

**COMPARAÇÃO ESTATÍSTICA DE DUAS SÉRIES DE  
MATERIAL PARTICULADO (MP<sub>10</sub>) NA CIDADE DE  
SÃO PAULO**

**FRANCIELLA MARQUES DA COSTA**

**2010**

**FRANCIELLA MARQUES DA COSTA**

**COMPARAÇÃO ESTATÍSTICA DE DUAS SÉRIES DE MATERIAL  
PARTICULADO (MP<sub>10</sub>) NA CIDADE DE SÃO PAULO**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de "Mestre".

Orientadora  
Profa. Dra. Thelma Sáfadi

LAVRAS  
MINAS GERAIS -BRASIL  
2010

**Ficha Catalográfica Preparada pela Divisão de Processos Técnicos da  
Biblioteca Central da UFLA**

Costa, Franciella Marques da.

Comparação estatística de duas séries de material particulado (MP<sub>10</sub>) na cidade de São Paulo / Franciella Marques da Costa. – Lavras : UFLA, 2010.

57 p. : il.

Dissertação (mestrado) – Universidade Federal de Lavras, 2010.

Orientador: Thelma Sáfyadi.

Bibliografia.

1. Séries temporais. 2. Testes de comparação. 3. Séries de valores máximos de poluição por MP<sub>10</sub>. I. Universidade Federal de Lavras. II. Título.

CDD - 519.55

**FRANCIELLA MARQUES DA COSTA**

**COMPARAÇÃO ESTATÍSTICA DE DUAS SÉRIES DE MATERIAL  
PARTICULADO (MP<sub>10</sub>) NA CIDADE DE SÃO PAULO**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de "Mestre".

APROVADA em 19 de fevereiro de 2010

Prof. Dr. Denismar Alves Nogueira

UNIFAL-MG

Prof. Dr. João Domingos Scalon

UFLA

Prof. Dr. Marcelo Angelo Cirillo

UFLA

Profa. Dra. Thelma Sáfadi

UFLA

(Orientadora)

LAVRAS  
MINAS GERAIS-BRASIL

*A minha querida família:  
meus pais, Cleuza e Orlando;  
meus irmãos, Kelle e Orlando Júnior;  
meus avós, Irene e Joaquim;  
(in memoriam), aos avós Maria Antônia e José  
e meu esposo, Ronivon.*

**DEDICO**

## AGRADECIMENTOS

A Deus, por iluminar e guiar a minha vida, dando-me forças para realizar meus objetivos.

Aos meus pais, imprescindíveis em minha vida e responsáveis por tudo que sou, agradeço pelo apoio, carinho e amor.

Aos meus irmãos, Kelle e Orlando Júnior, pelo carinho, incentivo e amizade.

Aos avós, Irene e Joaquim e, (*in memoriam*), aos avós Maria Antônia e José, pela atenção e carinho a mim dedicados.

Ao meu esposo, Ronivon, com quem compartilho todos os momentos de minha vida, por seu carinho, amor, compreensão e apoio incondicional para a realização deste sonho.

À Universidade Federal de Lavras e ao Departamento de Ciências Exatas, pela oportunidade de cursar o Mestrado.

Ao CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), pela concessão da bolsa de estudos.

À professora Dra. Thelma Sáfy, pela orientação.

Aos professores e demais funcionários do Departamento de Ciências Exatas.

Aos membros da banca examinadora, pelas sugestões e contribuições.

A todos os meus amigos, pelo apoio e carinho. Em especial, aos amigos: Lindson, Gabriella, Flávia, Edmilza, Patrícia, Nádia, Imaculada, Rosiana, Simone e Deyse.

Às amigas e companheiras de estudo do "Clube da Luluzinha": Flávia, Simone, Rosiana e Deyse, e também, aos demais colegas da Pós-Graduação em Estatística.

Ao Fábio, pela ajuda com a rotina no software R.

A todos aqueles, que de alguma forma , acreditaram e contribuíram para a realização deste sonho, meu sincero agradecimento.

## SUMÁRIO

	<b>Página</b>
LISTA DE FIGURAS . . . . .	i
RESUMO . . . . .	iv
ABSTRACT . . . . .	v
1 INTRODUÇÃO . . . . .	1
2 REFERENCIAL TEÓRICO . . . . .	3
2.1 Poluição do ar . . . . .	3
2.2 Séries Temporais . . . . .	5
2.2.1 Processos estocásticos . . . . .	6
2.2.1.1 Processos estocásticos estacionários . . . . .	8
2.2.2 Função de autocovariância e autocorrelação . . . . .	9
2.2.3 Decomposição clássica de uma série temporal . . . . .	11
2.2.4 Estacionariedade . . . . .	12
2.2.4.1 Teste do sinal (Cox-Stuart) . . . . .	13
2.2.5 Análise espectral . . . . .	14
2.2.6 Teste de Fisher . . . . .	16
2.2.7 Modelos autorregressivos de ordem $p$ (AR( $p$ )) . . . . .	16
2.2.7.1 Função de autocorrelação . . . . .	17
2.2.7.2 Função de autocorrelação parcial . . . . .	19
2.3 Teste de Box-Pierce . . . . .	21
2.4 Teste de Kolmogorov-Smirnov . . . . .	21
2.5 Testes para comparação de séries temporais . . . . .	22
2.5.1 Teste das Somas Acumuladas (Coates & Diggle, 1986) . . . . .	23
2.5.2 Teste de Igualdade das Funções de Autocorrelação (Quenouille, 1958) . . . . .	25
2.5.3 Método para comparação de séries temporais . . . . .	26

3 MATERIAL E MÉTODOS . . . . .	28
4 RESULTADOS E DISCUSSÃO . . . . .	30
4.1 Análise da série diária de valores máximos de poluição, por partículas inaláveis (MP <sub>10</sub> ), na Estação Centro, na cidade de São Paulo . . . . .	30
4.2 Análise da série diária de valores máximos de poluição, por partículas inaláveis (MP <sub>10</sub> ), na Estação Santo Amaro, na cidade de São Paulo . . . . .	35
4.3 Teste de Coates & Diggle (1986) aplicado às séries estacionárias da Estação Santo Amaro e da Estação Centro, ambas na cidade de São Paulo . . . . .	39
4.4 Teste de Quenouille (1958) aplicados às séries estacionárias da Estação Santo Amaro e da Estação Centro, ambas na cidade de São Paulo . . . . .	41
4.5 Comparação das séries estacionárias da Estação Santo Amaro e da Estação Centro, ambas na cidade de São Paulo . . . . .	46
5 CONCLUSÕES . . . . .	49
REFERÊNCIAS BIBLIOGRÁFICAS . . . . .	50
ANEXOS . . . . .	51

## LISTA DE FIGURAS

	<b>Página</b>
1 Um processo estocástico interpretado como uma família de variáveis aleatórias. . . . .	7
2 Um processo estocástico interpretado como uma família de trajetórias. . . . .	8
3 Série diária de valores máximos de poluição por $MP_{10}$ na Estação Centro, na cidade de São Paulo. . . . .	30
4 Série diária de valores máximos de poluição por $MP_{10}$ na Estação Centro, na cidade de São Paulo: amplitude vs média. . . . .	31
5 Série diária do $\ln$ dos valores máximos de poluição por ( $MP_{10}$ ) na Estação Centro, na cidade de São Paulo: amplitude vs média. . . . .	32
6 Série diária do $\ln$ dos valores máximos de poluição por $MP_{10}$ na Estação Centro, na cidade de São Paulo. . . . .	32
7 Periodograma da série diária do $\ln$ dos valores máximos de poluição por $MP_{10}$ na Estação Centro, na cidade de São Paulo. . . . .	33
8 Série estacionária de poluição por $MP_{10}$ na Estação Centro, na cidade de São Paulo. . . . .	34
9 Função de autocorrelação da série estacionária da Estação Centro, na cidade de São Paulo. . . . .	34
10 Série diária de valores máximos de poluição por $MP_{10}$ na Estação Santo Amaro, na cidade de São Paulo. . . . .	35
11 Série diária de valores máximos de poluição por $MP_{10}$ na Estação Santo Amaro, na cidade de São Paulo: amplitude vs média. . . . .	36
12 Série diária do $\ln$ dos valores máximos de poluição por $MP_{10}$ na Estação Santo Amaro, na cidade de São Paulo: amplitude vs média. . . . .	37
13 Série diária do $\ln$ dos valores máximos de poluição por $MP_{10}$ na Estação Santo Amaro, na cidade de São Paulo. . . . .	37

14	Periodograma da série diária do $\ln$ dos valores máximos de poluição por $MP_{10}$ na Estação Santo Amaro, na cidade de São Paulo. . . . .	38
15	Série estacionária de poluição por $MP_{10}$ na Estação Santo Amaro, na cidade de São Paulo. . . . .	39
16	Função de autocorrelação da série estacionária da Estação Santo Amaro, na cidade de São Paulo. . . . .	39
17	Periodograma da série estacionária de poluição por $MP_{10}$ na Estação Centro, na cidade de São Paulo. . . . .	40
18	Periodograma da série estacionária de poluição por $MP_{10}$ na Estação Santo Amaro, na cidade de São Paulo. . . . .	40
19	Função de autocorrelação da série estacionária da Estação Santo Amaro, na cidade de São Paulo. . . . .	42
20	Função de autocorrelação da série estacionária da Estação Centro, na cidade de São Paulo. . . . .	42
21	Função de autocorrelação comum estimada. . . . .	43
22	Função de autocorrelação parcial comum estimada $\hat{\Phi}(k)$ . . . . .	43
23	Função de autocorrelação da série residual da Estação Centro, na cidade de São Paulo. . . . .	44
24	Função de autocorrelação da série residual da Estação Santo Amaro, na cidade de São Paulo. . . . .	44
25	Função de autocorrelação parcial da série residual da Estação Centro, na cidade de São Paulo. . . . .	45
26	Função de autocorrelação parcial da série residual da Estação Santo Amaro, na cidade de São Paulo. . . . .	45
27	Gráfico da diferença entre as séries estacionárias da Estação Santo Amaro e da Estação Centro. . . . .	46

28	Função de autocorrelação da diferença entre as séries estacionárias da Estação Santo Amaro e da Estação Centro. . . . .	47
29	Função de autocorrelação parcial da diferença entre as séries estacionárias da Estação Santo Amaro e da Estação Centro. . . . .	47
30	Periodograma da diferença entre as séries estacionárias da Estação Santo Amaro e da Estação Centro. . . . .	48

## RESUMO

COSTA, Franciella Marques da. **Comparação estatística de duas séries de material particulado (MP<sub>10</sub>) na cidade de São Paulo**. 2010. 57 p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) - Universidade Federal de Lavras, Lavras, MG. \*

Os grandes centros urbanos vêm enfrentando problemas relacionados à poluição do ar. As partículas inaláveis (MP<sub>10</sub>) são um tipo de poluente atmosférico que afeta a maioria desses centros urbanos, entre os quais, a cidade de São Paulo. O MP<sub>10</sub> é considerado um indicador da qualidade do ar, o que ressalta a importância de estudá-lo. Assim, o objetivo deste trabalho consiste em verificar se as séries de valores máximos de poluição diários, por partículas inaláveis (MP<sub>10</sub>) observadas na estação Centro e na estação Santo Amaro, ambas localizadas em São Paulo, para o período de 01/01/2007 a 19/12/2009, estão sendo geradas pelo mesmo processo estocástico. Para a análise, foram utilizados o teste das somas acumuladas, o teste de igualdade das funções de autocorrelação e um procedimento de diferença de séries, concluindo que as duas séries analisadas não foram geradas pelo mesmo processo estocástico. Desta maneira, os dados de poluição obtidos em uma das estações em estudo não são suficientes para explicar a poluição na outra estação.

**Palavras-chave:** séries temporais, testes de comparação, séries de valores máximos de poluição, MP<sub>10</sub>

---

\*Orientadora: Profa. Dra. Thelma Sáfydi - UFLA

## ABSTRACT

COSTA, Franciella Marques da. **Statistical comparison of two series of particulate matter (PM<sub>10</sub>) in the city of São Paulo.** 2010. 57 p. Dissertation (Master in Statistics and Agricultural Experimentation) - Federal University of Lavras, Lavras, MG.\*

Large urban centres have had to handle problems related to the air pollution. The inhaled particles (PM<sub>10</sub>) are an atmospheric pollutant that affects most of such urban centres, among which, the city of São Paulo. The (PM<sub>10</sub>) is regarded as an indicator of the air quality, a characteristic that highlights the relevance in studying it. Thus, the aim of this study was to investigate whether a time series of maximum daily pollution values of inhaled particles (PM<sub>10</sub>) observed at the Centro and Santo Amaro stations – both located in the city of São Paulo - have been generated from the same stochastic process. The data were collected for the period from January, 1<sup>st</sup>, 2007 to December, 19<sup>th</sup>, 2009. The following statistical methods were applied: the test for the equality of autocorrelation functions, the cumulative sum test, and a procedure of time series subtraction. The conclusion drawn by this study was that the two data series analysed were not generated by the same stochastic process. Hence, the data on air pollution levels obtained in one of the stations studied are not enough to explain the same type of data obtained in the other one.

**Key words:** time series, comparison tests, series of maximum pollution values, PM<sub>10</sub>

---

\* Adviser: Profa. Dra. Thelma Sáfaci - UFLA

## 1 INTRODUÇÃO

A poluição atmosférica tem sido uma preocupação mundial. O crescente desenvolvimento urbano e industrial tem aumentado significativamente a emissão de poluentes, sendo os veículos automotores e as indústrias um dos principais responsáveis.

O estado de São Paulo é um grande centro urbano que vem sofrendo as consequências da poluição do ar. A Companhia Ambiental do Estado de São Paulo (CETESB) é um órgão responsável pelo monitoramento da poluição do ar no estado de São Paulo. A CETESB possui várias estações de monitoramento localizadas na região metropolitana da cidade de São Paulo, no interior e no litoral.

O  $MP_{10}$  é um dos poluentes atmosféricos que ocorrem com grande frequência, causando efeitos adversos ao meio ambiente, tais como danos à vegetação, contaminação do solo e deterioração da visibilidade. Além disso, ele pode causar prejuízo à saúde da população.

Utilizando métodos estatísticos e a partir de dados de poluição obtidos em uma determinada estação pode-se verificar se estes são suficientes para explicar o comportamento da poluição em outra estação viabilizando custo menores e reduzindo o tempo gasto na coleta e análise de dados. Para tais análises, podem ser utilizados testes de comparação de séries temporais, tais como o teste das Somas Acumuladas (Coates & Diggle, 1986), o teste de Igualdade das Funções de Autocorrelação (Quenouille, 1958) e um procedimento aplicado por Silva et al. (2000).

Dessa forma, este trabalho teve por objetivo comparar as séries de valores máximos de poluição diária observadas na Estação Centro e na Estação Santo Amaro, ambas localizadas em São Paulo. Esta comparação entre as duas séries

temporais foi realizada com o intuito de verificar se os dados de poluição por  $MP_{10}$  obtidos em uma das estações são suficientes para explicar a poluição por  $MP_{10}$  na outra estação.

Este trabalho está organizado em cinco seções, da seguinte forma:

- Primeira seção - Introdução;
- Segunda seção - Referencial Teórico, no qual são abordados alguns conceitos sobre poluição, séries temporais e os testes para comparação entre duas séries temporais estacionárias;
- Terceira seção - Material e métodos, que descreve as séries utilizadas na análise e os passos seguidos para realizar a comparação entre as séries em estudo;
- Quarta seção - Resultados e discussão, os testes para verificar se as séries em estudo foram geradas pelo mesmo processo estocástico são aplicados;
- Quinta seção - Conclusões, são apresentadas as conclusões referentes à comparação realizada entre as séries.

## 2 REFERENCIAL TEÓRICO

### 2.1 Poluição do ar

A poluição do ar é um problema grave enfrentado principalmente pelos grandes centros urbanos. É notável a preocupação da população quanto aos possíveis danos causados à saúde, em consequência da exposição à poluição do ar.

Segundo Gouveia et al. (2003), os níveis de poluição aos quais a população de São Paulo é submetida são suficientes para provocar agravos respiratórios e cardiovasculares em crianças e idosos.

Sáfadi & Peña (2007) , considerando o modelo fatorial dinâmico encontraram uma forte associação entre os poluentes atmosféricos (MP<sub>10</sub>), umidade e mortalidade por doenças respiratórias para a cidade de São Paulo, no período de 1994 a 1997.

O nível de poluição do ar é determinado pela quantificação das substâncias poluentes existentes no mesmo (Companhia Ambiental do Estado de São Paulo - CETESB, 2009). Segundo a resolução do Conselho Nacional do Meio Ambiente - CONAMA (1990) :

Entende-se como poluente atmosférico qualquer forma de matéria ou energia com intensidade e em quantidade, concentração, tempo ou características em desacordo com os níveis estabelecidos, e que tornem ou possam tornar o ar:

I - impróprio, nocivo ou ofensivo à saúde;

II - inconveniente ao bem-estar público;

III - danoso aos materiais, à fauna e flora;

IV - prejudicial à segurança, ao uso e gozo da propriedade e às atividades normais da comunidade.

Vale ressaltar que as condições meteorológicas influenciam na qualidade do ar, mesmo mantidas as quantidades de emissão, pois favorecem uma maior ou

menor diluição dos poluentes. Devido a esse fato, a qualidade do ar piora em relação aos parâmetros material particulado, dióxido de enxofre e monóxido de carbono, no período de inverno, quando as condições meteorológicas são desfavoráveis à dispersão e diluição dos poluentes na atmosfera (CETESB, 2009) .

O material particulado é composto por um conjunto de poluentes encontrados em suspensão na atmosfera, tais como fumaça, poeira e todo tipo de partículas de material líquido e sólido. As diversas fontes de emissão deste poluente são: processos industriais, veículos automotores, ressuspensão de poeira do solo, entre outros. A formação do material particulado também pode ocorrer na atmosfera, por meio de gases, tais como dióxido de enxofre, óxidos de nitrogênio e compostos orgânicos voláteis. Emitidos essencialmente em atividades de combustão, esses gases podem sofrer reações químicas no ar e transformar-se em partículas. O potencial das partículas para causar problemas à saúde está diretamente ligado ao seu tamanho. Quanto menor o seu tamanho, maiores os danos provocados (CETESB, 2010).

Conforme CETESB (2010), o material particulado pode ser classificado da seguinte forma:

- Partículas Totais em Suspensão: são aquelas cujo diâmetro aerodinâmico é menor que  $50 \mu m$  (micrômetro, equivale à milésima parte do milímetro). Parte destas partículas são inaláveis e podem causar danos à saúde, e a outra parte pode prejudicar a qualidade de vida da população como, por exemplo interferindo na estética do ambiente.
- Partículas Inaláveis ( $MP_{10}$ ): são aquelas em que o diâmetro aerodinâmico é menor que  $10 \mu m$ .
- Fumaça (FMC): resultante do processo de combustão.

As partículas inaláveis ( $MP_{10}$ ) podem ser classificadas em partículas inaláveis finas ( $MP_{2,5}$ ) que apresentam diâmetro aerodinâmico menor que  $2,5 \mu m$  e partículas inaláveis grossas possuindo diâmetro de  $2,5$  a  $10 \mu m$ .

As partículas inaláveis ( $MP_{10}$ ), causam problemas de visibilidade, além de estarem fortemente relacionadas a problemas de saúde, associadas principalmente, aos sistemas cardiovasculares e respiratórios (CETESB, 2008). Tal poluente atmosférico também provoca contaminação do solo e causa prejuízo a vegetação.

O poluente partículas inaláveis ( $MP_{10}$ ) é um dos poluentes considerados como indicadores da qualidade do ar devido aos efeitos desfavoráveis causados ao meio ambiente e por ocorrerem com muita frequência (CETESB, 2009).

A Companhia Ambiental do Estado de São Paulo (CETESB), ligada à Secretaria do Meio Ambiente, mantém uma rede de monitoramento da qualidade do ar no estado de São Paulo, sendo constituída por 45 estações de monitoramento automático, das quais 4 são móveis e outros 47 locais de amostragem da rede manual.

## 2.2 Séries Temporais

Também denominada série histórica ou cronológica, uma série temporal é um conjunto de observações coletadas sequencialmente ao longo do tempo. Entretanto, além do tempo, uma série pode estar em função de outra variável, por exemplo, volume, profundidade ou espaço. A dependência entre as observações adjacentes restringe a aplicabilidade de metodologias estatísticas clássicas, que dependem do pressuposto de independência entre os dados. Sendo assim, o procedimento mais usual a ser utilizado são as técnicas de séries temporais.

As observações ordenadas no tempo podem ser classificadas em séries discretas como, por exemplo, valores mensais de temperatura na cidade de Lavras -

MG, ou séries contínuas, como o registro de um eletrocardiograma. Para analisar uma série temporal contínua, pode-se transformá-la em uma série discreta, por meio de uma amostragem em intervalos de tempo iguais. Frequentemente, o valor da série em um determinado instante é obtido acumulando-se valores ou, então, pela média das observações, em intervalos de tempo iguais.

A análise de uma série temporal pode ser desenvolvida no domínio do tempo ou no domínio da frequência, sendo os modelos propostos, paramétricos e não paramétricos, respectivamente. No domínio da frequência, tem-se a análise espectral, que possui aplicação em diversas áreas do conhecimento.

Dados de séries temporais surgem em várias áreas do conhecimento tais como medicina, meteorologia, oceanografia, economia, epidemiologia, astronomia, entre outras.

A seguir, são apresentados alguns conceitos, definições e teoremas de séries temporais necessários ao desenvolvimento dos testes para a comparação de séries temporais, objetivo deste trabalho. Tais conceitos, definições e teoremas estarão de acordo com o exposto em Morettin & Tolo (2006).

### 2.2.1 Processos estocásticos

Os fenômenos estatísticos controlados por leis probabilísticas são denominados processos estocásticos. Séries temporais são trajetórias ou realizações de um processo estocástico.

**Definição:** Um processo estocástico é uma família  $Z = \{Z(t), t \in T\}$ , em que, para cada  $t \in T$ ,  $Z(t)$  é uma variável aleatória, sendo  $T$  um conjunto arbitrário.

A variável aleatória  $Z(t)$  é uma função de dois argumentos, ou seja,  $Z(t, \omega)$ , com  $t \in T$  e  $\omega \in \Omega$ . Deste modo, para cada  $t \in T$  fixado,  $Z(t, \omega)$  é uma variável

aleatória. Observa-se na Figura 1 um processo estocástico interpretado como uma família de variáveis aleatórias.

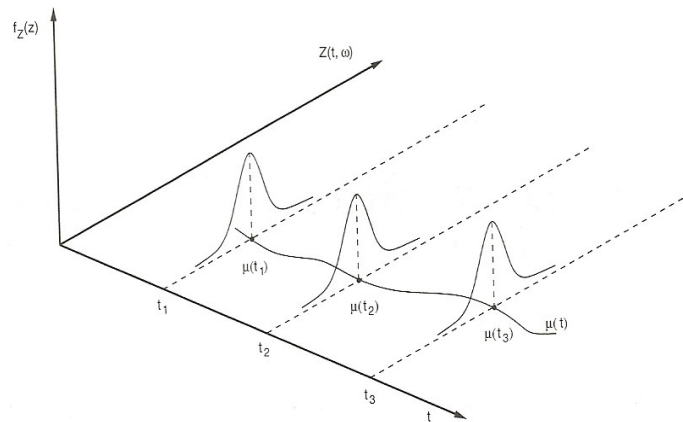


FIGURA 1 Um processo estocástico interpretado como uma família de variáveis aleatórias.

Fonte: Morettin & Tolo (2006).

Outra forma de interpretar um processo estocástico é fixando  $\omega \in \Omega$ . Desta maneira,  $Z(t, \omega)$ , é uma realização ou trajetória do processo estocástico, ou seja, uma série temporal. A Figura 2 mostra um processo estocástico interpretado como uma família de trajetórias.

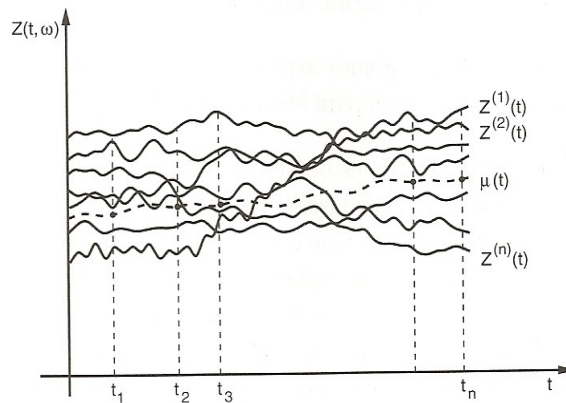


FIGURA 2 Um processo estocástico interpretado como uma família de trajetórias.

Fonte: Morettin & Toloí (2006).

**Definição:** A função de distribuição finito-dimensionais das variáveis aleatórias  $Z(t)$  é definida por:

$$F(z_1, \dots, z_n; t_1, \dots, t_n) = P\{Z(t_1) \leq z_1, \dots, Z(t_n) \leq z_n\}$$

para todo  $n \geq 1$  e  $t_1, t_2, \dots, t_n$  elementos quaisquer de  $T$ .

### 2.2.1.1 Processos estocásticos estacionários

"A idéia básica de estacionariedade é a de que as leis probabilísticas que regem o comportamento do processo não mudam ao longo do tempo. Assim, o processo está em equilíbrio estatístico" (Cryer & Chan, 2008, p. 16). Existem duas formas de estacionariedade: fraca (de segunda ordem ou ampla) e estrita (ou forte). Na sequência, serão definidos os processos estocásticos estritamente estacionários e fracamente estacionários, segundo Morettin & Toloí (2006).

**Definição:** Um processo estocástico  $Z = \{Z(t), t \in T\}$  é definido como estritamente estacionário se todas as distribuições finito-dimensionais continuam as

mesmas sob translações no tempo, isto é,

$$F(z_1, \dots, z_n; t_1 + \tau, \dots, t_n + \tau) = F(z_1, \dots, z_n; t_1, \dots, t_n), \quad (2.1)$$

para quaisquer  $t_1, \dots, t_n, \tau$  de  $T$ .

Na prática, é difícil avaliar a estacionariedade estrita, sendo assim, é comum que estacionariedade seja definida com base nos momentos de primeira e segunda ordem de uma série temporal. Dessa maneira, define-se a estacionariedade fraca.

**Definição:** Um processo estocástico  $Z = \{Z(t), t \in T\}$  é definido como fracamente estacionário se e somente se:

- i)  $E\{Z(t)\} = \mu(t) = \mu$ , constante, para todo  $t \in T$ ;
- ii)  $E\{Z^2(t)\} < \infty$ , para todo  $t \in T$ ;
- iii)  $\gamma(t_1, t_2) = Cov\{Z(t_1), Z(t_2)\}$  é uma função de  $|t_1 - t_2|$ .

### 2.2.2 Função de autocovariância e autocorrelação

A função de autocovariância de um processo estacionário real discreto, de média zero é definida por:

$$\gamma_\tau = E[Z_t Z_{t+\tau}]$$

em que  $\tau$ , denominado de lag, representa a defasagem no tempo.

A função de autocovariância mede a covariância entre as observações de uma série temporal em períodos diferentes. Essa função satisfaz as seguintes propriedades:

- i)  $\gamma_0 > 0$ ,
- ii)  $\gamma_{-\tau} = \gamma_\tau$ ,
- iii)  $|\gamma_\tau| \leq \gamma_0$ ,
- iv)  $\gamma_\tau$  é não negativa definida, no sentido de que:

$$\sum_{j=1}^n \sum_{k=1}^n a_j a_k \gamma_{\tau_j - \tau_k} \geq 0,$$

para quaisquer  $a_1, \dots, a_n \in \mathbb{R}$  e  $\tau_1, \dots, \tau_n \in \mathbb{Z}$ .

A estimativa da função de autocovariância  $\gamma_\tau$ , é dada por:

$$c_\tau = \frac{1}{N} \sum_{t=1}^{N-\tau} [(Z_t - \bar{Z})(Z_{t+\tau} - \bar{Z})], \quad \tau = 0, 1, \dots, N-1,$$

em que  $\bar{Z} = \frac{1}{N} \sum_{t=1}^N Z_t$  e  $c_{-\tau} = c_\tau$ .

**Definição:** A função de autocorrelação (fac) de um processo estacionário é definida por:

$$\rho_\tau = \frac{\gamma_\tau}{\gamma_0}, \quad \tau \in \mathbb{Z}$$

Para a função de autocorrelação são válidas as seguintes propriedades:

- i)  $\rho_0 = 1$
- ii)  $\rho_{-\tau} = \rho_\tau$
- iii)  $|\rho_\tau| \leq \rho_0 = 1$
- iv)  $\rho_\tau$  é não negativa definida

A estimativa da função de autocorrelação ( $\rho_\tau$ ) é dada por:

$$r_\tau = \frac{c_\tau}{c_0}, \quad \tau = 0, 1, \dots, N - 1$$

sendo  $c_\tau$  a estimativa da função de autocovariância. Tem-se  $r_{-\tau} = r_\tau$  e  $c_{-\tau} = c_\tau$ .

### 2.2.3 Decomposição clássica de uma série temporal

Um modelo de decomposição para séries temporais supõe que a série  $\{Z_t, t = 1, 2, \dots, N\}$  pode ser escrita como a soma de três componentes não observáveis, conforme expressão (2.2).

$$Z_t = T_t + S_t + a_t, \quad t = 1, 2, \dots, N \quad (2.2)$$

em que:

$T_t$  é uma componente de tendência,

$S_t$  é uma componente sazonal,

$a_t$  é uma componente aleatória com média zero e variância constante  $\sigma_a^2$ .

Se  $\{a_t\}$  for um ruído branco, então  $E(a_t a_s) = 0, s \neq t$ . Porém pode-se, eventualmente, relaxar esta suposição, tomando  $\{a_t\}$  como um processo estacionário.

O modelo (2.2) é chamado modelo aditivo. Sua utilização é apropriada quando a componente sazonal é independente das outras componentes. Mas se as amplitudes sazonais variam conforme a tendência, o modelo multiplicativo é o mais adequado e pode ser escrito da seguinte forma:

$$Z_t = T_t S_t a_t, \quad t = 1, 2, \dots, N \quad (2.3)$$

Aplicando-se logaritmo ao modelo (2.3), obtém-se um modelo aditivo, conforme pode ser visto a seguir:

$$\begin{aligned} \log(Z_t) &= \log(T_t S_t a_t) \Rightarrow \\ \log(Z_t) &= \log(T_t) + \log(S_t) + \log(a_t), \quad t = 1, 2, \dots, N \end{aligned}$$

#### 2.2.4 Estacionariedade

Segundo Morettin & Toloí (2006), uma série temporal é considerada estacionária quando ela se desenvolve no tempo aleatoriamente em torno de uma média constante, mostrando alguma forma de equilíbrio estável. Frequentemente, supõe-se que a série temporal é estacionária. Entretanto, na prática, a maioria das séries temporais encontradas não são estacionárias. Os componentes tendência e sazonalidade são formas de não estacionariedade.

A maioria das técnicas estatísticas utilizadas na análise de séries temporais partem do pressuposto de que a série é estacionária. Contudo, se a série não for estacionária, será necessário transformar os dados originais.

O procedimento mais simples utilizado para tornar uma série temporal estacionária consiste em tomar diferenças consecutivas na série original até obter uma série estacionária. Normalmente, uma ou duas diferenças são suficientes para retirar a tendência. Para retirar sazonalidade, deve-se fazer uma diferença com lag do tamanho da periodicidade da série em estudo.

A primeira diferença de uma série temporal  $Z(t)$  é definida por:

$$\Delta Z(t) = Z(t) - Z(t - 1)$$

A segunda diferença é dada por:

$$\Delta^2 Z(t) = \Delta[\Delta Z(t)] = \Delta[Z(t) - Z(t - 1)]$$

$$\Delta^2 Z(t) = Z(t) - 2Z(t-1) + Z(t-2)$$

Consequentemente, a  $n$ -ésima diferença de  $Z(t)$  é definida por:

$$\Delta^n Z(t) = \Delta[\Delta^{n-1} Z(t)]$$

No caso de algumas séries temporais, poderá ser necessário aplicar uma transformação não linear, como a logarítmica, para estabilizar a variância. Esta transformação deverá ser aplicada à série original. Para verificar se há necessidade de se fazer a transformação, utiliza-se o gráfico que apresenta a média no eixo das abscissas e a amplitude no eixo das ordenadas. Para o cálculo da média e amplitude, a série original deverá ser dividida em subconjuntos. Se o gráfico mostrar pontos espalhados em torno de uma reta paralela ao eixo das abscissas, é uma indicação de que não será necessário fazer a transformação. Porém, se a amplitude for diretamente proporcional à média, seria um indicativo de que a transformação logarítmica é adequada (Morettin & Toloí, 2006).

Na análise de séries temporais, a construção do gráfico é essencial, visto que ele revela características importantes, tais como tendência, sazonalidade, entre outras. Como a inspeção visual é muito subjetiva, existem testes de hipóteses estatísticos para verificar a presença das componentes tendência e sazonalidade. A seguir, são apresentados o teste do sinal de Cox-Stuart, a análise espectral e o teste de Fisher.

#### 2.2.4.1 Teste do sinal (Cox-Stuart)

O teste de Cox-Stuart é utilizado para verificar a existência de tendência em séries temporais. Serão testadas as seguintes hipóteses:

$$H_0 : P(Z_i < Z_{i+c}) = P(Z_i > Z_{i+c}), \quad \forall_i : \text{não existe tendência};$$

$$H_1 : P(Z_i < Z_{i+c}) \neq P(Z_i > Z_{i+c}), \quad \forall_i : \text{existe tendência}.$$

Conforme visto em Morettin & Tolo (2006), o procedimento para testar a existência de tendência é dado pelos seguintes passos:

1. Agrupam-se as observações em pares  $(Z_1, Z_{1+c}), (Z_2, Z_{2+c}), \dots, (Z_{N-c}, Z_N)$ , sendo que  $N$  é o número de observações da série e  $c = \frac{N}{2}$ , caso  $N$  seja par e  $c = \frac{N+1}{2}$ , se  $N$  for ímpar;
2. Associa-se a cada par  $(Z_i, Z_{i+c})$  um sinal de  $-$  ou  $+$ . Se  $Z_i < Z_{i+c}$ , associa-se um sinal de  $+$  e se  $Z_i > Z_{i+c}$ , utiliza-se o sinal de  $-$ . Desconsideram-se os pares em que  $Z_i = Z_{i+c}$ . Seja  $n$  o número de pares em que  $Z_i \neq Z_{i+c}$ ,
3. Seja  $T_2$  o número de pares com sinal  $+$ ;
4. Seja  $t$  o valor encontrado em uma tabela da distribuição binomial, com parâmetros  $p = \frac{1}{2}$  e  $n$ , caso  $n \leq 20$ . Se  $n > 20$  utiliza-se a distribuição normal com média  $np$  e variância  $npq$ ;
5. Se  $T_2 \geq n - t$ , rejeita-se  $H_0$ , ou seja, a série apresenta tendência.

### 2.2.5 Análise espectral

De acordo com Morettin & Tolo (2006), considere  $\{Z_t, t \in Z\}$  um processo estocástico estacionário de média zero e função de autocovariância obedecendo a uma condição de independência assintótica, no sentido em que, valores do processo bastante separados no tempo, sejam pouco dependentes, podendo ser expressos na forma:

$$\sum_{\tau=-\infty}^{\infty} |\gamma(\tau)| < \infty.$$

Sob essas condições, a função densidade espectral de  $Z_t$  é definida por:

$$f(\lambda) = \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} \gamma(\tau) e^{-i\lambda\tau}, \quad -\infty \leq \lambda \leq \infty$$

em que  $e^{i\lambda} = \cos \lambda + i \operatorname{sen} \lambda$  e  $i = \sqrt{-1}$ .

A função de densidade espectral ou espectro de  $Z_t$  é definida como a transformada de Fourier de  $\gamma(t)$ .

A função de densidade espectral  $f(\lambda)$  é periódica de período  $2\pi$ , sendo então suficiente considerar o intervalo  $[-\pi, \pi]$ . Como  $f(\lambda)$  é uma função par, pode-se representá-la no intervalo  $[0, \pi]$ .

Seja  $\{Z_t, t = 0, 1, \dots, N\}$  um processo estacionário com média zero. A transformada de Fourier finita dos valores  $(Z_1, Z_2, \dots, Z_N)$  é dada pela expressão:

$$d^{(N)}(\lambda) = \frac{1}{\sqrt{2\pi N}} \sum_{t=1}^N Z_t e^{-i\lambda t}, \quad -\infty < \lambda < \infty.$$

Se  $\lambda_j = \frac{2\pi j}{N}$  e  $-\lceil \frac{N-1}{2} \rceil \leq j \leq \lfloor \frac{N}{2} \rfloor$ , obtém-se a transformada de Fourier discreta:

$$\begin{aligned} d_j^{(N)} &= \frac{1}{\sqrt{2\pi N}} \sum_{t=1}^N Z_t e^{-i2\pi jt/N}, \\ &= \frac{1}{\sqrt{2\pi N}} \sum_{t=1}^N Z_t \cos\left(\frac{2\pi jt}{N}\right) + i \frac{1}{\sqrt{2\pi N}} \sum_{t=1}^N Z_t \operatorname{sen}\left(\frac{2\pi jt}{N}\right), \quad j = 0, 1, \dots, \lfloor \frac{N}{2} \rfloor. \end{aligned}$$

Suponha uma realização de um processo estacionário  $\{Z_t, t = 1, \dots, N\}$ .

Então,

$$I_j^{(N)} = |d_j^{(N)}|^2 = \frac{1}{2\pi N} \left| \sum_{t=1}^N Z_t e^{-i\lambda_j t} \right|^2$$

é denominado periodograma, cuja distribuição assintótica é apresentada pelo teorema a seguir.

**Teorema 1:** As ordenadas do periodograma  $I_j^{(N)}$  são variáveis aleatórias assintoticamente independentes e possuem distribuição assintótica múltipla de uma variável aleatória qui-quadrado, ou seja,

$$I_j^{(N)} \xrightarrow{D} \begin{cases} \frac{1}{2}f(\lambda_j)\chi_2^2, & j \neq 0, N/2, \\ f(\lambda_j)\chi_1^2, & j = 0, N/2. \end{cases}$$

A demonstraçãõ do teorema 1 pode ser encontrada em Priestley (1988).

### 2.2.6 Teste de Fisher

Para testar a existênciã de sazonalidade na sêrie temporal, utiliza-se o teste de Fisher (Morettin & Toloi, 2006). Serãõ testadas as hipóteses:

$H_0$  : não existe sazonalidade;

$H_1$  : existe sazonalidade.

Considere a estatística dada por:

$$g = \frac{\max I_p}{\sum_{p=1}^{N/2} I_p}$$

em que  $I_p$  é o valor do periodograma no período  $p$ , e  $N$  é o número de observações da sêrie.

A estatística do teste é dada por:

$$z_\alpha = 1 - \left(\frac{\alpha}{n}\right)^{\frac{1}{n-1}}$$

em que  $\alpha$  é o nível de significância do teste e  $n = \frac{N}{2}$ .

Se  $g > z_\alpha$ , rejeita-se  $H_0$ , isto é, a sêrie apresenta sazonalidade no período  $p$ .

### 2.2.7 Modelos autorregressivos de ordem $p$ (AR(p))

O modelo autorregressivo pertence à família dos modelos autorregressivos e de médias móveis (ARMA). Esta classe de modelos é bastante utilizada na análise de sêries temporais.

O processo autorregressivo de ordem  $p$ , denotado por AR( $p$ ) é definido por:

$$\tilde{Z}_t = \phi_1 \tilde{Z}_{t-1} + \phi_2 \tilde{Z}_{t-2} + \dots + \phi_p \tilde{Z}_{t-p} + a_t \quad (2.4)$$

em que,

$\tilde{Z}_t = Z_t - \mu$  e  $\mu$  é a média da série,

$\phi_i$  são parâmetros do modelo, com  $i = 1, \dots, p$ ,

$a_t$  é independente e identicamente distribuído, com média zero e variância constante, ruído branco.

Isolando  $a_t$ , tem-se:

$$a_t = \tilde{Z}_t - \phi_1 \tilde{Z}_{t-1} - \dots - \phi_p \tilde{Z}_{t-p} \quad (2.5)$$

O operador translação para o passado, denominado  $B$  é dado por,

$$B \tilde{Z}_t = \tilde{Z}_{t-1}, \quad B^m \tilde{Z}_t = \tilde{Z}_{t-m} \quad (2.6)$$

Substituindo a expressão (2.6) em (2.5) tem-se,

$$a_t = \tilde{Z}_t - \phi_1 B \tilde{Z}_t - \dots - \phi_p B^p \tilde{Z}_t = (1 - \phi_1 B - \dots - \phi_p B^p) \tilde{Z}_t$$

em que,

$$\phi(B) = (1 - \phi_1 B - \dots - \phi_p B^p)$$

é o operador autorregressivo estacionário de ordem  $p$ . Assim,

$$a_t = \phi(B) \tilde{Z}_t$$

### 2.2.7.1 Função de autocorrelação

Considere o processo autorregressivo de ordem  $p$  definido por:

$$\tilde{Z}_t = \phi_1 \tilde{Z}_{t-1} + \phi_2 \tilde{Z}_{t-2} + \dots + \phi_p \tilde{Z}_{t-p} + a_t$$

A função de autocovariância pode ser obtida multiplicando-se ambos os lados desta equação por  $\tilde{Z}_{t-j}$  e tomando a esperança. Deste modo, tem-se:

$$\begin{aligned} E(\tilde{Z}_t \tilde{Z}_{t-j}) &= \phi_1 E(\tilde{Z}_{t-1} \tilde{Z}_{t-j}) + \phi_2 E(\tilde{Z}_{t-2} \tilde{Z}_{t-j}) + \dots \\ &\quad + \phi_p E(\tilde{Z}_{t-p} \tilde{Z}_{t-j}) + E(a_t \tilde{Z}_{t-j}) \end{aligned}$$

Como  $E(\tilde{Z}_t \tilde{Z}_{t-j}) = \gamma_j$ ,  $E(a_t \tilde{Z}_{t-j}) = 0$  e  $j > 0$ , obtém-se a função de autocovariância dada pela expressão:

$$\gamma_j = \phi_1 \gamma_{j-1} + \phi_2 \gamma_{j-2} + \dots + \phi_p \gamma_{j-p}$$

Dividindo ambos os lados por  $\gamma_0 = Var(Z_t)$ , obtém-se a função de autocorrelação:

$$\rho_j = \phi_1 \rho_{j-1} + \phi_2 \rho_{j-2} + \dots + \phi_p \rho_{j-p} \quad (2.7)$$

Considerando  $j = 1, 2, \dots, p$  na equação (2.7), e lembrando que  $\rho_0 = 1$  e  $\rho_{-\tau} = \rho_\tau$ , obtém-se:

$$\begin{aligned} \rho_1 &= \phi_1 + \phi_2 \rho_1 + \dots + \phi_p \rho_{p-1} \\ \rho_2 &= \phi_1 \rho_1 + \phi_2 + \dots + \phi_p \rho_{p-2} \\ &\quad \vdots \\ \rho_p &= \phi_1 \rho_{p-1} + \phi_2 \rho_{p-2} + \dots + \phi_p \end{aligned}$$

que são chamadas de equações de Yule-Walker. Na forma matricial, tem-se:

$$\begin{bmatrix} 1 & \rho_1 & \dots & \rho_{p-1} \\ \rho_1 & 1 & \dots & \rho_{p-2} \\ \dots & \dots & \dots & \dots \\ \rho_{p-1} & \rho_{p-2} & \dots & 1 \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_p \end{bmatrix}.$$

Os coeficientes  $\phi_1, \phi_2, \dots, \phi_p$  do modelo AR(p) podem ser estimados utilizando-se as equações de Yule-Walker e substituindo as fac  $\rho_j$  por suas estimativas  $r_j$ .

### 2.2.7.2 Função de autocorrelação parcial

Segundo Box et al. (1994), a função de autocorrelação parcial (facp) ajuda na identificação da ordem do processo autorregressivo, que se adequa à série temporal em estudo.

Considere  $\phi_{kj}$  o  $j$ -ésimo coeficiente do modelo AR(k) e  $\phi_{kk}$  como sendo o último coeficiente.

Sabendo que,

$$\rho_j = \phi_{k1}\rho_{j-1} + \phi_{k2}\rho_{j-2} + \dots + \phi_{kk}\rho_{j-k}, \quad j = 1, \dots, k$$

logo, obtém-se as equações de Yule-Walker:

$$\begin{bmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{k-2} \\ \dots & \dots & \dots & \dots & \dots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \dots & 1 \end{bmatrix} \begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ \vdots \\ \phi_{kk} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{bmatrix}.$$

Resolvendo as equações de Yule-Walker para  $k = 1, 2, 3, \dots$  tem-se:

$$\phi_{11} = \rho_1$$

$$\phi_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}}$$

$$\phi_{33} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}}$$

a fórmula geral do  $\phi_{kk}$  é dada por:

$$\phi_{kk} = \frac{|P_k^*|}{|P_k|},$$

em que  $\phi_{kk}$  é chamada de função de autocorrelação parcial,  $P_k$  é a matriz de autocorrelação e  $P_k^*$  é a matriz de autocorrelação com a última coluna substituída pelo vetor de autocorrelação.

O processo autorregressivo de ordem  $p$  (AR( $p$ )) tem função de autocorrelação parcial (facp)  $\phi_{kk} \neq 0$ , para  $k \leq p$  e  $\phi_{kk} = 0$  para  $k > p$ .

Considerando que o processo é AR( $p$ ), tem-se:

$$Var(\hat{\phi}_{jj}) \simeq \frac{1}{N}, \quad j > p \quad (2.8)$$

Segundo Morettin & Tolo (2006), sob a hipótese de que o processo é AR( $p$ ) e se  $N$  for suficientemente grande, a função de autocorrelação parcial  $\phi_{kk}$  possui distribuição aproximadamente normal, com média zero e variância conforme equação (2.8). Deste modo,  $\phi_{kk}$  será significativamente diferente de zero se:

$$|\hat{\phi}_{kk}| > \frac{2}{\sqrt{N}}, \quad j > p$$

### 2.3 Teste de Box-Pierce

Um teste para as autocorrelações dos resíduos estimados foi sugerido por Box & Pierce (1970). Este teste utiliza as  $k$  primeiras autocorrelações dos resíduos  $\hat{r}_k$ . O teste de Box-Pierce (1970) é utilizado para verificar se os resíduos são um ruído branco, ou seja, independentes e identicamente distribuídos, com média zero e variância constante. A estatística do teste de Box-Pierce é dada por:

$$Q(k) = N(N + 2) \sum_{j=1}^k \frac{\hat{r}_j^2}{(N - j)}$$

Se  $Q > \chi_{k-p-q}^2$ , rejeita-se a hipótese de ruído branco. O valor  $k$  é o número de "lags",  $p$  é a parte autorregressiva do modelo e  $q$  a parte de médias móveis.

### 2.4 Teste de Kolmogorov-Smirnov

O teste de Kolmogorov-Smirnov verifica se uma amostra observada é proveniente de uma distribuição específica.

Segundo Bussab & Morettin (2003), a função de distribuição empírica é definida por:

$$F_N(x) = \frac{N(x)}{N}$$

em que  $N$  é o número de observações,  $x$  é um número real qualquer e  $N(x)$  é o número de observações menores ou iguais a  $x$ .

A hipótese a ser testada é dada por:

$$H_0 : F(x) = F_0(x), \text{ para todo } x,$$

$$H_1 : F(x) \neq F_0(x), \text{ para algum } x,$$

em que  $F_0(x)$  é a função de distribuição especificada e  $F(x)$  é a função de distribuição acumulada.

A estatística do teste é dada por:

$$D = \max_{1 \leq i \leq N} |F(x_i) - F_n(x_i)|$$

Se  $D$  for maior que o valor tabelado proveniente da distribuição em questão, rejeita-se  $H_0$ .

## 2.5 Testes para comparação de séries temporais

Em séries temporais, um problema importante é verificar se duas séries ou partes de uma mesma série foram geradas por um mesmo processo estocástico. Se isso ocorrer, pode-se analisar apenas uma ou parte de uma delas, dependendo do objetivo do pesquisador. Desse modo, em determinadas situações isso proporcionará um menor custo e redução do tempo gasto em coleta e análise de dados.

Para verificar se duas séries foram geradas pelo mesmo processo, Quenouille (1958) e Coates & Diggle (1986) propuseram o teste de igualdade das funções de autocorrelação e o teste das somas acumuladas, respectivamente.

Echeverry & Toloí (2000) aplicaram esses testes em séries de temperatura e salinidade, ambas medidas no fundo e na superfície da estação de monitoramento Boca de Caño Grande, da Ciénaga Grande de Santa Marta.

Os testes propostos por Quenouille (1958) e Coates & Diggle (1986) são testes univariados fundamentados na comparação da função de autocorrelação e na função de densidade espectral, respectivamente. Esses testes requerem que as séries temporais sejam estacionárias e univariadas. A seguir, estão descritos os testes para comparação de séries temporais.

Também será descrito um procedimento para comparar séries temporais utilizado por Silva et al. (2000) para comparar a série de índices de preços ao

consumidor de Lavras e a série do Distrito Federal.

### 2.5.1 Teste das Somas Acumuladas (Coates & Diggle, 1986)

Considere  $I_1(\lambda)$  e  $I_2(\lambda)$  como periodogramas das séries temporais independentes  $Z_1(t)$  e  $Z_2(t)$ , do teorema 1 tem-se que:

$$I_j(\lambda) \sim \frac{f_j(\lambda)\chi_2^2}{2}, \text{ com } 0 < \lambda < \pi. \quad (2.9)$$

em que  $f_j(\lambda)$  é a função de densidade espectral da série  $Z(t)$ .

Define-se as razões espectrais por:

$$J(\lambda) = \frac{I_1(\lambda)}{I_2(\lambda)} \quad \text{e} \quad u(\lambda) = \frac{f_1(\lambda)}{f_2(\lambda)} \quad 0 < \lambda < \pi$$

Da independência assumida entre  $Z_1(t)$  e  $Z_2(t)$  e de (2.9) tem-se que:

$$J(\lambda) = \frac{I_1(\lambda)}{I_2(\lambda)} \sim u(\lambda)F_{2,2}$$

e

$$\ln J(\lambda) \sim \text{logística} \{ \ln u(\lambda), 1 \}$$

em que  $F$  é a distribuição de Fisher-Snedecor e logística  $(\alpha, 1)$  denota a distribuição logística cuja função de distribuição é definida por:

$$F(x) = 1 + e^{-(x-\alpha)^{-1}}, \quad -\infty < x < \infty.$$

Note que, para  $Z \sim \text{logística}(\alpha, 1)$  tem-se

$$E(Z) = \alpha \text{ e } Var(Z) = \frac{\pi^2}{3}$$

Portanto,  $\ln J(\lambda)$  é um estimador não viesado, mas inconsistente para o log da razão espectral  $(\ln u(\lambda))$  e sua variância assintótica independe de  $\lambda$ .

Assintoticamente, os valores  $z_i = \ln(1 + J^{-1}(\lambda_i))$  ( $i = 1, \dots, m$ ), em que  $m = (N - 1)/2$ , são amostras aleatórias de uma distribuição exponencial. Deste modo, se  $c_j = \sum_{i=1}^j z_i$  então  $o_j = c_j/c_m$  é o vetor das estatísticas de ordem da distribuição uniforme.

Segundo Echeverry & Tolo (2000), o teste proposto por Coates & Diggle (1986) consiste em obter a estatística  $o_j$  e aplicar o teste de Kolmogorov-Smirnov para verificar o afastamento da distribuição uniforme (0,1).

Sejam  $f_1(\lambda)$  e  $f_2(\lambda)$  as funções de densidade espectral das séries temporais  $Z_1(t)$  e  $Z_2(t)$ , respectivamente. Serão testadas as hipóteses:

$$H_0 : f_1(\lambda) = f_2(\lambda) \text{ para todo } 0 < \lambda < \pi;$$

$$H_1 : f_1(\lambda) \neq f_2(\lambda) \text{ para algum } 0 < \lambda < \pi.$$

O procedimento de Coates & Diggle (1986) para testar as funções de densidade espectral é dado pelos passos seguintes:

1. Calcular os periodogramas  $I_1(\lambda_i)$  e  $I_2(\lambda_i)$  com  $i = 1, \dots, m$  das séries  $Z_1(t)$  e  $Z_2(t)$ , respectivamente.
2. Calcular a razão dos periodogramas,  $J(\lambda_i) = \frac{I_1(\lambda_i)}{I_2(\lambda_i)}$  com  $i = 1, \dots, m$ ;
3. Calcular  $z_i = \ln(1 + J^{-1}(\lambda_i))$  com  $i = 1, \dots, m$ ;
4. Calcular  $c_j = \sum_{i=1}^j z_i$  com  $j = 1, \dots, m$ ;
5. Obter as estatísticas  $o_j = \frac{c_j}{c_m}$  com  $j = 1, \dots, m$ ;
6. Comparar a distribuição dos  $o_j$  com a distribuição  $U(0, 1)$ , utilizando o teste de Kolmogorov-Smirnov. Se o *p-valor* for maior que  $\alpha$ , não rejeita-se a hipótese  $H_0$ , ao nível de significância  $\alpha$ .

Uma desvantagem deste teste é que ele pode variar dependendo da escolha da razão do periodograma. Segundo Coates & Diggle (1986), o teste é mais fraco para a escolha da razão do periodograma que produz um número maior de valores negativos de  $\ln J(\lambda)$ .

### 2.5.2 Teste de Igualdade das Funções de Autocorrelação (Quenouille, 1958)

Decidir se duas séries temporais ou duas seções de uma série temporal tem a mesma estrutura de correlação é um problema encontrado na prática. Quenouille (1958) propôs um teste para essa comparação. Esse teste é desenvolvido no domínio do tempo e não requer que as séries temporais utilizadas para comparação tenham o mesmo número de observações.

Considere  $\rho_1(j)$  e  $\rho_2(j)$  as funções de autocorrelação das séries  $Z_1(t)$  e  $Z_2(t)$ , respectivamente. Serão testadas as hipóteses:

$$H_0 : \rho_1(j) = \rho_2(j) \text{ para todo } j = \pm 1, \pm 2, \dots;$$

$$H_1 : \rho_1(j) \neq \rho_2(j) \text{ para algum } j = \pm 1, \pm 2, \dots$$

Segundo Echeverry & Toloi (2000), o método de Quenouille para testar as funções de autocorrelação pode ser descrito pelos passos seguintes:

1. Calcular a função de autocorrelação  $\hat{\rho}_1(j)$  e  $\hat{\rho}_2(j)$ , com  $j = 0, 1, \dots, J$ , das séries  $Z_1(t)$  e  $Z_2(t)$ , respectivamente;
2. Calcular a função de autocorrelação comum às duas séries, utilizando a fórmula,  $\hat{\rho}(j) = \frac{n_1 \hat{\rho}_1(j) + n_2 \hat{\rho}_2(j)}{n_1 + n_2}$ , em que  $n_1$  e  $n_2$  são o número de observações das séries  $Z_1(t)$  e  $Z_2(t)$ , respectivamente;
3. Calcular a função de autocorrelação parcial comum estimada ( $\hat{\Phi}(k)$ ), utili-

zando a função de autocorrelação comum ( $\hat{\rho}(j)$ );

4. Utiliza-se o  $\hat{\Phi}(k)$ , para identificar a ordem autorregressiva  $p$ ;
5. Resolvendo as equações de Yule-Walker, estimar os  $p$  coeficientes do modelo autorregressivo;
6. Ajustar para as séries  $Z_1(t)$  e  $Z_2(t)$  o modelo autorregressivo com os coeficientes obtidos no passo 5, obtendo assim as séries residuais  $\hat{a}_1$  e  $\hat{a}_2$ ;
7. Calcular as funções de autocorrelação parcial  $v_j$  e  $v'_j$  para as séries residuais  $\hat{a}_1$  e  $\hat{a}_2$ , respectivamente;
8. Obter a estatística do teste, dada por:

$$SQ = \sum_{j=1}^J \frac{(v_j - v'_j)^2}{\frac{1}{n_1-j} + \frac{1}{n_2-j}}$$

9. Se  $SQ > C_\alpha$ , em que  $C_\alpha$  é tal que  $P(\chi_j^2 > C_\alpha) = \alpha$ , rejeita-se  $H_0$  ao nível de significância  $\alpha$ .

### 2.5.3 Método para comparação de séries temporais

Além dos testes mostrados na seção 2.5.1 e seção 2.5.2, pode-se utilizar o procedimento encontrado em Silva et al. (2000) para comparar séries temporais. O procedimento é apresentado nos passos seguintes:

1. Primeiramente, efetua-se a diferença entre as duas séries temporais em estudo;
2. Na série obtida no passo 1, aplica-se o teste de Cox-Stuart (Morettin & Toloi, 2006) para verificar a existência de tendência;

3. Aplica-se o teste de Fisher (Morettin & Tolo, 2006) para verificar a existência de sazonalidade na série obtida no passo 1;
4. Utiliza-se o teste de Box & Pierce (1970) para verificar se os resíduos são ruído branco, ou seja, são independentes e identicamente distribuídos, com média zero e variância constante.

Se a série - - gerada pela diferença das duas séries - - não apresentar tendência e sazonalidade, isto é, for estacionária e se o resíduo for um ruído branco, conclui-se que as duas séries são iguais no período analisado.

### 3 MATERIAL E MÉTODOS

O conjunto de dados utilizados neste trabalho foi fornecido pela Companhia Ambiental do Estado de São Paulo (CETESB), através do website [www.cetesb.sp.gov.br](http://www.cetesb.sp.gov.br).

As séries temporais fornecidas pela CETESB são séries de valores horários. Porém, neste trabalho serão utilizadas as séries de valores máximos diários de MP<sub>10</sub>.

As séries temporais utilizadas foram:

- série de poluição, por partículas inaláveis (MP<sub>10</sub>), coletadas na Estação Centro localizada na R. da Consolação, 94 - Consolação - Biblioteca Mário de Andrade, no período de 01/01/2007 a 19/12/2009.
- série de poluição, por partículas inaláveis (MP<sub>10</sub>), observadas na Estação Santo Amaro localizada na R. Padre José Maria, 355 - Centro Educ. Esp. Mun. Joerg Bruder, no período de 01/01/2007 a 19/12/2009.

A metodologia aplicada para comparar as duas séries temporais será de acordo com os passos seguintes:

1. analisar o gráfico da amplitude em função da média, verificando se há necessidade de transformar os dados;
2. analisar as séries, através da representação gráfica, verificando indícios de tendência e sazonalidade;
3. aplicar o teste do sinal e o teste de Fisher, para verificar tendência e sazonalidade, respectivamente;

4. caso, a série apresente tendência e/ou sazonalidade, tomar diferenças para torná-la estacionária;
5. aplicar os métodos estatísticos para comparação de séries temporais.

Para realizar as análises estatísticas, será utilizado o software R (R Development Core Team, 2009) . A rotina usada na aplicação do teste de Quenouille (1958) será conforme Echeverry (1999), porém convertida da linguagem de S-Plus para R, anexo B.

## 4 RESULTADOS E DISCUSSÃO

As séries temporais utilizadas neste trabalho foram os valores máximos diários das séries de poluição do ar, por partículas inaláveis ( $MP_{10}$ ), coletadas na Estação Centro e na Estação Santo Amaro, ambas localizadas na cidade de São Paulo. O período analisado foi de 01/01/2007 a 19/12/2009.

### 4.1 Análise da série diária de valores máximos de poluição, por partículas inaláveis ( $MP_{10}$ ), na Estação Centro, na cidade de São Paulo

A análise da série em estudo foi iniciada a partir de uma inspeção visual de seu gráfico. Observa-se na Figura 3, a série diária de valores máximos de poluição por  $MP_{10}$  na Estação Centro na cidade de São Paulo.

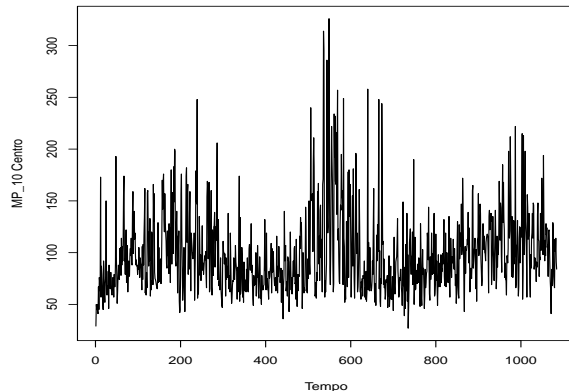


FIGURA 3 Série diária de valores máximos de poluição por  $MP_{10}$  na Estação Centro, na cidade de São Paulo.

A série foi dividida em grupos de 7 observações consecutivas. Para cada grupo foram calculadas a média e a amplitude, com o objetivo de verificar a necessidade ou não de uma transformação na série, para estabilizar a variância. O

resultado pode ser observado na Figura 4. Nota-se que a amplitude é diretamente proporcional à média, indicando a necessidade de uma transformação logarítmica para estabilizar a variância da série temporal em estudo.

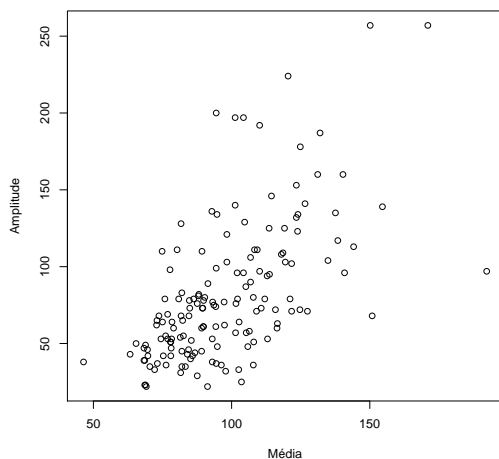


FIGURA 4 Série diária de valores máximos de poluição por  $MP_{10}$  na Estação Centro, na cidade de São Paulo: amplitude vs média.

Aplicou-se o logaritmo natural  $\ln$  aos dados originais. Na Figura 5, pode-se observar que a amplitude independe da média, indicando que a transformação foi adequada.

A Figura 6 apresenta o gráfico da série diária do  $\ln$  dos valores máximos de poluição por  $MP_{10}$  na Estação Centro.

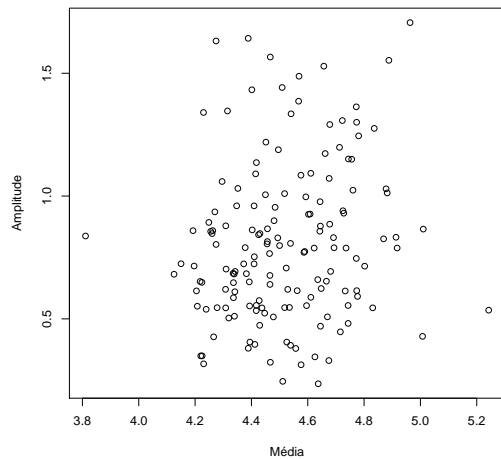


FIGURA 5 Série diária do  $\ln$  dos valores máximos de poluição por ( $MP_{10}$ ) na Estação Centro, na cidade de São Paulo: amplitude vs média.

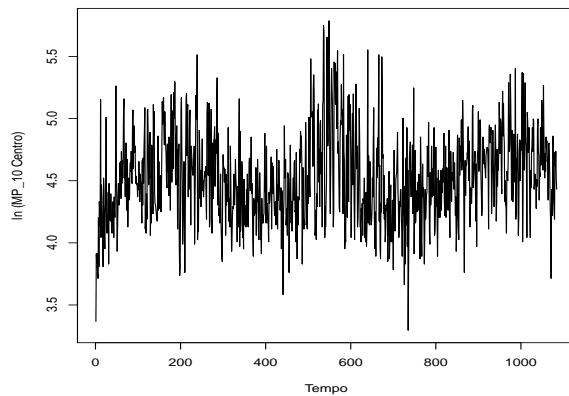


FIGURA 6 Série diária do  $\ln$  dos valores máximos de poluição por  $MP_{10}$  na Estação Centro, na cidade de São Paulo.

Aplicou-se o teste do sinal (Cox-Stuart) para verificar a existência de tendência na série transformada. Considerando que a série possui 1084 observações, tem-se  $c = 542$  e  $n = 537$ . O número de sinais positivos  $T_2$  é igual a

294 e  $t = 287,56$ . Como  $T_2 > 249,44$ , rejeita-se  $H_0$  ao nível de significância  $\alpha = 0,05$ , logo a série transformada apresenta tendência.

Observa-se no periodograma da série transformada (Figura 7) um pico aproximadamente no período de 365 dias. Aplicou-se o teste de Fisher para verificar a existência do efeito sazonal no período 365. Como  $g = 0,118 > z_{0,05} = 0,017$ , rejeita-se  $H_0$  ao nível de significância  $\alpha = 0,05$ , isto é, a série transformada apresenta sazonalidade no período analisado.

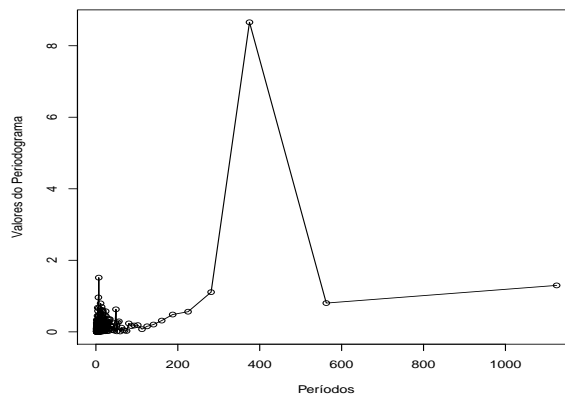


FIGURA 7 Periodograma da série diária do  $\ln$  dos valores máximos de poluição por  $MP_{10}$  na Estação Centro, na cidade de São Paulo.

Para tornar a série diária do  $\ln$  dos valores máximos de poluição por  $MP_{10}$  na Estação Centro, na cidade de São Paulo estacionária, aplicou-se uma diferença na série para remover tendência e uma diferença de 365 para retirar sazonalidade. Nas Figuras 8 e 9, observa-se a série estacionária e sua função de autocorrelação, respectivamente.

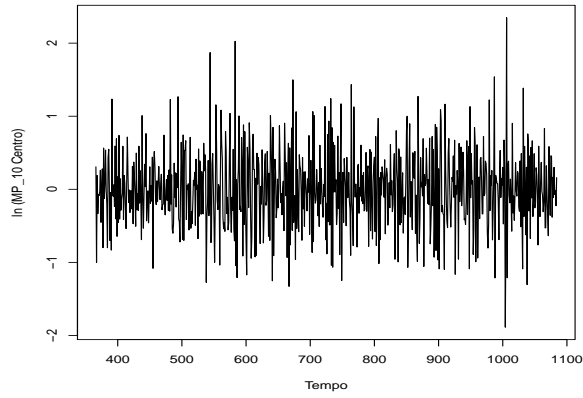


FIGURA 8 Série estacionária de poluição por  $MP_{10}$  na Estação Centro, na cidade de São Paulo.

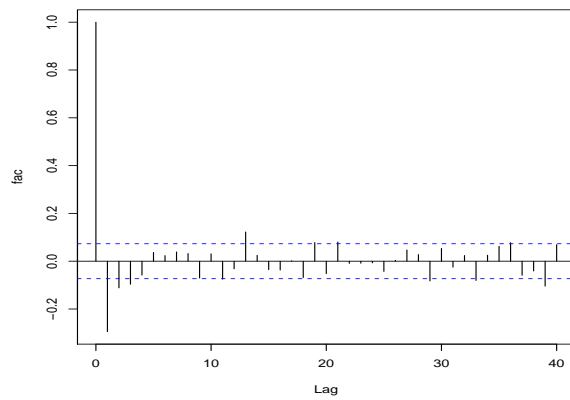


FIGURA 9 Função de autocorrelação da série estacionária da Estação Centro, na cidade de São Paulo.

#### 4.2 Análise da série diária de valores máximos de poluição, por partículas inaláveis (MP<sub>10</sub>), na Estação Santo Amaro, na cidade de São Paulo

Apresenta-se na Figura 10, o gráfico da série diária de valores máximos de poluição, por partículas inaláveis (MP<sub>10</sub>), na Estação Santo Amaro, na cidade de São Paulo, para o período de 01/01/2007 a 19/12/2009.

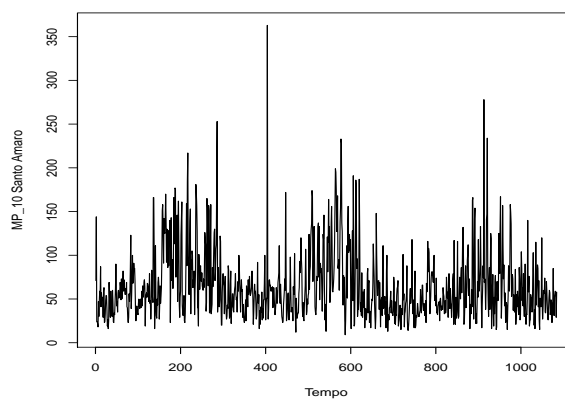


FIGURA 10 Série diária de valores máximos de poluição por MP<sub>10</sub> na Estação Santo Amaro, na cidade de São Paulo.

Inicialmente, a série foi dividida em grupos de 7 observações consecutivas. Para cada grupo, foram calculadas a média e a amplitude. A representação gráfica desse procedimento pode ser observada na Figura 11. O comportamento do gráfico indica que uma transformação logarítmica será necessária para estabilizar a variância da série temporal em estudo.

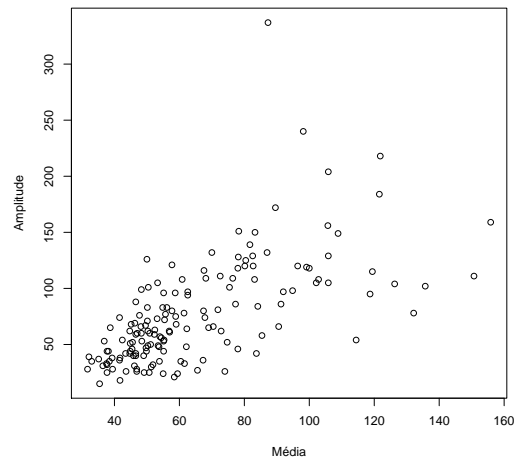


FIGURA 11 Série diária de valores máximos de poluição por  $MP_{10}$  na Estação Santo Amaro, na cidade de São Paulo: amplitude vs média.

Aplicou-se o logaritmo natural aos dados originais. Pode-se observar no gráfico da amplitude em função da média dos dados transformados, mostrado na Figura 12, que a amplitude independe da média, mostrando que a transformação foi apropriada. Na Figura 13, é mostrado o gráfico da série diária do  $\ln$  dos valores máximos de poluição por  $MP_{10}$  na Estação Santo Amaro.

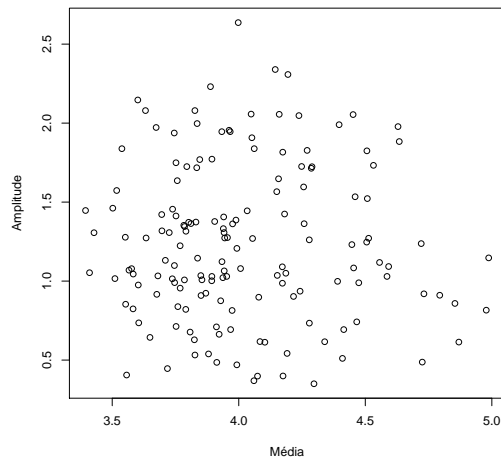


FIGURA 12 Série diária do  $\ln$  dos valores máximos de poluição por  $MP_{10}$  na Estação Santo Amaro, na cidade de São Paulo: amplitude vs média.

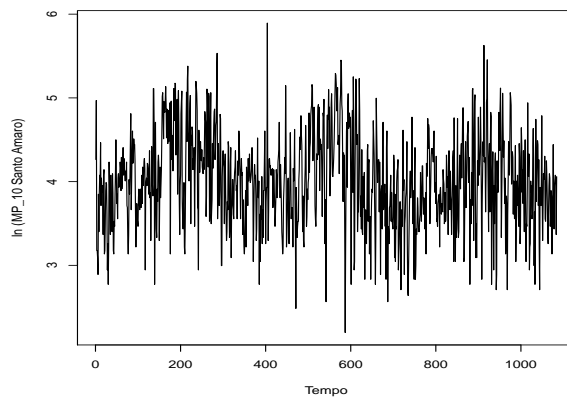


FIGURA 13 Série diária do  $\ln$  dos valores máximos de poluição por  $MP_{10}$  na Estação Santo Amaro, na cidade de São Paulo.

Para verificar se a série transformada apresenta tendência, foi aplicado o teste do sinal, em que foram obtidos  $c = 542$  e  $n = 539$ . O número de sinais positivos  $T_2$  é igual a 241 e  $t = 288,60$ . Como  $T_2 = 241 < 250,4$ , não rejeita-

se  $H_0$  ao nível de significância  $\alpha = 0,05$ . Portanto, a série transformada não apresenta tendência.

O periodograma da série transformada (Figura 14) mostra um pico aproximadamente no período de 365 dias. Aplicou-se o teste de Fisher para verificar sazonalidade no período 365. Como  $g = 0,101 > z_{0,05} = 0,017$ , rejeita-se  $H_0$ , ou seja, a série transformada apresenta sazonalidade no período estudado, ao nível de significância  $\alpha = 0,05$ .

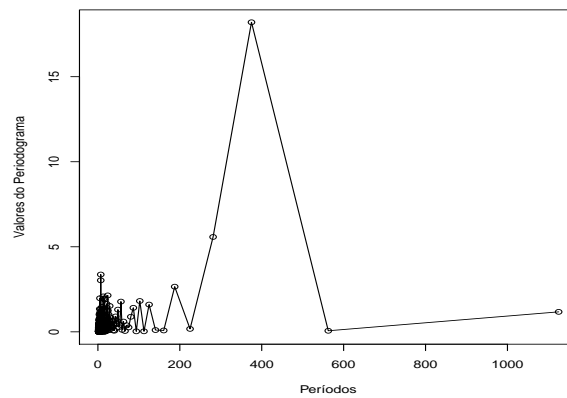


FIGURA 14 Periodograma da série diária do  $\ln$  dos valores máximos de poluição por  $MP_{10}$  na Estação Santo Amaro, na cidade de São Paulo.

Tomou-se uma diferença de 365 para retirar sazonalidade, deixando a série diária do  $\ln$  dos valores máximos de poluição por  $MP_{10}$  na Estação Santo Amaro na cidade de São Paulo estacionária. A série estacionária pode ser observada na Figura 15 e o gráfico da sua função de autocorrelação, na Figura 16.

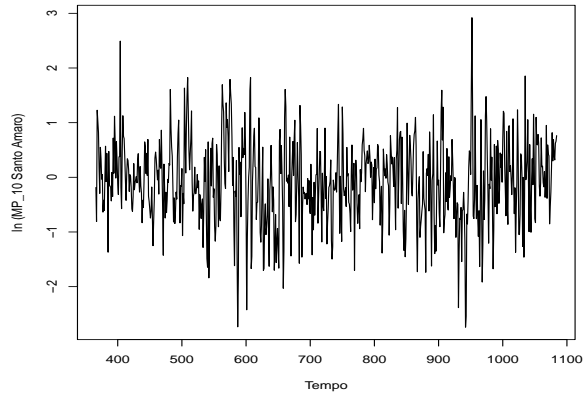


FIGURA 15 Série estacionária de poluição por  $MP_{10}$  na Estação Santo Amaro, na cidade de São Paulo.

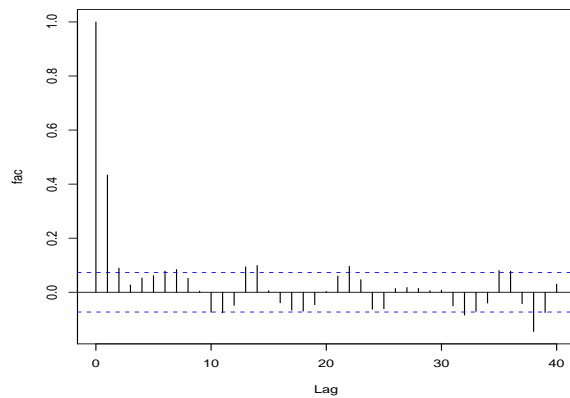


FIGURA 16 Função de autocorrelação da série estacionária da Estação Santo Amaro, na cidade de São Paulo.

#### 4.3 Teste de Coates & Diggle (1986) aplicado às séries estacionárias da Estação Santo Amaro e da Estação Centro, ambas na cidade de São Paulo

Na aplicação deste teste, serão utilizadas as séries estacionárias da Estação Santo Amaro e da Estação Centro, observadas de 01/01/2007 a 19/12/2009. O

teste de Coates & Diggle (1986) utiliza o periodograma das séries estacionárias que podem ser observados nas Figura 18 e 17.

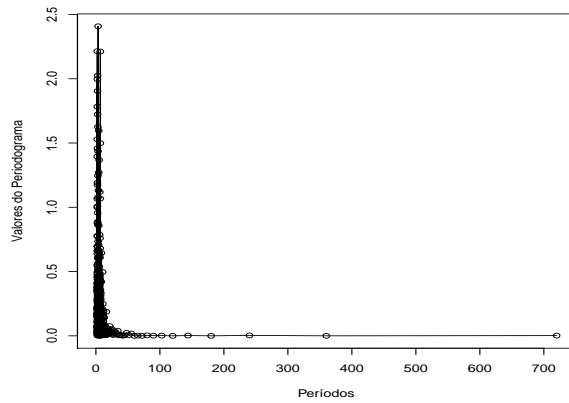


FIGURA 17 Periodograma da série estacionária de poluição por  $MP_{10}$  na Estação Centro, na cidade de São Paulo.

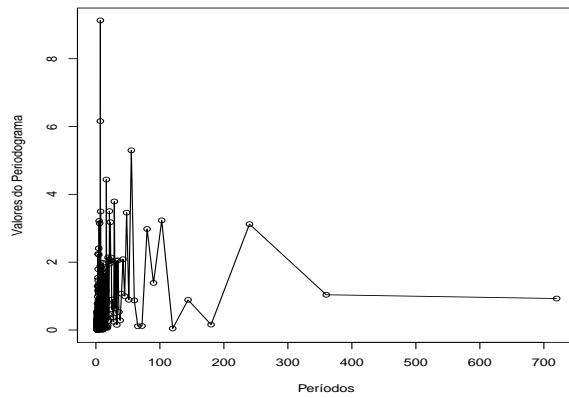


FIGURA 18 Periodograma da série estacionária de poluição por  $MP_{10}$  na Estação Santo Amaro, na cidade de São Paulo.

Foram obtidos os valores do periodograma da séries estacionárias em estudo, calculou-se a razão entre os periodogramas considerando o periodograma da

série estacionária da Estação Santo Amaro como numerador e, como denominador, o periodograma da série da Estação Centro. Desta forma, obteve-se um número menor de valores negativos para o  $\ln(J(\lambda))$ , melhorando a qualidade do teste. Os valores de  $z_i$ ,  $c_j$  e  $o_j$  foram calculados.

Foi utilizado o teste de Kolmogorov-Smirnov para comparar a estatística  $o_j$  do teste das somas acumuladas com distribuição  $U(0, 1)$ . O teste forneceu um  $p\text{-valor} = 2,2e^{-16}$ , levando a rejeitar a hipótese de igualdade das funções de densidade espectral, ao nível de significância  $\alpha = 0,01$ . Assim, conclui-se que as séries em estudo não são geradas pelo mesmo processo estocástico.

#### **4.4 Teste de Quenouille (1958) aplicados às séries estacionárias da Estação Santo Amaro e da Estação Centro, ambas na cidade de São Paulo**

Para a aplicação deste teste, foram usadas as séries estacionárias da Estação Santo Amaro e da Estação Centro, observadas no período de 01/01/2007 a 19/12/2009. O teste de Quenouille (1958) utiliza a função de autocorrelação (fac) das séries estacionárias, que podem ser observadas nas Figura 19 e 20.

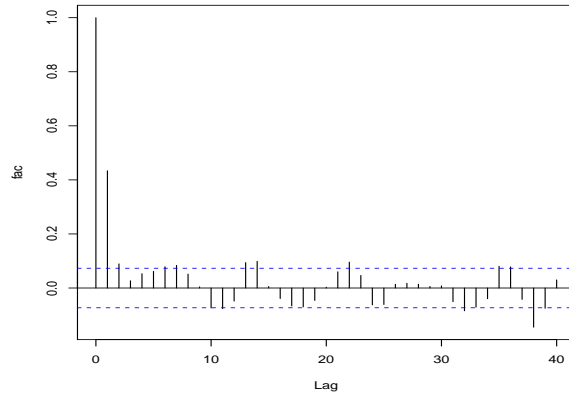


FIGURA 19 Função de autocorrelação da série estacionária da Estação Santo Amaro, na cidade de São Paulo.

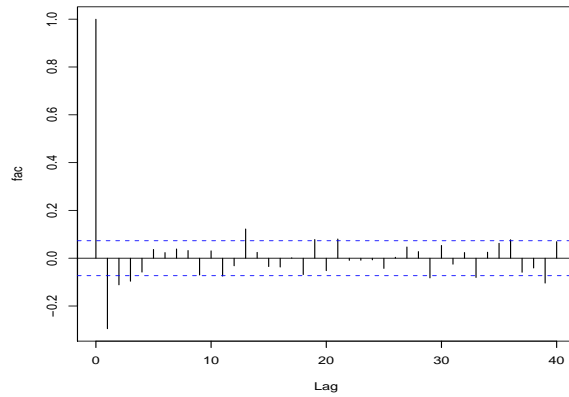


FIGURA 20 Função de autocorrelação da série estacionária da Estação Centro, na cidade de São Paulo.

A função de autocorrelação estimada comum e a função de autocorrelação parcial estimada comum  $\hat{\Phi}(k)$ , podem ser observadas nas Figuras 21 e 22. A Figura 22 mostrou que deve-se ajustar um modelo autorregressivo de ordem 11.

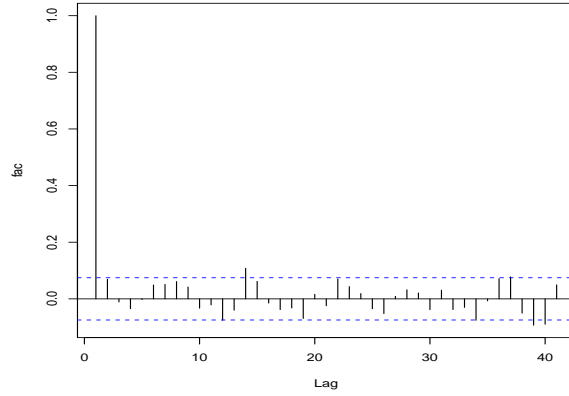


FIGURA 21 Função de autocorrelação comum estimada.

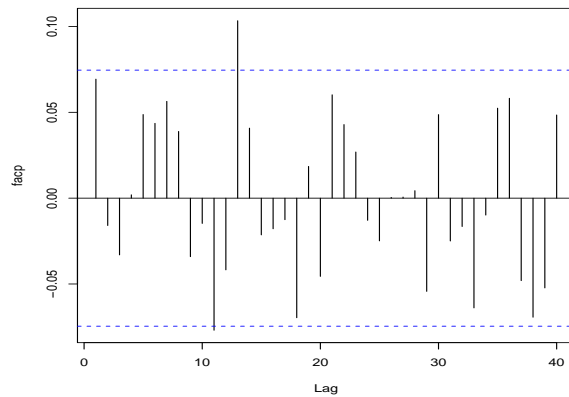


FIGURA 22 Função de autocorrelação parcial comum estimada  $\hat{\Phi}(k)$ .

O modelo ajustado AR(11) pode ser escrito como:

$$\begin{aligned}
 Z_t = & 0,0628Z_{t-1} - 0,0162Z_{t-2} - 0,0276Z_{t-3} + 0,0071Z_{t-4} \\
 & + 0,0513Z_{t-5} + 0,0430Z_{t-6} + 0,0531Z_{t-7} + 0,0385Z_{t-8} \\
 & - 0,0342Z_{t-9} - 0,0098Z_{t-10} - 0,0770Z_{t-11} + a_t
 \end{aligned} \tag{4.1}$$

Ajustou-se o modelo 4.1 às séries estacionárias da Estação Santo Amaro e da Estação Centro, obtendo as séries residuais  $\hat{a}_1$  e  $\hat{a}_2$ . As funções de autocorrela-

ção (fac) das séries residuais podem ser observadas nas Figuras 23 e 24.

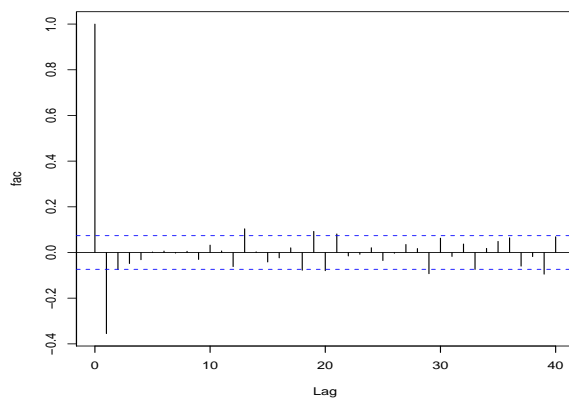


FIGURA 23 Função de autocorrelação da série residual da Estação Centro, na cidade de São Paulo.

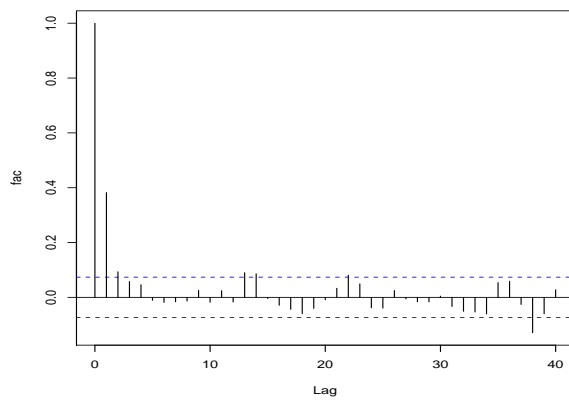


FIGURA 24 Função de autocorrelação da série residual da Estação Santo Amaro, na cidade de São Paulo.

Foram calculadas as funções de autocorrelação parcial (facp) das séries residuais, que podem ser observadas nas Figuras 25 e 26.

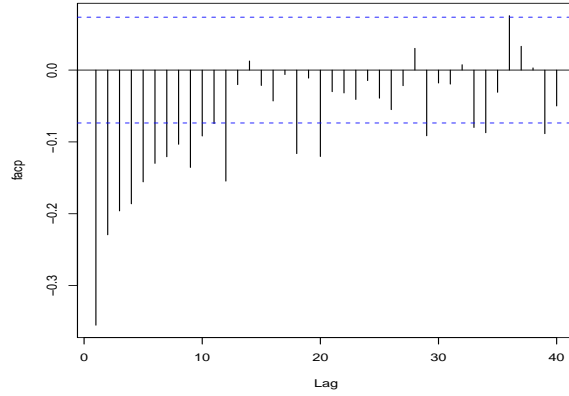


FIGURA 25 Função de autocorrelação parcial da série residual da Estação Centro, na cidade de São Paulo.

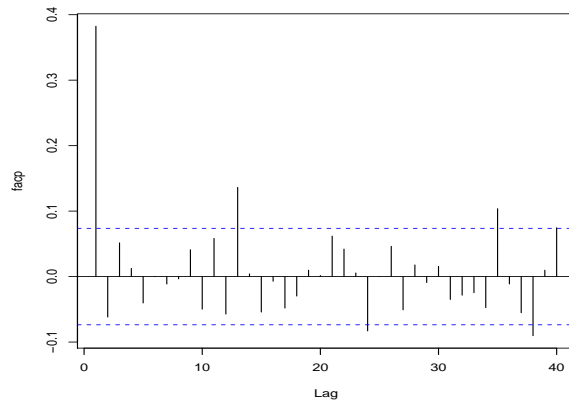


FIGURA 26 Função de autocorrelação parcial da série residual da Estação Santo Amaro, na cidade de São Paulo.

Calculou-se a estatística do teste de Quenouille (1958), obtendo  $SQ = 339,0173$ . Com o  $p\text{-valor} \cong 0$ , a hipótese de igualdade das funções de autocorrelação deve ser rejeitada. Deste modo, conclui-se que as duas séries não são geradas pelo mesmo processo estocástico.

#### 4.5 Comparação das séries estacionárias da Estação Santo Amaro e da Estação Centro, ambas na cidade de São Paulo

Inicialmente, foi realizada uma diferença entre as duas séries estacionárias, com o objetivo de comparar as séries da Estação Santo Amaro e da Estação Centro, ambas na cidade de São Paulo, no período de 01/01/007 a 19/12/2009. O gráfico da diferença entre as séries é mostrado na Figura 27.

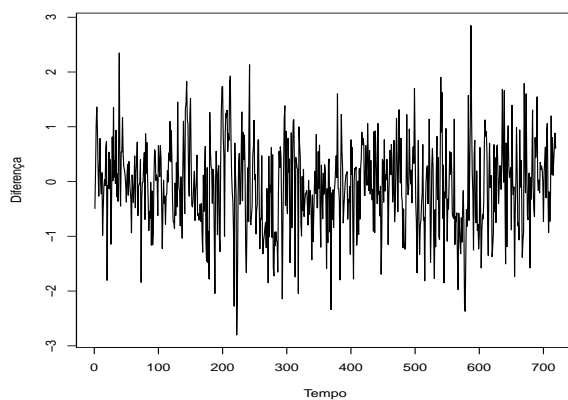


FIGURA 27 Gráfico da diferença entre as séries estacionárias da Estação Santo Amaro e da Estação Centro.

Observa-se nas Figuras 28 e 29, o gráfico da função de autocorrelação (fac) e autocorrelação parcial (facp) da série da diferença.

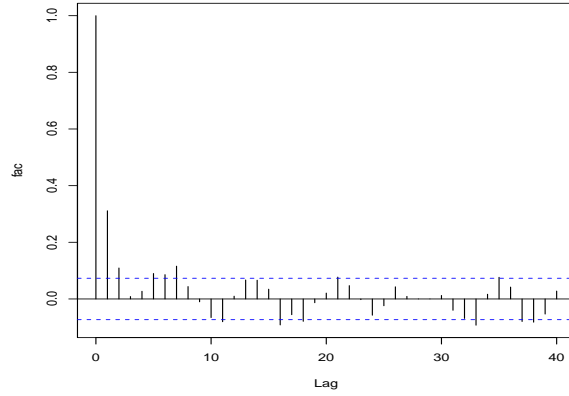


FIGURA 28 Função de autocorrelação da diferença entre as séries estacionárias da Estação Santo Amaro e da Estação Centro.

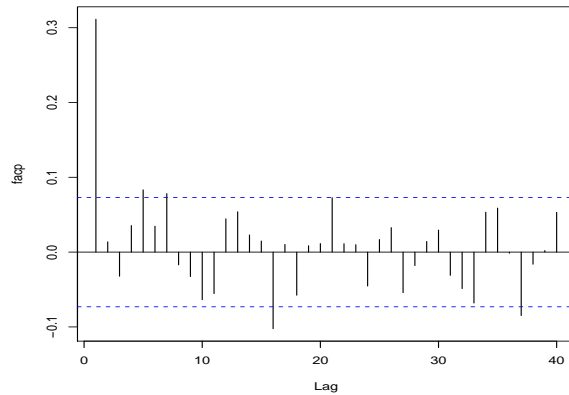


FIGURA 29 Função de autocorrelação parcial da diferença entre as séries estacionárias da Estação Santo Amaro e da Estação Centro.

Aplicou-se o teste do sinal, obtendo  $T_2 = 178$ ,  $n = 359$  e  $t = 195,08$ . Como  $T_2 = 178 > 163,92$ , conclui-se que a série da diferença apresenta tendência, ao nível de significância  $\alpha = 0,05$ . A Figura 30 apresenta o periodograma da série da diferença. O teste de Fisher foi aplicado para verificar sazonalidade de

aproximadamente 7 dias. Tendo como resultado  $g = 0,03 > z = 0,02$  logo, a série apresenta periodicidade semanal.

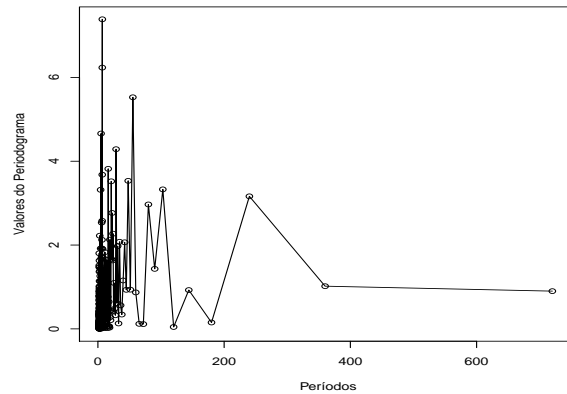


FIGURA 30 Periodograma da diferença entre as séries estacionárias da Estação Santo Amaro e da Estação Centro.

Para verificar se a série é um ruído branco, aplicou-se o teste de Box-Pierce. Rejeitou-se a hipótese de que a série é ruído branco, pois  $p\text{-valor} = 2,2e^{-16}$ .

Portanto, as séries temporais da Estação Santo Amaro e da Estação Centro em São Paulo não são geradas pelo mesmo processo estocástico, já que a série da diferença apresentou tendência, sazonalidade e a série não é um ruído branco.

## 5 CONCLUSÕES

Ao nível de significância  $\alpha = 0,01$ , as três metodologias aplicadas estão rejeitando a hipótese de que a série de valores máximos de poluição diários por  $MP_{10}$  na Estação Santo Amaro e na Estação Centro - - ambas localizadas na cidade de São Paulo - - estão sendo geradas pelo mesmo processo estocástico. Desta maneira, os dados de poluição obtidos em uma das estações em estudo não são suficientes para explicar a poluição na outra estação.

## REFERÊNCIAS BIBLIOGRÁFICAS

- BOX, G. E. P.; JENKINS, G. M.; REINSEL, G. C. **Time series analysis: forecasting and control**. New York: Prentice Hall, 1994. 598 p.
- BOX, G. E. P.; PIERCE, D. A. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. **Journal of the American Statistical Association**, New York, v. 65, n. 332, p. 1509–1526, Dec. 1970.
- BUSSAB, W. O.; MORETTIN, P. A. **Estatística básica**. São Paulo: Saraiva, 2003. 526 p.
- COATES, D. S.; DIGGLE, P. J. Tests for comparing two estimated spectral densities. **Journal of Time Series Analysis**, Clevedon, v. 7, n. 1, p. 7–20, Feb. 1986.
- COMPANHIA AMBIENTAL DO ESTADO DE SÃO PAULO. **Material particulado inalável fino (MP<sub>2,5</sub>) e grosso (MP<sub>2,5-10</sub>) na atmosfera da região metropolitana de São Paulo (2000 - 2006)**. São Paulo, 2008. Disponível em: <<http://www.cetesb.sp.gov.br/Ar/publicacoes.asp>>. Acesso em: 10 jan. 2010.
- COMPANHIA AMBIENTAL DO ESTADO DE SÃO PAULO. **Poluentes**. São Paulo, 2010. Disponível em: <[http://www.cetesb.sp.gov.br/Ar/ar\\_saude.asp](http://www.cetesb.sp.gov.br/Ar/ar_saude.asp)>. Acesso em: 10 jan. 2010.
- COMPANHIA AMBIENTAL DO ESTADO DE SÃO PAULO. **Relatório de qualidade do ar no estado de São Paulo 2008**. São Paulo, 2009. Disponível em: <<http://www.cetesb.sp.gov.br/Ar/publicacoes.asp>>. Acesso em: 10 jan. 2010.
- CONSELHO NACIONAL DO MEIO AMBIENTE. **Resolução n. 003**, de 28 de julho de 1990. Estabelece padrões de qualidade do ar determinando as concentrações de poluentes atmosféricos. Brasília, 1990. Disponível em: <<http://www.mma.gov.br/port/conama/res/res90/res0390.html>>. Acesso em: 15 dez. 2009.
- CRYER, J. D.; CHAN, K. S. **Time series analysis with applications in R**. Madison: Springer, 2008. 491 p.
- ECHEVERRY, G. E. S. **Métodos de comparação de séries temporais**. 1999. 131 p. Dissertação (Mestrado em Estatística) — Universidade de São Paulo, São Paulo.

ECHEVERRY, G. E. S.; TOLOI, C. M. C. Testes para comparação de séries temporais: uma aplicação a séries de temperatura e salinidade da água, medidas em profundidades diferentes. **Revista Brasileira de Estatística**, Rio de Janeiro, v. 61, n. 215, p. 51–80, 2000.

GOUVEIA, N.; MENDONÇA, G. A. S.; LEON, A. P.; CORREIA, J. E. M.; JUNGER, W. L.; FREITAS, C. U.; DAUMAS, R. P.; MARTINS, L. C.; GIUSSEPE, L.; CONCEIÇÃO, G. M. S.; MANERICH, A.; CUNHA-CRUZ, J. Poluição do ar e efeitos na saúde nas populações de duas grandes metrópoles brasileiras. **Epidemiologia e Serviços de Saúde**, Brasília, v. 12, n. 1, p. 29–40, jan./mar. 2003.

MORETTIN, P. A.; TOLOI, C. M. C. **Análise de séries temporais**. São Paulo: E. Blucher, 2006. 538 p.

PRIESTLEY, M. B. **Non-linear and non-stationary time series analysis**. New York: Academic, 1988. 237 p.

QUENOUILLE, M. H. The comparison of correlations in time-series. **Journal of the Royal Statistical Society**, London, v. 20, n. 1, p. 158–164, 1958.

R DEVELOPMENT CORE TEAM. **R: a language and environment for statistical computing**. Vienna: R Foundation for Statistical Computing, 2009. Software.

SÁFADI, T.; PEÑA, D. Bayesian analysis of dynamic factor models: an application to air pollution and mortality in São Paulo, Brazil. **Environmetrics**, Chichester, v. 19, n. 3, p. 582–601, Dec. 2007.

SILVA, R. B. V.; FERREIRA, D. F.; SÁFADI, T. Modelos de séries temporais aplicados à série dos índices de preços ao consumidor na região de Lavras, MG, no período de 1992 a 1999. **Organizações Rurais e Agroindustriais**, Lavras, v. 2, n. 2, p. 44–55, jul./dez. 2000.

## ANEXOS

ANEXO A Rotina para dividir a série em subconjuntos e fazer o gráfico da amplitude versus a média.

```
estats <- function(vs,n){
  ini <- 1
  fim <- n
  media <- NULL
  ampl <- NULL
  if ((length(vs)%n)==0){
for(i in 1:(length(vs)/n)){
    media[i] <- mean(vs[ini:fim],na.rm=TRUE)
    ampl[i]<-(max(vs[ini:fim],na.rm=TRUE)-min(vs[ini:fim],na.rm=TRUE))
    ini <- ini+n
    fim <- fim+n
  }
  media <- media[1:(length(vs)/n)]
  ampl <- ampl[1:(length(vs)/n)]
}
else{
for(i in 1:((length(vs)/n)+1)){
  media[i] <- mean(vs[ini:fim],na.rm=TRUE)
  ampl[i]<-(max(vs[ini:fim],na.rm=TRUE)-min(vs[ini:fim],na.rm=TRUE))
  ini <- ini+n
  fim <- fim+n
}
}
list(media=media,ampl=ampl)
}

dados <- read.table("maxsantoamaro.txt",h=T)
```

```
names(dados)
```

```
result <- estats((dados$MP10santoamaro),7)
```

```
result
```

```
plot(result$media ,result$ampl , xlab="Média" , ylab="Amplitude")
```

ANEXO B Rotina para aplicação do teste proposto por Quenouille (1958). Este programa será conforme Echeverry (1999), porém convertido da linguagem de S-Plus para R.

```
s1 <- read.table("centroestacionaria.txt")
s2 <- read.table("santoamaroestacionaria.txt")
dados2 <- data.frame(s1, s2)
colnames(dados2) <- c("sx", "sy")
n1 <- nrow(dados2)
n2 <- nrow(dados2)
P <- 40

tlag <- 40
sx <- acf(dados2$sx, lag.max=tlag, plot=FALSE)
sy <- acf(dados2$sy, lag.max=tlag, plot=FALSE)
acfco <- ((n1*sx$acf)+(n2*sy$acf))/(n1+n2)
acfc <- c(rep(0, tlag), acfco)
phi <- NULL
vetmat <- numeric(tlag)
vetmat[1]<- list(1)

library(matrixcalc)
teste <- acfco[, ,1][2:31]
matau <- as.matrix(diag(tlag))
matau[2:tlag,1]<- teste[1:(tlag-1)]
matau[1,2:tlag]<- teste[1:(tlag-1)]

lim1 <- 2
for(i in 1:tlag){
  matau[(i+1):tlag, i] <- teste[1:(tlag-i)]
  matau[i, (i+1):tlag] <- teste[1:(tlag-i)]
}
```

```

# Achar a função de autocorrelação parcial

for(i in 2:tlag){
  mat1 <- matrix(i*i,i,i)
  for(j in 1:i){
    lim1 <- tlag+2-j
    lim2 <- tlag+i+1-j
    mat1[j,] <- acfc[lim1:lim2]
  }
  matden <- mat1+t(mat1)
  matden <- matden-diag(i)
  vetmat[i] <- list(matden)
  denom <- det(matden)
  lim3 <- tlag+2
  lim4 <- tlag+i+1
  colu <- matrix(acfc[lim3:lim4],i,1)
  mat2 <- matden[,-c(i)]
  mat3 <- cbind(mat2,colu)
  num <- det(mat3)
  phi[i] <- num/denom
}

phi[1] <- acfco[2]
phi

write(phi,"fap.txt",ncol=40)

# Encontrando a ordem autorregressiva do modelo

n1 <- dim(dados2)[1]
n2 <- dim(dados2)[1]

```

```

maximo <- max(n1, n2)
inter <- 2*(1/sqrt(maximo))
facp <- read.table("facp.txt")
V <- c(seq(1, 40, by=1))
plot(V, facp, type="h", xlab="Lag", ylab="facp")
abline(h=0)
abline(h=-inter, lty=2, col="blue")
abline(h=inter, lty=2, col="blue")

p <- 11 # ordem autorregressiva do modelo

# Calculando os coeficientes do modelo AR(p)

lim1 <- tlag+2
lim2 <- tlag+1+p
r <- matrix(acfc[lim1:lim2], p, 1)
R <- vetmat[[p]]
phipar <- solve(R)%*%r
phipar

# Ajustando o modelo autorregressivo a cada série

p1 <- p+1
np1 <- n1-p
np2 <- n2-p
x <- dados2$sx[p1:n1]
y <- dados2$sy[p1:n2]

xmat <- matrix(0, np1, p)
ymat <- matrix(0, np2, p)

lim1 <- p1

```

```

lim2 <- n1
lim3 <- n2

for(i in 1:p){
  lim1 <- lim1-1
  lim2 <- lim2-1
  lim3 <- lim3-1
  xmat[,i] <- dados2$sx[lim1:lim2]
  ymat[,i] <- dados2$sy[lim1:lim3]
}

res1 <- x-xmat%*%phipar
res2 <- y-ymat%*%phipar

# Obtendo a estatística SQ do teste de Quenouille (1958)

fap1 <- acf(res1,P,type="partial",plot=FALSE)
fap2 <- acf(res2,P,type="partial",plot=FALSE)

s <- fap1$lag

fap1 <- fap1$acf
fap2 <- fap2$acf

denom2 <- (1/(n1-s))+(1/(n2-s))
numer2 <- (fap1-fap2)^2
sq <- sum(numer2/denom2)
pv1 <- 1-pchisq(sq,P)
sq
pv1

```