

ASSOCIAÇÕES SIGNIFICATIVAS ENTRE DADOS DE TEMPERATURA DO AR E DE SAÚDE EM CIDADES LITORÂNEAS E NÃO LITORÂNEAS

Joselene Marques

Joseph Harari

RESUMO

Este estudo visa determinar associações entre dados de saúde de cidades do Estado de São Paulo (capital, três cidades do interior e uma cidade do litoral) e seus respectivos dados de temperatura do ar. Os dados de saúde, correspondentes ao número de internações ocorridas (casos registrados), por local de residência, para doenças cardiovasculares e respiratórias (CID-I e CID-J), foram extraídos do banco de dados de saúde do DATASUS (Departamento de Informática do Sistema Único de Saúde do Brasil). Os dados de temperatura do ar foram obtidos dos modelos operacionais do NCEP/NCAR (Centro Nacional de Previsão Ambiental/Centro Nacional de Pesquisa Atmosférica dos EUA) e das estações meteorológicas do INMET (Instituto Nacional de Meteorologia, do Brasil). Os dados analisados são referentes ao período de 1º de janeiro de 2008 a 31 de dezembro de 2016 (exceto para a cidade de Bauru, que só tinha dados a partir de 2011). Com a análise das séries temporais (por meio do tratamento estatístico dos dados e da aplicação dos métodos ARIMA, SARIMA, ARIMAX e SARIMAX), procurou-se determinar se há variações nos parâmetros de saúde devido às condições ambientais

e climáticas das cidades analisadas, ou seja, se há associações estatísticas entre registros de internações por CID-I e CID-J e dados de temperatura do ar. As análises demonstram que há falhas nos registros de informações das internações, que possuem comportamento periódico semanal (provavelmente em função de ausências de registros em finais de semana). Os modelos apresentaram dependência da diferenciação das séries, ou seja, eram funções diretas do operador de atraso das séries. O estudo das correlações cruzadas mostrou de forma clara que há correlação considerável entre os registros de internações CID-I e CID-J e os dados de temperatura do ar média, máxima e mínima de cada cidade, sendo essas correlações maiores para dados de saúde e de temperatura defasados.

Palavras-chave: Temperatura do Ar; Doenças Respiratórias; Doenças Cardíacas; ARIMA; SARIMA; ARIMAX; SARIMAX.

1. INTRODUÇÃO

Este capítulo é referente à Tese de Doutorado “Associações significativas entre dados de temperatura e de saúde em cidades litorâneas e não litorâneas”, defendida no PROCAM em 2020.

A referida tese teve como motivação inicial os estudos nas últimas décadas sobre as mudanças climáticas que vêm afetando e assolando o mundo e que contribuíram para a criação do Painel Intergovernamental sobre Mudanças Climáticas (IPCC), cujo principal objetivo é sintetizar e divulgar informações científicas sobre as mudanças climáticas associadas a atividades humanas (IPCC, 2019).

Nos estudos referentes a mudanças climáticas (HARARI; BRITO, 2005; JACOBI, 2014, 2015; HARARI; SIEGLE, 2012), diversos autores analisaram a ação humana sobre a natureza, promovendo alterações como o aumento da temperatura média da Terra, com diversos desdobramentos, como mudanças nos ecossistemas, a elevação do nível médio do mar, a migração de populações, a incidência e o aumento de determinadas doenças em certas regiões (BARCELLOS et al., 2009).

Ademais, em muitas regiões do planeta, as condições ambientais, sociais e econômicas vêm sofrendo modificações, muitas vezes abruptas, por conta de processos de urbanização e crescimento das cidades sem uma efetiva preocupação com o meio ambiente. Esses processos de urbanização também acabam alterando as condições do tempo e do clima e influenciam o meio ambiente, trazendo problemas de qualidade para a vida humana (BAKONYI et al., 2004; MARCÍLIO; HAJAT; GOUVEIA, 2013; ALEIXO; NETO, 2014; PARAISO; GOUVEIA, 2015; SON et al., 2015).

Estudos sobre a relação entre o clima e a saúde pública, levando em conta mudanças ambientais, se encontram em Conceição et al. (2015). A relação temperatura do ar e saúde é discutida em Gouveia, Freitas, Martins e Marcílio (2006), embora o estudo fosse focado na influência da poluição e das mudanças ambientais na saúde, já que, associado a isso, temos a variação das temperaturas e do clima; nesse estudo, os autores salientam que mesmo com os notáveis avanços obtidos nas últimas décadas em direção a um ambiente com ar mais limpo, especialmente nos países desenvolvidos, os atuais níveis de poluição experimentados pela maior parcela da população humana continuam a ser danosos à saúde.

As pesquisas sobre a relação entre clima e saúde têm sido desenvolvidas em muitos aspectos, como nos estudos de Souza e Moraes (2003), que apresentam correlações de mudanças climáticas com doenças e epidemias em diferentes regiões do Brasil, e Moraes e Gomes (2009), que analisaram como alterações climáticas podem influenciar os ciclos dos vetores causadores de doenças (mosquitos). Silva, Ribeiro e Santana (2014) demonstram a relevância do estudo das relações do clima com a saúde, enfatizando as mudanças climáticas globais, seus efeitos e as vulnerabilidades da sociedade.

O presente estudo procura verificar se há associações entre dados de saúde com dados de temperatura do ar no período de 2008 a 2016, por métodos estatísticos, em escalas de tempo sazonais e de longo prazo.

2. HIPÓTESE E OBJETIVO

A hipótese inicial deste trabalho é que há associações estatísticas entre os registros de internações por CID-I e CID-J e a temperatura do ar, em escalas temporais de dias, semanas e meses, nas cidades de São Paulo, Santos, Bauru, São Carlos e Presidente Prudente, no período de 2008 a 2016.

Dessa forma, o objetivo central é estudar as associações estatísticas entre o número de pessoas internadas com problemas cardíacos e respiratórios (doentes dos tipos I e J) e dados de temperatura do ar nos locais e nos períodos de interesse citados.

3. MATERIAIS

Os dados de temperatura do ar foram extraídos do NCEP/NCAR (Centro Nacional de Previsão Ambiental/Centro Nacional de Pesquisa Atmosférica, dos EUA) e do INMET (Instituto Nacional de Meteorologia, do Brasil), de forma que nos casos em que não havia dados suficientes no NCEP/NCAR utilizavam-se ou

complementavam-se os dados com os dados das estações do INMET. Vale destacar que os dados do INMET que foram utilizados são dados obtidos pelas estações próximas ou na região das cidades consideradas, de modo que o valor utilizado é uma média da região que a cidade ocupa.

Dados de saúde foram obtidos do DATASUS (Departamento de Informática do Sistema Único de Saúde do Brasil), sendo referentes a número de casos de internações, por local de residência, para dois tipos de doenças do CID-10 (Classificação Internacional de Doenças) – CID-I e CID-J, correspondentes a doenças cardiovasculares e respiratórias, respectivamente.

A região estudada é formada por cinco municípios do Estado de São Paulo: São Paulo (capital), Santos (no litoral), Bauru, São Carlos e Presidente Prudente (no interior do Estado), proporcionando, assim, uma comparação entre a capital, o interior e o litoral do Estado. O período de análise é de 1º de janeiro de 2008 a 31 de dezembro de 2016 (excetuando-se a análise feita para Bauru, que só tinha dados a partir de 2011). Na escolha das cidades para o estudo, foram consideradas características sociais, políticas e econômicas importantes, como o IDH (Índice de Desenvolvimento Humano), de modo que as comparações fossem mais significativas.

A pesquisa procurou correlacionar os dados de saúde com os dados de temperatura do ar, para determinar se há associações entre eles a partir de modelos estatísticos. Os dados de temperatura do ar foram retirados dos arquivos do NCEP/NCAR e do INMET, e os dados de saúde foram obtidos do DATASUS (2002), filtrando os dados de doentes internados e/ou atendidos nas cidades analisadas com problemas cardíacos e respiratórios (tipos I e J), no período de 2008 a 2016.

4. MÉTODOS

4.1. Séries temporais e processos estocásticos

Uma série temporal é, basicamente, uma sequência de observações tomada ao longo de um certo período, geralmente a intervalos de tempo iguais. Um processo estocástico pode ser definido como uma variável ou uma coleção de variáveis aleatórias ordenadas no tempo. Num processo estocástico, dados aleatórios formam uma série temporal, de modo que, a cada instante, há um valor para cada variável aleatória (por exemplo, temperatura, pressão, umidade etc.). Por seu turno, uma variável aleatória de um processo estocástico (temperatura, por exemplo) assume valores numéricos e tem um resultado que é determinado por um experimento. O desafio de um cientista é, portanto, tendo acesso apenas à série temporal,

buscar compreender um processo estocástico desconhecido. Quanto melhor for essa compreensão, melhor será a modelagem e, conseqüentemente, a previsão de observações futuras (MORETTIN; TOLOI, 2006).

Os dados de séries temporais geralmente não são independentes, especialmente se os intervalos de amostragem são pequenos. As observações em instantes próximos costumam ser mais parecidas que observações em instantes mais afastados (como na temperatura do ar medida a cada hora) – ou seja, as medidas geralmente são mais próximas das medidas dos dias mais próximos, seja no passado ou no futuro. Um processo estocástico fica caracterizado se definimos a função da distribuição conjunta das variáveis aleatórias para qualquer período de observações. Portanto, diz-se que um processo estocástico é conhecido se a estrutura probabilística do processo é conhecida (o que, em geral, requer observar muitas realizações); isso significa conhecer a média, a variância, a autocovariância e autocorrelação.

Um processo estocástico é estacionário quando as propriedades estatísticas de qualquer sequência finita são semelhantes às de outra sequência finita, ou seja, a distribuição das variáveis não se modifica ao longo do tempo. Portanto, num processo estacionário, os momentos de primeira e segunda ordem (média e variância) são constantes, e a covariância depende somente da separação no tempo.

Na tese, utiliza-se o conceito de ruído branco, que é um processo estocástico em que todas as variáveis aleatórias seguem uma distribuição normal de média zero e variância constante, e as covariâncias são nulas.

A diferenciação em estatística é uma transformação aplicada aos dados de uma série temporal a fim de tornar essa série estacionária. Na diferenciação, as diferenças entre observações consecutivas são calculadas, de modo a remover as mudanças nos níveis da série temporal, eliminando tendências e, conseqüentemente, estabilizando a média da série temporal. Pode ser necessário, às vezes, diferenciar os dados d vezes para obter uma série temporal estacionária, processo referido como diferenciação de ordem d . Um exemplo de uma diferenciação de segunda ordem numa série temporal $X(t)$ para obtenção do valor modelado $Y(t)$ é dado por

$$\Delta X(t) = X(t) - X(t - 1) \quad (1)$$

$$Y(t) = \Delta X(t) - \Delta X(t - 1) = (X(t) - X(t - 1)) - (X(t - 1) - X(t - 2)) \quad (2)$$

$$Y(t) = X(t) - 2 \times X(t - 1) + X(t - 2) \quad (3)$$

Além de diferenciação, dois operadores são importantes na análise de séries temporais, envolvendo processos autorregressivos e médias móveis.

Dizemos que X_t é um processo autorregressivo de ordem p e escrevemos $X_t \sim \text{AR}(p)$ se pudermos escrever o processo na seguinte forma:

$$Y_t = \emptyset_0 + \emptyset_1 X_{t-1} + \dots + \emptyset_p X_{t-p} + \varepsilon_t \quad (4)$$

em que $\emptyset_0, \emptyset_1, \dots, \emptyset_p$, são parâmetros reais e ε_t é o ruído.

Um processo de médias móveis com $q = 2n + 1$ valores e pesos w_i aplicado a uma série temporal X_t é representado por

$$Y_t = \frac{[\sum_{-n}^{+n} w_i X_{t+i}]}{[\sum_{-n}^{+n} w_i]} \quad (5)$$

Os tipos de modelos utilizados na análise dos dados do presente trabalho foram: modelos autorregressivos (AR), modelos de médias móveis (MA), modelos autorregressivos e de médias móveis (ARMA), modelos autorregressivos integrados de médias móveis (ARIMA), modelos autorregressivos integrados sazonais de médias móveis (SARIMA), modelos autorregressivos integrados de médias móveis com variável auxiliar (ARIMAX) e modelos autorregressivos integrados de médias móveis sazonais com variável auxiliar (SARIMAX).

Modelos autorregressivos AR de ordem p , denotados como AR(p), representam as variáveis com uma dependência linear de p valores anteriores. Modelos de médias móveis MA de ordem q , denotados como MA(q), representam as variáveis como médias móveis de q valores. Modelos ARMA(p, q) consideram p componentes autorregressivos e q componentes de médias móveis. Modelos ARIMA(p, d, q) são utilizados nos casos em que o critério de estacionariedade da série temporal não é atendido; nesse caso, deve-se aplicar a diferenciação na série até que se torne estacionária; os parâmetros desse modelo, sendo p, d, q interpretados como: p é o número de parâmetros do tipo autorregressivo, d é o número de diferenciações que a série requer até se tornar estacionária e q é o número de parâmetros de média móvel. Modelos SARIMA(p, d, q)(P, D, Q) m correspondem aos modelos ARIMA(p, d, q), com componentes sazonais incorporados, em que m se refere ao número de períodos da sazonalidade e (P, D, Q) se referem aos termos de autorregressão, diferenciação e média móvel para a parte sazonal do modelo.

Modelos ARIMAX e SARIMAX são modelos ARIMA e SARIMA que fazem uso de uma ou mais variáveis auxiliares, de modo que os coeficientes desses modelos são dados pelos coeficientes autorregressivos integrados e de média móvel em conjunto com variáveis auxiliares.

4.2. Análise de resíduos

Depois de um modelo ter sido ajustado a uma série temporal, deve-se verificar se o modelo fornece uma descrição adequada dos dados. Para verificar o comportamento dos resíduos, temos:

$$\text{resíduo} = \text{observação} - \text{valor ajustado.} \quad (6)$$

Se o modelo tiver um “bom” ajuste, espera-se que os resíduos se distribuam aleatoriamente em torno de zero, com variância aproximadamente constante, e sejam não correlacionados (devem corresponder a um ruído branco).

Existem diversas maneiras de verificar a adequação de um modelo, analisando a hipótese de ruído branco para os seus resíduos; no presente trabalho foram utilizadas 5 formas de se fazer essa verificação:

- 1) Akaike (1973) estabeleceu um critério de informação que serve para averiguar a qualidade relativa dos modelos estatísticos para dados em estudo (é também denominado Critério de Informação de Akaike, AIC). No decorrer do processo de construção dos modelos ou de sua estimação, o AIC estima a qualidade de cada modelo, em que o modelo preferido é aquele com o menor valor AIC. O AIC sugere escolher o modelo cujas ordens p e q minimizam o critério definido.
- 2) Gráfico temporal: pode revelar a presença de dados discrepantes, efeitos de autocorrelação ou padrões cíclicos, além de revelar se a série parece ser estacionária ou não.
- 3) Gráfico da função de autocovariância (f. a. c.) dos resíduos, para verificar as características de ruído branco; a função de autocorrelação parcial (f. a. c. p.) é uma medida da correlação entre as observações de uma série temporal que são separadas por k unidades de tempo ($y(t)$ e $y(t-k)$) após o ajuste para a presença de todos os outros termos de menor separação ($y(t-1)$, $y(t-2)$, ..., $y(t-k-1)$).
- 4) Teste de estatística de Ljung-Box: esse é um teste que analisa os valores da série dos resíduos, para verificar se são independentes, sendo, assim, uma verificação da hipótese de ruído branco.
- 5) Histograma: mostra a distribuição dos erros.

4.3. Previsão e ajuste

O critério de precisão usado foi o Erro relativo, definido como a razão do Desvio Padrão Absoluto da Média (MAD) pela média da série. Desse modo, o erro relativo, ao ser multiplicado por 100%, corresponde a uma porcentagem de erro médio definida pelo MAD.

A vantagem do erro relativo é que ele pode ser utilizado para definir a precisão de um modelo, mesmo para séries com valores baixos ou nulos. A desvantagem é que, ainda assim, para séries com valores baixos ou nulos, pequenos erros acabam fornecendo um erro relativo considerável. Dessa forma, existem modelos com um erro relativo grande, mas que ainda assim são precisos.

A função correlação cruzada (f.c.c. ou CCF, do inglês Cross Correlation Function) é uma medida de similaridade entre duas séries temporais, levando em conta um atraso aplicado a uma delas. Essa medida leva, portanto, em consideração a similaridade das séries, com ou sem defasagens entre elas.

4.4. Associações estatísticas

Associações estatísticas são qualquer tipo de associação ou relação entre um tipo de dado com outro, sendo definidas por algum tipo de análise estatística. Dado isso, a fim de provar que existem associações estatísticas e uma conexão entre o número de internações CID-I e CID-J com a temperatura do ar, foram feitas diversas análises:

- vi) Análise da f.c.c. entre o número de internações CID-I e CID-J e os dados de temperatura do ar, verificando se há alta correlação entre os dois tipos de dados;
- vii) Análise secundária da f.c.c., para verificar uma possível relação com explicação física entre internações e temperatura, verificando se há uma maior correlação com dados do passado do que com dados do futuro – se há uma relação física, certamente há uma dependência dos dados do passado e, com isso, uma maior correlação com eles;
- viii) Análise da precisão dos modelos ARIMAX e SARIMAX com variáveis auxiliares de temperatura, a fim de verificar se há um aumento na precisão do ajuste e das previsões, visto que, se há uma relação entre internações CID-I e CID-J com a temperatura do ar, a inclusão de dados de temperatura nos modelos pode aumentar sua precisão.

4.5. Implementação computacional

Os modelos das séries temporais foram obtidos com o auxílio computacional do Software R (R *Development Core Team* 2008); para estimar os parâmetros de interesse dos modelos da classe ARIMA, foi necessário instalar pacotes adicionais (*packages*) de comandos TSA (*Time Series Analysis*) (CRYER; CHAN, 2012).

O R é um programa com parte gráfica muito eficiente para gerar e plotar gráficos das representações de séries, funções de autocorrelação, autocorrelações parciais etc. O R possui funções específicas para plotar gráficos de ajuste e para fazer previsões. A fim de fazer um único gráfico com o que se nomeou como ajuste e previsão, inicialmente foram feitos o ajuste e a previsão de forma separada. A seguir, agregaram-se os dados referentes às duas séries, para gerar uma série nova e, assim, plotar um novo gráfico com o ajuste e a previsão juntos, de forma a se ter uma visão analítica melhor via gráfico de ajuste.

5. RESULTADOS E DISCUSSÕES

5.1. Análise gráfica dos dados

Inicialmente, foi feita a análise descritiva dos dados de temperatura do ar e de internações. Para cada cidade estudada – São Paulo, Santos, Bauru, Presidente Prudente e São Carlos –, foram considerados 3 tipos de análises: diária, semanal e mensal. Entretanto, as correlações com análises mensais foram muito baixas, e as análises com séries semanais e diárias foram muito mais significativas.

Foram inicialmente utilizados dados de temperaturas média, máxima e mínima. Entretanto, embora esses dados tenham sido utilizados na modelagem das séries das internações CID-I e CID-J, a análise se focou nas temperaturas médias diária e semanal apenas, pelos seguintes motivos:

- i) Não foi possível obter uma relação de temperatura máxima ou temperatura mínima mensal ou semanal com o total de internações num mês ou numa semana, respectivamente; essa relação seria “ao acaso”, dado que as temperaturas máxima e mínima não apresentam um padrão definido do período em que elas ocorrem. Não é lógico supor que uma temperatura máxima irá afetar o mês todo ou uma semana completa;
- ii) A análise gráfica da temperatura teve como objetivo avaliar os ciclos anuais de temperatura e como esses ciclos se relacionam com as internações CID-I e ID-J;

- iii) As correlações entre internações e os dados mensais de temperatura foram relativamente mais baixas que as correlações entre internações e temperatura nas análises diárias e semanais;
- iv) As séries de temperatura diária e semanal foram utilizadas como variável auxiliar para a modelagem das associações;
- v) As análises diária e semanal têm muito mais dados que a mensal.

Vale destacar que, caso as suposições expostas não fossem confirmadas, certamente outras análises seriam feitas; mas, como elas foram confirmadas, as análises foram limitadas às séries médias diárias e semanais.

De maneira geral, em todas as cidades, notou-se que nos últimos anos houve uma variação na amplitude térmica: as temperaturas médias aumentaram e os máximos e mínimos foram mais acentuados. É interessante destacar que, pela análise visual das séries temporais, nos últimos anos as variações nos padrões das temperaturas pareceram estar mais relacionadas com as variações nos padrões das internações.

5.2 Modelagens ARIMA, ARIMAX, SARIMA e SARIMAX

As séries temporais dos dados de saúde I e J e de temperatura média são referentes ao período de 2008 a 2016. Entretanto, todas as modelagens ARIMA, ARIMAX, SARIMA e SARIMAX, com dados de saúde e de temperatura, são datadas de 2008 a 2015, a fim de utilizar os modelos para previsão de 2016, comprovando a precisão desses modelos para um intervalo além daquele em que a série foi modelada (previsões futuras).

As análises e modelagens ARIMA, ARIMAX, SARIMA e SARIMAX foram divididas em diárias e semanais.

Em geral, as séries diárias são visivelmente periódicas, mas não estacionárias. Observa-se, também, que nos últimos 4 anos há um relativo aumento no número de registros entre o fim de dezembro e o começo de janeiro. Esse efeito pode ser facilmente explicado pela falta de registros em dezembro e janeiro nos primeiros anos, por serem períodos de férias de funcionários. Entretanto, esse efeito foi minimizado nos últimos anos, provavelmente devido a maior automação de serviços e maior controle na disciplina de trabalho. De fato, vale destacar que em 2015 e 2016, em todas as cidades, é notório o aumento dos registros de internações em dezembro e janeiro em relação aos anos anteriores.

Dadas as séries temporais de interesse, fez-se a análise dos gráficos de auto-correlação f.a.c. e f.a.c.p. delas para o período de 2008 até 2015, de forma que se

modelaram todos os ARIMA (p,d,q) possíveis com parâmetros p, d, q. Em geral, os gráficos de f.a.c. e f.a.c.p. mostraram periodicidade 7. Logo, para os modelos ARIMA (p,d,q), devido ao custo computacional e periodicidade 7 verificada, não foram testados modelos com $p, q > 6$.

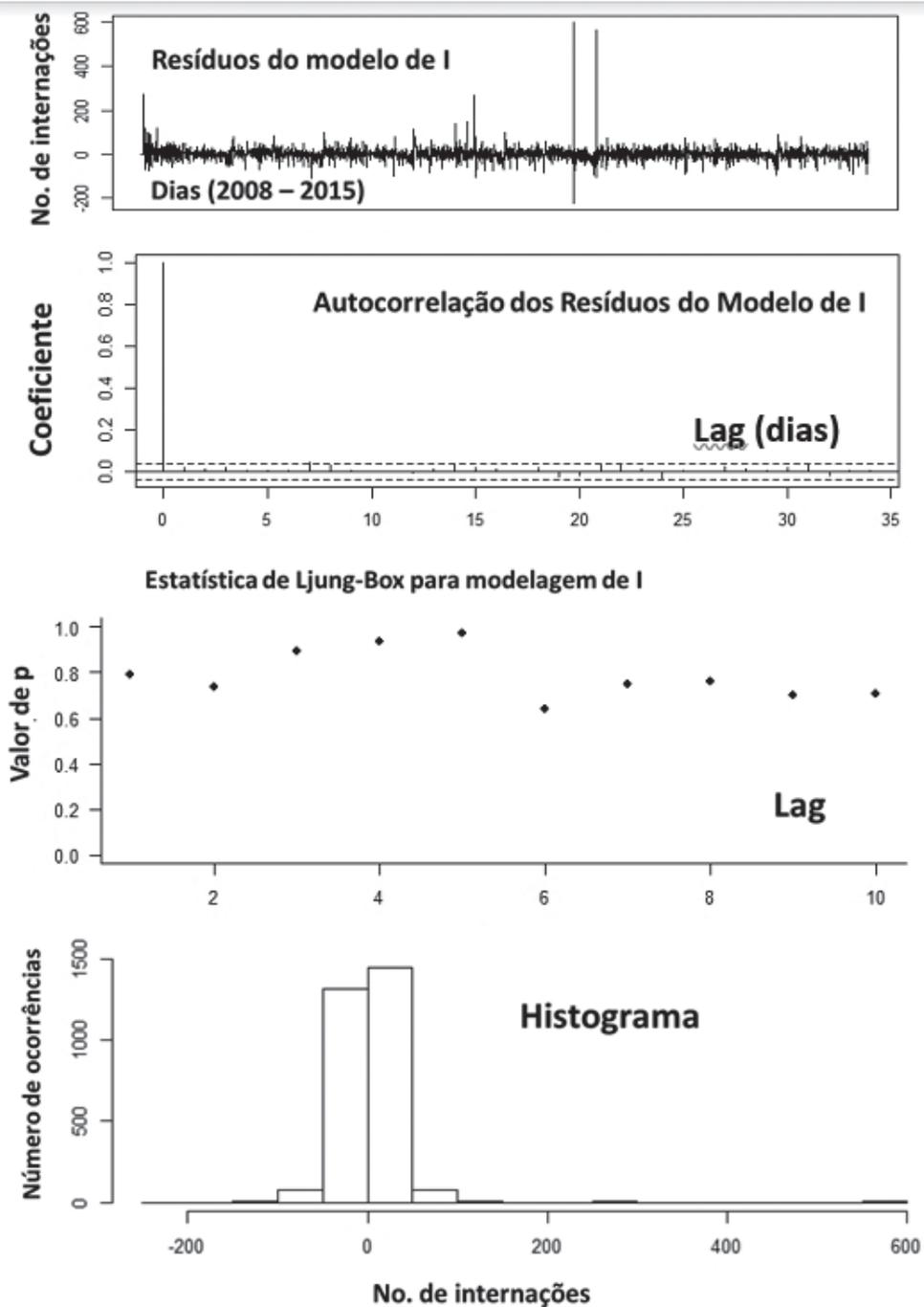
Após essa análise, também se analisaram os gráficos das diferenças das interações CID-I e CID-J e, em todos os casos, o gráfico das primeiras diferenças pareceram ter mostrado séries estacionárias, o que fez com que se usasse $d = 1$ para todos os casos.

Escolhidos os valores para p, d, q, os modelos ARIMA foram simulados de forma que o valor do Critério de Informação de Akaike (AIC) e a estimativa da variância do erro (σ^2) definiram os melhores modelos (menor AIC e menor estimativa de variância).

Após escolher o melhor modelo ARIMA (p,d,q), foram incorporadas as componentes sazonais, de forma que, para tanto, foram simulados os modelos SARIMA (p,d,q) (p1, 1, q1). Após isso, foram analisados novamente os critérios AIC e a estimativa de σ^2 , o que permitiu, por fim, obter os melhores modelos SARIMA (p,d,q) (p1, 1, q1).

Feita a escolha dos modelos ARIMA (p,d,q) e SARIMA (p,d,q) (p_1, d_1, q_1) foram feitas as análises de seus resíduos para identificar resíduos do tipo ruído branco e, assim, aplicar esses modelos. Aqui vale notar que essa análise serve como um filtro, excluindo modelos que possam ter sido selecionados como bons pelos critérios adotados, mas que na verdade continham erros. Na Figura 1, é apresentado um exemplo dessa análise.

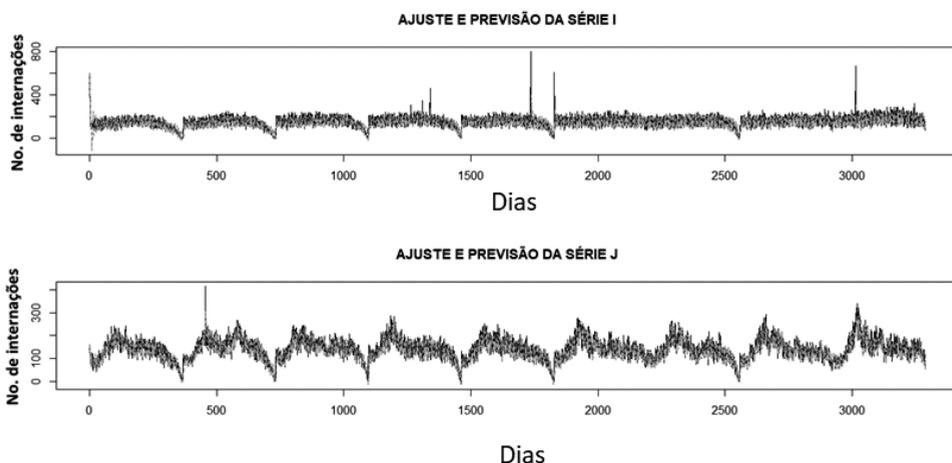
Figura 1 – Análises de Resíduos I (São Paulo – série diária)
modelo SARIMA (6,1,6) (2,1,5) 7.



Fonte: autoria própria.

O gráfico de resíduos mostra se eles são estacionários ou não, pelo formato do gráfico, e suas oscilações em torno de zero. O gráfico f.a.c. dos resíduos mostra se as correlações entre os dados da série são baixas, sendo, assim, um ruído branco. No gráfico da Estimativa de Ljung-Box, caso todos os valores sejam maiores que um certo limiar (que, no exemplo, sequer aparece, por ser muito pequeno), constata-se que os resíduos são independentes. E o histograma confirma a distribuição aleatória dos resíduos. Assim, da análise desses gráficos, pode-se inferir se o resíduo do modelo é ruído branco e, sendo confirmado que o resíduo é ruído branco, então foi obtido um bom modelo.

Figura 2 – Séries temporais diárias de internações de CID I e J para São Paulo. Em verde: Ajuste (2008 a 2015) e Previsão (2016). Em preto: as respectivas séries temporais diárias. Modelo Doença I (acima) – SARIMA (6, 1, 6) (2,1,5) 7 – Modelo Doença J (abaixo) – SARIMA (4, 1, 5) (4,1,2) 7.



Fonte: autoria própria.

Confirmada a hipótese de ruído branco para os resíduos dos modelos, o procedimento a seguir consistiu no ajuste e na previsão utilizando os modelos SARIMA. A Figura 2 apresenta previsões diárias para São Paulo.

Na maioria dos casos diários, observou-se que é possível ver graficamente, com nitidez, as superposições dos dados das séries das doenças I e J (em preto) com as séries de ajuste e previsão (em verde). Mas só a análise gráfica não é suficiente para provar que a superposição das séries indica que o modelo é de boa precisão. Por essa razão, pesquisadores (MORETTIN, 2011; MORETTIN; TOLOI, 2006; MONTGOMERY et al., 2015) defendem a teoria de que, além do teste gráfico, deve-se fazer o mapeamento da função de autocorrelação e do histograma dos

resíduos antes de produzir as previsões. Além disso, há o MAD definido anteriormente, que pode ser utilizado para checar a precisão dos modelos. Com efeito, devido ao fato de as séries diárias de registros de doentes CID-I e CID-J terem termos nulos, escolheu-se o MAD para verificar a precisão dos modelos.

Feita a análise temporal, foram verificadas as correlações com temperatura média, temperatura máxima, temperatura mínima e variação de temperatura diária. Na maioria dos casos, verificou-se que há alta correlação entre os dados de temperatura média e os registros de internações das doenças CID-I e CID-J. Vale notar que, em todos os casos, as doenças respiratórias apresentaram correlação com a variação de temperatura maior. Dessa forma, baseando-se nas correlações cruzadas, implementaram-se novos modelos para cada série temporal de doença CID-I e CID-J: modelos SARIMAX. Com base nesses modelos, foram utilizados os modelos SARIMA já selecionados para escolher os melhores modelos SARIMAX; então, novamente confirmada a hipótese de ruído branco, progrediu-se para a etapa de ajuste e previsão.

Após os ajustes e as previsões, foram calculados os parâmetros MAD, sendo então possível comparar os modelos com a temperatura como variável auxiliar e os modelos apenas com componentes temporais. Nas análises diárias, em alguns casos, houve uma melhora de precisão nos modelos com temperatura média como variável auxiliar; no entanto, até mesmo para as doenças respiratórias, não se pode afirmar que em todos os casos a temperatura melhorou a precisão do modelo.

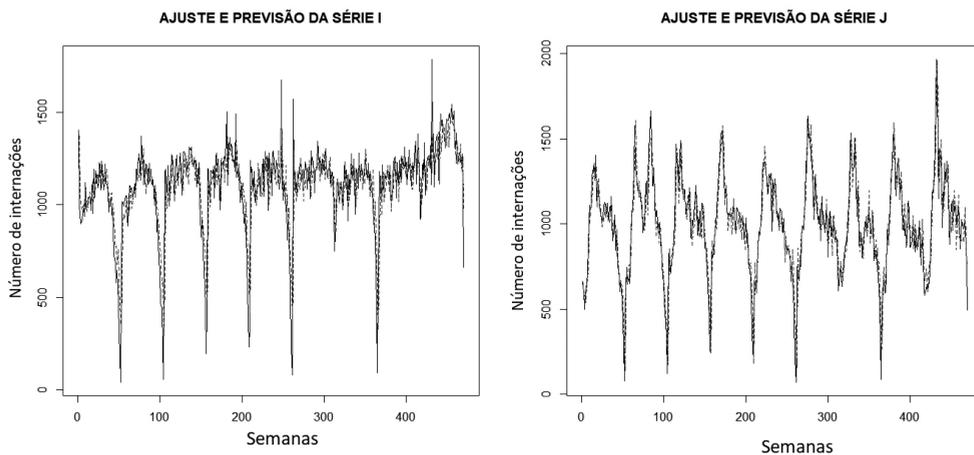
Após a análise diária, foram feitas as análises com as séries semanais, pelo mesmo procedimento feito nas análises diárias, porém com menos dados (devido à amostragem semanal no lugar da diária), o que possibilitou o teste de modelos ARIMA com $p, q > 6$. A seguir, foram feitas análises semanais análogas às análises diárias. Também vale destacar que a precisão melhorou para as análises semanais em relação às diárias; de fato, a análise semanal pareceu mais coerente.

Para as análises semanais, não se notou sazonalidade evidente, de forma que não foram utilizados modelos SARIMA e SARIMAX nessas análises, apenas ARIMA e ARIMAX – aqui vale destacar que, mesmo não colocando componentes sazonais, os modelos semanais foram mais precisos que os diários. De fato, a análise semanal pareceu mais coerente e, na maioria das análises semanais, a implementação de modelos ARIMAX com a temperatura como variável auxiliar melhorou a precisão dos modelos.

Na Figura 3 se encontram exemplos de modelos ARIMAX, para as doenças CID-I e CID-J em séries semanais, para a cidade de São Paulo; esses modelos

foram os que apresentaram a maior precisão entre todos os modelos, provavelmente devido à melhor qualidade de dados.

Figura 3 – Séries temporais semanais de internações de CID I e J para São Paulo. Em vermelho: Ajuste (2008 a 2015) e Previsão (2016). Em preto: as respectivas séries temporais semanais. X = temperatura média. Modelo Doença I (à esquerda) – ARIMAX (11, 1, 9) – Modelo Doença J (à direita) – ARIMAX (11, 1, 10).



Fonte: autoria própria.

Finalizando o presente trabalho, será apresentado um estudo comparativo da qualidade dos modelos nas cidades analisadas, considerando apenas as séries semanais, em forma de tabelas dos valores de MAD e do erro relativo: a Tabela 1 apresenta os resultados dos modelos ARIMA para CID-I; e as Tabelas 2 e 3 têm os resultados dos modelos ARIMA e ARIMAX para CID-I e CID-J. Essas tabelas demonstram, claramente, os melhores resultados para São Paulo, certamente em função da maior qualidade dos seus dados e, possivelmente, pelo fato de apresentar valores mais elevados nas séries temporais, em relação aos valores das séries de outras cidades. Ademais, as tabelas demonstram a melhora da precisão dos modelos com variável auxiliar de temperatura.

Tabela 1 – MAD e erro relativo – CID-I para séries semanais de todas as cidades.

MODELOS	CIDADE	MAD	ERRO RELATIVO
ARIMA (5,1,11)	Presidente Prudente	4,8910	0,1546
ARIMA (10, 1, 10)	Santos	4,0857	0,1014
ARIMA (6,1,6)	São Paulo	101,53	0,0914
ARIMA (11,1,9)	São Carlos	3,5437	0,1123
ARIMA (7,1,8)	Bauru	4,5590	0,1140

Fonte: autoria própria.

Tabela 2 – MAD e erro relativo – CID-I para séries semanais de todas as cidades.

CIDADES	MODELOS	VARIÁVEL AUXILIAR (X)	MAD	ERRO RELATIVO
SÃO PAULO	ARIMA (11,1,9)	-	101,53	0,0914
	ARIMAX (11,1,9)	temperatura média	101,31	0,0912
	ARIMAX (11,1,9)	temperatura média, temperatura máxima, temperatura mínima	100,09	0,0901
PRESIDENTE PRUDENTE	ARIMA (5,1,11)	-	4,8910	0,1546
	ARIMAX (5,1,11)	temperatura média	4,8689	0,1539
	ARIMAX (5,1,11)	temperatura média, temperatura máxima, temperatura mínima	4,8721	0,1540
BAURU	ARIMA (7,1,8)	-	4,5590	0,1140
	ARIMAX (7,1,8)	temperatura média	4,3870	0,1097
	ARIMAX (7,1,8)	temperatura média, temperatura máxima, temperatura mínima	4,6509	0,1163
SÃO CARLOS	ARIMA (11,1,9)	-	3,5437	0,1123
	ARIMAX (11,1,9)	temperatura média	3,5184	0,1115
	ARIMAX (11,1,9)	temperatura média, temperatura máxima, temperatura mínima	3,4680	0,1099
SANTOS	ARIMA (10, 1, 10)	-	4,0857	0,1014
	ARIMAX (10, 1, 10)	temperatura média	4,0776	0,1012
	ARIMAX (10, 1, 10)	temperatura média, temperatura máxima, temperatura mínima	4,0413	0,1003

Fonte: autoria própria.

Tabela 3 – MAD e erro relativo – CID-J para séries semanais de todas as cidades.

CIDADES	MODELOS	VARIÁVEL AUXILIAR (X)	MAD	ERRO RELATIVO
SÃO PAULO	ARIMA (11,1,10)	-	83,41	0,0834
	ARIMAX (11,1,10)	temperatura média	77,41	0,0774
	ARIMAX (11,1,10)	temperatura média, temperatura máxima, temperatura mínima	77,11	0,0771
PRESIDENTE PRUDENTE	ARIMA (5,1,11)	-	3,7455	0,1557
	ARIMAX (5,1,11)	temperatura média	3,7166	0,1545
	ARIMAX (5,1,11)	temperatura média, temperatura máxima, temperatura mínima	3,6661	0,1524
BAURU	ARIMA (5,1,11)	-	4,8910	0,1546
	ARIMAX (5,1,11)	temperatura média	4,8689	0,1539
	ARIMAX(5,1,11)	temperatura média, temperatura máxima, temperatura mínima	4,8721	0,1540
SÃO CARLOS	ARIMA (11,1,9)	-	3,2266	0,1332
	ARIMAX (11,1,10)	temperatura média	3,1518	0,1302
	ARIMAX (11,1,10)	temperatura média, temperatura máxima, temperatura mínima	3,1615	0,1306
SANTOS	ARIMA (11, 1, 11)	-	3,6831	0,1034
	ARIMAX (11, 1, 11)	temperatura média	3,4693	0,0974
	ARIMAX (11, 1, 11)	temperatura média, temperatura máxima, temperatura mínima	3,4587	0,0971

Fonte: autoria própria.

6. CONCLUSÕES

Conclui-se que há associações estatísticas entre número de internações e temperatura do ar, uma vez que há grande correlação entre internações de doenças CID I e J e dados de temperatura do ar nas análises diárias e semanais (em que as correlações são maiores).

Nota-se que há uma melhora na precisão dos modelos quando se incluem dados de temperatura como variável auxiliar. No entanto, há casos em que não há melhora na precisão com a inclusão de dados de temperatura como variável

auxiliar; porém, há explicações cabíveis, que não excluem a possibilidade de haver associação entre internações e temperatura, para todos os casos.

As autocorrelações, apresentadas nos gráficos de f.a.c. e f.a.c.p., mostram que o número de internações no mês atual sofre influência do próprio mês no ano anterior (ciclo anual, relacionado às estações do ano); e mostram também que há um comportamento periódico semanal (parece haver um padrão no comportamento das pessoas, que preferem ser internadas ou buscam socorro médico em certos dias da semana).

Além da associação estatística, a análise das correlações cruzadas mostra que há a possibilidade de não só haver correlação entre internações e a temperatura do ar, mas também haver alguma conexão causal indireta, por exemplo: mudanças nos padrões de temperaturas afetam o número de internações CID-I e CID-J. Desse modo, pode-se afirmar que variações no comportamento da temperatura, nas cidades estudadas, podem significar uma variação no padrão das internações CID-I e CID-J (nesse caso, vale enfatizar a variação da temperatura, ou seja, não se fala em temperaturas médias altas ou baixas, e sim em quanto a temperatura variou a cada dia).

Dado que se observa uma maior correlação entre internações CID-J e temperatura do ar que entre internações CID-I e temperatura ar, em todas as análises feitas de todas as cidades, tem-se que as internações por doenças respiratórias são, possivelmente, mais afetadas por variações nos padrões de comportamento da temperatura do ar que as internações por doenças cardíacas.

Considerando que as correlações cruzadas entre dados com defasagens menores foram maiores, pode-se dizer que, na associação estatística encontrada, havendo relação entre número de internações e temperatura do ar, então essa relação deve ser uma relação de memória curta, ou seja, uma relação na qual padrões antigos não afetam tanto quanto padrões mais recentes.

Um estudo comparativo da qualidade dos modelos nas cidades analisadas, considerando apenas as séries semanais, demonstra, claramente, os melhores resultados para São Paulo, certamente em função da maior qualidade dos seus dados e, possivelmente, pelo fato de apresentar valores mais elevados nas séries temporais em relação aos valores das séries de outras cidades. Outro resultado comparativo de interesse é que, excetuando a cidade de São Paulo, a modelagem da cidade litorânea (Santos) teve o menor erro relativo, enquanto a cidade mais interior (Presidente Prudente) teve o maior erro relativo; entretanto, ainda não é possível interpretar esse resultado comparativo entre o litoral e o interior, de modo que a inclusão de análises de outras cidades poderá levar a uma interpretação definitiva.

Por fim, a inclusão de dados de temperatura como variáveis auxiliares parece aumentar a precisão dos modelos nas melhores análises diárias e nas análises semanais, embora não de forma muito drástica.

REFERÊNCIAS BIBLIOGRÁFICAS

ALEIXO, N. C. R.; NETO, J. L. S. Condicionantes climáticos e interações por pneumonia: estudo de caso em Ribeirão Preto/SP. **Revista do Departamento de Geografia – USP**, v. 27, pp. 1-20, 2014. Disponível em: <http://www.revistas.usp.br/rdg/article/view/85430>. Acesso em: mar. 2015.

AKAIKE, H. Information theory and the maximum likelihood principle. In: B. N. PETROV; F. CSÄKI (eds.). **2nd International Symposium on Information Theory**. Akademiai Kiadó, Budapest, 1973.

BARCELLOS, C. et al. Mudanças climáticas e ambientais e as doenças infecciosas: cenários e incertezas para o Brasil. **Epidemiol. Serv. Saúde**, Brasília, 18(3): pp. 285-304, jul.-set. 2009. Disponível em: <http://scielo.iec.gov.br/pdf/ess/v18n3/v18n3a11.pdf>. Acesso em: jun. 2016.

BAKONYI, S. M. C.; DNNI-OLIVEIRA, I. M.; MARTINS, L. C.; BRAGA, A. L. F. Poluição atmosférica e doenças respiratórias em crianças na cidade de Curitiba, PR. **Rev. Saúde Pública**, 38(5): pp. 695-700, 2004. Disponível em: http://www.scielo.br/scielo.php?script=sci_abstract&pid=S0034-89102004000500012&lng=en&nrm=iso&tlng=pt. Acesso em: jun. 2016.

CONCEIÇÃO, R. S. DA et al. A temperatura do ar e sua relação com algumas doenças respiratórias em Vitória da Conquista BA. **Revista Eletrônica Geoaraguaia Barra do Garças MT**, v. 5, n. 2, pp. 69-81, jul.-dez. 2015. Disponível em: <http://periodicoscientificos.ufmt.br/ojs/index.php/geo/article/view/4929>. Acesso em: dez. 2016.

DATASUS – BRASIL – MINISTÉRIO DA SAÚDE. DATASUS Trajetória 1991-2002. **Ministério da saúde – Departamento de Informática do SUS** – Brasília – DF, 2002. Disponível em: http://bvsms.saude.gov.br/bvs/publicacoes/trajetoria_datusus.pdf. Acesso em: dez. 2017.

CID 10 – **Classificação Estatística Internacional de Doenças e Problemas Relacionados à Saúde**. São Paulo, 1998. Disponível em: https://www.cremesp.org.br/pdfs/cid10_ultimaversaodisponivel_2012.pdf. Acesso em: dez. 2017.

CRYER, J. D.; CHAN, K. S. **Time Series Analysis: With Applications in R.** (Springer Texts in Statistics). 2nd. ed., Kindle Edition. Springer Texts in Statistics, ISBN-10: 0387759581.

GOUVEIA, N.; FREITAS, C. U.; MARTINS, L. C.; MARCILIO, I. O. Hospitalizações por causas respiratórias e cardiovasculares associadas à contaminação atmosférica no Município de São Paulo, Brasil. **Caderno de Saúde Pública**. v. 22, n. 12, Rio de Janeiro, 2006. Disponível em: <http://www.scielo.br/pdf/csp/v22n12/15.pdf>. Acesso em: dez. 2015.

HARARI, J.; BRITO, C. R. O Núcleo Brasileiro de Pesquisas Ambientais e Saúde (NBPAS) de Hoje. **Safety, Health and Environmental World Congress SHEWC**, 2005. Disponível em: <http://copec.eu/congresses/cbpas2005/proc/CBPAS.pdf>. Acesso em: jul. 2015.

HARARI, J.; SIEGLE, E. Analysis of meteorological and oceanographic conditions off the coastal region of Bahia State Brazil. **Safety, Health and Environmental World Congress SHEWC**, São Paulo – SP – Brasil. v. 1. pp. 36-40, 2012. Disponível em: <http://copec.eu/congresses/shewc2012/proc/works/009.pdf>. Acesso em: abr. 2015.

INMET – INSTITUTO NACIONAL DE METEOROLOGIA. Disponível em: http://www.inmet.gov.br/portal/index.php?r=home/page&page=sobre_inmet. Acesso em: dez. 2017.

IPCC. **Painel Intergovernamental sobre Mudanças Climáticas 2018**. Sumário para formuladores de política. Versão em português publicada pelo MCTIC – Brasil, 2019.

JACOBI, P. R. Mudanças climáticas e ensino superior: a combinação entre pesquisa e educação. **Educar em Revista**, Curitiba, Brasil, Edição Especial n. 3, pp. 57-72. Editora UFPR, 2014. Disponível em: <https://revistas.ufpr.br/educar/article/view/38107>. Acesso em: nov. 2016.

JACOBI, P. R. et al. **Temas atuais em mudanças climáticas: para os ensinos fundamental e**

médio. São Paulo: IEE USP, Disponível em: http://www.incline.iag.usp.br/data/arquivos_download/TEMAS_ATUAIS_EM_MUDANCAS_CLIMATICAS_on_line.pdf. Acesso em: out. 2016.

MARCÍLIO, I.; HAJAT, S.; GOUVEIA, N. Forecasting Daily Emergency Department Visits Using Calendar Variables and Ambient Temperature Readings. **Academic Emergency Medicine**, v. 20, pp. 769-777, 2013. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/24033619>. Acesso em: jul. 2015.

MASTER – LABORATÓRIO DO DEPARTAMENTO DE CIÊNCIAS ATMOSFÉRICAS. Disponível em: <http://www.master.iag.usp.br/index.php?pi=N>. Acesso em: dez. 2017.

MORAES, R. M.; GOMES, R. G. S. Alterações climáticas e suas influências sobre as doenças transmitidas por vetores. **XIV Safety, Health and Environment World Congress (SHEWC'2009)**, 2009. Disponível em: http://www.de.ufpb.br/~ronei/Gomes2009AlteracoesClimaticas_Influencias.pdf. Acesso em: maio 2016.

MORETTIN, P. A. **Econometria financeira – Um curso em séries temporais financeiras**. 2. ed. São Paulo: Blucher, 2011.

MORETTIN, P. A.; TOLOI, C. M. C. **Análise das séries temporais**. São Paulo: Blucher, 2006.

MONTGOMERY, D. C.; JENNINGS, C. L.; KULAHCI, M. **Introduction to Time Series Analysis and Forecasting**. 2 ed. New Jersey e Canadá: Wiley, 2015.

MINISTÉRIO DA SAÚDE. Política Nacional de Informação em Saúde – Proposta Versão 2.0. **Secretaria Executiva - Departamento de Informação e Informática do SUS**, Brasília – DF, 2004. Disponível em: http://bvsmis.saude.gov.br/bvs/publicacoes/PoliticaInformacaoSaude29_03_2004.pdf. Acesso em: dez. 2017.

NCEP/NCAR – ESRL – **PHYSICAL SCIENCES DIVISION**. Disponível em: <https://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis.html>. Acesso em: dez. 2017.

PARAISO, M. L. S.; GOUVEIA, N. Riscos à saúde devido à queima prévia da palha de cana-de-açúcar no Estado de São Paulo, Brasil. *Revista Brasileira de Epidemiologia*, São Paulo, v. 18, n. 3, pp. 691-701, 2015. Disponível em: http://www.scielo.br/pdf/rbepid/v18n3/pt_1415-790X-rbepid-18-03-00691.pdf. Acesso em: out. 2017.

SILVA, E. N.; RIBEIRO, H.; SANTANA, P. Clima e saúde em contextos urbanos: uma revisão da literatura. **Revista bibliográfica de geografia y ciencias sociales**, Universidad de Barcelona, v. XIX, n. 1092, 2014. Disponível em: <http://www.ub.edu/geocrit/b3w-1092.htm>. Acesso em: dez. 2015.

SON J. Y.; GOUVEIA, N.; BRAVO M. A.; FREITAS C. U.; BELL M. L. The impact of temperature on mortality in a subtropical city: effects of cold, heat, and heat waves in São Paulo, Brazil. **International Journal of Biometeorology**, v. 1,

pp. 1-9, 2015. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/25972308>. Acesso em: dez. 2016.

SOUZA, I. C. A.; MORAES, R. M. Análise Temporal de Dados de Saúde Pública do Estado da Paraíba. Período 1998 a 2001. In: III Congresso Brasileiro de Pesquisas Ambientais e Saúde. 2003, Santos. **Anais do Congresso Brasileiro de Pesquisas Ambientais e Saúde (CBPAS'2003)**. 2003. pp. 176-180.