

Modelos MLG e MAG para análise da associação entre poluição atmosférica e marcadores de morbi-mortalidade: uma introdução baseada em dados da cidade de São Paulo

GLM and GAM model for analyzing the association between atmospheric pollution and morbidity-mortality markers: an introduction based on data from the city of São Paulo

Gleice Margarete de Souza Conceição

Laboratório de Poluição Atmosférica Experimental
Departamento de Patologia
Faculdade de Medicina
Universidade de São Paulo.

Endereço para correspondência/correspondence to:
Av. Dr. Arnaldo 455, São Paulo, SP, 01246-903, Brasil.
gleice@usp.br

Paulo Hilário Nascimento Saldiva

Laboratório de Poluição Atmosférica Experimental
Departamento de Patologia
Faculdade de Medicina
Universidade de São Paulo

Julio da Motta Singer

Departamento de Estatística,
Instituto de Matemática e Estatística
Universidade de São Paulo

Apoio:

Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP)
Laboratório de Investigação Médica do Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo (LIM05-HCFMUSP)

Resumo

Este estudo, descreve e compara duas classes de modelos - os Modelos Lineares Generalizados (MLG) e os Modelos Aditivos Generalizados (MAG) - que podem ser utilizadas para avaliar a associação entre poluição atmosférica e marcadores de morbi-mortalidade. Enfoca os MAG como uma alternativa para a modelagem de relações não lineares não especificadas, e mostra que essa classe de modelos constitui uma boa opção para representar tanto a sazonalidade quanto a relação entre o despecho e as variáveis meteorológicas. Como exemplo de aplicação é avaliada a associação entre mortalidade em idosos e poluição atmosférica na cidade de São Paulo no período de 1994 a 1997. Os dados de mortalidade foram obtidos do Programa de Aprimoramento das Informações de Mortalidade (PRO-AIM) e as concentrações diárias de poluentes (PM_{10} , SO_2 , CO, e ozônio) foram obtidas da Companhia de Tecnologia de Saneamento Ambiental (CETESB). Informações acerca da temperatura e umidade relativa do ar foram obtidas do Instituto Astronômico e Geofísico da Universidade de São Paulo (IAG-USP). As duas classes de modelos produziram resultados coerentes, mas os modelos estatisticamente mais sofisticados tiveram mais poder para detectar efeitos significantes. Foram observadas associações entre mortalidade e os níveis de CO, SO_2 e, em menor escala, PM_{10} . As associações observadas foram dose-dependente e evidentes após um curto período de exposição.

Palavras chave: Modelos lineares. Modelos aditivos generalizados. Poluição do ar. Morbidade. Mortalidade. Idoso.

Abstract

In this study, we have described and compared two classes of models - the Generalized Linear Models (GLM) and the Generalized Additive Models (GAM) - which can be used to evaluate the association between atmospheric pollution and morbidity and mortality counts. We consider GAM as an alternative for modeling unspecified nonlinear relationships, and show that this class of models constitute a good option to represent both seasonality and the relation between the outcome and the meteorological variables. As an illustration, we evaluate the association between mortality in elderly people and air pollution in São Paulo, Brazil for the period ranging from 1994 to 1997. The two modeling classes have produced coherent results, but the most sophisticated models were more powerful to detect significant effects. Associations between mortality and the levels of CO, SO₂ and, in to a lesser extent, PM₁₀ were observed.

Keywords: Linear models. Generalized additive models. Air pollution. Morbidity. Mortality. Aged.

Introdução

Modelos estatísticos constituem ferramentas extremamente úteis para resumir e interpretar dados. Em particular, eles podem facilitar a avaliação da forma e da intensidade de associações de interesse em estudos epidemiológicos.

Por exemplo, consideremos estudos em que o objetivo é avaliar os efeitos da poluição atmosférica sobre a saúde dos habitantes de grandes centros urbanos. A variável resposta, nesses estudos, geralmente é alguma contagem de eventos que representam danos à saúde, como o número de óbitos ou o número de internações por determinada causa respiratória. A escolha da concentração de alguns gases como NO_x, SO₂ ou CO ou material particulado como candidatas a variáveis explicativas também é bastante comum. Além disso, há evidências tanto teóricas quanto empíricas de que as possíveis variáveis respostas citadas acima são fortemente influenciadas por fatores sazonais e pelas condições climáticas, como a temperatura e a umidade do ar¹⁻⁵, por exemplo. Entretanto, a maneira como esses fatores exercem sua influência sob o desfecho não é tão óbvia. As relações entre as diversas variáveis intervenientes podem não apresentar o mesmo comportamento ao longo de todos os valores do seu domínio e, mesmo se o apresentarem, a intensidade da associação pode não ser constante; por exemplo, o número de óbitos pode variar linearmente numa determinada faixa de valores de temperatura, e quadrática ou exponencialmente em outras.

Dada a complexidade das relações de interesse, a escolha de modelos apropriados para a análise se reveste de bastante importância. Por exemplo, modelos de regressão linear servem para investigar se uma variável resposta Y está associada com variáveis explicativas X_1, X_2, \dots, X_N , mas este tipo de modelo avalia esta associação apenas sob a ótica linear, que nem sempre é aquela que rege os fenômenos considerados. Além disso, é importante ter em mente que as suposições usualmente empregadas na análise

dificilmente corresponderão à realidade de modo exato, por mais sofisticado que seja o modelo em questão.

Outros aspectos a serem considerados nos estudos que avaliam os efeitos da poluição utilizando séries de contagens são a distribuição da variável resposta, a presença de tendência e sazonalidade, variáveis de confundimento (geralmente temperatura e umidade), a existência de defasagem entre o aumento da poluição e a ocorrência do evento, a presença de sobredispersão e a possível existência autocorrelação entre as observações medidas ao longo do tempo, entre outros.

Uma metodologia bastante empregada na tentativa de solucionar os problemas citados envolve o ajuste de Modelos Lineares Generalizados⁶. Nesse caso, o controle da sazonalidade é realizado por intermédio do emprego de variáveis indicadoras para estações do ano^{7,8}, meses do ano ou meses de estudo^{4,5,9-11}, ou da inclusão de termos de seno e co-seno de diferentes frequências¹² etc.

Pretendemos apresentar aqui uma classe de modelos intitulada “Modelos Aditivos Generalizados” (MAG) como uma alternativa para a modelagem de relações não lineares que não apresentam uma forma definida. Esse tipo de modelo baseia-se em funções não paramétricas, denominadas curvas de alisamento, em que a forma da associação é definida pelos próprios dados, conforme a metodologia descrita por Hastie e Tibshirani¹³, e amplamente utilizada para este fim por Schwartz^{2,3,14}.

Como exemplo de aplicação, analisaremos dados de um estudo em que o interesse recai nos efeitos da poluição atmosférica sobre a saúde dos habitantes da cidade de São Paulo. Mostraremos que tais modelos podem constituir boas opções para representar tanto a sazonalidade quanto a relação entre o número de óbitos e as variáveis climáticas, como temperatura e umidade.

Na seção II, descreveremos tanto os modelos lineares generalizados quanto os modelos aditivos generalizados que essencialmente constituem uma generalização dos primeiros e introduziremos alguns conceitos básicos sobre curvas de alisamento. A

seção III contém uma aplicação prática dos modelos descritos e alguns comentários sobre as vantagens e desvantagens observadas em cada abordagem.

Descrição dos modelos

Nos modelos discutidos aqui, a **variável resposta** ou a **variável de interesse** é uma contagem (por exemplo, o número diário de óbitos ou de internações hospitalares). Essa resposta é supostamente influenciada pela concentração de poluentes na atmosfera, temperatura e umidade entre outros, que são as **variáveis explicativas**.

Um modelo bastante simples, amplamente utilizado na análise desse tipo de dados, é o modelo de regressão linear gaussiana⁶. Este modelo é interessante por sua simplicidade, interpretabilidade e boas propriedades dos estimadores de seus parâmetros. Como o número de óbitos é uma contagem, nem sempre as suposições de normalidade e homocedasticidade dos erros inerentes a esses modelos são satisfeitas.

O Modelo Linear Generalizado (MLG)

É possível utilizar métodos análogos àqueles desenvolvidos para o modelo de regressão linear gaussiana, em situações em que a variável resposta obedece a outras distribuições que não a Normal, ou em que a relação entre a variável resposta e as variáveis explicativas não é linear. Isto se deve, em parte, ao conhecimento de que muitas das boas propriedades da distribuição Normal são partilhadas por uma larga classe de distribuições chamada de **família exponencial de distribuições**. Muitas distribuições conhecidas pertencem a essa família, como a própria Normal, a Poisson, a Binomial, a Gama etc⁶.

Os modelos lineares generalizados foram primeiramente apresentados por Nelder e Wedderburn¹⁵ e englobam os modelos de regressão linear simples e múltipla, regressão logística, regressão de Poisson e muitos outros, como modelos log-lineares para dados categorizados. Os ingredientes básicos

desses modelos são:

1. K valores independentes Y_1, \dots, Y_K de uma variável resposta que segue uma distribuição da família exponencial, com valor esperado $E(Y_i) = \mu_i$;
2. K vetores $x_i = (1 \ x_{i1} \ x_{i2} \ \dots \ x_{ip})^t$, $i=1, \dots, K$, contendo os valores das p variáveis explicativas;
3. uma função monotônica e diferenciável g, chamada de função de ligação, tal que $g(\mu_i) = x_i^t \beta$, $i=1, \dots, K$ (I) com $\beta = (\beta_1 \ \beta_2 \ \dots \ \beta_p)$ representando o vetor de parâmetros a serem estimados.

Se g é a função identidade, isto é, se $g(\mu_i) = \mu_i$, então $\mu_i = E(Y_i) = x_i^t \beta$, e o modelo resultante, com algumas suposições adicionais, é o modelo de regressão linear gaussiana citado anteriormente.

Se g é a função logarítmica e Y_i tem distribuição de Poisson, o modelo resultante é o modelo de regressão de Poisson, comumente utilizado para avaliar efeitos da poluição^{4,5,7-11}, entre outras aplicações.

Cada β_j em (I) determina o “efeito” da variável X_j em $g(\mu_i)$. No modelo de regressão de Poisson, o valor $e^{\Delta \beta_j}$ equivale ao risco relativo de ocorrência do evento de interesse (por exemplo, óbito ou internação), quando há um aumento Δ unidades em X_j .

O vetor de parâmetros β em (I) pode ser estimado pelo método de **máxima verossimilhança**, e os cálculos envolvem um procedimento iterativo⁶.

O modelo aditivo generalizado (MAG)

O modelo aditivo generalizado é uma extensão do modelo linear generalizado, em que o termo $x_i^t \beta = \sum_j x_{ij} \beta_j$ é substituído por $\sum_j f_j(x_{ij})$, com $f_j(x_{ij})$ denotando uma função não paramétrica (i.e. cuja forma não é especificada) estimada através de curvas de alisamento¹³. Com essa substituição, não é necessário assumir uma relação linear entre $g(\mu_i)$ e as variáveis explicativas, como no MLG. De fato, não é necessário nem mesmo conhecer previamente a forma dessa relação, mas é possível estimá-la a partir de um conjunto de dados. Essa função estimada ($\hat{f}_j(x_{ij})$), também chamada de curva alisada, em muitas situações, nada mais é do que algum tipo de média dos valores Y_i na vizinhança de um dado valor x_j . A curva alisada permite então descrever a forma, e mesmo revelar possíveis não linearidades nas relações estudadas, uma vez que não apresenta a estrutura rígida de uma função paramétrica.

Os procedimentos de estimação para os MAG são semelhantes àqueles adotados na estimação sob os MLG, valendo-se basicamente de uma combinação do algoritmo “Scoring” de Fisher⁶ com métodos de alisamento.

Com fins ilustrativos, vamos descrever um procedimento de alisamento. A Figura 1 contém o diagrama de dispersão do número diário de óbitos por causas respiratórias em idosos, $y = (y_1 \ y_2 \ \dots \ y_N)^t$, para o período de 1994 a

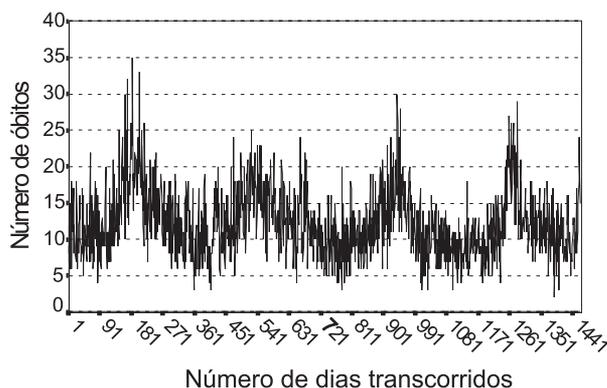


Figura 1 - Número diário de óbitos por causas respiratórias em idosos para o período 1994-1997.

Figure 1 - Daily deaths due to respiratory causes in the elderly, for the 1994-1997 period

1997. Para maior praticidade, vamos fazer $x_i = (x_1, x_2, \dots, x_N)^t$ igual ao número de dias transcorridos desde o início do estudo, isto é, $x_i = i$. Desejamos ajustar uma função que represente a relação entre y e x , isto é, $f(x)$ tal que

$$y_i = f(x_i) + \varepsilon_i.$$

No caso mais simples, vamos assumir $E(\varepsilon_i) = 0$, $\text{Var}(\varepsilon_i) = \sigma^2$, $i=1, \dots, N$, e que os erros são independentes.

Um alisador é uma função de x e y com o mesmo domínio de x , definida para todo ponto x_0 ou, algumas vezes, apenas para os x_i da amostra. Para cada valor x_i o alisador associa um valor $f(x_i)$ cuja estimativa $\hat{f}(x_i)$ pode ser obtida. Os valores desta função devem, por definição, ser mais “suaves” do que os valores de y , ou seja, devem ter menor variabilidade do que os valores de y . Uma estimativa bastante simples para $f(x_i)$ é a média dos valores de y correspondentes a valores próximos de x_i ; essa estimativa é usualmente chamada de média móvel. Esses “valores próximos” podem ser definidos a partir de um conjunto que inclui o próprio x_i

além de k pontos à sua direita e k pontos à sua esquerda, quando existirem. Formalmente, $\hat{f}(x_i) = \text{média}_{j \in V^s(x_i)}$, onde $V^s(x_i)$ é uma vizinhança próxima simétrica de x_i , isto é, $V^s(x_i) = \{\max(i-k, 1), \dots, i-1, i, i+1, \dots, \min(i+k, N)\}$. A Figura 2-b apresenta a função f estimada para $k = 90$.

Hastie e Tibshirani¹³ apresentam uma série de alisadores, alguns mais refinados do que a média móvel, como os “cubic smoothing splines” ou o “locally weighted running line smoother (loess)”. Estimativas desses alisadores para os dados da Figura 1 estão apresentadas na Figura 2.

As duas considerações mais importantes para a escolha do alisador envolvem o tipo de média dos valores de y a ser calculada e o tamanho da vizinhança (k). Este último, às vezes chamado de parâmetro de alisamento, determina o compromisso entre o viés e a variância do alisador. Intuitivamente, quanto maior o tamanho da vizinhança, mais lisa será a curva (Figura 3) e, em geral, maior o viés. Isto pode ser facilmente visualizado no caso da média móvel. Este alisador pode ser escrito como

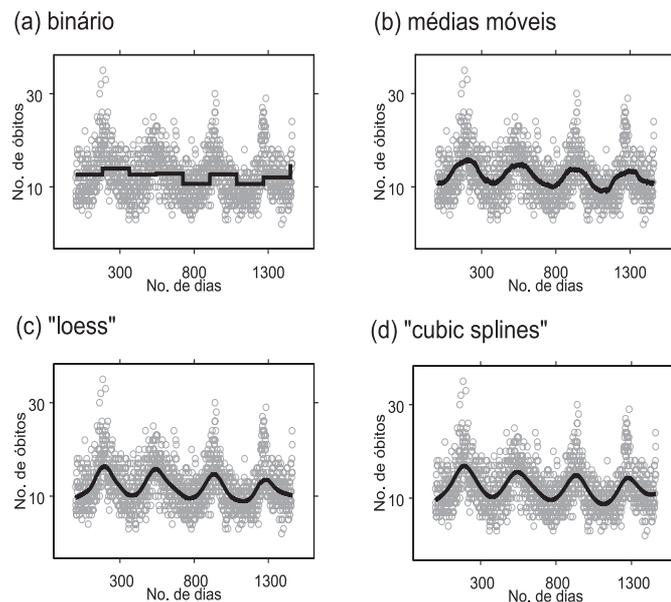


Figura 2 - Curvas de alisamento do número de óbitos por causas respiratórias em idosos em função do número de dias transcorridos.

Figure 2 - Smoothing curves for respiratory-related deaths in the elderly as a function of the number of days.

$$\hat{f}_k(x_i) = \sum_{j \in V_k^+(x_i)} \frac{y_j}{2k+1} \quad (\text{ii})$$

e tem esperança

$$E[\hat{f}_k(x_i)] = \sum_{j \in V_k^+(x_i)} \frac{f(x_j)}{2k+1} \quad (\text{iii})$$

e variância

$$\text{Var}[\hat{f}_k(x_i)] = \frac{\sigma^2}{2k+1}. \quad (\text{iv})$$

Maiores valores de k em (iv) produzem menor variância mas, em geral, aumentam o viés porque a esperança em (iii) envolverá mais termos com valores de f diferentes de $f(x_i)$. Analogamente, menores valores de k geram maior variância, mas devem produzir menos viés.

A Figura 3 apresenta o alisador “loess” aplicado ao número de óbitos respiratórios em idosos em função do número de dias transcorridos para diferentes parâmetros de

alisamento (ou tamanhos de vizinhança). Neste caso, como o objetivo é modelar a sazonalidade e uma possível tendência, devemos definir um parâmetro que apreenda apenas o padrão sazonal e a tendência. É importante que o alisador não capture a variabilidade diária do número de óbitos, porque esta pode ser explicada pelas demais variáveis, inclusive por aquelas que representam a poluição. Esta idéia também vale para a definição dos parâmetros de alisamento do número de óbitos em função das variáveis climáticas.

Exemplo

Descrição dos dados

Foram obtidos do Programa de Aprimoramento das Informações de Mortalidade no Município de São Paulo (PRO-AIM) informações referentes a todos os óbitos ocorri-

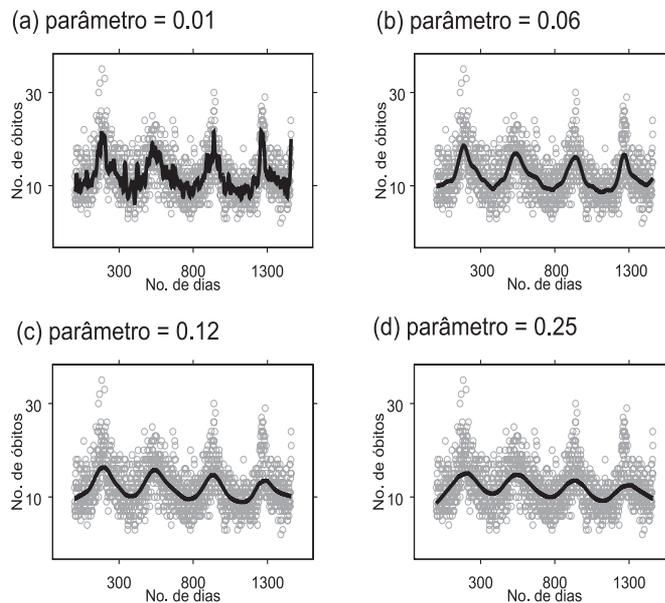


Figura 3 - Curvas de alisamento do número de óbitos respiratórios em idosos em função do número de dias transcorridos, para diferentes parâmetros de alisamento, utilizando o alisador “loess”. O parâmetro de alisamento, no caso do “loess”, corresponde à proporção de dias em cada vizinhança. Por exemplo, uma vizinhança com 180 dias corresponde a um parâmetro de alisamento aproximadamente igual a 0.12.

Figure 3 - Smoothing curves for respiratory-related deaths in the elderly as a function of the number of days for different smoothing parameters, utilizing the “loess” smoother. The smoothing parameter, for “loess”, corresponds to the ratio of days in each neighborhood. For example, a 180-day neighborhood corresponds to a smoothing parameter of roughly 0.12.

dos na cidade de São Paulo, nos anos de 1994 a 1997, por causas respiratórias, para pessoas maiores de 65 anos. A causa básica da morte foi definida segundo a Classificação Internacional de Doenças No. 9 (CID-9) até o ano de 1995; a partir de 1996 a classificação passou a obedecer as regras da CID-10.

A Companhia de Tecnologia de Saneamento Ambiental (CETESB) forneceu informações sobre as concentrações diárias dos poluentes PM_{10} ($\mu\text{mg}/\text{m}^3$), SO_2 ($\mu\text{mg}/\text{m}^3$), CO (ppm), O_3 ($\mu\text{mg}/\text{m}^3$) e NO_2 ($\mu\text{mg}/\text{m}^3$). Uma vez que as concentrações de NO_2 foram medidas em menos da metade dos dias, este poluente foi excluído da análise.

Foram obtidas do Instituto Astronômico e Geofísico da Universidade de São Paulo (IAG-USP) a temperatura mínima e máxima ($^{\circ}\text{C}$) e a umidade relativa do ar média e mínima (%) por dia.

Análise

Para fins comparativos, foram ajustados modelos das classes MLG e MAG. Em ambos os casos, a variável de interesse foi o número diário de óbitos de pessoas maiores de 65 anos. Os modelos foram construídos iterativamente, sendo que a estratégia de análise consistiu em tentar explicar ao máximo a variabilidade da mortalidade por meio dos controles para sazonalidade, tendência, clima etc., definindo um modelo básico, antes de adicionar as concentrações de poluentes.

Nos MLG, foram utilizados, como variáveis explicativas, 12 indicadores para meses do ano, 4 para anos de estudo e 7 para dias da semana (a fim de controlar a sazonalidade e a tendência), temperatura, umidade, o número de óbitos por outras causas que não as respiratórias. Foram ajustados modelos com diferentes especificações para as variáveis meteorológicas (envolvendo combinações de termos lineares e polinômios de segundo grau) e com diferentes defasagens (foram consideradas as observações do dia corrente, de dois dias anteriores e médias móveis de dois dias). A escolha do "melhor" conjunto de variáveis meteorológicas e da estrutura de defasagem

mais adequada baseou-se no critério de informação de Akaike¹³, na significância dos coeficientes das variáveis meteorológicas do modelo em questão e no nível descritivo do teste da razão de verossimilhanças que compara o modelo em questão com o modelo sem as variáveis meteorológicas.

Nos MAG, os indicadores para meses e anos foram substituídos por uma curva alisada do número de óbitos em função do número de dias transcorridos. Os polinômios envolvendo as medidas de temperatura e umidade foram substituídos por curvas alisadas do número de óbitos em função dessas variáveis. Foram utilizadas diferentes defasagens para as variáveis meteorológicas e o critério para escolha do "melhor" modelo foi similar ao descrito anteriormente. O alisador utilizado foi o "loess". O parâmetro de alisamento foi definido após a investigação de gráficos de alisamento como os da Figura 3.

Uma vez que as variáveis indicadoras para dias da semana não foram significativas, foram retiradas dos modelos.

Uma vez definidos os modelos com o "melhor" conjunto de controles para sazonalidade e clima, foram incluídas as concentrações de poluentes, considerando as medidas do dia corrente, defasagens até 3 dias e médias móveis até 7 dias.

Também foram ajustados modelos contendo variáveis indicadoras para categorias das concentrações de poluentes em lugar das variáveis contínuas, com o objetivo de melhor visualizar o efeito da poluição à medida que a sua concentração aumenta. Os limites superiores de cada categoria foram os quintis. Assim, a primeira categoria contém os 20% dos dias menos poluídos do período de estudo, a segunda categoria contém os 20% seguintes, e assim por diante. A última categoria contém os dias mais poluídos do período de estudo. Fixando a primeira categoria como referência, a exponenciação do coeficiente de cada variável indicadora produz o risco relativo de mortalidade em cada categoria em relação à primeira. Isto permite avaliar se existe um limite mínimo para a concentração de cada poluente, abaixo do

qual não há grandes alterações no risco de mortalidade.

Finalmente, o modelo MAG foi reajustado com a utilização de um método de estimação de “quasi-verossimilhança”, segundo o modelo de Zeger para análise de séries de contagens¹⁶, com o objetivo de controlar uma possível autocorrelação ou sobredispersão nos resíduos dos modelos. Foram incluídos parâmetros autoregressivos até a ordem 7, permanecendo no modelo apenas os termos significantes.

A análise foi realizada com o auxílio do programa S-plus 4¹⁷.

Resultados

O parâmetro de alisamento para o número de óbitos em função do tempo foi fi-

xado em 0.12 (Figura 3). Isto significa que a vizinhança em cada ponto contém 12% das observações, o que corresponde a 175 pontos (aproximadamente 6 meses).

Os parâmetros de alisamento para a temperatura e para a umidade (médias móveis de dois dias) foram fixados em 0.5.

As Tabelas 1 e 2 mostram os modelos MLG e MAG com o “melhor” conjunto de controles para sazonalidade e clima. As Figuras 4 a 6 apresentam as curvas de alisamento estimadas para o número de óbitos em função das variáveis explicativas no “melhor” MAG.

Na Tabela 1, os coeficientes das variáveis indicadoras para meses e anos indicam um aumento significativo da mortalidade nos meses de inverno, além de uma redução ao longo dos anos. Este comportamento pode

Table 1 - Estimativas para os parâmetros do GLM, antes da inclusão dos poluentes.

Table 1 - Estimates for GLM parameters, before including pollutants

Variáveis explicativas	coeficiente	erro padrão
intercepto	2.671**	.146
meses		
janeiro		
fevereiro	.000	.044
março	.093*	.043
abril	.128**	.046
maio	.330**	.050
junho	.563**	.054
julho	.533**	.053
agosto	.341**	.054
setembro	.315**	.051
outubro	.170**	.047
novembro	.129**	.045
dezembro	.060	.043
anos		
1994		
1995	-.021	.022
1996	-.138**	.023
1997	-.154**	.023
variáveis meteorológicas		
temperatura - média móvel 2 dias	-.045**	.015
temperatura ²	.002**	.001
umidade - média móvel 2 dias	-.004**	.001
óbitos não-respiratórios	.004**	.001
AIC	1608.67	
nível descritivo (p) #	< 0,01	

* p<.05; **p<.01

teste que compara o modelo em questão com o modelo sem as variáveis meteorológicas

test that compares the model being assessed with the model without meteorology variables

Tabela 2 - Estimativas para os parâmetros do MAG, antes da inclusão dos poluentes.
Table 2 - Estimates for GAM parameters, before including pollutants.

Variáveis explicativas	coeficiente	erro padrão
intercepto	2.523*	.099
f (para o No. de dias transcorridos)	-2.521*	-8.083
variáveis meteorológicas		
f (temperatura – média móvel 2 dias)	-1.586*	-4.784
umidade – média móvel 2 dias	-.005*	.001
óbitos não-respiratórios	.006*	.001
AIC		1625.53
nível descritivo (p) #		< 0,01

*p<.01

teste que compara o modelo em questão com o modelo sem as variáveis meteorológicas

test that compares the model being assessed with the model without meteorology variables.

ser claramente visualizado na Figura 4. O alisador apreende a sazonalidade de modo satisfatório, bem como a tendência de decréscimo da mortalidade ao longo dos anos, sem capturar a variabilidade diária da mortalidade.

A Figura 5 mostra a curva de alisamento ajustada para o número de óbitos respiratórios de idosos em função da temperatura sob o “melhor” MAG. Convém interpretar esta figura para valores de temperatura en-

tre 10 e 20 graus Celsius, onde é maior o número de observações. Pode-se dizer que para temperaturas mais baixas (na faixa de 10 a 16 graus Celsius) o número de óbitos decresce à medida que a temperatura aumenta; para temperaturas a partir de 16 graus Celsius, o número de óbitos parece não sofrer alterações consideráveis.

A Figura 6 sugere que a relação entre o número de óbitos e a umidade deve ser linear. O modelo que inclui apenas um termo

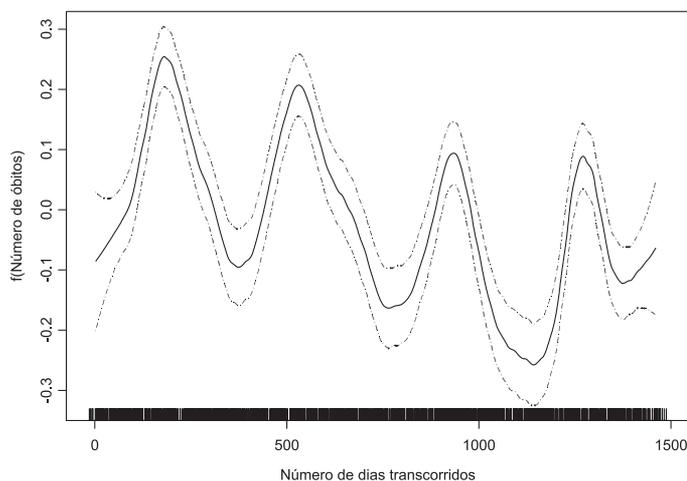


Figura 4 - Curva alisada do número de óbitos respiratórios em idosos em função do número de dias transcorridos no MAG. As linhas pontilhadas são uma aproximação para o intervalo de confiança da curva de alisamento. Cada barra vertical no eixo das abscissas representa uma observação.

Figure 4 - Smoothing curve of respiratory-related deaths in the elderly as a function of the number of days in GAM. The dotted lines are an approximation for the confidence interval of the smoothing curve. Each vertical bar in the x axis represents an observation.

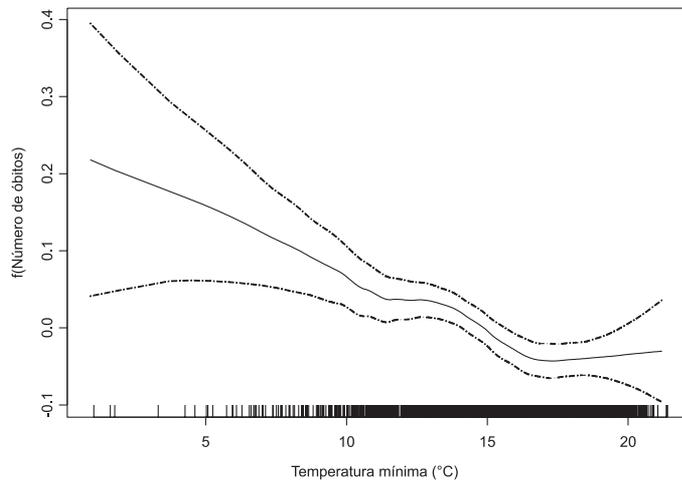


Figura 5 - Curva alisada do número de óbitos respiratórios em idosos em função da temperatura no MAG.

Figure 5 - Smoothing curve of respiratory - related deaths in the elderly as a function of temperature in GAM.

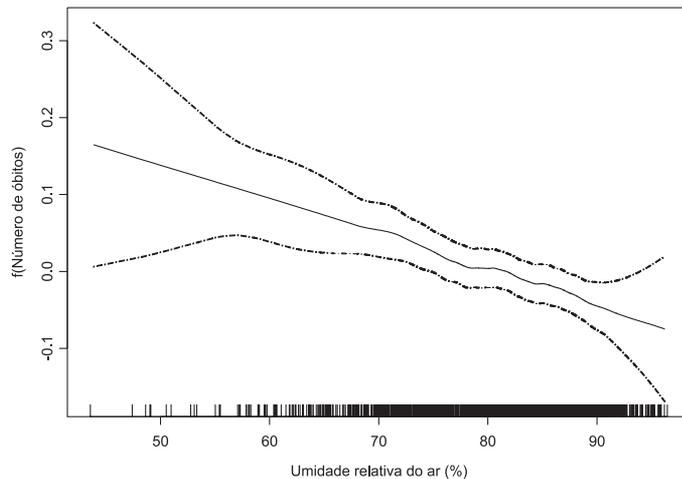


Figura 6 - Curva alisada do número de óbitos respiratórios em idosos em função da umidade relativa do ar no MAG.

Figure 6 - Smoothing curve of respiratory related deaths in the elderly as a function of relative air humidity in GAM.

linear para a umidade foi re-estimado sem a utilização de um alisador.

A Tabela 3 contém os coeficientes e respectivos erros padrão estimados sob os modelos MLG e MAG para cada poluente, quando incluídos separadamente ou em conjunto no modelo. Em geral, as estimativas da magnitude do efeito da poluição foram maiores sob os MAG; este modelo também produziu os menores erros padrão. Sob

os MLG, nenhum poluente explicou a mortalidade de forma significativa. Sob os MAG, houve efeitos significativos para CO , SO_2 e PM_{10} . Quando todos os poluentes foram incluídos simultaneamente no modelo, apenas o SO_2 permaneceu significativo. O Ozônio não se mostrou associado à mortalidade em nenhum dos modelos ajustados. Após o controle da autocorrelação e da sobredispersão, o PM_{10} deixou de ser significativo.

Tabela 3 - Coeficiente de regressão (e erro padrão) para cada poluente, quando incluídos separadamente ou juntos no modelo controlado para sazonalidade, clima e número de óbitos por causas não respiratórias.

Table 3 - Regression coefficient (and standard error) for each pollutant, when included alone or together, in the model controlled for seasonality, climate and number of deaths due to non-respiratory causes.

Poluentes	Separados		Juntos	
	MLG	MAG	MLG	MAG
PM ₁₀ ^a	.0000 (.0005)	.0009** (.0004)	-.0005 (.0007)	-.0009 (.0006)
SO ₂ ^a	.0020* (.0012)	.0051*** (.0009)	.0030* (.0017)	.0069*** (.0013)
CO ^a	-.0019 (.0067)	.0177*** (.0054)	-.0008 (.0097)	.0115 (.0086)
O ₃ ^b	.0001 (.0003)	-.0002 (.0002)	.0001 (.0003)	-.0002 (.0003)

* p < 0.10 ** p < 0.05 *** p < 0.01

^a média móvel de cinco dias/^a *mobile five-day mean*

^b concentração do dia corrente/^b *concentration of the current day*

A Figura 7 apresenta o risco relativo de mortalidade para categorias com diferentes níveis de concentração de poluentes, sob os modelos MLG e MAG. Embora o padrão das curvas seja semelhante, um comportamento dose-resposta tornou-se evidente apenas sob os MAG (convém lembrar que a poluição não foi significativamente associada à mortalidade sob os MLG).

Comentários finais

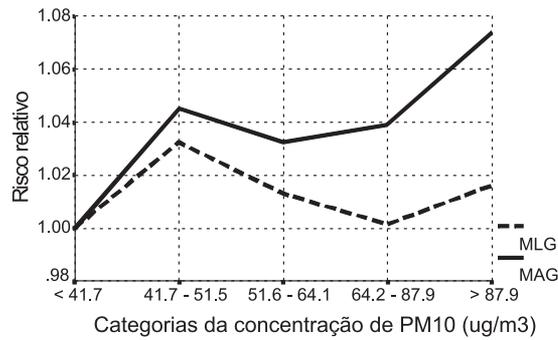
As duas classes de modelos apresentadas produziram resultados coerentes, mas os modelos estatisticamente mais sofisticados tiveram mais poder para detectar efeitos significativos que foram, como esperado, de pequena magnitude. Isto pode ter ocorrido porque os MAG incorporam informações acerca da sazonalidade, tendência e da relação entre mortalidade e variáveis meteorológicas de um modo menos restritivo. A abordagem que utiliza variáveis indicadoras impõe a restrição de que o número de óbitos ao longo de todo o mês é constante e só vai se alterar no primeiro dia do mês seguinte, quando deverá permane-

cer constante novamente até o final do mês e assim por diante. Do mesmo modo, as variáveis indicadoras para anos de estudo têm o objetivo de eliminar a tendência do número de óbitos, mas este procedimento impõe uma restrição semelhante. A abordagem baseada em curvas de alisamento permite que o padrão sazonal seja definido pelos próprios dados, sem a imposição de uma estrutura rígida e, talvez, menos fidedigna.

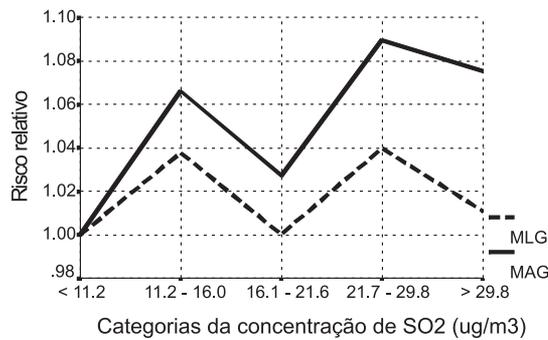
O ajuste de curvas de alisamento também possibilita avaliar a relação entre o número de óbitos e a temperatura, sem a necessidade de definir previamente a forma desta relação. A curva ajustada sugere que o número de óbitos aumenta consideravelmente na presença de baixas temperaturas, mas não na presença de temperaturas muito altas.

A desvantagem da utilização de curvas de alisamento é que os coeficientes estimados correspondentes a essas curvas nos modelos de regressão não são interpretáveis, e a estimação do risco relativo de mortalidade associado a variações na temperatura envolve procedimentos computacionais intensos e que ainda não foram implemen-

(a) PM₁₀



(b) SO₂



(c) CO

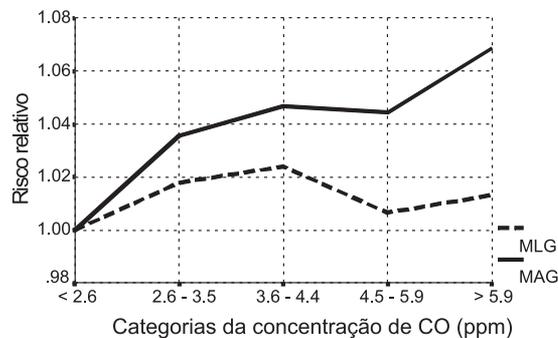


Figura 7 - Risco relativo de mortalidade de idosos por causas respiratórias segundo as concentrações de cada poluente.

Figure 7 - Relative risk of respiratory related mortality according to concentrations of each pollutant. Tabela 1. Estimativas para os parâmetros do MLG, antes da inclusão dos poluentes.

tados nos pacotes estatísticos atuais. Entretanto, sempre que houver interesse direto na estimação do risco, os alisadores podem ser utilizados em uma primeira abordagem, com o objetivo de sugerir ao pesquisador que tipo de função paramétrica (polinômios

de 2º grau, 3º grau etc.) ou variável indicadora seria mais adequada.

De um modo geral, as curvas estimadas para o risco relativo de mortalidade em função das categorias da concentração de poluentes apresentaram um padrão seme-

lhante. O padrão de comportamento do efeito da poluição não parece depender da classe de modelo adotado, mas os MAG mostraram-se mais eficientes para detectar associações.

Aparentemente, os MAG são modelos mais parcimoniosos, ou seja, necessitam de um menor número de variáveis explicativas, o que justificaria o fato de detectarem um maior número de associações. Por exemplo, excetuando a concentração de poluentes, foram consideradas 18 variáveis explicativas nos MLG, enquanto que apenas 4 foram requeridas pelos MAG. Na verdade, toda curva de alisamento está associada a um determinado número de graus de liberdade¹³. O número de graus de liberdade para as curvas de alisamento do número de óbitos em função do número de dias transcorridos foi aproximadamente 14 e, portanto, equivalente ao número de graus de liberdade correspondentes às variáveis indicadoras para meses e anos.

Alguns estudos avaliaram a relação entre poluição atmosférica e mortalidade nos Estados Unidos, valendo-se de diferentes métodos estatísticos e realizaram uma análise comparativa de tais métodos. Entre estes podemos citar Schwartz e colaboradores¹⁸ e Kinney e colaboradores¹⁹. Entretanto, estas análises envolveram basicamente o PM₁₀.

Schwartz e colaboradores¹⁸ avaliaram a sensibilidade da associação entre mortalidade total e PM₁₀ em Birmingham, comparando modelos de regressão de Poisson com controle da autocorrelação e da sobredispersão, modelos de regressão robusta (que são ba-

sicamente modelos de regressão linear múltipla ponderada²⁰ e modelos de regressão de Poisson utilizando alisadores. Foi observada uma associação significativa entre mortalidade e PM₁₀, sendo que essencialmente os mesmos resultados foram obtidos nas diferentes análises empregadas. Nenhum outro poluente foi investigado.

Kinney e colaboradores¹⁹ avaliaram a sensibilidade da associação entre mortalidade total e as concentrações de PM₁₀ em Los Angeles, comparando o risco relativo associado às concentrações do PM₁₀ em três classes de modelos: regressão linear múltipla, regressão log-linear (regressão linear múltipla com o logaritmo do número de óbitos como variável resposta) e regressão de Poisson. Diferentes alternativas para controles da sazonalidade e especificações para variáveis meteorológicas foram consideradas. O risco relativo associado ao PM₁₀ não se mostrou sensível às classes de modelos utilizadas mas, na maioria dos modelos, os riscos estimados foram apenas marginalmente significantes.

Embora este trabalho tenha um enfoque de cunho metodológico, é importante ressaltar que os resultados observados reiteraram a hipótese de que a poluição urbana pode levar à morte. Foi observada uma associação entre mortalidade em idosos por causas respiratórias e poluição atmosférica, representada pelos níveis de CO, SO₂ e, em menor escala, PM₁₀, na cidade de São Paulo, para o período de 1994 a 1997. Essa associação é dose-dependente e evidente após um curto período de exposição.

Referências

1. Schwartz, J, Spix C, Touloumi G, Bachárová L, Barumamdzadeh T, Tertre A et al. Methodological issues in studies of air pollution and daily counts of deaths or hospital admissions. *J Epidemiol Community Health* 1996; 50 Suppl 1:S3-S11.
2. Schwartz J, Dockery DW, Neas LM. Is daily mortality associated specifically with fine particles? *J Air Waste Manage Assoc* 1996; 46:927-39
3. Schwartz J. Air pollution and hospital admissions for respiratory disease. *Epidemiology* 1996; 7:20-8.
4. Saldiva PHN, Pope III CA, Schwartz J, Dockery DW, Lichtenfels AJFC, Salge JM et al. Air pollution and mortality in elderly people: a time-series study in São Paulo, Brazil. *Arch Environ Health* 1995; 50:159-63.

5. Lin CA, Martins MA, Farhat SCL, Pope III CA, Conceição GMS, Anastácio VM et al. Air pollution and respiratory illness of children in São Paulo, Brazil. ***Pediatr Perinat Epidemiol*** 1999; 13:475-88.
6. Dobson AJ. *An introduction to generalized linear models*. 2.ed. New York:Chapman & Hall; 1990.
7. Schwartz J, Dockery DW. Particulate air pollution and daily mortality in Steubenville, Ohio. ***Am J Epidemiol*** 1992;135:12-9.
8. Saldiva PHN, Lichtenfels AJFC, Paiva PSO, Barone IA, Martins MA, Massad E et al. Association between air pollution and mortality due to respiratory diseases in children in São Paulo, Brazil: a preliminar report. ***Environ Res*** 1994;65:218-25.
9. Schwartz J. Air pollution and hospital admissions for the elderly in Detroit, Michigan. ***Am J Respir Crit Care Med*** 1994;150:648-55.
10. Pereira LAA, Loomis D, Conceição GMS, Braga ALF, Arcas RM, Kishi H et al. Association between air pollution and intrauterine mortality in São Paulo, Brazil. ***Environ Health Perspect*** 1998; 106:325-29.
11. Braga ALF, Conceição GMS, Pereira LAA, Kishi H, Pereira JCR, Andrade MF et al. Air pollution and pediatric respiratory admissions in São Paulo, Brazil. ***J Environ Med*** 1999; 1:95-102.
12. Katsouyanni K, Schwartz J, Spix C, Touloumi G, Zmirou D, Zanobetti A et al. Short term effects of air pollution on health: an European approach using epidemiologic time series data: the APHEA protocol. ***J Epidemiol Community Health*** 1996; 50 Suppl 1:S12-S18.
13. Hastie T, Tibshirani R. *Generalized additive models*. London:Chapman & Hall, 1990.
14. Schwartz J. Nonparametric smoothing in the analysis of air pollution and respiratory illness. ***Can J Stat*** 1994; 22: 471-87.
15. Nelder JA, Wedderburn RWM. *Generalized linear models*. ***J R Stat Soc A*** 1972; 135:370-84.
16. Zeger SL. A regression model for times series of counts. ***Biometrika*** 1988; 75:621-29.
17. Mathsoft. *S-PLUS 4 user's guide*. Seattle: Data Analysis Products Division; 1997.
18. Schwartz J, Slater D, Larson TV, Pierson WE, Koenig J. Particulate air pollution and hospital emergency room visits for asthma in Seattle. ***Am Rev Respir Dis*** 1993; 147:826-31.
19. Kinney PL, Ito K; Thurston GD. A sensitivity analysis of mortality/PM₁₀ associations in Los Angeles. ***Inhal Toxicol*** 1995;7:59-69.
20. Neter J, Kutner MH, Nachtsheim CJ, Wasserman W. *Applied linear statistical models*. 4.ed. Chicago: Times Mirror Higher Education Group, 1996.