Dear editors of the Journal of Exposure Science and Environmental Epidemiology,

Please find the enclosed manuscript entitled "Temporal Association of Atmoshperic Fine Particulate and Mortalities in Wake County, North Carolina." As you are certainly aware, air pollution poses many risks to public health, and there are numerous difficulties both in measuring exposure and connecting that exposure to its consequences at the general population level. Powerful but overcomplicated spatiotemporal models like that of Choi, Fuentes, and Reich (2009) are published in journals on computational and theoretical statistics; my goal with this paper is to demonstrate a similar but simpler model that may be employed by your audience of statistically sophisticated epidemiologists.

Sincerely,
K.A. Flagg
Deptartment of Mathematical Sciences
Montana State University, Bozeman

Temporal Association of Atmoshperic Fine Particulate and Mortalities in Wake County, North Carolina

Kenny Flagg

December 12, 2016

Abstract

Fine particulate matter ($PM_{2.5}$) is one component of air pollution that can potentially impact public health, but an individual's exposure difficult to measure. The total concentration of particulate in the atmosphere is one possible proxy. I used Poisson generalized linear models and generalized linear autoregressive moving average models to study the association between daily average $PM_{2.5}$ concentration and daily mortality counts in Wake County, North Carolina, for the year 2014. I used a pseudo-Fourier decomposition to separate the $PM_{2.5}$ time series into several different timescales. I did not find evidence of temporal dependence in the mortality counts after accounting for $PM_{2.5}$ concentration, nor did I find any evidence of an association between mortality count and $PM_{2.5}$ concentration on any timescale. There are known mechanisms by which fine particulate can affect health, so this result implies that atmospheric $PM_{2.5}$ concentration is an inadequate metric of exposure and that further study is needed.

1 Introduction

Atmospheric particulate is one aspect of air pollution that is of major concern to regulators and policymakers because particles released into the air in one location can disperse through space and time and be inhaled by people in distant locations. Breathing fine particulate, or $PM_{2.5}$ (particles with diameter less than 2.5 μm in diameter), is associated with diverse health problems (Charlesworth, De Miguel, and Ordóñez 2011). The chemical makeup of $PM_{2.5}$ depends on many factors, but it is natural to ask whether the overall amount of particulate in the atmosphere is associated with the number of disease occurrances or mortalities. Several statistical models have been proposed to estimate this association. Choi, Fuentes, and Reich (2009) decomposed the atmoshperic $PM_{2.5}$ concentration time series into five different timescales and used a multi-stage Bayesian approach to model the relationship between $PM_{2.5}$ and daily mortality counts across the state of North Carolina in the year 2001. They aggregated by county and accounted for spatiotemporal correlation, finding a weak association between mortality and $PM_{2.5}$ concentration on the two longest timescales in most counties.

The complexity of a Bayesian spatiotemporal model is not needed in every situation, so in this paper I discuss a simpler approach to analyze data for a single location, using Poisson generalized linear models (GLMs) and generalized linear autoregressive moving average (GLARMA) models fit by maximum likelihood. I apply this model to the 2014 $PM_{2.5}$ and mortality data from Wake County, North Carolina, an urbanized county that includes Raleigh, home of the state capitol and North Carolina State University.

2 Data

The mortality data were collected by the NC State Center for Health Statistics and are available for download from the University of North Carolina Dataverse (SCHS 2016). The year 2014 is the most recent year with data available. The dataset contains records of all known deaths in North Carolina and of North Carolina residents who died elsewhere, including each individual's sex, race, and cause(s) of death; I use the subset of 5,239 non-accidental deaths that occurred in Wake County. For simplicity in illustrating the model, I will ignore the sex, race, and cause of death variables in the analysis.

The response variable in this analysis is the daily count of mortalities in Wake County in 2014, which is a time series of 365 observations. The mean is 14.4 deaths per day, the standard deviation is 3.87 deaths per day, and the distribution is approximately symmetric (Figure 1). There are no blatantly obvious temporal trends in the mortality counts (Figure 2).

The $PM_{2.5}$ data are available from the United States Environmental Protection Agency website (EPA 2016). The values in the dataset are the daily averages of hourly total $PM_{2.5}$ concentration measurements in $\mu g/m^3$. These were recorded by a BAM-1020 continuous particulate monitor (Met One Instruments, Inc., `http://www.metone.com/docs/bam1020_datasheet.pdf`) located at East Millbrook Middle School in Raleigh, about 7 miles northeast of the state capitol.

The $PM_{2.5}$ measurements for May 31 and December 3 are missing, so there are a total of 363 observations. The available daily average $PM_{2.5}$ concentrations have a mean of 11.2 $\mu g/m^3$ and a standard deviation of 4.42 $\mu g/m^3$. The $PM_{2.5}$ time series shows lots of short-term variation on the scale of days and weeks, but is less variable in the summer (June through Semptember) than in the rest of the year (Figure 3). On the scale of months, the concentration wanders up and down slightly with no clear pattern; daily average fine particulate concentration tended to be highest in February through May, with the maximum value of 31.1 $\mu g/m^3$ occurring on May 30, and lowest in August and September, the minimum being 1.82 $\mu g/m^3$ September 8.

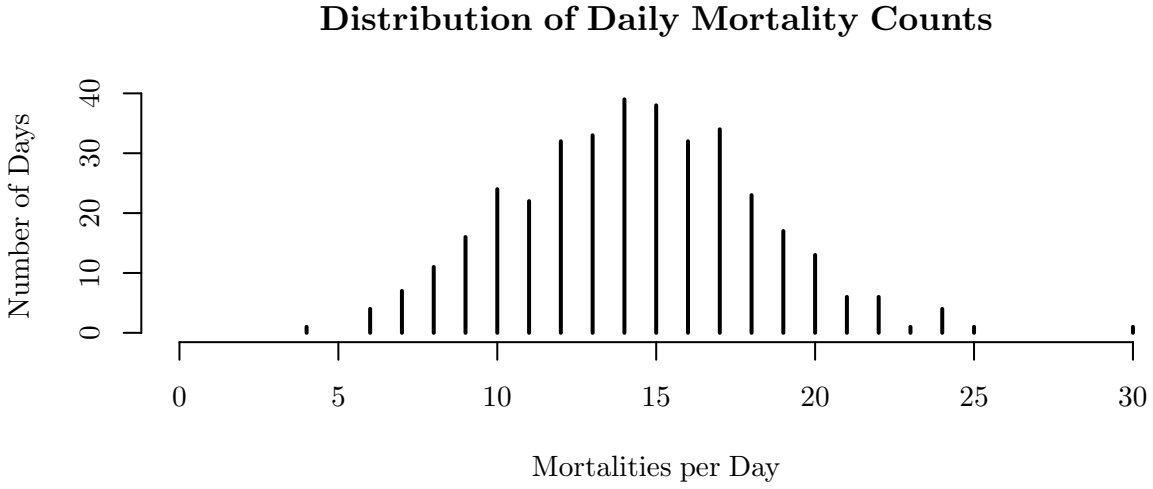

Figure 1: Distribution of daily mortality counts in Wake County in 2014. Mean = 14.4, sd = 3.87, min = 4, max = 30.

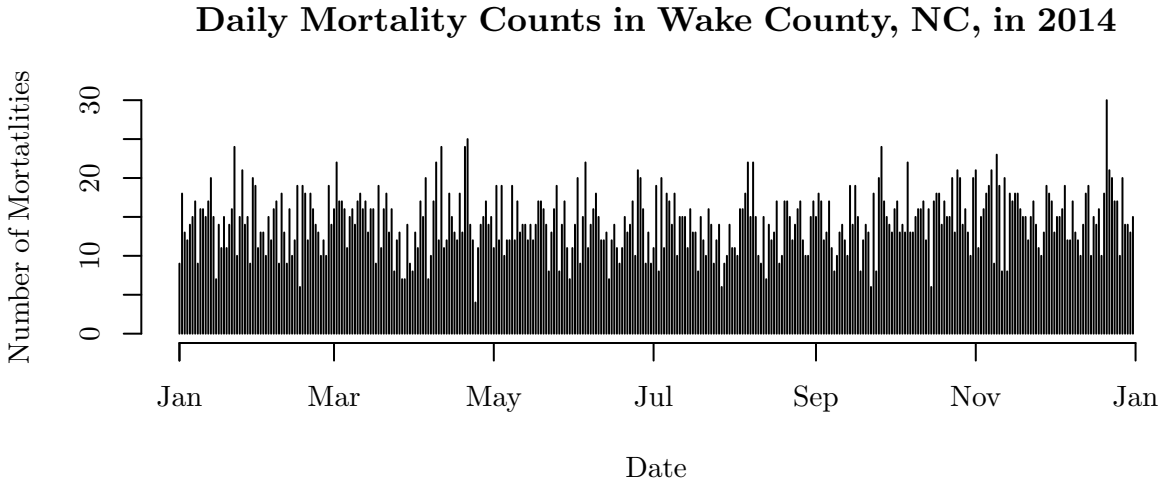

Figure 2: Daily counts of recorded non-accidental deaths in Wake County in 2014.

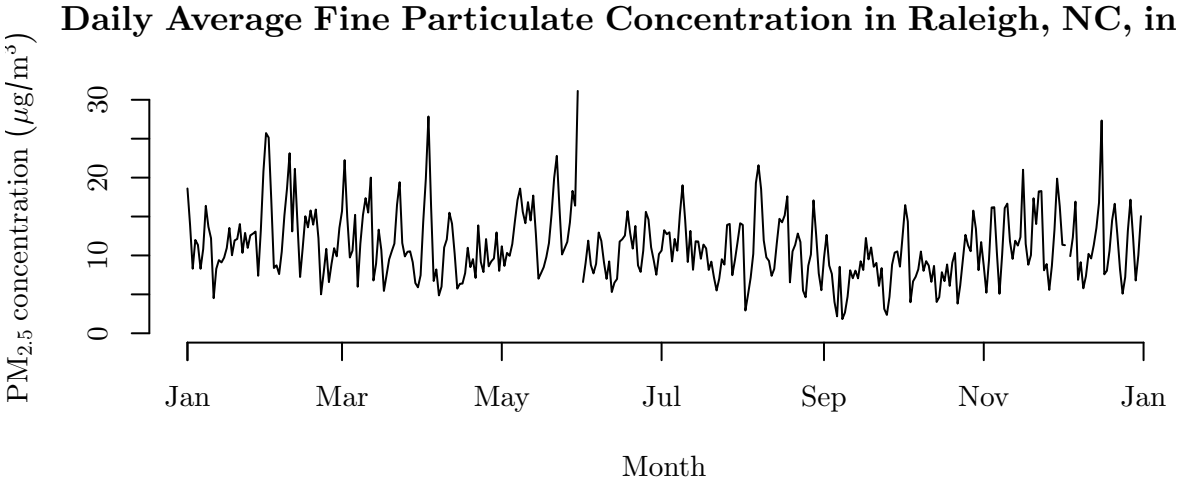

Figure 3: Daily average $PM_{2.5}$ concentration in Raleigh, NC, in 2014. Note the breaks in the line at May 31 and December 3, when $PM_{2.5}$ concentration was not recorded.

3 Methods

3.1 Model

I implemented the analysis in R (R Core Team 2016) and used the `dplyr` package (Wickham and Francois 2016) to help format the data. For clarity and reproducibility, I present my R code in Appendix A. I began by fitting the Poisson GLM,

$$y_t \sim \text{Poisson}(\mu_t);$$

$$\log(\mu_t) = \beta_0 + \beta_1 \text{PM}_{\text{year}} + \beta_2 \text{PM}_{\text{month}} + \beta_3 \text{PM}_{\text{2weeks}} + \beta_4 \text{PM}_{\text{1week}} + \beta_5 \text{PM}_{\text{days}}$$

where the PM_* represent the $PM_{2.5}$ time series decomposed into the following timescales, as in Choi, Fuentes, and Reich.

- PM_{year}: cycles with period longer than a month
- PM_{month}: cycles with period between two weeks and one month
- $\text{PM}_{\text{2weeks}}$: cycles with period between one week and two weeks
- PM_{1week}: cycles with period between 3.5 days and one week
- PM_{days}: cycles with period shorter than 3.5 days

Decomposing the $PM_{2.5}$ series in this way allows the effects of short-term and longer-term fluxuations in fine particulate concentration to be estimated separately; see Section 3.2 for details of how the decomposition is done.

I omitted May 31 and December 3 from the analysis because of the missing $PM_{2.5}$ values. After fitting the GLM, I investigated temporal dependency in the Pearson residuals $((y_t - \widehat{\mu}_t)/\sqrt{\widehat{\mu}_t})$ by examining the sample autocorration function (ACF), partial autocorrelation function (PACF), and extended autocorrelation function (EACF), available in the `TSA` package (Chan and Ripley 2012). If autocorrelation was present, I would use the `glarma` package (Dunsmuir and Scott 2015) to fit a GLARMA model with the same mean structure as above, but with an appropriate correlation structure placed on the Pearson residuals.

3.2 Fourier Series Decomposition

Mapping a time series to cyclic components of different periods is known as a spectral analysis, and is typically accomplished via the Fourier transform. This is available in R, but is not implemented for unevenly-spaced or missing observations. However, we can accomplish a similar decomposition by solving a system of linear equations.

The Fourier series representation of the $PM_{2.5}$ concentration is

$$\mathrm{PM}_{2.5}(t) = a_0 + \sum_{j=1}^{m} \left(a_j \cos\left(\frac{2\pi j}{365} t\right) + b_j \sin\left(\frac{2\pi j}{365} t\right) \right)$$

where t is the time in days and j is the period in days of the jth term. In a theoretical setting, this would be a continuous function defined for all real numbers t, and m would be taken to approach infinity, making $\{a_j\}$ and $\{b_j\}$ infinite sequences. However, the data are a finite set of 363 measurements at different times, so it is possible to can solve for 363 variables. Thus, setting $m = 181$ and substituting in the available times and $PM_{2.5}$ values yields an exact representation of the observed $PM_{2.5}$ time series.

After substituting t into the equation, the Fourier series has the form of a linear model. I used R's `lm` function as a convenient tool for computing the coefficients. The five components described in Section 3.1 are constructed by summing the Fourier series terms with the desired periods.

Separate plots of the components appear in Figure 4. Note that the constant term a_0 was not included in any of these components, so they are all centered near zero. They capture changes in $PM_{2.5}$ concentration on different timescales, with the "more than one month" component containing gradual changes across months, and the other components appearing increasingly noisy as the period decreases.

The components have spikes around December 3, which are likely artifacts of approximating the spectral decomposition from incomplete data. I fit least-squares approximations of the Fourier series for several smaller values of m; using $m = 175$ smoothed away the spikes, but this had a

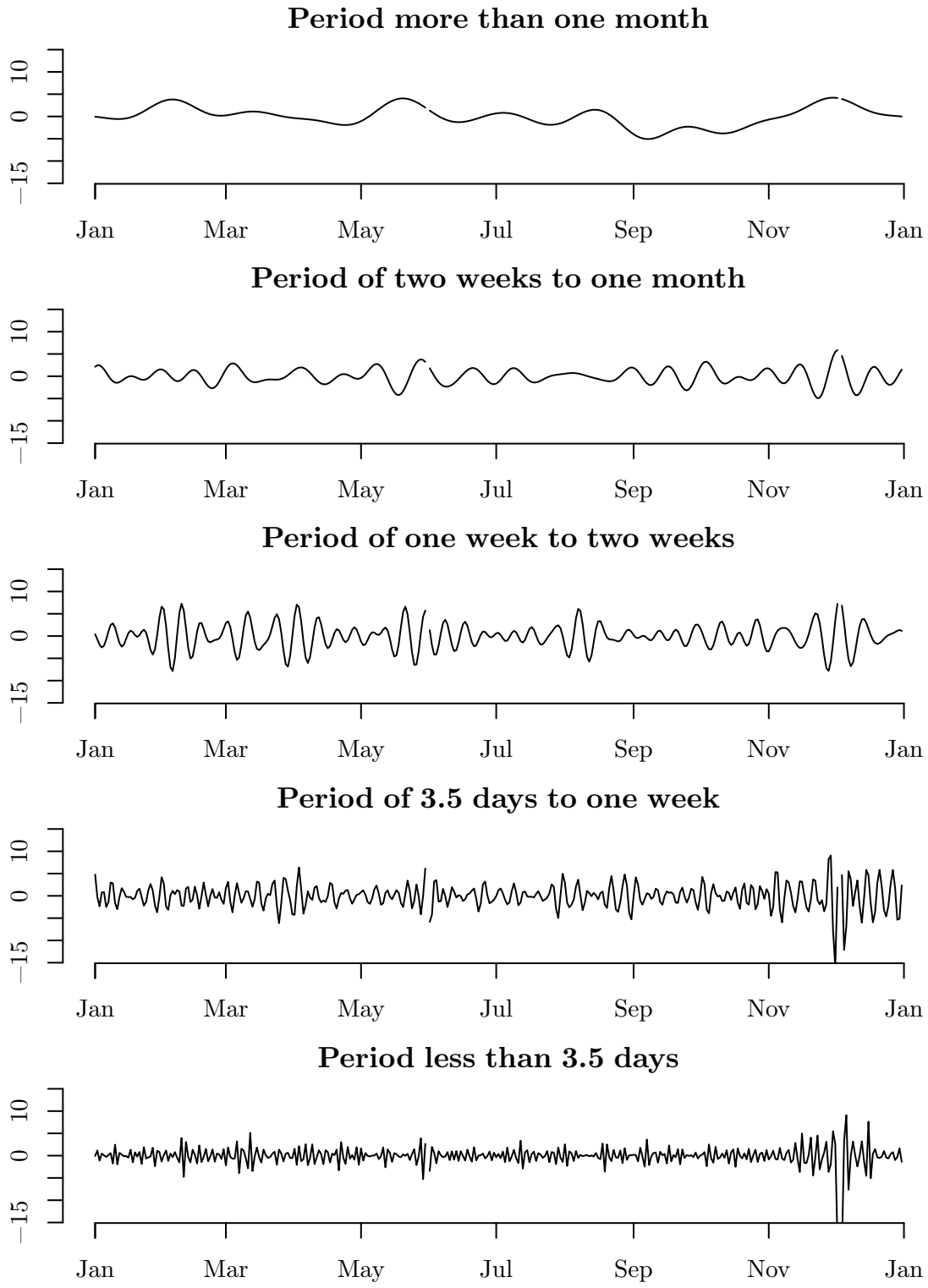

Figure 4: Time series plots of the $PM_{2.5}$ series decomposed into five different timescales.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	2.6652	0.0139	192.2609	<0.0001
More than one month	0.0029	0.0065	0.4535	0.6502
Two weeks to one month	0.0064	0.0082	0.7798	0.4355
One week to two weeks	0.0081	0.0051	1.5973	0.1102
3.5 days to one week	0.0005	0.0053	0.0952	0.9242
Less than 3.5 days	0.0061	0.0062	0.9833	0.3254

Table 1: Coefficient estimates from the GLM without an autocorrelation structure. The residual deviance is 382.82 on 357 degrees of freedom.

negligible effect on the GLM coefficient estimates, predicted values, and residual deviance. As the smoothing did not improve the model fit, I left m at 181 and used the components with the spikes.

Note that the spectral decomposition can be used to impute $PM_{2.5}$ concentration values at the dates where the concentration was not measured. There is uncertainty in the imputation process that should be accounted for in the analysis; quantifying that uncertainty is beyond the scope of this paper so I do not use the imputed values.

4 Results

The estimated GLM coefficients appear in Table 1. Each of the $PM_{2.5}$ components has a small z-statistic and a large p-value, giving little to no evidence that changes in daily average $PM_{2.5}$ concentration on any timescale are associated with changes in mean number of daily mortalities, after controlling for changes on other timescales.

When modeling time series data, it is crucial to examine the residuals to ensure that temporal dependency is accounted for, because a failure to do so can result in underestimating the standard errors. Figure 5 shows the Pearson residuals from the GLM plotted against the fitted values and against time. There is one minor outlier, but there are no trends or patterns of changing variability in the residuals. It appears the model describes the data adequately.

I also examined the sample ACF and PACF of the Pearson residuals (Figure 6). The correlations are small at all lags, with the ACF exceeding the 95% confidence bounds of $\pm 2/\sqrt{363} = \pm 0.105$ only at lags of 11, 42, and 50 days, and the PACF exceeding the bounds only at a of 11 days. The pattern in these plots is consistent with independent random noise with no true autocorrelation. The sample EACF (Figure 7) is a tool that can help choose a reasonable ARMA structure; red cells in the plot correspond to models that do not account for all the autocorrelation present. The ARMA(0,0) cell at the top left is not red, implying that the GLM assuming independent Pearson residuals is appropriate.

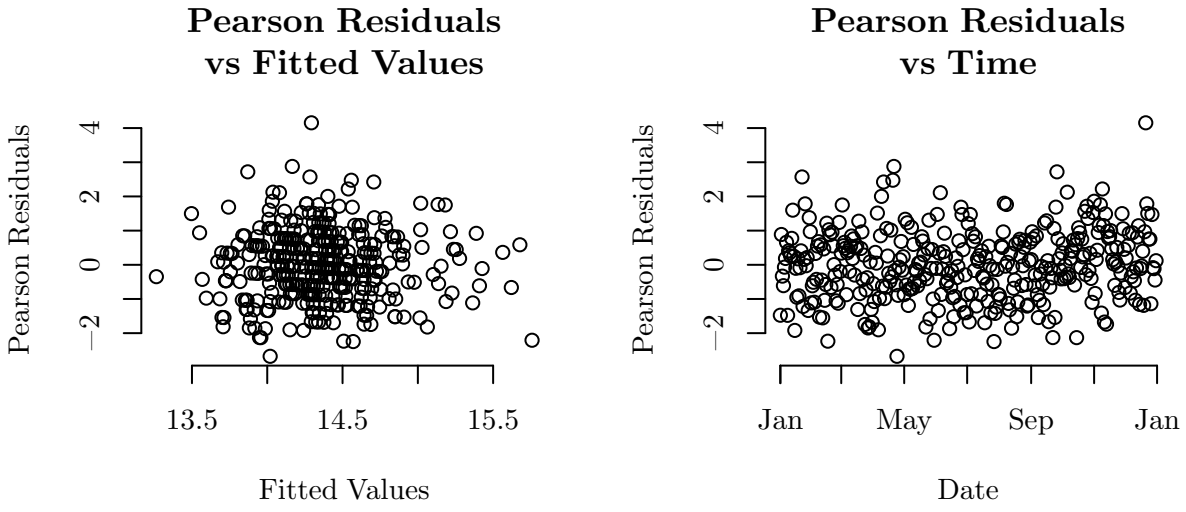

Figure 5: Diagnostic plots of the GLM Pearson residuals.

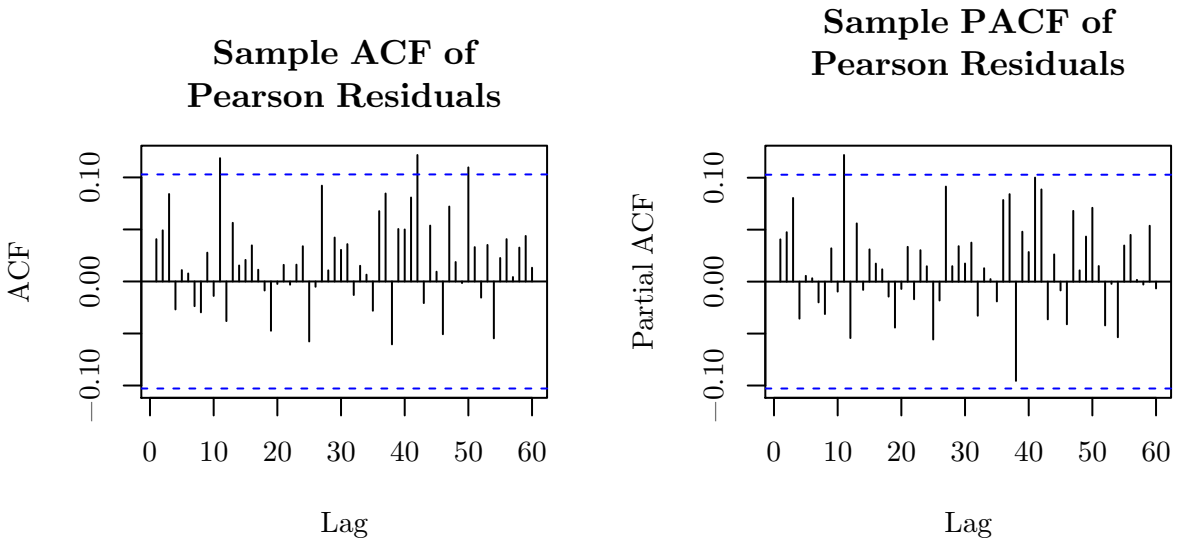

Figure 6: Sample ACF and partial ACF of the GLM Pearson residuals. The patterns are consistent with random noise.

To illustrate the use of GLARMA models, I fit a model with an MA(11) structure to capture the spike in the ACF at lag 11. The `glarma` function in R could not fit the model with missing data, so I used the estimated spectrum to impute $PM_{2.5}$ values. The estimated model coefficients appear in Table 2. Again, on all timescales, the $PM_{2.5}$ terms have small z-statistics and large p-values providing no evidence of associations. Table 3 presents the estimated moving average parameters. The 3rd order term has a z-statistic of 1.88 with a p-value of 0.0599, and the 11th order term has z-

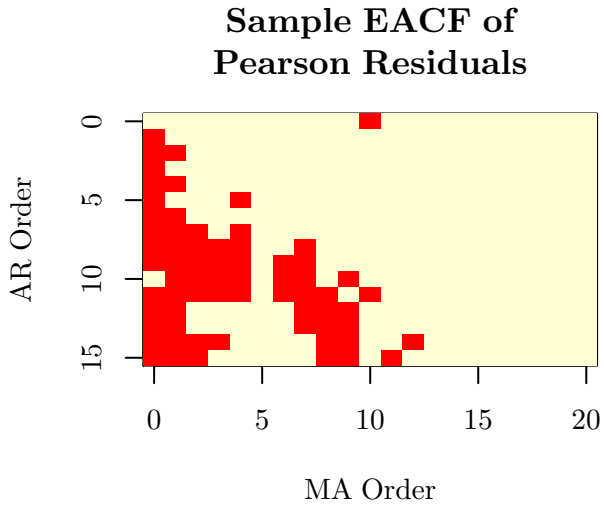

Figure 7: Sample extendend ACF of the GLM Pearson residuals. Red cells indicate ARMA models that fail to accurately characterize the temporal dependence in the residuals. This plot suggests that an ARMA(0,0) model (with no autocorrelation structure) is appropriate.

	Estimate	Std.Error	z-ratio	Pr(>|z|)
(Intercept)	2.6626	0.0177	150.6877	0.0000
More than one month	0.0023	0.0079	0.2916	0.7706
Two weeks to one month	0.0049	0.0084	0.5813	0.5610
One week to two weeks	0.0063	0.0055	1.1323	0.2575
3.5 days to one week	-0.0013	0.0046	-0.2778	0.7812
Less than 3.5 days	0.0006	0.0046	0.1321	0.8949

Table 2: Linear predictor coefficient estimates from the GLARMA model with an MA(11) autocorrelation structure. The residual deviance is 364.96 on 348 degrees of freedom.

statistic 2.8 with p-value = 0.0051. However, I selected this model after examining (and implicitely testing for) many autocorrelation terms, so it is likely that these small p-values are spurious. The model summary also includes a likelihood ratio test for testing all of the autocorrelation parameters against the null hypothesis of no autocorrelation; this has a statistic of $\chi^2_{11} = 13.8$ with p-value = 0.2451, which provides no evidence that the MA(11) structure is needed in the model. The appropriate model for these data is the simpler GLM.

	Estimate	Std.Error	z-ratio	Pr(>|z|)
θ_1	0.0172	0.0137	1.2574	0.2086
θ_2	0.0088	0.0137	0.6457	0.5185
θ_3	0.0257	0.0137	1.8816	0.0599
θ_4	-0.0150	0.0138	-1.0856	0.2777
θ_5	0.0069	0.0138	0.5003	0.6168
θ_6	-0.0012	0.0139	-0.0861	0.9313
θ_7	-0.0058	0.0139	-0.4162	0.6773
θ_8	-0.0053	0.0138	-0.3825	0.7021
θ_9	0.0156	0.0136	1.1408	0.2540
θ_{10}	-0.0118	0.0138	-0.8568	0.3915
θ_{11}	0.0394	0.0141	2.8026	0.0051

Table 3: Moving average parameter estimates from the GLARMA model with an MA(11) autocorrelation structure. The likelihood ratio statistic for the autocorrelation terms is $\chi^2_{11} = 13.8$ with p-value = 0.2451.

5 Discussion and Conclusion

The daily average fine particulate concentration measurements on five different timescales were not strongly associated with daily mortality counts for Wake County, North Carolina, in 2014. This does not mean that fine particulate has no relationship with mortality; it means that the atmospheric concentration on the day of death is a poor proxy for exposure. More detailed studies are needed that use more targeted measurements of fine particulate exposure and account for characteristics of the individuals.

My approach has several limitations. As mentioned above, individual information such as age and sex would allow more specific inferences. Also, I used one location to represent the whole county. Choi et al. averaged $PM_{2.5}$ predictions over the whole spatial area, which is a better measurement of the overall amount of fine particulate. Their model also allows prediction for unmeasured locations and times, so a discrete Fourier transform can be used to compute the spectrum of predicted $PM_{2.5}$ concentrations.

I demonstrated the application of a simpler GLARMA model. Unfortunately, the software is not well developed. There are several R options for fitting GLMs with random effects and correlation structures — `glmmPQL` in the `MASS` package (Venables and Ripley 2002) and `gamm` in the `mgcv` package (Wood 2000) to name a few — but the `glarma` package is the only option I know of for fitting GLMs with correlation structures but without random effects. My hope is that the R community will continue to develop GLARMA methods so they can be used by the broader scientific community.

A R Code

```
## SETTING UP DATA

library(dplyr)

# Make a vector of all dates in 2014.
dates <- as.POSIXct(seq(0, by = 24 *60 * 60, length.out = 365),
                    origin = '2014-01-01 05:00:00', tz = 'EST')


## MORTALITY DATA

# Loads a data frame unhelpfully named x.
load('mortality/deaths2014.RData')

# Filter doum to the data we need. Henderson County only, nonaccidental deaths.
# Accidental deaths have codes starting with letters S-Y.
wakeDeaths <- x %>%
  filter(CNTYOCC == 'Wake', !grepl('[S-Y]', ACMECOD)) %>%
  select(DHTDATE, SEX, RACER, HISP) %>%
  mutate(Date = DHTDATE, Sex = SEX,
         Race = ifelse(RACER == 'White', 'White',
                ifelse(RACER == 'Black or African American', 'Black',
                'Other/Unknown')),
         Hispanic = ifelse(HISP == 'Non-Hispanic' | HISP == 'Unknown',
                           'Non-Hispanic/Unknown', 'Hispanic')) %>%
  select(Date, Sex, Race, Hispanic)

# Get the daily mortality counts.
wakeCounts <- wakeDeaths %>%
  group_by(Date) %>%
  summarise(Deaths = n()) %>%
  ungroup


## PM2.5 DATA

# Read all FRM/FEM PM2.5 for the US in 2014.
all88101 <- read.csv('pm25/daily_88101_2014.csv')
# Met One BAM-1020 Mass Monitor w/VSCC - Beta Attenuation (method code 170)
# measurements made on all days except May 31 and December 3.

# Filter down to Wake County, NC, method 170.
wake88101 <- all88101 %>%
  filter(State.Name == 'North Carolina',
         County.Name == 'Wake',
         Method.Code == 170)


## COMBINED DATA FRAME

wake <- data.frame(Date = dates,
                   PM2.5 = rep(as.numeric(NA), 365),
                   Deaths = rep(as.integer(0), 365),
```

```
                  row.names = dates)

# Insert PM2.5 values and mortality counts by date.
wake[as.character(wake88101$Date.Local), 'PM2.5'] <- wake88101$Arithmetic.Mean
wake[wakeCounts$Date, 'Deaths'] <- wakeCounts$Deaths


## FIGURE 1

plot(table(`Deaths` = wake$Deaths), bty = 'n', xaxt = 'n', xlim = c(0, 30),
     xlab = 'Mortalities per Day', ylab = 'Number of Days',
     main = 'Distribution of Daily Mortality Counts')
axis(1)


## FIGURE 2

plot(Deaths ~ Date, data = wake, type = 'h', bty = 'n', ylim = c(0, 30),
     main = 'Daily Mortality Counts in Wake County, NC, in 2014',
     xlab = 'Date', ylab = 'Number of Mortatlities')


## FIGURE 3

plot(PM2.5 ~ Date, data = wake, type = 'l', bty = 'n', ylim = c(0, 30),
     main = 'Daily Average Fine Particulate Concentration in Raleigh, NC, in 2014',
     xlab = 'Month', ylab = expression(PM[2.5]*' concentration '*(mu*g/m^3)))


## FOURIER DECOMPOSITION

pm2.5series <- ts(wake$PM2.5, start = 0, frequency = 365)

# Number of Fourier terms to compute.
m <- 181

# h is the "model matrix" of harmonic functions of time such that
# t(h) %*% beta is a matrix representation of the Fourier series.
library(TSA)
h <- cbind(`(Intercept)` = 1, harmonic(pm2.5series, m))

# Use lm to find the Fourier coefficients.
pm2.5coefs <- coef(lm(pm2.5series ~ harmonic(pm2.5series, m)))

# Construct the four components.
# 0 < Frequency < 12.
pm_year <- h[,c(2:12, m + 2:12)] %*% pm2.5coefs[c(2:12, m + 2:12)]

# 12 <= Frequency < 26.
pm_month <- h[,c(13:26, m + 13:26)] %*% pm2.5coefs[c(13:26, m + 13:26)]

# 26 <= Frequency < 52.
pm_2week <- h[,c(27:52, m + 27:52)] %*% pm2.5coefs[c(27:52, m + 27:52)]

# 52 <= Frequency < 104.
```

```
pm_1week <- h[,c(53:104, m + 53:104)] %*% pm2.5coefs[c(53:104, m + 53:104)]

# 104 <= Frequency <= 181.
pm_day <- h[,c(105:(m+1), m + 105:(m+1))] %*% pm2.5coefs[c(105:(m+1), m + 105:(m+1))]

# Put the components into the data frame and put in NAs.
wake$`Less than 3.5 days` <- ifelse(is.na(pm2.5series), NA, pm_day)
wake$`3.5 days to one week` <- ifelse(is.na(pm2.5series), NA, pm_1week)
wake$`One week to two weeks` <- ifelse(is.na(pm2.5series), NA, pm_2week)
wake$`Two weeks to one month` <- ifelse(is.na(pm2.5series), NA, pm_month)
wake$`More than one month` <- ifelse(is.na(pm2.5series), NA, pm_year)


## FIGURE 4

par(mfrow = c(5, 1), mar = c(2.1, 4.1, 2.1, 2.1), cex = 1)
plot(`More than one month` ~ Date, data = wake, type = 'l', bty = 'n',
     ylim = c(-14, 14), xlab = '', ylab = '', main = 'Period more than one month')
plot(`Two weeks to one month` ~ Date, data = wake, type = 'l', bty = 'n',
     ylim = c(-14, 14), xlab = '', ylab = '', main = 'Period of two weeks to one month')
plot(`One week to two weeks` ~ Date, data = wake, type = 'l', bty = 'n',
     ylim = c(-14, 14), xlab = '', ylab = '', main = 'Period of one week to two weeks')
plot(`3.5 days to one week` ~ Date, data = wake, type = 'l', bty = 'n',
     ylim = c(-14, 14), xlab = '', ylab = '', main = 'Period of 3.5 days to one week')
plot(`Less than 3.5 days` ~ Date, data = wake, type = 'l', bty = 'n',
     ylim = c(-14, 14), xlab = '', ylab = '', main = 'Period less than 3.5 days')


## GLM FIT

wakeGLM <- glm(Deaths ~ `More than one month` + `Two weeks to one month` +
                 `One week to two weeks` + `3.5 days to one week` +
                 `Less than 3.5 days`, data = wake, family = poisson)
wakeGLMsummary <- summary(wakeGLM)
coefdf <- data.frame(wakeGLMsummary$coefficients)
colnames(coefdf) <- colnames(wakeGLMsummary$coefficients)
coefdf$`Pr(>|z|)` <- format.pval(coefdf$`Pr(>|z|)`, digits = 4, eps = 0.0001)


## TABLE 1

library(xtable)
xtable(coefdf, digits = 4, align = 'crrrr', label = 'glmsummary',
       caption = paste0('Coefficient estimates from the GLM without an
                        autocorrelation structure. The residual deviance is ',
                        signif(wakeGLMsummary$deviance, 5), ' on ',
                        wakeGLMsummary$df.residual, ' degrees of freedom.'))


## FIGURE 5

par(mfrow = c(1, 2))
plot(residuals(wakeGLM, type = 'pearson') ~ fitted(wakeGLM), bty = 'n',
     main = 'Pearson Residuals\nvs Fitted Values',
     xlab = 'Fitted Values', ylab = 'Pearson Residuals')
```

```
plot(residuals(wakeGLM, type = 'pearson') ~ wake$Date[!is.na(wake$PM2.5)],
     bty = 'n', main = 'Pearson Residuals\nvs Time',
     xlab = 'Date', ylab = 'Pearson Residuals')


## FIGURE 6

library(TSA)
par(mfrow = c(1, 2), cex = 1)
acf(residuals(wakeGLM, type = 'pearson'), lag.max = 60,
    main = 'Sample ACF of\nPearson Residuals')
pacf(residuals(wakeGLM, type = 'pearson'), lag.max = 60,
    main = 'Sample PACF of\nPearson Residuals')


## FIGURE 7

image(0:20, 0:15, t(eacf(residuals(wakeGLM, type = 'pearson'),
                         ar.max = 15, ma.max = 20)$symbol == 'o'),
      ylim = c(15.5, -0.5), xlab = 'MA Order', ylab = 'AR Order',
      main = 'Sample EACF of\nPearson Residuals')


## GLARMA FIT

library(glarma)
X <- model.matrix(Deaths ~ pm_year + pm_month + pm_2week + pm_1week + pm_day, data = wake)
wakeGLARMA <- glarma(wake$Deaths, X, type = 'Poi', phiLags = NULL, thetaLags = 1:11)
wakeGLARMAsummary <- summary(wakeGLARMA)
rownames(wakeGLARMAsummary$coefficients1) <- rownames(wakeGLMsummary$coefficients)
rownames(wakeGLARMAsummary$coefficients2) <- paste0('\\(\\theta_{', 1:11, '}\\)')

## TABLE 2

xtable(wakeGLARMAsummary$coefficients1, digits = 4, align = 'crrrr',
       label = 'glarmasummary1',
       caption = paste0('Linear predictor coefficient estimates from the
                        GLARMA model with an MA(11) autocorrelation structure.
                        The residual deviance is ',
                        signif(wakeGLARMAsummary$deviance, 5), ' on ',
                        wakeGLARMAsummary$df.residual, ' degrees of freedom.'))


## TABLE 3

xtable(wakeGLARMAsummary$coefficients2, digits = 4, align = 'crrrr',
       label = 'glarmasummary2',
       caption = paste0('Moving average parameter estimates from the
                        GLARMA model with an MA(11) autocorrelation structure.
                        The likelihood ratio statistic for the autocorrelation
                        terms is \\(\\chi^{2}_{11} = ',
                        signif(wakeGLARMAsummary$likTests['LR Test', 'Statistic'], 3),
                        '\\) with p-value = ',
                        signif(wakeGLARMAsummary$likTests['LR Test', 'p-value'], 4), '.'))
```

References

Chan, Kung-Sik and Brian Ripley (2012). *TSA: Time Series Analysis*. R package version 1.01. URL: https://CRAN.R-project.org/package=TSA.

Charlesworth, S, E De Miguel, and A Ordóñez (2011). "A review of the distribution of particulate trace elements in urban terrestrial environments and its application to considerations of risk". In: *Environmental Geochemistry and Health* 33.2, pp. 103–123.

Choi, Jungsoon, Montserrat Fuentes, and Brian J Reich (2009). "Spatial–temporal association between fine particulate matter and daily mortality". In: *Computational Statistics and Data Analysis* 53.8, pp. 2989–3000.

Dunsmuir, William T. M. and David J. Scott (2015). "The glarma Package for Observation-Driven Time Series Regression of Counts". In: *Journal of Statistical Software* 67.7, pp. 1–36.

R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: https://www.R-project.org/.

State Center for Health Statistics (2016). *North Carolina Vital Statistics – Deaths 2014*. URL: http://dx.doi.org/10.15139/S3/12361.

U.S. Environmental Protection Agency (2016). *$PM_{2.5}$ – Local Conditions, 2014 Daily Summary Data*. Parameter code 88101. URL: https://aqsdr1.epa.gov/aqsweb/aqstmp/airdata/download_files.html.

Venables, W. N. and B. D. Ripley (2002). *Modern Applied Statistics with S*. Fourth. New York: Springer.

Wickham, Hadley and Romain Francois (2016). *dplyr: A Grammar of Data Manipulation*. R package version 0.5.0. URL: https://CRAN.R-project.org/package=dplyr.

Wood, S. N. (2000). "Modelling and smoothing parameter estimation with multiple quadratic penalties". In: *Journal of the Royal Statistical Society (B)* 62.2, pp. 413–428.