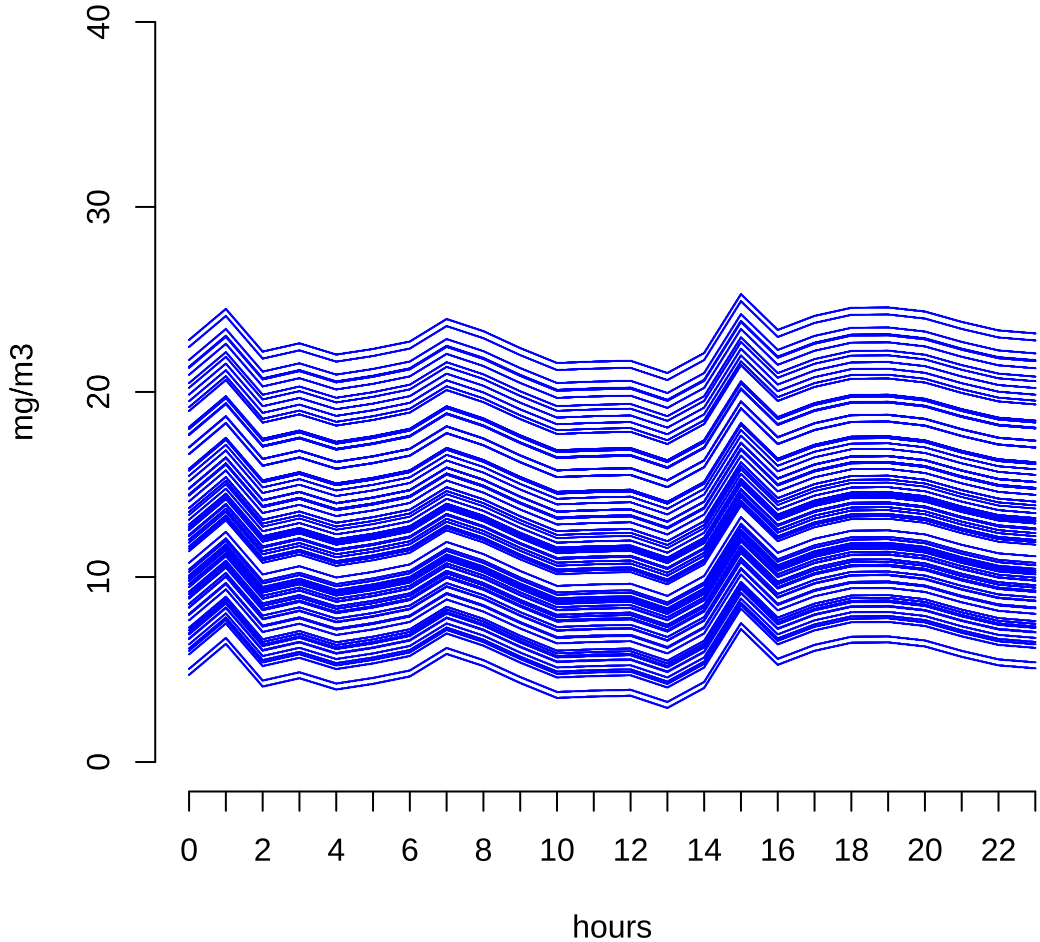

FIGURE 7. PM Curves, Smoothed

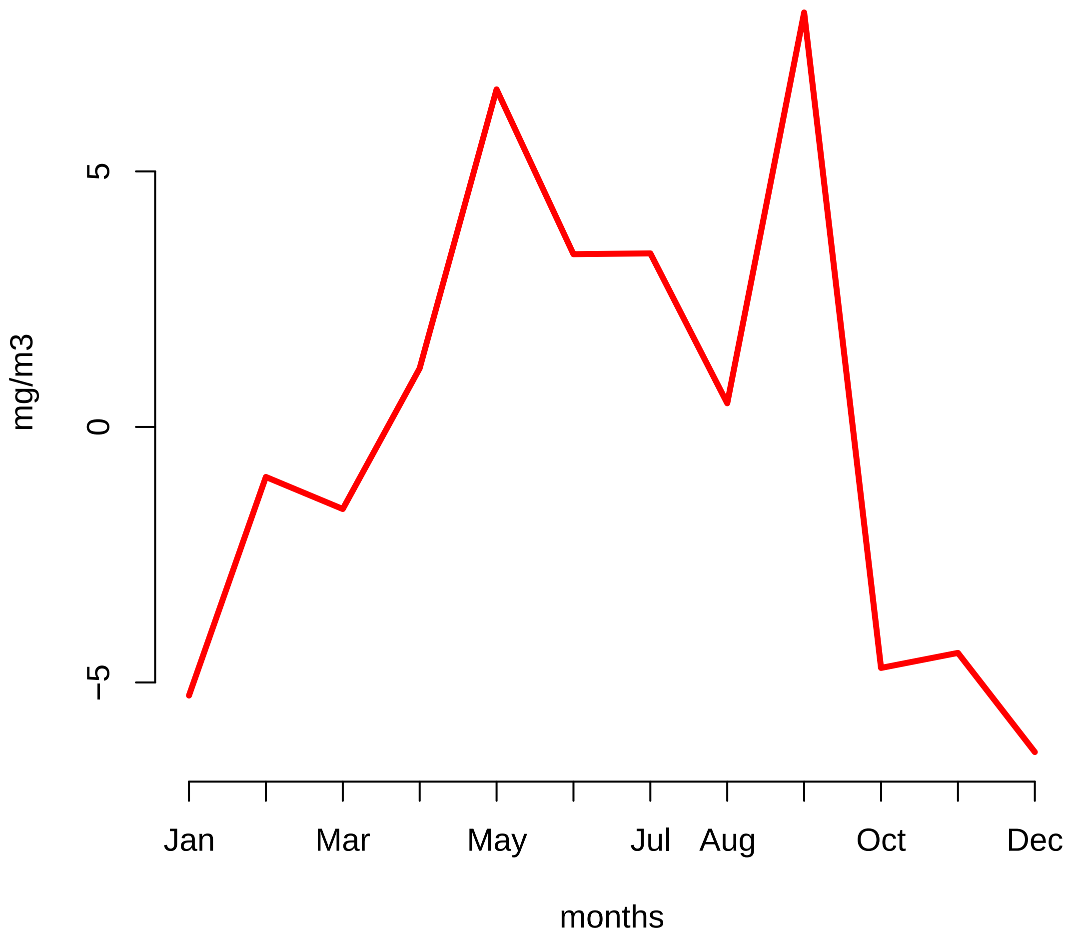

FIGURE 8. PM Effect of Month

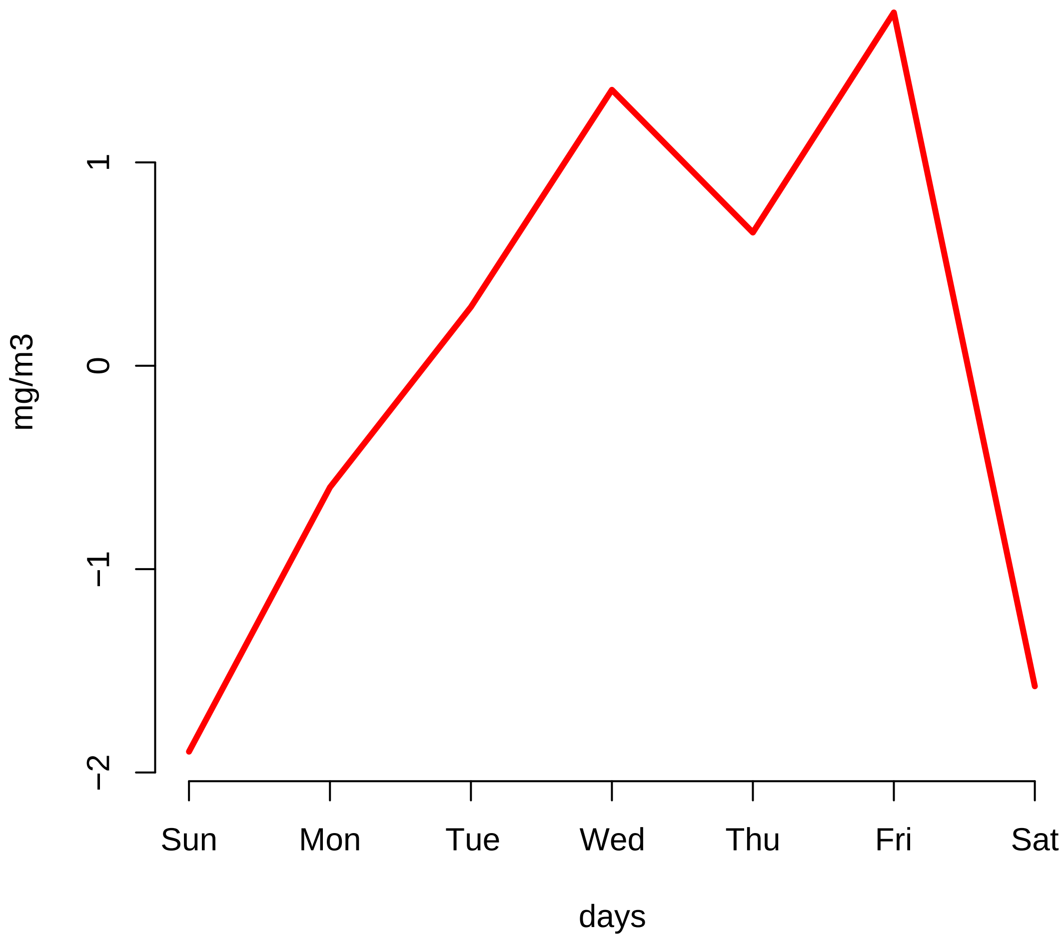

FIGURE 9. PM Effect of Weekday

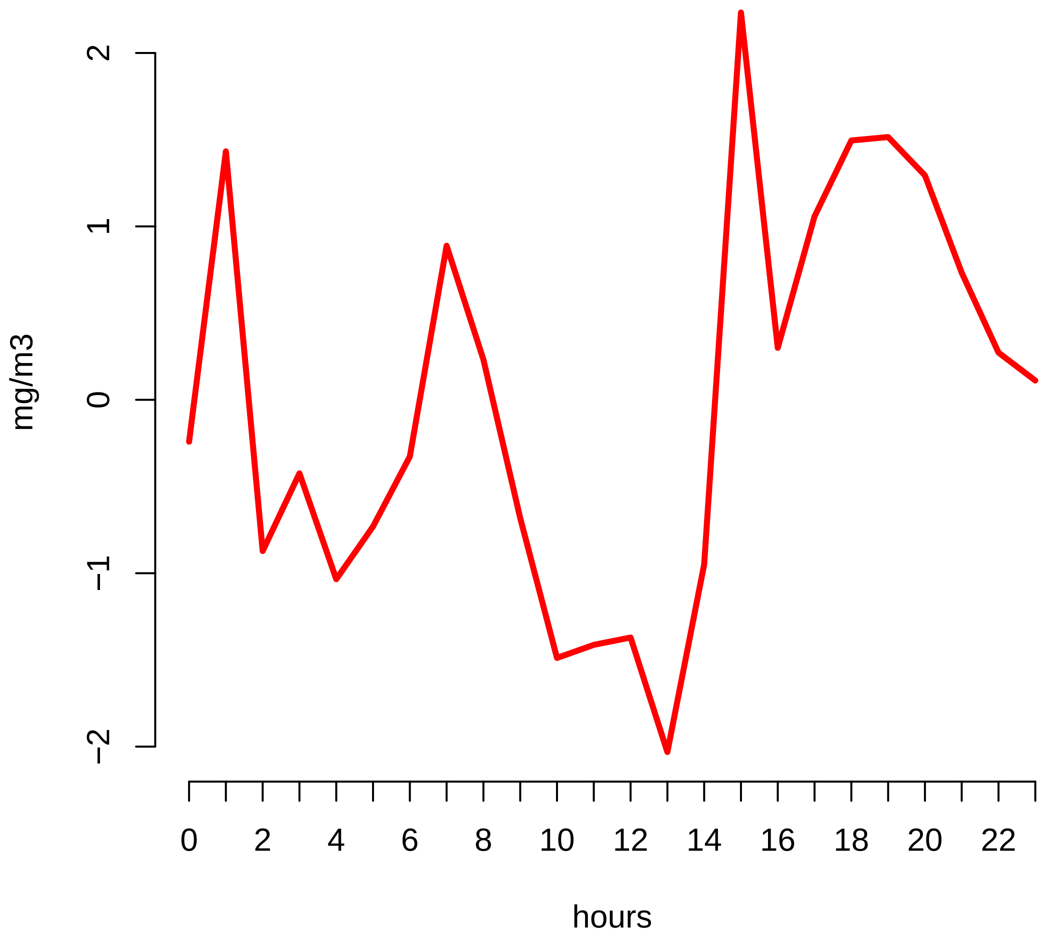

FIGURE 10. PM Effect of Hour

CENTER FOR ENVIRONMENTAL STATISTICS, DEPARTMENT OF STATISTICS, UNIVERSITY OF CALIFORNIA, LOS ANGELES, CA 90095-1554

E-mail address, Jan de Leeuw: deleeuw@stat.ucla.edu

URL, Jan de Leeuw: http://gifi.stat.ucla.edu