



UNICAMP

UNIVERSIDADE ESTADUAL DE CAMPINAS
Faculdade de Engenharia Química

WESLEY HELENO PRIETO

**APLICAÇÃO DA TEORIA DO CAOS NA IDENTIFICAÇÃO DA
DINÂMICA E PREVISÃO DE SÉRIES TEMPORAIS DE
POLUIÇÃO ATMOSFÉRICA**

**APPLICATION OF CHAOS THEORY IN THE IDENTIFICATION
OF THE DYNAMICS AND PREDICTION OF ATMOSPHERIC
POLLUTION TIME SERIES**

CAMPINAS
2018

WESLEY HELENO PRIETO

**APLICAÇÃO DA TEORIA DO CAOS NA IDENTIFICAÇÃO DA
DINÂMICA E PREVISÃO DE SÉRIES TEMPORAIS DE
POLUIÇÃO ATMOSFÉRICA**

**APPLICATION OF CHAOS THEORY IN THE IDENTIFICATION
OF THE DYNAMICS AND PREDICTION OF ATMOSPHERIC
POLLUTION TIME SERIES**

*Tese de Doutorado apresentada à
Faculdade de Engenharia Química da
Universidade Estadual de Campinas como
requisito para a obtenção do título de
Doutor em Engenharia Química.*

*Thesis presented to the School of Chemical
Engineering of the University of Campinas
as part of the in partial fulfillment of the
requirements for the degree of Doctor in
Chemical Engineering.*

Orientador: Prof. Dr. Marco Aurélio Cremasco

**ESTE TRABALHO CORRESPONDE À
VERSÃO FINAL DA TESE DE
DOUTORADO DEFENDIDA PELO
ALUNO WESLEY HELENO PRIETO E
ORIENTADA PELO PROF. DR.
MARCO AURÉLIO CREMASCO**

CAMPINAS
2018

Agência(s) de fomento e nº(s) de processo(s): CAPES, 33003017034P8

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca da Área de Engenharia e Arquitetura
Rose Meire da Silva - CRB 8/5974

P933a Prieto, Wesley Heleno, 1988-
Aplicação da teoria o caos na identificação da dinâmica e previsão de séries temporais de poluição atmosférica / Wesley Heleno Prieto. – Campinas, SP : [s.n.], 2018.

Orientador: Marco Aurélio Cremasco.
Tese (doutorado) – Universidade Estadual de Campinas, Faculdade de Engenharia Química.

1. Caos. 2. Predição. 3. Inteligencia artificial. 4. Lyapunov, Expoentes de. 5. Poluição atmosférica. I. Cremasco, Marco Aurélio, 1962-. II. Universidade Estadual de Campinas. Faculdade de Engenharia Química. III. Título.

Informações para Biblioteca Digital

Título em outro idioma: Application of chaos theory in the identification of the dynamics and prediction of atmospheric pollution time series

Palavras-chave em inglês:

Chaos

Prediction

Artificial Intelligence

Lyapunov, Expoent

Atmospheric pollution

Área de concentração: Engenharia Química

Titulação: Doutor em Engenharia Química

Banca examinadora:

Marco Aurélio Cremasco [Orientador]

Edson Tomaz

Gustavo Paim Valença

Mônica Lopes Aguiar

Roberto Guardani

Data de defesa: 10-12-2018

Programa de Pós-Graduação: Engenharia Química

Tese de Doutorado defendida por Wesley Heleno Prieto e aprovada em 10 de Dezembro de 2018 pela banca examinadora constituída pelos doutores:

Prof. Dr. Marco Aurélio Cremasco – Orientador
FEQ/UNICAMP

Prof. Dr. Edson Tomaz
FEQ/UNICAMP

Prof. Dr. Gustavo Paim Valença
FEQ/UNICAMP

Profa. Dra. Mônica Lopes Aguiar
DEQ/UFSCar

Prof. Dr. Roberto Guardani
DEQ/USP

A Ata da defesa com as respectivas assinaturas dos membros encontra-se no SIGA/Sistema de Fluxo de Tese e na Secretaria do Programa da Unidade.

Aos meus pais, irmão, avós e sobrinho, os verdadeiros pilares que sustentam minha vida, a razão do meu fôlego, os motivadores da minha jornada!

AGRADECIMENTOS

Nenhum projeto bem sucedido é executado individualmente. Um simples incentivo já é mais do que suficiente para motivar agradecimentos! Nesta Tese não faltaram incentivos, ajudadores, mãos amigas que juntas teceram fio a fio este trabalho. Palavras não conseguem expressar toda minha gratidão, mas aqui dedico um pequeno espaço para demonstrar o quão grato sou a todos vocês.

Aos meus *pais*, incríveis e iluminados, sempre se dedicaram em me encaminhar nas veredas da retidão, tornando-me um ser humano melhor. Vocês são o meu orgulho, a minha inspiração!

Aos meus queridos *mestres*, os guardiões do conhecimento, essenciais para minha formação!

Ao colega *Victor de Backer Moura*, cuja contribuição no que tange à metodologia preditiva foi vital para o desenvolvimento desta Tese.

Ao *Prof. Dr. Edson Tomaz* por sua crucial contribuição no encaminhamento desta Tese e por todas as orientações sobre o tema da poluição atmosférica.

Ao muito estimado *Prof. Dr. Marco Aurélio Cremasco. Mestre*, suas orientações para minha acadêmica estão estampadas nesta Tese, mas seus ensinamentos para minha vida estão cravados em meu coração! Muito obrigado por tudo!!

A *Faculdade de Engenharia Química da Universidade Estadual de Campinas* que me forneceu todo subsídio necessário para o desenvolvimento deste trabalho.

A *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)* pelo apoio financeiro.

Voici mon secret. Il est très simple: on ne voit bien qu'avec le coeur. L'essentiel est invisible pour les yeux.

Antoine de Saint-Exupéry

RESUMO

A complexa dinâmica envolvida na poluição atmosférica promove a necessidade de metodologias de análise que representem, com maior fidelidade, suas características físicas. Nas últimas décadas, diversas pesquisas demonstraram os efeitos nocivos dos contaminantes atmosféricos à saúde humana, fauna e flora, desta forma explicitando a urgência do conhecimento das causas, padrões e mecanismos deste fenômeno. Os métodos clássicos empregados para este fim, frequentemente, negligenciam as interações não lineares presentes na evolução temporal dos poluentes. Nesta vanguarda, esta Tese aplica a teoria caótica determinista no estudo de séries temporais de concentração de material particulado inalável (MP_{10}) e óxidos de nitrogênio (NO_x) fornecidas por cinco estações de monitoramento pertencente à Companhia Ambiental do Estado de São Paulo (CETESB) nos anos de 1998 a 2017. Objetiva-se também a aplicação do modelo preditivo dos vizinhos próximos (k -NN) na predição da concentração horária de MP_{10} e NO_x dos primeiros dois meses (1440 pontos) de 2018. Foram obtidos os atratores no espaço de fases e, a partir de sua topografia, foram determinados os valores dos invariantes dimensão de correlação (D_2), entropia de Kolmogorov (K_{ML}) e máximo expoente de Lyapunov (λ). Com base nessas grandezas, observou-se a natureza não linear, aperiódica e determinística dos dados avaliados, demonstrando que a teoria do caos é uma alternativa viável no estudo de poluição atmosférica. Realizou-se também a associação da abordagem k -NN com a Teoria do Caos no que tange a determinação da dimensão de imersão. Por meio do modelo preditivo, verificou-se que é possível realizar previsões dos poluentes destacados, principalmente para previsões de 24 horas, sendo estas informações úteis que podem ser aplicadas na tomada de decisão e elaboração de políticas públicas.

Palavras-chave: *caos; predição; entropia de Kolmogorov; expoente de Lyapunov; dimensão de correlação; poluição atmosférica; NO_x ; MP_{10} ; k -NN.*

ABSTRACT

The complex dynamics involved in atmospheric pollution promotes the need for analysis methodologies that represent, with greater fidelity, its physical characteristics. In recent decades, several researches have demonstrated the harmful effects of atmospheric contaminants on human health, fauna and flora, thus explaining the urgency of knowing the causes, patterns and mechanisms of this phenomenon. The classical methods employed for this purpose often neglect the nonlinear interactions present in the temporal evolution of pollutants. In this vanguard, this thesis applies the deterministic chaotic theory in the study of time series of concentration of inhalable particulate matter (PM_{10}) and nitrogen oxides (NO_x) provided by five monitoring stations belonging to the Environmental Company of São Paulo State (CETESB) in years from 1998 to 2017. The forecasting model of nearest neighbors (k -NN) is also used to predict the hourly concentration of PM_{10} and NO_x in the first two months (1440 points) of 2018. The attractors were obtained in the phase space and, from their topography, the values of the invariants correlation dimension (D_2), Kolmogorov entropy (K_{ML}) and maximum Lyapunov exponent (λ) were determined. Based on these quantities, the nonlinear, aperiodic and deterministic nature of the evaluated data was observed, demonstrating that chaos theory is a viable alternative in the study of atmospheric pollution. The association of the k -NN approach with the Chaos Theory with regard to the determination of the immersion dimension was also carried out. Through the predictive model, it was verified that it is possible to carry out forecasts of the pollutants highlighted, mainly for 24-hour forecasts, and this useful information can be applied in decision making and public policy making.

Keywords: *chaos; forecasting; Kolmogorov entropy; Lyapunov exponent; correlation dimension; atmospheric pollution; NO_x ; MP_{10} ; k -NN.*

NOMENCLATURA

Letras latinas

$C(\omega)$ - função de correlação [-];

D – distância absoluta [L];

D_2 – dimensão de correlação [-];

e_t – ruído branco no instante t .

f_s – frequência de amostragem [Hz];

I = dimensão de imersão [-];

K - entropia de Kolmogorov [bits/s];

k – número de vizinhos próximos [-];

K_{ML} - entropia de Kolmogorov calculada pelo método da verossimelhança [bits/s]

MAE = erro absoluto médio;

MP_{10} – material particulado inalável;

$[MP_{10}]$ – concentração de material particulado inalável [$\mu\text{g}/\text{m}^3$];

$[NO_x]$ – concentração de óxidos de nitrogênio [ppb];

n – tamanho da amostra;

p – número de observações consideradas;

$PC1$ – principal componente 1;

$PC2$ – principal componente 2;

$RMSE$ = raiz do erro quadrático médio;

t – tempo [s];

x_i – elemento da amostra;

\bar{x} – média amostral;

z_t – ponto na série no momento do tempo t .

Letras gregas

β – variável do sistema de Lorenz [-];

ε – número de Rayleigh [-];

θ – tempo de atraso [s];

δ – nível inicial do modelo;

λ – expoente de Lyapunov [-];

σ – desvio padrão amostral;

τ_S – passo no tempo [-];

φ – número de Prandtl [-];

φ_i = i-ésimo coeficiente autoregressivo.

LISTA DE FIGURAS

Figura 2.1 – Estação de Congonhas (arquivo fotográfico QUALAR CETESB: http://qualar.cetesb.sp.gov.br).....	38
Figura 2.2 – Estação do Ibirapuera (arquivo fotográfico QUALAR CETESB: http://qualar.cetesb.sp.gov.br).....	39
Figura 2.3 – Estação de Guarulhos (arquivo fotográfico QUALAR CETESB: http://qualar.cetesb.sp.gov.br).....	40
Figura 2.4 – Estação de Cubatão – Centro (arquivo fotográfico QUALAR CETESB: http://qualar.cetesb.sp.gov.br).....	42
Figura 2.5 – Estação do Centro de São Paulo (arquivo fotográfico QUALAR CETESB: http://qualar.cetesb.sp.gov.br).....	43
Figura 2.6 – Atrator de Lorenz (Lorenz, 1963).....	46
Figura 2.7 – Metodologia da análise caótica (Prieto, 2014).....	51
Figura 2.8 – Variação do valor da integral de correlação ($C(r)$) com o raio (r) para dimensão de imersão (I) variando de 2 até 10, para a concentração de O_3 (Raga e Le Moyne, 1996).....	52
Figura 2.9 – Variação da dimensão de correlação (D) com a dimensão de imersão (I) para a concentração de O_3 (Raga e Le Moyne, 1996).....	52
Figura 2.10 – Cálculo do máximo expoente de Lyapunov (Yu <i>et al.</i> , 2011).....	53
Figura 2.11 – Variação da dimensão de correlação com a dimensão de imersão (Lee e Lin, 2008).....	54
Figura 2.12 – Método dos vizinhos mais próximos (Mulak e Talhar, 2015).....	70
Figura 3.1 – Estações de monitoramento escolhidas para o estudo (CETESB, 2014).....	75
Figura 3.2 – Fluxograma de obtenção das séries temporais (o autor, 2018).....	76
Figura 3.3 – Tela de aquisição de dados do sistema QUALAR (QUALAR CETESB: http://qualar.cetesb.sp.gov.br).....	77
Figura 3.4 – Comparativo entre a série bruta e o mesmo conjunto de dados após a retirada de células vazias e zeradas (o autor, 2018).....	78
Figura 3.5 – Fluxograma da análise do Caos determinístico e do método preditivo (o autor, 2018).....	79
Figura 3.6 – Fluxograma detalhado das etapas de análise adotadas nesta Tese (o autor, 2018).....	81

Figura 3.7 – Convergência da dimensão de correlação e obtenção da dimensão de imersão (o autor, 2018).....	82
Figura 3.8 – Metodologia de obtenção do máximo expoente de Lyapunov (o autor, 2018).....	83
Figura 3.9 – Integração da Teoria do Caos com a abordagem <i>k-NN</i> (o autor, 2018).....	84
Figura 4.1 – Variação da concentração anual de MP ₁₀ e NO _x na estação do Ibirapuera dos anos de 1998 à 2017 (o autor, 2018).....	87
Figura 4.2 – Séries temporais e atratores de concentração anual MP ₁₀ obtidas na estação de monitoramento do Ibirapuera nos anos de 1998 a 2014 (o autor, 2018).....	89
Figura 4.3 – Séries temporais e atratores de concentração anual NO _x obtidas na estação de monitoramento do Ibirapuera nos anos de 1998 a 2017 (o autor, 2018).....	93
Figura 4.4 – Variação da concentração anual de MP ₁₀ e NO _x na estação de Congonhas dos anos de 1998 à 2017 (o autor, 2018).....	102
Figura 4.5 – Séries temporais e atratores de concentração anual de MP ₁₀ obtidas na estação de monitoramento de Congonhas nos anos de 1998 a 2017 (o autor, 2018).....	104
Figura 4.6 – Séries temporais e atratores de concentração anual de NO _x obtidas na estação de monitoramento de Congonhas nos anos de 1998 a 2017 (o autor, 2018).....	108
Figura 4.7 – Variação da concentração anual de MP ₁₀ e NO _x na estação do Centro de São Paulo desde 1998 até o último ano disponibilizado pela plataforma (o autor, 2018).....	115
Figura 4.8 – Séries temporais e atratores de concentração anual de MP ₁₀ obtidas na estação de monitoramento do Centro de São Paulo nos anos de 1998 a 2010 (o autor, 2018).....	117
Figura 4.9 – Séries temporais e atratores de concentração anual de NO _x obtidas na estação de monitoramento do Centro de São Paulo nos anos de 1998 a 2005 (o autor, 2018).....	119
Figura 4.10 – Variação da concentração anual de MP ₁₀ na estação de Guarulhos de 1998 a 2009 (o autor, 2018).....	124
Figura 4.11 – Séries temporais e atratores de concentração anual de MP ₁₀ obtidas na estação de monitoramento de Guarulhos nos anos de 1998 a 2009 (o autor, 2018).....	127
Figura 4.12 – Variação da concentração anual de MP ₁₀ e NO _x na estação de Cubatão de 1998 a 2017 (o autor, 2018).....	131
Figura 4.13 – Séries temporais e atratores de concentração anual de MP ₁₀ obtidas na estação de monitoramento de Cubatão nos anos de 1998 a 2009 (o autor, 2018).....	134

Figura 4.14 – Séries temporais e atratores de concentração anual de NO _x obtidas na estação de monitoramento de Cubatão nos anos de 1998 a 2017 (o autor, 2018).....	138
Figura 4.15 – Série de treino da concentração horária de MP ₁₀ e NO _x dos anos de 1998 a 2017 na estação do Ibirapuera (o autor, 2018).....	145
Figura 4.16 – Atrator reconstruído a partir da série de treino da concentração horária de MP ₁₀ para os anos de 1998 a 2017 (Ibirapuera) (o autor, 2018).....	146
Figura 4.17 – Atrator reconstruído a partir da série de treino da concentração horária de NO _x para os anos de 1998 a 2017 (Ibirapuera) (o autor, 2018).....	147
Figura 4.18 – Séries temporais da concentração de MP ₁₀ e NO _x obtidas na estação de monitoramento do Ibirapuera para os meses de janeiro e fevereiro de 2018 (o autor, 2018).....	148
Figura 4.19 – Previsão <i>out of sample</i> da concentração horária de MP ₁₀ nos primeiros dois meses de 2018 na estação de Congonhas utilizando o algoritmo <i>k-NN</i> (o autor, 2018).....	149
Figura 4.20 – Previsão <i>out of sample</i> da concentração horária de NO _x nos primeiros dois meses de 2018 na estação de Congonhas utilizando o algoritmo <i>k-NN</i> (o autor, 2018).....	150
Figura 4.21 – Série de treino da concentração horária de MP ₁₀ e NO _x dos anos de 1998 a 2017 na estação de Congonhas (o autor, 2018).....	157
Figura 4.22 – Atrator reconstruído a partir da série de treino da concentração horária de MP ₁₀ para os anos de 1998 a 2017 (Congonhas) (o autor, 2018).....	158
Figura 4.23 – Atrator reconstruído a partir da série de treino da concentração horária de NO _x para os anos de 1998 a 2017 (Congonhas) (o autor, 2018).....	158
Figura 4.24 – Séries temporais da concentração de MP ₁₀ e NO _x obtidas na estação de monitoramento de Congonhas para os meses de janeiro e fevereiro de 2018 (o autor, 2018).....	160
Figura 4.25 – Previsão <i>out of sample</i> da concentração horária de MP ₁₀ nos primeiros dois meses de 2018 na estação de Congonhas utilizando o algoritmo <i>k-NN</i> (o autor, 2018).....	161
Figura 4.26 – Previsão <i>out of sample</i> da concentração horária de NO _x nos primeiros dois meses de 2018 na estação de Congonhas utilizando o algoritmo <i>k-NN</i> (o autor, 2018).....	162
Figura 4.27 – Série de treino da concentração horária de MP ₁₀ (de 1998 a 2009) e NO _x	

(de 1998 a 2000) na estação do Centro de São Paulo (o autor, 2018).....	167
Figura 4.28 – Atrator reconstruído a partir da série de treino da concentração horária de MP ₁₀ para os anos de 1998 a 2009 (Centro de São Paulo) (o autor, 2018).....	169
Figura 4.29 – Atrator reconstruído a partir da série de treino da concentração horária de NO _x para os anos de 1998 a 2000 (Centro de São Paulo) (o autor, 2018).....	169
Figura 4.30 – Séries temporais da concentração de MP ₁₀ (ano de 2010) e NO _x (ano de 2001) obtidas na estação de monitoramento do Centro de São Paulo para os meses de janeiro e fevereiro dos anos relativos as últimas aquisições de cada um dos poluentes (o autor, 2018).....	170
Figura 4.31 – Previsão <i>out of sample</i> da concentração horária de MP ₁₀ nos primeiros dois meses de 2010 na estação do Centro de São Paulo utilizando o algoritmo <i>k-NN</i> (o autor, 2018).....	171
Figura 4.32 – Previsão <i>out of sample</i> da concentração horária de NO _x nos primeiros dois meses de 2000 na estação do Centro de São Paulo utilizando o algoritmo <i>k-NN</i> (o autor, 2018).....	172
Figura 4.33 - Série de treino da concentração horária de MP ₁₀ dos anos de 1998 a 2009 na estação de Guarulhos (o autor, 2018).....	178
Figura 4.34 – Atrator reconstruído a partir da série de treino da concentração horária de MP ₁₀ para os anos de 1998 a 2008 (Guarulhos) (o autor, 2018).....	179
Figura 4.35 – Série temporal da concentração de MP ₁₀ obtida na estação de monitoramento de Guarulhos para os meses de janeiro e fevereiro de 2009 (o autor, 2018).....	180
Figura 4.36 – Previsão <i>out of sample</i> da concentração horária de MP ₁₀ nos primeiros dois meses de 2009 na estação de Guarulhos utilizando o algoritmo <i>k-NN</i> (o autor, 2018).....	181
Figura 4.37 – Séries de treino da concentração horária de MP ₁₀ e NO _x de 1998 a 2017 na estação de Cubatão – Centro (o autor, 2018).....	185
Figura 4.38 – Atrator reconstruído a partir da série de treino da concentração horária de MP ₁₀ para os anos de 1998 a 2017 (Cubatão - Centro) (o autor, 2018).....	186
Figura 4.39 – Atrator reconstruído a partir da série de treino da concentração horária de NO _x para os anos de 1998 a 2017 (Cubatão - Centro) (o autor, 2018).....	186
Figura 4.40 – Séries temporais da concentração de MP ₁₀ e NO _x obtidas na estação de monitoramento de Cubatão para os meses de janeiro e fevereiro de 2018 (o autor, 2018)..	187

Figura 4.41 – Previsão <i>out of sample</i> da concentração horária de MP ₁₀ nos primeiros dois meses de 2018 na estação de Cubatão - Centro utilizando o algoritmo <i>k-NN</i> (o autor, 2018).....	188
Figura 4.42 – Previsão <i>out of sample</i> da concentração horária de NO _x nos primeiros dois meses de 2018 na estação de Cubatão - Centro utilizando o algoritmo <i>k-NN</i> (o autor, 2018).....	189
Figura 4.43 – Série de treino (com lacunas) da concentração horária de MP ₁₀ de 1998 a 2017 na estação de Congonhas (o autor, 2018).....	200
Figura 4.44 – Série de treino (com lacunas) da concentração horária de NO _x de 1998 a 2017 na estação de Cubatão (o autor, 2018).....	200
Figura 4.45 – Série de treino da concentração horária de MP ₁₀ de 1998 a 2017 na estação de Congonhas, com lacunas preenchidas pelo método <i>k-NN</i> (o autor, 2018).....	201
Figura 4.46 – Série de treino da concentração horária de NO _x de 1998 a 2017 na estação de Cubatão, com lacunas preenchidas pelo método <i>k-NN</i> (o autor, 2018).....	201
Figura 4.47 – Atrator reconstruído a partir da série de treino da concentração horária de MP ₁₀ , com lacunas preenchidas por meio da metodologia <i>k-NN</i> , para os anos de 1998 a 2017 (Congonhas) (o autor, 2018).....	203
Figura 4.48 – Atrator reconstruído a partir da série de treino da concentração horária de NO _x , com lacunas preenchidas por meio da metodologia <i>k-NN</i> , para os anos de 1998 a 2017 (Cubatão - Centro) (o autor, 2018).....	203
Figura 4.49 – Previsão <i>out of sample</i> da concentração horária de MP ₁₀ nos primeiros dois meses de 2018 na estação de Congonhas utilizando o algoritmo <i>k-NN</i> e série de treino interpolada (o autor, 2018).....	205
Figura 4.50 – Previsão <i>out of sample</i> da concentração horária de NO _x nos primeiros dois meses de 2018 na estação de Cubatão utilizando o algoritmo <i>k-NN</i> e série de treino interpolada (o autor, 2018).....	206
Figura A1 – Série temporal do sistema de Lorenz para a componente <i>x</i> versus o tempo (s) (o autor, 2018).....	233
Figura A2 – Solução do atrator de Lorenz (o autor, 2018).....	233
Figura A3 – Série temporal de Hénon para a componente <i>x</i> versus o tempo (s) (o autor, 2018).....	234
Figura A4 – Solução do mapa de Hénon (o autor, 2018).....	235
Figura B1: Série histórica de concentração de MP ₁₀ de 1998 a 2017 obtida na estação de	

Congonhas com lacunas preenchidas por meio do método ARIMA (o autor, 2018).....	236
Figura B2: Séries temporais resultantes das previsões da concentração de MP ₁₀ na estação de Congonhas para <i>k</i> variando de 2 a 10 utilizando metodologia ARIMA no preenchimento dos dados faltantes (o autor, 2018).....	239
Figura C1 – Preenchimento das séries temporais utilizando o método dos vizinhos mais próximos, <i>k-NN</i> (o autor, 2018).....	241
Figura D1 – Comparação das previsões <i>out of sample</i> da concentração horária de MP ₁₀ com as séries originais nos primeiros dois meses de 2018 na estação do Ibirapuera utilizando o algoritmo <i>k-NN</i> para <i>k</i> variando de 2 a 10 (o autor, 2018).....	243
Figura D2 – Comparação das previsões <i>out of sample</i> da concentração horária de NO _x com as séries originais nos primeiros dois meses de 2018 na estação do Ibirapuera utilizando o algoritmo <i>k-NN</i> para <i>k</i> variando de 2 a 10 (o autor, 2018).....	244
Figura D3 – Comparação das previsões <i>out of sample</i> da concentração horária de MP ₁₀ com as séries originais nos primeiros dois meses de 2018 na estação de Congonhas utilizando o algoritmo <i>k-NN</i> para <i>k</i> variando de 2 a 10 (o autor, 2018).....	245
Figura D4 – Comparação das previsões <i>out of sample</i> da concentração horária de NO _x com as séries originais nos primeiros dois meses de 2018 na estação de Congonhas utilizando o algoritmo <i>k-NN</i> para <i>k</i> variando de 2 a 10 (o autor, 2018).....	247
Figura D5 – Comparação das previsões <i>out of sample</i> da concentração horária de MP ₁₀ com as séries originais nos primeiros dois meses de 2010 na estação do Centro de São Paulo utilizando o algoritmo <i>k-NN</i> para <i>k</i> variando de 2 a 10 (o autor, 2018).....	248
Figura D6 – Comparação das previsões <i>out of sample</i> da concentração horária de NO _x com as séries originais nos primeiros dois meses de 2000 na estação do Centro de São Paulo utilizando o algoritmo <i>k-NN</i> para <i>k</i> variando de 2 a 10 (o autor,2018).....	249
Figura D7 – Comparação das previsões <i>out of sample</i> da concentração horária de MP ₁₀ com as séries originais nos primeiros dois meses de 2009 na estação de Guarulhos utilizando o algoritmo <i>k-NN</i> para <i>k</i> variando de 2 a 10 (o autor, 2018).....	250
Figura D8 – Comparação das previsões <i>out of sample</i> da concentração horária de MP ₁₀ com as séries originais nos primeiros dois meses de 2018 na estação de Cubatão - Centro utilizando o algoritmo <i>k-NN</i> para <i>k</i> variando de 2 a 10 (o autor, 2018).....	252
Figura D9 – Comparação das previsões <i>out of sample</i> da concentração horária de NO _x com as séries originais nos primeiros dois meses de 2018 na estação de Cubatão - Centro utilizando o algoritmo <i>k-NN</i> para <i>k</i> variando de 2 a 10 (o autor, 2018).....	253

LISTA DE TABELAS

Tabela 2.1 – <i>Classificação das estações em termos de uso do solo e população exposta (CETESB, 2004)</i>	36
Tabela 2.2 – <i>Classificação das estações em termos de representatividade das estações (CETESB, 2004)</i>	37
Tabela 2.3 – <i>Expoentes de Lyapunov característicos a cada tipo de atrator (Monteiro, 2011)</i>	49
Tabela 2.4 – <i>Diferentes modelos, técnicas e aplicações de previsão da poluição do ar encontrados em literatura (Bai et al., 2018)</i>	65
Tabela 2.5 – <i>Áreas de aplicação da metodologia k-NN (o autor, 2018)</i>	69
Tabela 4.1 – <i>Médias anuais e desvios padrão da concentração de MP₁₀ e NO_x na estação do Ibirapuera (o autor, 2018)</i>	88
Tabela 4.2 – <i>Invariantes do Caos para os poluentes MP₁₀ e NO_x avaliados na estação de monitoramento do Ibirapuera dos anos de 1998 a 2017 (o autor, 2018)</i> ...	98
Tabela 4.3 – <i>Médias anuais e desvios padrão da concentração de MP₁₀ e NO_x na estação de Congonhas (o autor, 2018)</i>	103
Tabela 4.4 – <i>Invariantes do Caos para os poluentes MP₁₀ e NO_x avaliados na estação de monitoramento de Congonhas dos anos de 1998 a 2017 (o autor, 2018)</i> ..	114
Tabela 4.5 – <i>Médias anuais e desvios padrão da concentração de MP₁₀ e NO_x na estação do Centro de São Paulo (o autor, 2018)</i>	116
Tabela 4.6 – <i>Invariantes do Caos para os poluentes MP₁₀ e NO_x avaliados na estação de monitoramento do Centro de São Paulo dos anos de 1998 a 2010 (o autor, 2018)</i>	122
Tabela 4.7 – <i>Médias anuais e desvios padrão da concentração de MP₁₀ na estação de Guarulhos (o autor, 2018)</i>	125
Tabela 4.8 – <i>Invariantes do Caos para o MP₁₀ avaliados na estação de monitoramento de Guarulhos dos anos de 1998 a 2009 (o autor, 2018)</i>	130
Tabela 4.9 – <i>Médias anuais e desvios padrão da concentração de MP₁₀ e NO_x na estação de Cubatão (o autor, 2018)</i>	132
Tabela 4.10 – <i>Invariantes do Caos para os poluentes MP₁₀ e NO_x avaliados na estação de monitoramento de Cubatão dos anos de 1998 a 2017 (o autor, 2018)</i>	143
Tabela 4.11 - <i>Comparação entre as séries simuladas e reais para a concentração</i>	

<i>de MP₁₀ – Ibirapuera (o autor, 2018)</i>	153
Tabela 4.12 - <i>Comparação entre as séries simuladas e reais para a concentração de NO_x – Ibirapuera (o autor, 2018)</i>	154
Tabela 4.13 - <i>Comparação entre as séries simuladas e reais para a concentração de MP₁₀ – Congonhas (o autor, 2018)</i>	165
Tabela 4.14 - <i>Comparação entre as séries simuladas e reais para a concentração de NO_x – Congonhas (o autor, 2018)</i>	166
Tabela 4.15 - <i>Comparação entre as séries simuladas e reais para a concentração de MP₁₀ – Centro de São Paulo (o autor, 2018)</i>	176
Tabela 4.16 - <i>Comparação entre as séries simuladas e reais para a concentração de NO_x – Centro de São Paulo (o autor, 2018)</i>	177
Tabela 4.17 - <i>Comparação entre as séries simuladas e reais para a concentração de MP₁₀ – Guarulhos (o autor, 2018)</i>	183
Tabela 4.18 - <i>Comparação entre as séries simuladas e reais para a concentração de MP₁₀ – Cubatão - Centro (o autor, 2018)</i>	193
Tabela 4.19 - <i>Comparação entre as séries simuladas e reais para a concentração de NO_x – Cubatão - Centro (o autor, 2018)</i>	194
Tabela 4.20 - <i>Melhor valor de k para cada uma das estações e poluentes (o autor, 2018)</i>	196
Tabela 4.21 - <i>Previsões para k = 3 mantendo a dimensão de imersão de Grassberger e Procaccia (1983) (o autor, 2018)</i>	196
Tabela 4.22 - <i>Previsões para k = 3 e I = 13 (o autor, 2018)</i>	198
Tabela 4.23 – <i>Análise do Caos para as séries de treino originais e com preenchimento por meio da metodologia k-NN para as estações Congonhas (MP₁₀) e Cubatão (NO_x) (o autor, 2018)</i>	204
Tabela 4.24 - <i>Comparação entre as séries simuladas e reais para a concentração de MP₁₀ – Congonhas, utilizando a série de treino interpolada (o autor, 2018)</i>	209
Tabela 4.25 - <i>Comparação entre as séries simuladas e reais para a concentração de NO_x – Cubatão, utilizando a série de treino interpolada (o autor, 2018)</i>	210
Tabela A1 – <i>Comparação dos valores de entropia de Kolmogorov, dimensão de correlação e máximo expoente de Lyapunov obtidos por meio do BRChaos com os dados de literatura (o autor, 2018)</i>	235
Tabela B1: <i>Resultados das previsões da concentração de MP₁₀ na estação de</i>	

Congonhas utilizando a série histórica com lacunas preenchidas por meio do método ARIMA para k variando de 2 a 10 nas situações de 60, 30, 15, 7, 2 e 1 dia de previsão (o autor, 2018)..... 238

SUMÁRIO

CAPÍTULO I

INTRODUÇÃO	24
1.1 OBJETIVOS.....	26

CAPÍTULO II

REVISÃO BIBLIOGRÁFICA	28
2.1 POLUIÇÃO ATMOSFÉRICA.....	28
2.2 POLÍTICAS PÚBLICAS PARA O CONTROLE DA POLUIÇÃO ATMOSFÉRICA NO ESTADO DE SÃO PAULO.....	30
2.3 ESTAÇÕES DE MONITORAMENTO DA CETESB.....	35
2.3.1 Estação de Congonhas.....	37
2.3.2 Estação do Ibirapuera.....	39
2.3.3 Estação de Guarulhos	40
2.3.4 Estação de Cubatão – Centro.....	41
2.3.5 Estação do Centro de São Paulo.....	42
2.4 TEORIA DO CAOS.....	43
2.4.1 Atratores (<i>Bacia de Atração</i>).....	45
2.4.2 Dimensão de Correlação.....	47
2.4.3 Entropia de Kolmogorov.....	48
2.4.4 Expoentes de Lyapunov.....	49
2.5 APLICAÇÃO DA TEORIA CAÓTICA DETERMINISTA NO ESTUDO DA POLUIÇÃO ATMOSFÉRICA.....	53
2.6 MÉTODOS DE PREVISÃO DE SÉRIES TEMPORAIS.....	55
2.6.1 Métodos estatísticos.....	55
2.6.1.1 Médias móveis.....	56
2.6.1.2 Suavização exponencial simples.....	57

2.6.1.3 Suavização exponencial de Holt.....	58
2.6.1.4 Suavização exponencial de Holt-Winters.....	59
2.6.1.5 Modelos AR, ARMA, ARIMA.....	60
2.6.1 <i>Métodos de aprendizado de máquina</i>	62
2.6.2.1 <i>k-NN</i> – Atual cenário da predição de séries temporais de poluentes atmosféricos.....	64
2.6.2.1.1 – O método <i>k-NN</i> e a Teoria do Caos.....	71
 CAPÍTULO III	
MATERIAIS E MÉTODOS	74
 CAPÍTULO IV	
RESULTADOS E DISCUSSÕES	86
4.1 ESTAÇÃO DO IBIRAPUERA.....	86
4.2 ESTAÇÃO DE CONGONHAS.....	102
4.3 ESTAÇÃO DO CENTRO DE SÃO PAULO.....	115
4.4 ESTAÇÃO DE GUARULHOS.....	124
4.5 ESTAÇÃO DE CUBATÃO.....	131
4.6 PREVISÕES.....	144
4.6.1 <i>Estação de Ibirapuera</i>	145
4.6.2 <i>Estação de Congonhas</i>	156
4.6.3 <i>Estação do Centro de São Paulo</i>	167
4.6.4 <i>Estação de Guarulhos</i>	178
4.6.5 <i>Estação de Cubatão - Centro</i>	184
4.6.6 <i>Generalização de k (vizinhos próximos) e I (dimensão de imersão)</i>	195
4.6.7 <i>Proposta de melhoria das séries de treino</i>	199

CAPÍTULO V

CONCLUSÕES..... 212

CAPÍTULO VI

REFERÊNCIAS BIBLIOGRÁFICAS..... 215

APÊNDICE A..... 232

APÊNDICE B..... 236

APÊNDICE C..... 241

APÊNDICE D..... 243

CAPÍTULO I

INTRODUÇÃO

Cada vez mais se tem verificado o aumento da preocupação de órgãos ambientais em monitorar e controlar os níveis de poluentes atmosféricos, pois a relação entre o aumento da concentração destes compostos na atmosfera e a incidência de diversos problemas de saúde da população é evidente (Martin e Bradley, 1960; Saldiva *et al.*, 1995; Duchiadé, 1992; Braga *et al.*, 1999), sendo a poluição um grave problema de saúde pública. Nas metrópoles, a frota de veículos automotores e as crescentes emissões de poluentes tóxicos pelas chaminés industriais, conforme mencionado por Derisio (1992), provocam altas concentrações de substâncias nocivas que são responsáveis pela baixa visibilidade e diversos problemas respiratórios (Amdur *et al.*, 1991; Dockery *et al.*, 1993; Pope *et al.*, 2002). Com o advento da Revolução Industrial pouco se fez para controlar ou estudar esses efeitos, sendo os episódios de inversão térmica, relacionados a eventos meteorológicos, os responsáveis por impulsionar a comunidade científica a averiguar, certificar e relacionar os índices de mortalidade em centros urbanos à poluição atmosférica (Dockery e Pope, 1994). Movidos pela crescente necessidade de controle destas emissões, pesquisadores do mundo inteiro, muitas vezes financiados por agências governamentais, esforçam-se na compreensão da dispersão das diversas categorias de poluentes e no desenvolvimento de indicadores de qualidade do ar. Distintas abordagens, considerando as combinações lineares das interações entre os poluentes atmosféricas, foram aplicadas (Hellebust *et al.*, 2010; Guo, 2011) e, desta forma, negligenciaram as complexas interações não lineares presentes no fenômeno.

Na intenção de estudar a dispersão de poluentes atmosféricos, alguns trabalhos aplicam a *Análise Fatorial Dinâmica* (DFA) como ferramenta para decompor uma série temporal multivariada em combinações lineares dos fatores majoritários e suas variações temporais comuns associadas (Yu *et al.*, 2013). Contudo, analisar as características

estocásticas e suas relações associadas não permite avaliar a inteireza de todas as, já solidificadas, relações dinâmicas não lineares presentes nos sistemas atmosféricos (Lorenz, 1963; Tsonis, 1992). Sendo assim, a análise de Caos determinístico surge como metodologia alternativa no entendimento destes fenômenos.

Nesta tendência, a presente Tese aplica a Teoria do Caos determinístico a dados de variação temporal de concentração horária dos poluentes atmosféricos MP_{10} (material particulado inalável) e NO_x objetivando, inicialmente, identificar características não lineares e posteriormente realizar a predição das séries temporais, associando, portanto, a teoria caótica determinista à um modelo preditivo que utiliza aprendizado de máquina (Inteligência Artificial).

Caos é o termo que designa a teoria de sistemas dinâmicos aplicada na investigação dos fenômenos cuja evolução temporal se enquadra entre acontecimentos determinísticos e estocásticos. Fenômenos temporais são, costumeiramente, estudados no domínio do tempo e da frequência, porém, aquelas que apresentam não linearidades, aperiodicidade a longo prazo, sensibilidade às condições iniciais e evolução determinística, características definitivas de um processo inerentemente caótico, exigem, muitas vezes, uma avaliação no domínio do espaço à luz da teoria caótica determinística.

Historicamente, o termo Caos aplicado aos fenômenos dinâmicos é obra de James York e Tien-Yien Li que, em 1975, publicaram o artigo “Period three implies chaos” Neste cenário, nasce a Teoria do Caos objetivando preencher a lacuna existente entre os sistemas periódicos, portanto determinísticos, e os aleatórios - estocásticos. São processos caóticos determinísticos aqueles que apresentam sensibilidade a pequenas perturbações, aperiodicidade e não linearidade (Savi, 2017), condições que perpetram incerteza na determinação de seu estado futuro (de Paula, 2010). A identificação destas características não é trivial, sendo necessário lançar mão de um conjunto de ferramentas matemáticas que permitam avaliar o grau de imprevisibilidade e complexidade do processo. Após a obtenção de uma série temporal reconstrói-se, em um espaço de fases, um atrator (também conhecido como bacia de atração), via de regra estranho, e de suas características topológicas abstrai-

se a entropia de Kolmogorov, dimensão de correlação e expoentes de Lyapunov, invariantes característicos que determinam o quão caótica é a série avaliada.

Após o início da década de 1960, quando Lorenz demonstrou a existência de determinismo caótico em um sistema de previsão meteorológica (Lorenz, 1963), foi que esta modalidade de análise de sinais passou a ser considerada para situações em que as metodologias convencionais (abordagem espectral, estatística, entre outras) não fossem apropriadas. Neste tocante, a poluição atmosférica se enquadra como um fenômeno subjacentemente elegível à aplicação deste método.

1.1 OBJETIVOS

O objetivo principal desta Tese é o de avaliar, à luz da Teoria do Caos determinístico, a presença de padrões existentes em dados de poluição atmosférica e, por meio de um modelo de inteligência artificial (abordagem *kNN*), realizar a predição de séries temporais obtidas em cinco estações de monitoramento pertencentes à Companhia Ambiental do Estado de São Paulo (CETESB):

- Ibirapuera
- Centro (São Paulo)
- Cubatão – Centro
- Guarulhos – Paço Municipal
- Congonhas

As estações foram escolhidas de tal maneira a cobrir importantes centros poluidores da região metropolitana de São Paulo.

As séries temporais analisadas foram diretamente exportadas da plataforma eletrônica da Companhia de Tecnologia de Saneamento Ambiental do Estado de SP (CETESB) - <http://www.cetesb.sp.gov.br/ar/qualidade-do-ar/32-qualar>. Todas as análises foram realizadas no intervalo de 1998 a 2017.

Os poluentes estudados foram a concentração de material particulado inalável (MP_{10}) e óxidos de nitrogênio (NO_x). Inicialmente foram obtidas as séries temporais dos poluentes e posteriormente foram reconstruídos os atratores, calculados os invariantes entropia de Kolmogorov, dimensão de correlação e o máximo expoente de Lyapunov e, por último, todos os dados foram avaliados no que diz respeito às não linearidades presentes na evolução temporal e também no que tange as características locais de onde estão instaladas as estações. A última etapa foi a realização de previsões das concentrações de MP_{10} e NO_x nos primeiros dois meses do ano de 2018.

A partir dos sinais de concentração de poluentes, foram utilizados os seguintes procedimentos para se alcançar os objetivos desta Tese:

1. Aprimorar o código BRChaos, incluindo o cálculo do máximo expoente de Lyapunov;
2. Aplicar a Teoria do Caos no estudo da evolução temporal da concentração anual de MP_{10} e NO_x em cinco estações da Companhia Ambiental do Estado de São Paulo;
3. Depois de determinados os parâmetros caóticos, utilizá-los em um modelo preditivo.
 - 3.1 – associar, por meio da dimensão de imersão e correlação, a Teoria do Caos a um modelo preditivo de aprendizado de máquina (Inteligência Artificial);
 - 3.2 – avaliar estatisticamente a qualidade da predição.

CAPÍTULO II

REVISÃO BIBLIOGRÁFICA

Neste capítulo são abordados os conceitos necessários para a compreensão do conteúdo desta Tese de Doutorado. Na primeira parte são apresentados os principais aspectos da Teoria do Caos, definindo-se atratores, atratores estranhos, entropia de Kolmogorov, dimensão de correlação e expoentes de Lyapunov. Em seguida, apresenta-se uma revisão geral de trabalhos que utilizaram a Teoria do Caos no estudo da poluição atmosférica e também o atual cenário dos modelos preditivos, bem como a associação do método de predição k-NN com a Teoria do Caos no que tange o cálculo da dimensão de imersão. Apresenta-se também a questão da poluição atmosférica como um grave problema de saúde pública, suas consequências e as principais políticas governamentais que objetivam a redução dos níveis dos poluentes atmosféricos.

2.1 POLUIÇÃO ATMOSFÉRICA

Pode-se relacionar a poluição do ar com toda e qualquer emissão antropogênica, biogênica e geogênica que, de forma adversa, altera a composição química atmosférica (Daly e Zannetti, 2007). Dois marcos das emissões atmosféricas foram a Revolução Industrial e o desenvolvimento e popularização do automóvel, ambos acarretando em alterações significativas na composição química da atmosfera pelo lançamento de diversos compostos nocivos, a curto e longo prazo, diretamente no ar por meio das chaminés industriais e escapamentos automotivos. Desta forma, postula-se que a poluição do ar teve início no momento em que o homem começou a queimar combustíveis (Daly e Zannetti, 2007).

Os poluentes atmosféricos são divididos em duas categorias, a saber:

- *Poluentes primários;*
- *Poluentes secundários.*

São considerados poluentes primários àqueles lançados, por determinada fonte, diretamente na atmosfera. Os exemplos mais comuns são o monóxido de carbono, fuligem, óxidos de nitrogênio, óxidos de enxofre, hidrocarbonetos, aldeídos entre outros.

Definem-se poluentes secundários como todos os compostos formados, dadas as necessárias condições reacionais, a partir dos primários, ou seja, estes não são emitidos por nenhuma fonte, mas produzidos no próprio ambiente atmosférico. Como exemplo, pode-se citar a decomposição de óxidos de nitrogênio pela radiação ultravioleta oriunda do Sol na formação de ozônio e nitratos de peroxiacetila.

Muitos estudos têm demonstrado os efeitos negativos da poluição do ar na saúde de cidadãos residentes em centros urbanos, como a Região Metropolitana de São Paulo (Braga *et al.*, 2001; Martins *et al.*, 2002). As concentrações de particulados inaláveis (MP₁₀) merecem especial atenção neste cenário, pois, os efeitos de seus níveis atmosféricos, podem ocasionar desde tosse seca, cansaço, ardor nos olhos, nariz e garganta até sérios riscos de manifestação de doenças respiratórias e cardiovasculares e aumento das mortes prematuras em pessoas de grupos sensíveis (crianças, idosos e acometidos de doenças do aparelho respiratório e cardíaco) (CETESB, 2014). São considerados particulados inaláveis (MP₁₀) todas as partículas dispersas na atmosfera com diâmetro menor que 10 µm e não correspondem a nenhum poluente específico.

O Estado de São Paulo abriga mais de 40 milhões de habitantes, sendo que metade deste número corresponde aos moradores da região metropolitana. O estado conta ainda com a maior frota de veículos do país e contribui com cerca de 30% de todo o produto interno bruto (PIB) brasileiro (Vieira-Filho *et al.*, 2015), contando, para tanto, com regiões altamente industrializadas como as de Cubatão e do ABCM (Santo André, São Caetano do Sul, São Bernardo do Campo e Mauá). Desta feita, é essencial o conhecimento dos níveis, causas e consequências da poluição atmosférica no estado.

Vieira-Filho *et al.* (2015) avaliaram a influência do local das fontes emissoras (Cubatão e cidade de São Paulo) bem como da topografia em dados de qualidade do ar e na composição da água da chuva. O estudo mostrou que a qualidade do ar e da chuva em Cubatão são bem inferiores e bastante preocupantes quando comparados à região metropolitana de São Paulo. Os autores associaram esta conclusão ao fato de que as baixas velocidades do vento associadas à topografia local e circulação da brisa do mar, tornam Cubatão um local de baixa dispersão, portanto, alta concentração, de poluentes atmosféricos.

Carvalho *et al.* (2015) se propuseram a estudar os níveis de qualidade do ar, na região metropolitana de São Paulo, em face da implementação do Programa Brasileiro de Controle de Emissões Veiculares em 1986. O estudo constatou uma tendência anual de redução de todos os poluentes, exceto o ozônio, porém é imperativo que esses níveis sejam ainda mais reduzidos. Os autores finalizam recomendando a implementação de melhores sistemas de transporte público, restrição de padrões para os veículos que funcionam com gasolina, etanol e diesel e incentivos econômicos para o uso de tecnologias limpas, demonstrando assim a necessidade da continuidade de estudos que objetivam interpretar as causas e efeitos da poluição atmosférica.

2.2 POLÍTICAS PÚBLICAS PARA O CONTROLE DA POLUIÇÃO ATMOSFÉRICA NO ESTADO DE SÃO PAULO

Segundo dados da Organização Mundial da Saúde em 2015, após monitoramento de 193 cidades do mundo, São Paulo se enquadra como a 6^a metrópole mais poluída (toneladas/ano) do Planeta, perdendo apenas para Pequim (China), Cairo (Egito), Jacarta (Indonésia) e Los Angeles (EUA), sendo que em primeiro lugar encontra-se Cidade do México, localizada em uma região cuja geografia dificulta muito a dispersão dos poluentes atmosféricos, causando, diariamente, 17 mortes.

Desde a Revolução Industrial, a poluição atmosférica, causada, em parte, pelas chaminés industriais, é indicativo do progresso de uma cidade (Cedec, 1997), contudo, na

Região Metropolitana de São Paulo, diferente de quase todos os grandes centros urbanos, após a migração dos polos industriais para áreas menos custosas, a fonte mais contundente de poluentes são os veículos automotores (Menezes *et al.*, 2015). Os gases emitidos pelos escapamentos dos 15,3 milhões de automóveis, estimados somente no Estado de São Paulo (PCPV 2017 – 2019, 2017), contribuem para o aumento do efeito estufa, destruição da camada de ozônio, diversos danos às mais variadas espécies e também alterações climáticas (Menezes *et al.*, 2015). Somente no estado paulista, a estimativa, alarmante, é de que no ano de 2015, a poluição do ar tenha causado 11.200 mortes precoces, associadas, principalmente, à inalação de particulados, sendo que, no mundo todo, a OMS estima a perda de cerca de 8 milhões de vidas pela poluição do ar. Um olhar cuidadoso nos relatórios CETESB revela que os níveis anuais de particulados inaláveis em muitas das estações de monitoramento do estado, superam de 2 a 5 vezes os valores recomendados pela OMS (Vormittag e Saldiva, 2015).

Sendo assim, fica evidente a necessidade de políticas públicas para o controle e monitoramento da poluição atmosférica. Neste contexto, observa-se a Resolução CONAMA Nº 491 de 19/11/2018 que dispõe sobre os padrões de qualidade do ar e o Decreto Paulista 59.113 de 23/04/2013 que, além de determinar novos padrões de qualidade do ar, ainda estabelece algumas providências. Sob o prisma das resoluções internacionais, tanto o Decreto Paulista 2013 quanto a Resolução CONAMA Nº 491 apresentam níveis de tolerância a, por exemplo, MP_{10} , muito superiores ao sugerido pela OMS em 2005. Para a Organização Mundial da Saúde as médias máximas diárias e anuais de MP_{10} devem ser, respectivamente, de 50 e 20 $\mu\text{g}/\text{m}^3$. Já na CONAMA Nº 491 os níveis máximos são 120 $\mu\text{g}/\text{m}^3$ (média diária) e 40 (média anual) ao passo que no Decreto Paulista 2013 os valores vigentes são, também, 120 $\mu\text{g}/\text{m}^3$ e 40 $\mu\text{g}/\text{m}^3$ para a média diária e anual (Marques e Santos, 2012). Se forem observadas outras agências reguladoras ao redor do mundo a Agência de Proteção ao Meio Ambiente dos Estados Unidos da América estipula que a média de MP_{10} não pode ultrapassar mais que uma vez no ano o valor de 150 $\mu\text{g}/\text{m}^3$, já a Comissão Europeia do Meio Ambiente preconiza a concentração anual de 50 $\mu\text{g}/\text{m}^3$ para o MP_{10} (Marques e Santos, 2012).

Dentre as iniciativas públicas, tanto federais quanto estaduais/municipais, pode-se salientar o Programa de Controle da Poluição do Ar por Veículos Automotores (PROCONVE), implementado em 1980 pelo Governo Federal Brasileiro que especificou mudanças tecnológicas nos veículos automotores objetivando reduzir cada vez mais as emissões de poluentes (Carvalho *et al.*, 2015). Em 2009, outro programa de iniciativa federal foi a redução da concentração de enxofre presente nos combustíveis (Cortes, Souza e Cuevas, 2015). No que diz respeito ao Estado de São Paulo, salienta-se o programa de inspeção veicular que teve início em 2009 com término em 2014 e, ainda, em 2007, a proibição da circulação de caminhões de 2ª a 6ª feira das 5h às 21h e aos sábados das 10h às 14h. Essa iniciativa promoveu redistribuição das emissões com menor concentração durante o período diurno e também obrigou que cargas, antes transportadas em veículos pesados, fossem transportadas em veículos menores e, portanto, menos poluentes (Cortes, Souza e Cuevas, 2015). Ainda neste tocante, o rodízio veicular não se destaca como resolução impactante na redução de poluentes atmosféricos (Han, Yang, & Wang, 2010), pois, devido ocorrer em períodos mais curtos, as emissões são apenas proteladas por algumas horas (Cortes, Souza e Cuevas, 2015), surtindo efeito real apenas na fluidez do tráfego.

Em respeito ao que preconiza a Resolução 418/2009 do Conselho Nacional de Meio Ambiente, o Governo do Estado de Paulo, por meio da Secretaria de Meio Ambiente, lançou em 2017 o Plano de Controle de Poluição Veicular 2017 – 2019. Nele, observam-se as principais ações de controle e recomendações de políticas públicas para redução da emissão e melhoria da qualidade do ar, a saber:

- *Inspeção ambiental de veículos*: é uma inspeção técnica periódica, associada ao licenciamento anual do automóvel, em que são avaliados, entre outras coisas, os níveis de emissão do motor. Trata-se de importante ferramenta no controle de emissões de veículos que teve início no ano de 2008 e, por problemas licitatórios, foi suspensa em 2014. Mesmo prevista no Código de Trânsito Brasileiro e com implementação nacional agendada para o ano de 2019, conforme pautado pelo Conselho Nacional de Trânsito (CONTRAN) pela Resolução Nº 716 de 30/11/2017, o Departamento Nacional de Trânsito operou a suspensão

da resolução sob a justificativa da necessidade de debates “mais profundos de modo a aplicar da melhor maneira possível, com o mínimo de transtornos à população”.

- *Fiscalização de fumaça dos veículos diesel com a Escala de Ringelmann*: utiliza a “Escala de Ringelmann Reduzida” para monitorar a emissão de fumaça preta acima do limite legal.

- *Operação inverno*: trata-se da intensificação das ações de cunho fiscalizatório e educativo no período de inverno (junho a agosto), uma vez que esta é a época do ano cujas condições meteorológicas desfavorecem a dispersão dos poluentes, principalmente os materiais particulados, devido ao fenômeno da inversão térmica. Esta é uma medida adotada desde 1984 com resultados expressivos.

- *Fiscalização com uso do opacímetro*: muito mais rigoroso que a Escala de Ringelmann, o opacímetro é um instrumento óptico que mede a quantidade de fumaça emitida por um veículo a diesel. No relatório sobre as atividades da Operação de Inverno 2005, a CETESB demonstrou que, quando esta tecnologia é empregada, cerca de 40% da frota apresenta desconformidades.

- *Fiscalização de ARLA 32*: o ARLA 32 é uma solução de 32% de ureia em água desmineralizada utilizada junto com o sistema de Redução Catalítica Seletiva (SCR) na redução química de óxidos de nitrogênio resultantes da combustão. O ARLA 32 é frequentemente fraudado pelos proprietários dos veículos que utilizam essa tecnologia, fazendo com que a fumaça emitida pelo automóvel exceda os limites permitidos por legislação.

- *Expansão do Programa de Melhoria da Manutenção de Veículos a Diesel – PMMVD*: trata-se de uma rede de sete controladoras que visam realizar serviços de acordo com os padrões de qualidade, realizando a aplicação das medições de opacidade, orientação aos proprietários de veículos movidos a diesel e etc.

- *Incentivo à gestão ambiental de frotas e garagens*: sugere, basicamente, que a CETESB realize estudos de modo a incentivar que empresas que possuam grandes frotas de veículos e/ou atuem no ramo do transporte adotem práticas de gestão ambiental. Para tal, o Governo do Estado deve disponibilizar a Cartilha de Gestão Ambiental de Frotas e Garagens, com orientações para empresas que tenham interesse na sustentabilidade.

- *Aperfeiçoamentos do PROCONVE e PROMOT*: o Programa de Controle da Poluição do Ar por Veículos Automotores (PROCONVE) e o Programa de Controle da Poluição do Ar por Motociclos e Veículos Similares (PROMOT), ambos de iniciativa do Governo Federal, instituídos, respectivamente em 1986 e 2003, têm por objetivo a redução tanto de poluentes atmosféricos quanto de ruídos dos veículos automotores comercializados no Brasil. Entre as principais funções destes programas está a determinação de limites de emissão de poluentes e o controle da poluição do ar.

- *Laboratórios de emissões veiculares*: a CETESB possui, atualmente, uma estrutura bem ampla com capacidade para atender às demandas do PROCONVE por meio de dois laboratórios (um na cidade de SP, inaugurado em 1977 e, desde 2017, outro na cidade de São Bernardo do Campo). As metas estipuladas para essas duas instalações são de obtenção de diversas certificações e reconhecimento junto ao Inmetro para a realização, oficial, de ensaios de emissão de poluentes por veículos automotores novos (lançamentos), principalmente os veículos pesados movidos a diesel.

- *Capacitação*: o Governo do Estado, por meio da Escola Superior da CETESB, possui estrutura completa para oferecer periodicamente cursos variados na área de controle e poluição ambiental. A meta estipulada é de oferta anual dos cursos de “Emissão Veicular”, “Fiscalização de veículos em uso” e “Fiscalização de fumaça preta”.

Por meio das ações, adotadas pelo estado, nota-se que a necessidade de fiscalização e atualização dos parâmetros de controle da poluição atmosférica é constante. Desta forma, a academia contribui de forma decisiva na elaboração e desenvolvimento de novas formas de mensuração, análise e controle dos dados de variação/concentração destes poluentes.

2.3 ESTAÇÕES DE MONITORAMENTO DA CETESB

Desde sua fundação, a Companhia Ambiental do Estado de São Paulo (CETESB) edificou-se como a maior agência ambiental do Brasil e tornou-se referência para instituições internacionais, como a OMS (Organização Mundial da Saúde) e ONU (Organização das Nações Unidas) (CETESB, 2018). Atualmente, a rede de monitoramento de qualidade do ar dispõe, no Estado de SP, de 62 estações automáticas (61 fixas e uma móvel) e mais 27 pontos de monitoramento manual. Desta feita, a seguir, serão caracterizadas as estações objeto deste estudo.

A classificação ou caracterização de uma dada estação de monitoramento é o conjunto de informações que, por meio delas, é possível caracterizar a qualidade do ar da região em que a estação está inserida (CETESB, 2004). Dados como as fontes influenciadoras, a população exposta e a área de abrangência, são vitais para descrever adequadamente a estação. As Tabelas 2.1 e 2.2 foram extraídas do relatório “Caracterização das Estações da Rede Automática de Monitoramento da Qualidade do Ar na RMSP - Estação Congonhas” de 2004 e demonstram as classificações baseadas no uso do solo, população e representatividade das estações.

As classificações expostas nas Tabelas 2.1 e 2.2 são gerais de modo a considerar uma rede de monitoramento completa, mas cada poluente apresenta características individuais no que tange os limites legais, fontes e dispersão, predicados decisivos na determinação dos locais e escalas de representatividade ideais ao monitoramento. Isso quer dizer que, muitas vezes, um dado local não é oportuno para a medição de todo e qualquer tipo de poluente devido às características das fontes emissoras. Por exemplo, não tem sentido realizar medições de ozônio nas proximidades de vias de tráfego, uma vez que o O₃ é consumido quimicamente pelo NO emitido pelos veículos (CETESB, 2004).

Tabela 2.1 – Classificação das estações em termos de uso do solo e população exposta (CETESB, 2004).

Característica da estação	Descrição
Comercial	Mede a exposição da população em áreas urbanas centrais, áreas de comércio, com grande movimentação de pedestres e veículos.
Residencial	Mede a exposição da população em bairros residenciais e áreas suburbanas das cidades.
Industrial	Em áreas onde as fontes industriais têm grande influência nas concentrações observadas, tanto em longo prazo quanto para avaliação de picos de concentração.
Urbana/concentração de fundo (<i>background</i>)	Em áreas urbanas, localizadas não próximas de fontes específicas, representa as concentrações de fundo da área urbana como um todo.
Próxima de vias de tráfego (veicular)	Localizada próxima de uma via de tráfego, mede a influência da emissão dos veículos que circulam na via (rua, estrada etc.).
Rural	Mede as concentrações em áreas rurais, deve estar situada o mais distante possível de fontes veiculares, industriais e urbanas.
Ambiente fechado (<i>indoor</i>)	Mede as concentrações em ambientes domésticos e de trabalho (exceto ambientes ocupacionais).

Tabela 2.2 – *Classificação das estações em termos de representatividade das estações (CETESB, 2004).*

Escala de Representatividade	Área de Abrangência
Microescala	Concentrações abrangendo áreas de dimensão de poucos metros até 100 metros.
Média escala	Concentrações para blocos de áreas urbanas (poucos quarteirões com características semelhantes), com dimensões entre 100 e 500 metros.
Escala de bairro	Concentrações para áreas da cidade (bairros), com atividade uniforme, com dimensões de 500 a 4.000 metros.
Escala urbana	Concentrações de cidades ou regiões metropolitanas, da ordem de 4 a 50 km.
Escala regional	Concentrações geralmente de uma área rural, de geografia razoavelmente uniforme e de dimensões de dezenas a centenas de quilômetros.
Escalas nacional e global	Concentrações de um país e do planeta como um todo, respectivamente.

2.3.1 Estação de Congonhas

A estação de Congonhas está localizada na Alameda dos Tupiniquins, 1571 – Aeroporto (E.M. Prof. J. C. da Silva Borges), na cidade de São Paulo, altitude de 760 m, sob as coordenadas de Latitude 23°36'29" S (UTM (23K) 330 337) e Longitude 46° 39'37" W (UTM (23K) 7 387 309). Os parâmetros e poluentes monitorados nesta estação são: MP₁₀, SO₂, CO, NO_x e O₃. A Figura 2.1 traz a fotografia da estação de Congonhas.

Figura 2.1 – Estação de Congonhas (arquivo fotográfico QUALAR CETESB:
<http://qualar.cetesb.sp.gov.br>).



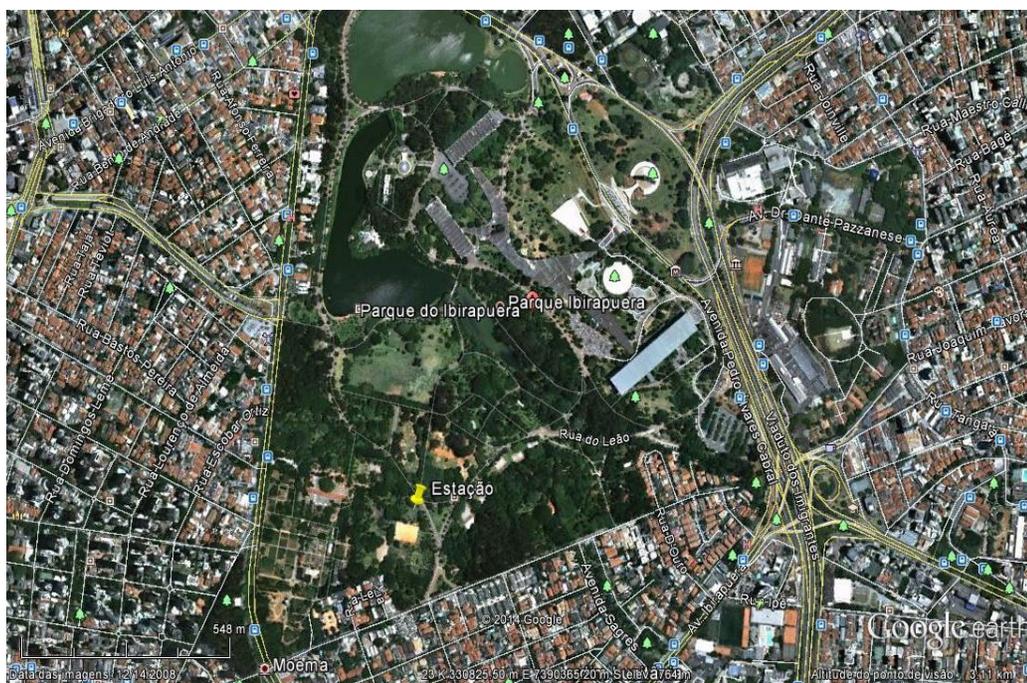
Em termos das características locais, a estação está instalada na Escola Municipal Professor J. C. da Silva Borges, uma localização predominantemente residencial e comercial, a poucos metros de vias de alto tráfego como a Avenida dos Bandeirantes e Alameda dos Tupiniquins. Outra informação importante é a proximidade com o Aeroporto de Congonhas (cerca de 400 metros). A sonda de amostragem capta amostras a 3,5 m e em seu entorno são verificadas algumas árvores.

Quanto às características das fontes emissores nos arredores da estação, não se verifica a presença de fontes fixas com potencial representativo, pois a área se encontra delimitada pelas restrições da lei de zoneamento municipal (CETESB, 2004). As poucas fontes fixas observadas são pizzarias, restaurantes e posto de abastecimento (CETESB, 2004). Em relação às fontes móveis, tem-se com a mais impactante o intenso tráfego das avenidas dos Bandeirantes, 23 de Maio, Ruben Berta, Moreira Guimarães, Washington Luiz e Alameda dos Tupiniquins.

2.3.2 Estação do Ibirapuera

A estação do Ibirapuera está localizada na Avenida Manoel da Nóbrega, 1985, na cidade de São Paulo, dentro do Parque do Ibirapuera (Setor 25), altitude de 750 m, sob as coordenadas de Latitude 23°34'55" S (UTM (23K) 330 592) e Longitude 46° 39'25" W (UTM (23K) 7 390 026). Os parâmetros e poluentes monitorados nesta estação são: fumaça, partículas totais em suspensão, MP_{2,5}, MP₁₀, SO₂, CO, NO_x, O₃, umidade, temperatura, velocidade e direção do vento, pressão, radiação global e radiação UVA. A altura de captação das amostras é 3,5 m. A Figura 2.1 traz a fotografia da estação do Ibirapuera

Figura 2.2 – Estação do Ibirapuera (arquivo fotográfico QUALAR CETESB:
<http://qualar.cetesb.sp.gov.br>).



No que se refere às características locais onde está instalada a estação do Ibirapuera, verifica-se uma região de área verde, urbanizada e de predominância residencial. Nos entornos, não são verificadas grandes vias de tráfego automotivo e, portanto, poucas ou quase nenhuma fonte veicular. Salienta-se como a via de tráfego mais intensa a Avenida

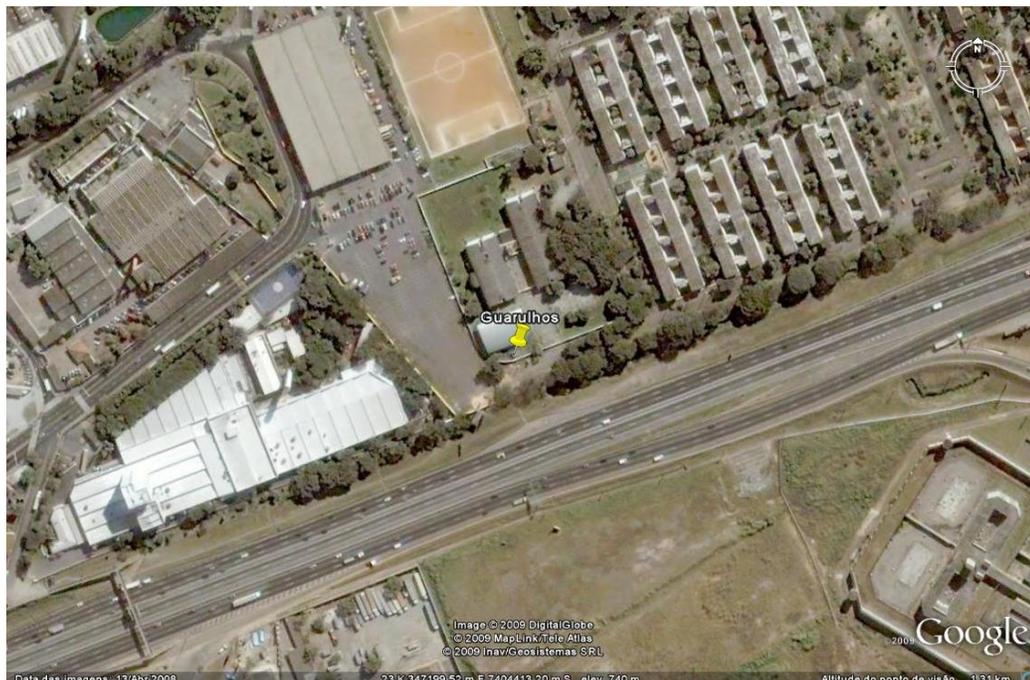
Quarto Centenário, localizada a 250 metros do ponto de amostragem. Fontes fixas também não são encontradas na região. De fato, a referida estação é bastante atípica em relação às demais encontradas em todo Estado.

O Parque do Ibirapuera é caracterizado por ser uma grande área verde ao redor de extensos corpos d'água. Não diferente, em volta da estação, notam-se diversas árvores plantadas nos bosques do parque e um campo de futebol de chão batido.

2.3.3 Estação de Guarulhos

A estação de Guarulhos está localizada na Rua Professora Maria Del Pilar Munhoz Bononato, s/nº, E. E. Francisco Antunes Filho, Parque Cecap, Guarulhos, altitude de 740 m, sob as coordenadas de Latitude 23°27'44" S (UTM (23K) 347 182) e Longitude 46°29'45" W (UTM (23K) 7404 407). A Figura 2.3 traz a fotografia da vista aérea da estação de Guarulhos.

Figura 2.3 – Estação de Guarulhos (arquivo fotográfico QUALAR CETESB: <http://qualar.cetesb.sp.gov.br>).



Nesta estação estão disponíveis apenas dados históricos de MP_{10} , SO_2 , velocidade do vento. A Figura 2.3 traz a fotografia da vista aérea da estação de Guarulhos.

A estação se localiza em uma região de intenso tráfego de veículos leves e pesados, estando próxima à Rodovia Presidente Dutra, interligação com a Rodovia Hélio Smidt e a Avenida Presidente Tancredo Neves. São observadas algumas unidades industriais em seu entorno, diversos estacionamentos, Shopping Center e o Aeroporto Internacional de São Paulo. Ao redor do ponto de amostragem, verifica-se a presença de algumas árvores e uma região descampada.

2.3.4 Estação de Cubatão – Centro

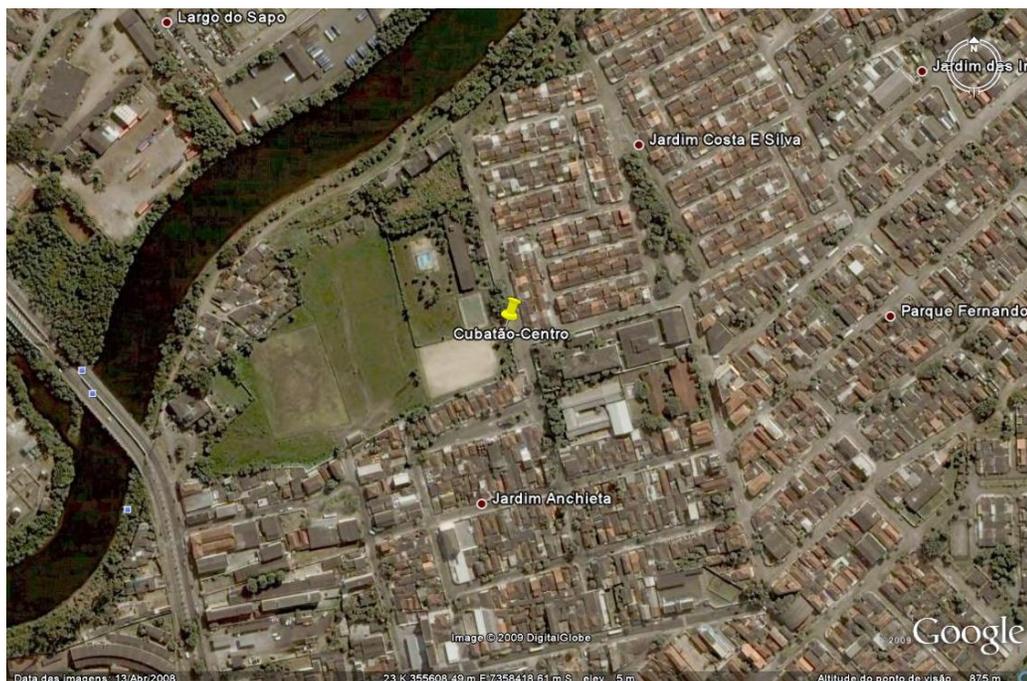
A estação de Cubatão Centro está localizada na Rua Salgado Filho, 121, Parque Fernando Jorge, Cubatão, altitude de 4 m, sob as coordenadas de Latitude $23^{\circ}52'48''$ S (UTM (23K) 355 580) e Longitude $46^{\circ}25'05''$ W (UTM (23K) 7 358 443).

Os parâmetros e poluentes monitorados nesta estação são: Benzeno, MP_{10} , NO, NO_x , NO_2 , O_3 , SO_2 , Tolueno, direção e velocidade dos ventos, pressão, umidade e temperatura. A Figura 2.4 traz a fotografia aérea da estação de Cubatão – Centro.

Nos entornos do bairro onde está localizada a estação, observa-se uma região cuja geografia é bastante acidentada, com montes e uma região verde bastante acentuada e há também a presença do Rio Cubatão, contudo, verifica-se a existência de ruas de alto fluxo de automóveis pesados e leves (Avenida 9 de Abril, Avenida Tiradentes).

Trata-se de uma área central da cidade, com residências, empresas como supermercados, padarias, restaurantes, bares, escritórios e pizzarias. A cidade abriga diversos parques industriais.

Figura 2.4 – Estação de Cubatão – Centro (arquivo fotográfico QUALAR CETESB: <http://qualar.cetesb.sp.gov.br>).

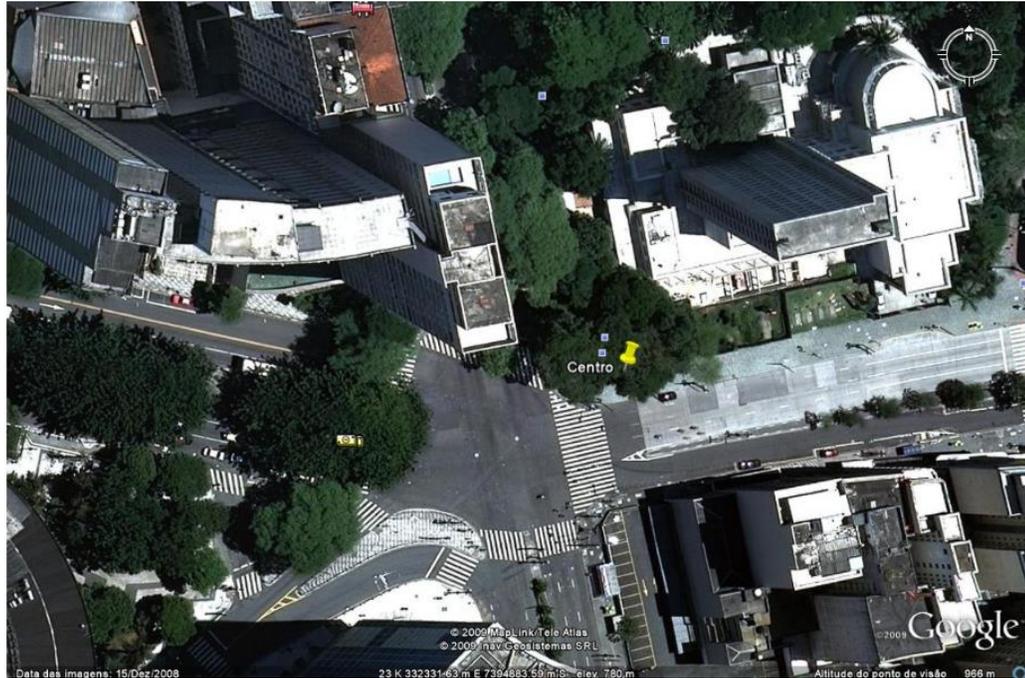


2.3.5 Estação do Centro de São Paulo

A estação do Centro de São Paulo está localizada na Rua da Consolação, 94, Biblioteca Municipal Mário de Andrade, Consolação, São Paulo, altitude de 780 m, sob as coordenadas de Latitude 23°32'50" S (UTM (23K) 332 354) e Longitude 46°38'31" W (UTM (23K) 7 394 876). Nesta estação estão disponíveis apenas dados históricos de: MP₁₀, SO₂, CO, NO_x, NO₂ e NO. A Figura 2.5 traz a fotografia da vista aérea da estação do Centro de São Paulo.

Esta estação está localizada em uma região considerada mista, residencial e comercial, tendo em seu entorno diversas vias de intenso tráfego automotivo de veículos leves, a Rua da Consolação, Avenida São Luis, Avenida Ipiranga e outras. No que diz respeito à atividade industrial local, é uma região onde não se verifica a presença de indústrias, apenas diversos prédios comerciais e empresas como padaria, restaurantes, pizzarias e etc.

Figura 2.5 – Estação do Centro de São Paulo (arquivo fotográfico QUALAR CETESB: <http://qualar.cetesb.sp.gov.br>).



2.4 TEORIA DO CAOS

As primeiras menções ao conceito de “Chaos” datam de aproximadamente 4.000 anos e associavam a palavra ao conceito de primitivo, “inicial originário”, sendo seu entendimento relacionado, na maioria das vezes, à desordem. Os gregos antagonizaram o Caos com o vocábulo “Cosmos” (κόσμος) que, na sua essência, indica a presença de ordem e disciplina (Gutiérrez, 2011).

Muito embora as definições filosóficas remetam a um estado de desorganização, o termo Caos foi aplicado por James York e Tien-Yien Li, em 1975, para descrever os fenômenos aleatórios, cuja dinâmica não linear pode ser explicada por leis deterministas criando assim o termo “Teoria do Caos”. A aparente casualidade dos sistemas dinâmicos passou a ser encarada como disfarce para um padrão de organização que se tornou alvo dos investigadores desta teoria, que demonstraram diversos sistemas cuja viabilização puramente determinista não é possível. Os pilares do determinismo estão assentados sob o

fato de que os eventos atuais são determinados por acontecimentos anteriores criando, uma vez preservadas as condições iniciais, uma conexão conhecida entre passado e futuro. Se considerado um sistema em que imperam as leis do Caos o resultado final será imprevisível não podendo ser obtido por nenhuma lei precedente (Gleick, 1987).

Assim, a análise de Caos é aplicada a processos dissipativos, com evolução temporal não linear, aperiódica a longo prazo, determinístico e extremamente sensível às condições iniciais (Savi, 2004). Um fenômeno que possua estas características certamente apresentará, durante sua evolução temporal, uma contração no volume do espaço de fases e suas linhas de fluxo dependerão das condições iniciais (Ruelle e Takens, 1971) definindo o chamado atrator estranho, atrator caótico ou bacia de atração.

Muito embora o estudo da dinâmica não linear por meio da avaliação do espaço de fases em Mecânica Celeste seja mérito de Henri-Poincaré, foi Edward Lorenz, em 1963, quem constatou a sensibilidade às condições iniciais de um modelo para previsão do tempo a longo prazo (Lorenz, 1963). O sistema representado pelas Equações 2.1 é chamado Sistema de Lorenz e foi proposto com o objetivo de descrever a transição de turbulência fraca para convecção natural (Choudhury, 1997).

As constantes φ e ε são respectivamente os números de Prandtl e Rayleigh, $\beta = 4/(1+a^2)$, as variáveis x , y e z possuem significados físicos, sendo x uma quantificadora do movimento convectivo, y é a diferença térmica entre o fluido ascendente e descendente e z refere-se à variação da temperatura na direção vertical.

$$\begin{cases} \frac{dx}{dt} = \varphi(x - y) \\ \frac{dy}{dt} = \varepsilon x - y - xz \\ \frac{dz}{dt} = xy - \beta z \end{cases} \quad (2.1)$$

Após as constatações de Lorenz, o interesse pela dinâmica não linear aumentou, pois se verificou a aplicação do *Caos Determinístico* em áreas como astronomia na investigação dos ciclos de manchas no Sol (Letellier *et al.*, 2006), na detecção de sinais biológicos como o eletroencefalograma e magnetoencefalografia (Stam, 2005; Dafilisa *et al.*, 2009), no estudo de predição biológica do crescimento de tumores (Andrey, 1989; El-Gohary, 2008), na identificação da evolução temporal de poluentes atmosféricos (Yu *et al.*, 2013), em hidrologia (Sivakumar *et al.*, 1999), criptografia (Wong *et al.*, 2009), na análise de séries de tempo financeiros (Erbano, 2004), na investigação de processos químicos e fluidomecânicos (van den Bleek e Schouten, 1993b; van den Bleek *et al.*, 2002; Li *et al.*, 1997; Cremasco *et al.*, 2011; Castilho, 2012; Villegas *et al.*, 2012; Moura, 2013; Prieto, 2014; Prieto e Cremasco, 2015) entre outros diversos exemplos. Com este advento, a difícil tarefa de identificação e caracterização dos sistemas caóticos passou a ser prioridade, pois se acreditava que no futuro estes fenômenos poderiam ser controlados por meio de parâmetros de controle. Desta forma, a técnica de análise de sinais com características não lineares, porém determinísticas, passou a contar com um conjunto de métodos que auxiliam no reconhecimento e estudo destes sistemas aplicando um conjunto de invariantes como a entropia de Kolmogorov-Sinai (Ruelle, 1989) e dimensão de correlação (McCauley, 1990), além dos expoentes de Lyapunov (Wolf *et al.*, 1985) e das características topológicas dos atratores no espaço de fases.

2.4.1 Atratores (Bacia de Atração)

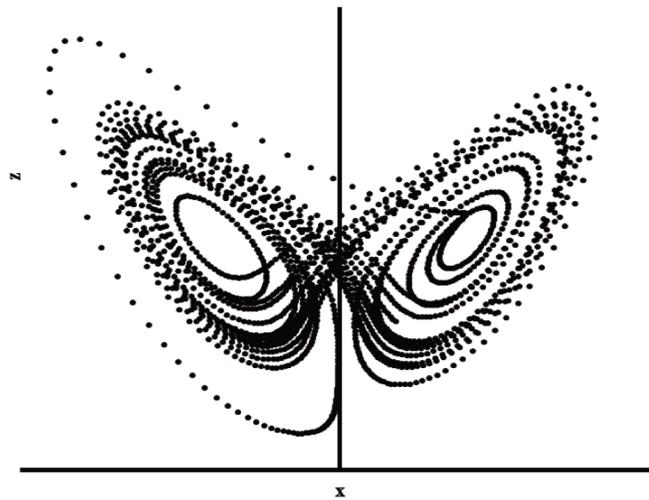
Atrator (também chamado de bacia de atração) é o nome dado à região do espaço de fases para o qual evoluem, passado um tempo infinito, as soluções (ou trajetórias) de um determinado sistema (Vander Stappen, 1996) caracterizando um evento dinâmico (Briongos e Soler, 2004). Define-se bacia de atração como sendo o conjunto de todas as possíveis soluções que convergem para o mesmo atrator (Ferrara e Prado, 1994). De posse do atrator, é possível determinar, por meio de suas características topológicas, o estado em que se encontra o sistema em estudo (Briongos e Soler, 2004). Para um fenômeno caótico, sua solução está inserida em um espaço de n -dimensões, com linhas de fluxo aperiódicas que

jamais se cruzam e nunca retornam a um estado anterior, definindo o que se conhece como atrator estranho.

Um exemplo típico e pioneiro deste tipo de atrator foi o de Lorenz, apresentado na Figura 2.6. Sua reconstrução foi realizada utilizando-se os valores de 10, 28 e $8/3$ respectivamente para os parâmetros ρ , σ e β do sistema de Lorenz.

As curvas, que se assemelham as asas de uma borboleta, descrevem a complexa interação determinística e não periódica (Tsonis, 1992) entre as variáveis dinâmicas do modelo, deixando claro que, com o aumento da temperatura, o fluido é transportado e as soluções se concentram à direita, porém quando ocorre uma parada no movimento rotacional as trajetórias migram para a esquerda (Gleick, 1991).

Figura 2.6 – Atrator de Lorenz (Lorenz, 1963).



O número de dimensões (n) de um atrator é chamado dimensão de imersão e está diretamente relacionado aos graus de liberdade do sistema no espaço real (van den Bleek e Schouten, 1993a). É importante ressaltar que na medida em que ocorre a evolução das soluções, muitas vezes as linhas de fluxo preenchem apenas uma fração limitada do espaço, dando ao atrator uma característica fractal. Uma grandeza adimensional denominada

dimensão de correlação (D_2) caracteriza a relação espacial entre os pontos do atrator e quantifica sua homogeneidade no espaço de fases (van den Bleek e Schouten, 1993a).

Antes de qualquer abordagem que vise a caracterização do Caos, é passo primordial a reconstrução do atrator para que, somente depois desta etapa, se possam abstrair quaisquer informações ou propriedades de uma determinada série temporal. Embora não seja trivial a obtenção do atrator estranho (Ferrara e Prado, 1994), Broomhead e King (1986) desenvolveram uma metodologia, especialmente aplicável a séries temporais experimentais, para a reconstrução do atrator no espaço de fases que utiliza as propriedades da matriz de covariância para esboçar o grau de independência de cada variável, obtendo variáveis não relacionadas. Tal método ficou conhecido como *decomposição em valores singulares* (SVD) e, a partir da seleção e separação das frequências características dos ruídos da série original, permite a reconstrução de um atrator representativo da dinâmica real do fenômeno.

As seções a seguir abordarão as definições e métodos de obtenção dos invariantes caóticos dinâmicos Entropia de Kolmogorov-Sinai (K) e expoentes de Lyapunov (λ) e o invariante estático dimensão de correlação, parâmetros de extrema importância na caracterização do Caos.

2.4.2 Dimensão de Correlação

Vander Stappen (1996) define a dimensão de correlação como a medida da complexidade global do sistema por meio da quantificação do grau de heterogeneidade do atrator. Uma das metodologias clássicas de obtenção da dimensão de correlação (D_2) utiliza-se do algoritmo proposto por Grassberger e Procaccia (1983). Neste método, a probabilidade de que dois pontos do atrator estejam inseridos em um hipercubo de lado ω é aproximada pela probabilidade de que estes mesmos pontos estejam espaçados a uma distância menor ou igual a ω , sendo assim:

$$D_2 = \lim_{\omega \rightarrow 0} \frac{\log C(\omega)}{\log(\omega)} \quad (2.2)$$

em que $C(\omega)$ é uma função de correlação da forma:

$$C(\omega) = \frac{1}{N^2} \lim_{N \rightarrow \infty} \sum_{\substack{i,j=1 \\ i \neq j}} \phi[\omega - (\vec{x}_i - \vec{x}_j)] \quad (2.3)$$

Nesta função, N é o número de pontos do atrator, \vec{x}_i é o vetor associado à função de Heaviside ($\phi = 1$ se $x \geq l$ ou $\phi = 0$ se $x < 0$) e ω é comprimento do lado dos hipercubos que recobrem os pontos do atrator. Este método é também conhecido como *Método da Integral de Correlação*, e se consagrou devido sua a facilidade de implementação.

2.4.3 Entropia de Kolmogorov

A entropia de Kolmogorov (K) solidificou-se como uma das grandezas matemáticas mais utilizadas na caracterização de processos caóticos. Análogo ao conceito termodinâmico clássico, a entropia de Kolmogorov também quantifica o grau de desorganização ou imprevisibilidade de um sistema, sendo que para $K > 0$ o sistema é considerado caótico; $K \rightarrow \infty$ para sistemas estocásticos e $K = 0$ para sistemas determinísticos. A sua dinâmica, todavia, baseia-se na quantificação da taxa de perda de informações ao longo do atrator ou na previsibilidade de rota de certo ponto arbitrário (Castilho, 2011). Um método para estimar K é aproximá-lo à entropia de Verossimilhança (K_{ML}) (Schouten *et al.*, 1994), em que:

$$K \approx K_{ML} = \frac{1}{\tau_S} \ln \left(1 - \frac{1}{b} \right) \quad (2.4)$$

O passo de tempo (τ_S) é obtido pela frequência de amostragem (f_S) como $\tau_S = 1/f_S$ e o parâmetro “ b ” é definido como uma sequência de cubos ($i_1, i_2, i_3, \dots, i_b$) necessários para recobrir o atrator reconstruído no espaço de fases. Este método é largamente utilizado no

cálculo da entropia de Kolmogorov para sinais experimentais, pois exibe uma menor influência de ruídos.

2.4.5 Expoentes de Lyapunov

Os *expoentes de Lyapunov* constituem-se uma ferramenta importante na descrição de sistemas dinâmicos, em especial a dinâmica caótica. Esta nomenclatura é atribuída a Oseledec (1968) que provou o *Teorema Ergódico Multiplicativo (TEM)*, abrindo caminho para a análise de processos dinâmicos por meio do cálculo das médias geométricas de processos ergódicos (Cherif, 2011). Os expoentes de Lyapunov são uma consequência direta do *TEM* de Oseledec e desempenham papel fundamental na análise de Caos. O significado destes expoentes está associado à previsibilidade de um sistema, sendo o maior expoente responsável por quantificar a sensibilidade às condições iniciais. Teoricamente, podem existir tantos expoentes de Lyapunov quanto o número de dimensões associadas ao espaço de fases (Tsonis, 1992; Cherif, 2011). Outra característica deste invariante é que em caso de dinâmica caótica existirá pelos menos um expoente positivo (Moon, 1992; Mullin, 1993; Monteiro, 2011).

A Tabela 2.3 apresenta a correspondência de cada tipo de atrator com a combinação dos expoentes de Lyapunov para um sistema tridimensional:

Tabela 2.3 – *Expoentes de Lyapunov característicos a cada tipo de atrator (Monteiro, 2011).*

Condição	Expoentes de Lyapunov	Atrator
(i)	$\lambda_1 < 0, \lambda_2 < 0, \lambda_3 < 0$	Ponto de equilíbrio
(ii)	$\lambda_1 < 0, \lambda_2 < 0, \lambda_3 = 0$	Ciclo limite
(iii)	$\lambda_1 < 0, \lambda_2 = 0, \lambda_3 = 0$	Toro bidimensional
(iv)	$\lambda_1 > 0, \lambda_2 = 0, \lambda_3 < 0$	Atrator estranho

Na condição (i) o volume de condições iniciais se contrai no espaço de fases e a trajetória converge para um único ponto. O expoente nulo presente na condição (ii) é

relativo à direção da trajetória fechada. Na condição (iii) a evolução temporal ocorre sobre uma superfície e, por fim, para a situação (iv), um expoente positivo explicita a dependência às condições iniciais, o que representa a trajetória é nulo e o terceiro é negativo e maior, em módulo, do que o primeiro, descrevendo assim um sistema dissipativo.

2.5 APLICAÇÃO DA TEORIA CAÓTICA DETERMINISTA NO ESTUDO DA POLUIÇÃO ATMOSFÉRICA

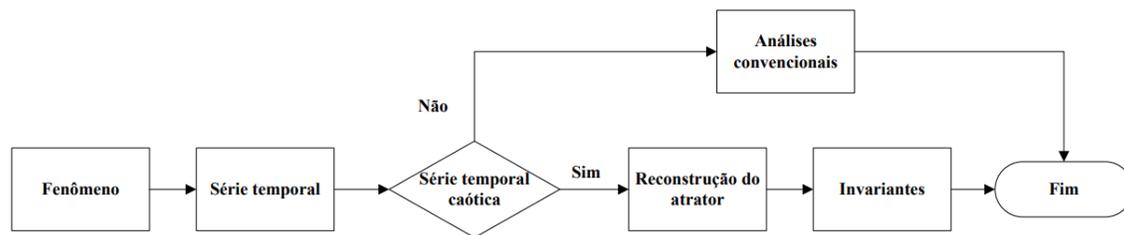
Uma das formas de se estudar a poluição atmosférica é por meio da avaliação da variação média da concentração em períodos como dias, semanas, meses ou anos. Dependendo do objetivo da investigação, como a simples obtenção da curva de variação anual ou mensal do poluente, esse valor médio já é suficiente, porém em casos em que se deseja estudar profundamente o fenômeno da dispersão de poluentes, as características da série temporal de cada ponto são importantes. Neste sentido, emergem as análises de séries temporais.

Uma série temporal é um conjunto ordenado, que apresenta variação no tempo, sem a necessidade de que haja homogeneidade no espaçamento e, tão pouco, qualquer tipo de simetria, sendo que os instantes no tempo estão interligados. Uma série temporal pode ser classificada como determinística (a evolução temporal deste tipo de sistemas é linear e sua trajetória pode ser prevista a curto e longo prazo) ou estocástica (sistemas completamente imprevisíveis em um dado intervalo de tempo (Castilho, 2011)).

A aplicação da análise de Caos à meteorologia teve início com o trabalho pioneiro de Lorenz em 1963 e, posteriormente, outros autores, como Stringer (1989), utilizaram esta teoria na engenharia. Com a divulgação do “novo método” de análise de sinais outros pesquisadores iniciaram a implementação desta metodologia em diversas áreas do conhecimento. Independentemente do fenômeno físico estudado, o procedimento de análise por meio da Teoria do Caos é o mesmo. A Figura 2.7 traz o fluxograma da metodologia para obtenção dos parâmetros de Caos. Uma vez que se identifica o objeto do estudo, deve-

se proceder com a obtenção de uma série temporal relativa à variável de interesse que represente o fenômeno. Esta evolução temporal pode ser medida a partir de uma dada instrumentação adequada ou simulada teoricamente. Depois de obtida a série no tempo, passa-se à reconstrução do atrator em um espaço de fases e depois se realiza o cálculo, a partir das características topológicas do atrator, dos invariantes do Caos determinístico.

Figura 2.7 – Metodologia da análise caótica (Prieto, 2014).



No campo da poluição atmosférica, poucos trabalhos são encontrados, merecendo especial destaque àqueles publicados por Raga e Le Moyne (1996), Lin e Lee (2008), Yu *et al.* (2011), Yu *et al.* (2013), Saeed *et al.* (2017). Alguns desses autores utilizaram a dimensão de correlação na tentativa de encontrar o número de variáveis regentes da dinâmica dos sistemas estudados.

Raga e Le Moyne (1996) aplicaram o método da integral de correlação na identificação dos graus de liberdade presentes em 13 estações de monitoramento da cidade do México, México, para séries temporais de O₃, NO₂, SO₂ e CO em 1993. As Figuras 2.8 e 2.9 trazem um exemplo da metodologia admitida pelos autores para uma das estações consideradas. Na primeira (2.8), observa-se a convergência da integral de correlação com o raio e na segunda (2.9) é possível verificar a saturação da dimensão de correlação com a variação da dimensão de imersão. Na estação em questão, é possível observar que a dinâmica do sistema é regida por 6 variáveis. Os autores ultimam dizendo que as regiões ocidentais e centrais da cidade apresentam menores valores de dimensão de correlação e que isso se deve ao possível fato de serem áreas de menor circulação.

Figura 2.8 – Variação do valor da integral de correlação ($C(r)$) com o raio (r) para dimensão de imersão (I) variando de 2 até 10, para a concentração de O_3 (Raga e Le Moyne, 1996).

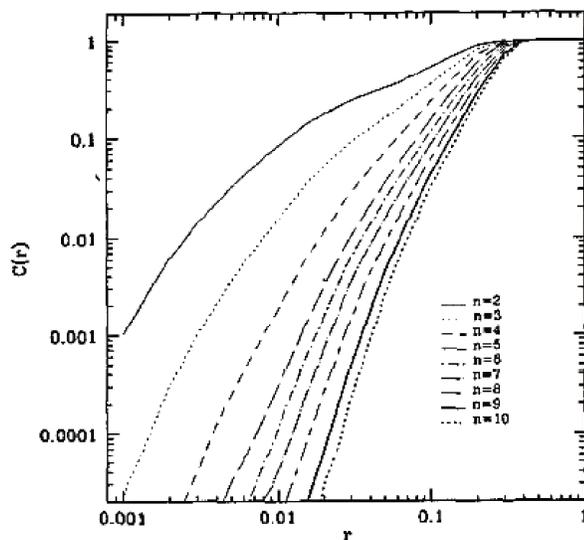
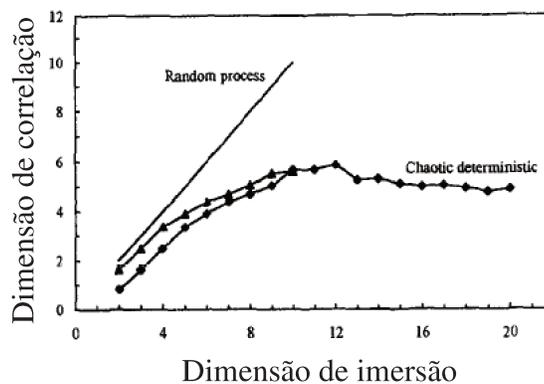


Figura 2.9 – Variação da dimensão de correlação (D) com a dimensão de imersão (I) para a concentração de O_3 (Raga e Le Moyne, 1996).

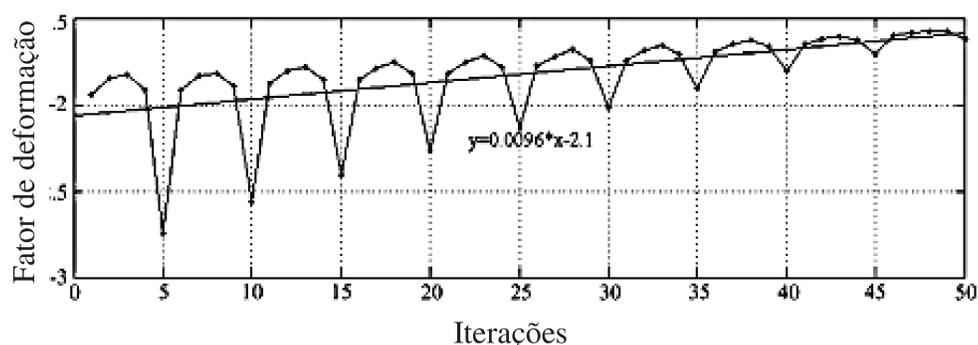


No estudo publicado por Yu *et al.* (2011) é possível encontrar a aplicação da análise do Caos na avaliação do índice de poluição do ar (API) em um período de 10 anos da cidade de Lanzhou (China). De forma análoga ao realizado por Raga e Le Moyne (1996), os autores avaliaram a dinâmica do sistema por meio da dimensão de correlação e também utilizaram o expoente de Lyapunov. Os autores encontraram, para a dimensão de correlação, um valor da ordem de 3,5, indicando, após a saturação, dinâmica caótica. Na Figura 2.10, apresenta-se o cálculo do máximo expoente de Lyapunov. Pelo coeficiente

angular da reta interpolada, obteve-se $\lambda_i = 0,0096$ o que levou os autores à confirmação da presença do Caos determinístico, pois $\lambda_i > 0$ (vide Tabela 2.1).

Após estas constatações, o artigo sugere o que seriam as quatro variáveis regentes da dinâmica local: (1) a topografia diferenciada da bacia do vale, (2) as emissões poluentes, (3) condições atmosféricas e (4) a estrutura de consumo energético.

Figura 2.10 – Cálculo do máximo expoente de Lyapunov (Yu *et al.*, 2011).



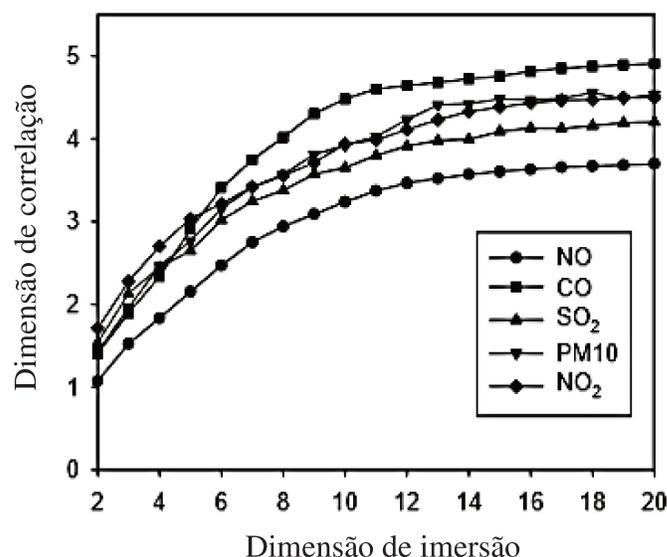
No que diz respeito à análise de Caos, Yu *et al.* (2011) e Yu *et al.* (2013) concluem que esta metodologia é adequada para avaliar séries temporais de poluentes atmosféricos, pois por meio dela, mais especificamente do cálculo da dimensão de correlação, é possível obter resultados muito próximos daqueles postulados por análises clássicas como DFA, porém o Caos determinístico considera de forma ampla as relações não lineares associadas ao fenômeno com a vantagem de um custo computacional muito menor daquele exigido por metodologias tradicionais. Os resultados obtidos por esses autores revelam a complexidade dinâmica dos sistemas em estudo e levaram a uma reflexão a respeito dos mecanismos internos destes fenômenos que poderiam, contrastados com estudos anteriores, explicar a lei de evolução temporal dos poluentes atmosféricos e o que levaria descoberta das principais causas da poluição do ar (Yu *et al.*, 2011)

Lee e Lin (2008) utilizaram a dimensão de correlação, máximo expoente de Lyapunov e a entropia de Kolmogorov para identificar características caóticas em séries temporais de concentração de alguns poluentes atmosféricos (NO, CO, SO₂, MP₁₀) na estação de monitoramento da qualidade do ar de Chung-Shan, Taipei (Taiwan), de Janeiro

de 1998 à Dezembro de 1998. A Figura 2.11 apresenta a variação da dimensão de correlação com dimensão de imersão.

Figura 2.11 – Variação da dimensão de correlação com a dimensão de imersão

(Lee e Lin, 2008).



Os autores observaram um aumento da dimensão de correlação com o incremento da dimensão de imersão até o ponto claro ponto de saturação, indicativo de dinâmica determinística. Na saturação, os valores da dimensão de correlação para NO, CO, SO₂, MP₁₀, e NO₂ foram, respectivamente 3,42, 4,71, 3,98, 4,32, e 4,25. Lee e Lin (2008) concluem que, as baixas dimensões de correlação, indicam que as séries temporais analisadas apresentam comportamento caótico de baixa dimensionalidade. Mantendo a mesma linha de raciocínio dos trabalhos anteriores, os autores asseveram a relação direta da dimensão de correlação com o número de variáveis regentes da dinâmica do sistema, comentando a dificuldade de relacionar tais graus de liberdade devido à complexidade das interações de centenas ou milhares de fontes na área em que as séries temporais foram obtidas, variáveis atmosféricas, diluição e reações químicas, interação biológica entre outros.

No que diz respeito à análise pelo máximo expoente de Lyapunov, Lee e Lin (2008) foram breves e simplesmente comentaram a evidente presença de Caos devido aos valores

encontrados, em todos os casos, serem positivos. A mesma constatação foi publicada por Saeed *et al.* (2017) quando analisaram a presença de Caos em séries temporais de MP_{10} e $MP_{2,5}$ obtidas em seis estações de monitoramento da capital do Paquistão (Islamabad). De forma análoga, também foram observadas características inerentemente caóticas na análise da entropia de Kolmogorov, uma vez que apenas valores positivos e finitos foram observados. Os autores realizaram, tanto para o expoente de Lyapunov quanto para dimensão de correlação e entropia de Kolmogorov, um estudo bem detalhado da influência da dimensão de imersão nos valores destes invariantes.

Por fim, Lee e Lin (2008) concluem afirmando que a abordagem caótica determinista pode ser uma nova perspectiva no estudo da variação temporal da concentração de poluentes atmosféricos, contudo eles alegam que são necessários mais estudos que correlacionem as características fractais com parâmetros estatísticos tradicionais.

2.6 MÉTODOS DE PREVISÃO DE SÉRIES TEMPORAIS

Um dos objetivos desta Tese é a predição de séries temporais de poluição atmosférica e, como toda evolução temporal, apresenta peculiaridades que necessitam ser exploradas pelos mais variados métodos preditivos disponíveis. Neste sentido, diversos processos foram desenvolvidos para a predição destas evoluções. A presente seção apresentará as metodologias consagradas para este fim.

2.6.1 Métodos estatísticos

Os métodos estatísticos partem da premissa de que, uma vez conhecida a natureza da distribuição dos dados, é perfeitamente possível propor um modelo de previsão (Parmezan e Batista, 2016). Eles podem ser divididos em dois grupos, os modelos Autoregressivos Integrados de Médias Móveis (ARIMA) e os modelos de suavização exponencial (Morettin e Tolo, 2006).

Basicamente, os modelos ARIMA envolvem três procedimentos estatísticos, a saber: autorregressão, integração e médias móveis (Box *et al.*, 2015). A componente autoregressiva denota a relação entre as observações, ou seja, avalia o quanto e como certa observação, dentro da série temporal, influencia a próxima informação (Montgomery *et al.*, 2015; Parmezan e Batista, 2016). A integração indica o número de vezes que uma variável deve ser diferenciada para se tornar estacionária. Já o procedimento de médias móveis avalia as informações futuras que não tem relação com os valores passados da série temporal (Parmezan e Batista, 2016).

Em relação aos modelos de suavização exponencial, estes realizam a decomposição da série temporal em componentes (tendência, nível e sazonalidade) e realiza a suavização de seus valores por meio da atribuição de diferentes pesos que decaem exponencialmente com o tempo (Gardner, 1985). Uma vez suavizados, os componentes são recompostos por meio de uma estrutura multiplicativa ou aditiva de modo a realizar as previsões (Parmezan e Batista, 2016).

A seguir serão descritos mais detalhadamente as metodologias estatísticas.

2.6.1.1 – Médias móveis

Trata-se de uma técnica simples que aplica um modelo de ordem r , em que a predição é obtida, a partir dos r últimos dados temporais da série, por meio de uma média aritmética (Parmezan e Batista, 2016). A Equação 2.5 é definição do modelo de médias móveis.

$$z_{t+1} = \frac{z_t + z_{t-1} + z_{t-2} + \dots + z_{t-r+1}}{r} \quad (2.5)$$

O parâmetro r é definido como o número de observações da média (z_{t+1}) (Montgomery *et al.*, 2015). Este valor é constante, explorando estrutura de autocorrelação dos resíduos da predição do período presente com aqueles ocorridos no passado (Parmezan e Batista, 2016). Sendo assim, o valor de r deve aumentar de acordo com as distorções ou

flutuações aleatórias apresentadas pela evolução temporal, com efeito, na previsão de uma série com comportamento mais determinístico, menos aleatório, deve-se utilizar baixos valores de r . Uma discussão mais detalhada sobre a escolha deste parâmetro pode ser encontrada na obra de Morettin e Toloi (2006).

Pode-se citar como desvantagem do método das médias móveis, a baixa acurácia quando aplicado a fenômenos com tendência ou sazonalidades, pois nesta metodologia a capacidade preditiva está diretamente relacionada à adição de novas informações e desconsideração das passadas. Outro fato que imprime erros no uso do método é que o valor de r é igual e completamente desatrelado de eventos passados. Para corrigir este problema, o método da Suavização Exponencial Simples é uma boa alternativa.

2.6.1.2 – Suavização Exponencial Simples

Este é considerado um método flexível, com boa precisão e simples. Nele, cada ponto da série é atrelado um valor diferente (Parmezan e Batista, 2016) que cresce exponencialmente com o passar do tempo, fazendo com que informações recentes tenham maior peso e influência na predição (Hyndman *et al.*, 2002; Montgomery *et al.*, 2015). A Equação 2.6 é uma generalização deste modelo. L é a estimativa no instante t , z_t é o último valor observado e a é uma constante de suavização ($0 < a < 1$) relacionada à evolução temporal dos dados.

$$L_t = az_t + a(1 - a)z_{t-1} + a(1 - a)^2z_{t-2} + \dots + a(1 - a)^{m-1}z_1 \quad (2.6)$$

De forma prática, durante as etapas preliminares da previsão, escolhe-se $L_1 = z_1$, ou seja, o valor inicial ajustado é igual ao primeiro valor da própria série temporal, fazendo com o ajuste só comece realmente a partir da segunda observação. Em linhas gerais, a “qualidade” da predição está diretamente atrelada à escolha da constante de suavização. Neste sentido, Morettin e Toloi (2006) comentam que se deve escolher $a \sim 0$ quando se intenta a obtenção do componente de tendência. Pequenos valores de a fazem com que a previsão seja mais influenciada por observações antigas. Já para situações em que o

interesse reside na própria predição, então se deve escolher um valor próximo da unidade, de modo a trazer mais influência de observações recentes. Desta forma, o conhecimento dos objetivos e da própria natureza da série é importante, pois influenciará diretamente na escolha da constante de suavização, sendo, portanto, uma das desvantagens deste método (Hyndman *et al.*, 2002). Outra desvantagem é que ao ser aplicado às séries com comportamento linear crescente ou decrescente, o método subestima ou superestima os valores reais. Neste sentido, cita-se o método da suavização exponencial de Holt que contorna este problema (Holt, 2004).

2.6.1.3 – Suavização Exponencial de Holt

Em princípio, o algoritmo de suavização exponencial de Holt se assemelha à suavização exponencial descrita na seção 2.6.1.2, contudo se verifica que na metodologia de Holt, emprega-se, além da tradicional constante de suavização (a), outro parâmetro suavizante (B) para auxiliar a modelagem da série temporal. As Equações 2.7, 2.8 e 2.9 trazem o esboço matemático da metodologia.

$$L_t = az_t + a(1 - a)(L_{t-1} + T_{t-1}) \quad (2.7)$$

$$T_t = B(L_t - L_{t-1}) + (1 - B)T_{t-1} \quad (2.8)$$

$$z_{t+h} = L_t + hT_t \quad (2.9)$$

As constantes a e B estão no intervalo $[0,1]$ e os componentes relativos ao nível e tendência são, respectivamente, obtidos por meio das Equações 2.7 e 2.8 (Parmezan e Batista, 2016). Na Equação 2.9, a variável z_{t+h} é a predição do valor z no instante $t+h$, sendo h o horizonte da predição (Parmezan e Batista, 2016). Como em qualquer método de suavização exponencial, as equações empregadas promovem uma modificação nas estimativas prévias no momento em que uma nova observação é calculada. Sobre os valores iniciais, Morettin e Toloi (2006) sugerem $T_1 = z_2 - z_1$ e $L_1 = z_1$. Devido à estrutura de autoaprendizagem do algoritmo, a escolhas das condições para o início das previsões não impactam as projeções (Parmezan e Batista, 2016), contudo, ainda se deve ter bastante

cautela na seleção das constantes amortizadoras, pois estas sim são decisivas no desempenho preditivo do método.

2.6.1.4 – Suavização Exponencial Sazonal de Holt-Winters

As metodologias de Holt-Winters são indicadas para fenômenos em que são observadas, simultaneamente, as componentes de sazonalidade e tendência. Eventos com esta descrição apresentam padrões cíclicos com repetição em intervalos relativamente constantes (Montgomery *et al.*, 2015). Os modelos deste grupo são divididos em duas categorias: aditivos e multiplicativos (Winters, 1960). O rigor matemático destes modelos pode ser observado nas Equações 2.10, 2.11, 2.12 e 2.13 para o modelo multiplicativo e nas Equações 2.14, 2.15, 2.16 e 2.17 para o aditivo.

$$L_t = a \frac{z_t}{S_{t-s}} + (1 - a)(L_{t-1} + T_{t-1}) \quad (2.10)$$

$$T_t = B(L_t - L_{t-1}) + (1 - B)T_{t-1} \quad (2.11)$$

$$S_t = y \frac{z_t}{L_t} + (1 - y)S_{t-s} \quad (2.12)$$

$$z_{t+h} = (L_t + hT_t)S_{t-s+h} \quad (2.13)$$

$$L_t = a(z_t - S_{t-s}) + (1 - a)(L_{t-1} + T_{t-1}) \quad (2.14)$$

$$T_t = B(L_t - L_{t-1}) + (1 - B)T_{t-1} \quad (2.15)$$

$$S_t = y(z_t - L_t) + (1 - y)S_{t-s} \quad (2.16)$$

$$z_{t+h} = L_t + hT_t + S_{t-s+h} \quad (2.17)$$

A escolha de um ou outro modelo dependerá do padrão sazonal da série. A metodologia multiplicativa se aplica às séries temporais que apresentam tendência e sazonalidade multiplicativa, em outras palavras, aquelas evoluções em que a amplitude da variação sazonal aumenta com o acréscimo do nível médio da série (Parmezan e Batista, 2016). Já o modelo aditivo se aplica às séries temporais que apresentam tendência e sazonalidade aditiva, logo, evoluções em que a diferença entre o maior e o menor valor estudado dentro das estações permanece constante no tempo (Parmezan e Batista, 2016).

Aqui, os parâmetros a , B e y , são as constantes de suavização no intervalo $[0,1]$, z_{t+h} é a predição de z para o período $t + h$ e s é a quantidade de observações um período sazonal (Parmezan e Batista, 2016). Similar ao observado nos outros métodos de suavização, os valores iniciais, em algum período do passado, de L (nível), T (tendência) e S (sazonalidade) precisam ser fornecidos. As Equações 2.18, 2.19, 2.20 e 2.21 demonstram uma maneira simples de se estimar essas condições iniciais aplicando o nível e tendência no mesmo período sazonal. É importante ressaltar que as Equações 2.18 e 2.19 são aplicadas para ambas as metodologias, já as equações 2.20 e 2.21, aplicam-se, respectivamente, para a multiplicativa e aditiva. Novamente, aqui as dificuldades residem na determinação das constantes de suavização.

$$L_S = \frac{1}{S} (z_1 + z_2 + \dots + z_S) \quad (2.18)$$

$$T_S = \frac{1}{S} \left(\frac{z_{S+1} - z_1}{S} + \frac{z_{S+2} - z_2}{S} + \dots + \frac{z_{S+S} - z_S}{S} \right) \quad (2.19)$$

$$S_1 = \frac{z_1}{L_S}, S_2 = \frac{z_2}{L_S}, \dots, S_S = \frac{z_S}{L_S} \quad (2.20)$$

$$S_1 = z_1 - L_S, S_2 = z_2 - L_S, \dots, S_S = z_S - L_S \quad (2.21)$$

2.6.1.5 – Modelos AR, ARMA, ARIMA

Os modelos autoregressivos integrados de médias móveis (ARIMA) de ordem (p, d, q) , como o próprio nome sugere, são a combinação dos procedimentos estatísticos de autorregressão, integração e médias móveis (Box *et al.*, 2015). Não é obrigatória a aplicação simultânea destes procedimentos, de modo que a utilização separada dá nome, a depender da ordem utilizada, a outros dois métodos, o modelo autoregressivo de ordem p (AR(p)) e o autoregressivo de médias móveis de ordem (p,q) (ARMA(p,q)).

No modelo AR(p) o valor previsto da série é obtido por meio de um conjunto linear de valores passados e uma sequência de valores aleatórios (ou ruído branco) (Parmezan e

Batista, 2016). A Equação 2.22 traz o modelo $AR(p)$ e a Equação 2.23 demonstra o cálculo do nível inicial do modelo (δ).

$$z_t = \delta + \sum_{i=1}^p \varphi_i z_{t-i} + e_t \quad (2.22)$$

$$\delta = \mu(1 - \varphi_1 - \varphi_2 - \dots - \varphi_p) \quad (2.23)$$

As variáveis presentes na Equação 2.22 e 2.23 são:

p = número de observações consideradas;

z_t = ponto na série no momento do tempo (t);

φ_i = i -ésimo coeficiente autorregressivo;

δ = nível inicial do modelo;

μ = média do processo estacionário;

e_t = ruído branco no instante t .

De modo a aumentar o número de aplicações às séries temporais, o modelo $ARMA(p,q)$ (Equação 2.24) incorpora os procedimentos autorregressivos de ordem p com o modelo de médias móveis de ordem q (Shumway e Stoffer, 2011) para realizar a estimativa de um processo parte autorregressivo parte médias móveis.

$$z_t = \delta + \sum_{i=1}^p \varphi_i z_{t-i} + \sum_{i=1}^q \theta_i e_{t-i} + e_t \quad (2.24)$$

Na Equação 2.24 o termo θ_i é o i -ésimo valor de ponderação usado (Parmezan e Batista, 2016). A principal vantagem do modelo $ARMA$ é o custo computacional, pois o este utiliza menos termos do que o exigido pelos métodos de médias móveis e AR .

A aplicação dos métodos AR e ARMA é possível quando a série objeto do estudo é estacionária, porém, para um caso oposto a esta hipótese, os elementos da séries temporal podem ser transformados por meio de um procedimento de diferenciação, assegurando, todavia, a estacionariedade dos dados. Neste sentido, o modelo ARMA é acrescido de um termo de integração, culminando, portanto, no modelo autorregressivo integrado de médias móveis de ordem (p,d,q) , chamado ARIMA (p,d,q) (Parmezan e Batista, 2016). Esta metodologia é também conhecida como método de Box-Jenkins e é aplicada na previsão de informações futuras de uma série temporal (de Smith, 2015). A Equação 2.25 traz a forma geral do modelo.

$$I'_t = \delta + \sum_{i=1}^{p+d} \varphi_i I'_{t-i} + \sum_{i=1}^q \theta_i e_{t-i} + e_t \quad (2.25)$$

A estratégia para confecção do ARIMA é baseada em um ciclo iterativo, no qual a escolha da estrutura do modelo baseia-se nos próprios dados em estudo (Morettin e Toloi, 2004). Na Equação 2.25, o termo $I'_t = \Delta^d z_t = \Delta(\Delta^{d-1} z_t)$, sendo d o grau do operador de diferença. Em linhas gerais, a suposição base do ARIMA é que a d -ésima diferença entre os pontos da série pode ser representado por um processo estacionário que, por sua vez, é representado por um modelo ARMA (Parmezan e Batista, 2016).

2.6.2 Métodos de Aprendizado de Máquina

Métodos de aprendizado de máquina, diferentemente dos estatísticos, não necessitam do conhecimento antecipado da distribuição dos dados para realizar a previsão. Isto ocorre pelo fato de eles não dependerem de parâmetros para a modelagem do evento. São, via de regra, metodologias mais simples e com excelente desempenho (Parmezan e Batista, 2016). Islam e Sivakumar (2002) dividem os métodos de aprendizado de máquina em duas categorias: os locais e os globais.

As abordagens locais incluem algoritmos adaptados que adicionam à predição a informação temporal associada ao conjunto de dados, por meio de um particionamento da

série em subsequências de valores, mais próximos ou mais importantes que o valor atual, e posterior combinação destas informações de modo a produzir o valor futuro. Um dos métodos consagrados que aplicam essa abordagem é o algoritmo do vizinho mais próximo.

Já em relação às abordagens globais, estas constroem modelos partindo de um procedimento de treinamento que recebe como entrada as informações da série (Parmezan e Batista, 2016). ANN (redes neurais artificiais) e SVM (máquinas de suporte vetorial) são exemplos desse nicho.

Redes neurais artificiais objetivam simular o cérebro humano no que tange o processamento de informação. Desta forma, o sistema é composto por neurônios, unidades de processamento, que se comunicam por meio de conexões sinápticas (Zhang *et al.*, 1998). Diferente dos métodos tradicionais, as ANN são auto-adaptativas com aprendizado que se baseia em dados experimentais. Elas apresentam a capacidade de aprender com exemplo e capturar relações lineares e não lineares (Pommeranzenbaum, 2014). Podem trabalhar com dados ruidosos, incompletos e ainda imprecisos, sendo, portanto, uma metodologia eficaz na predição de séries temporais.

Em relação às máquinas de suporte vetorial (SVM), estas se baseiam na Teoria do Aprendizado Estatístico (Vapnik, 1999) e podem auxiliar a solução de diferentes problemas como regressão e classificação. De estrutura bastante semelhante a das redes neurais, a diferença reside na forma de aprendizado. No que diz respeito às ANN, estas trabalham minimizando o erro do modelo sobre os dados de treinamento, já as SVM se baseiam na minimização do risco estrutural, buscando, portanto, o menor erro de treino minimizando um limite superior para o erro de generalização do modelo (Parmezan e Batista, 2016). Devido sua capacidade de resolver tarefas de regressão, é um excelente método na predição de séries temporais (Ristanoski *et al.*, 2013).

Na presente Tese, o algoritmo aplicado para previsão de séries temporais é aquele baseado na abordagem do vizinho mais próximo (k-NN) que se enquadra como um método de inteligência artificial que usa aprendizado de máquina. Desta feita, a seção a seguir traz

esta metodologia bem como um breve panorama dos métodos aplicados à predição de séries temporais de poluentes atmosféricos.

2.6.2.1 – k-NN – Atual cenário da predição de séries temporais de poluentes atmosféricos

É absolutamente clara a necessidade de métodos de previsão das emissões atmosféricas, uma vez que os danos causados pela poluição do ar se estendem não somente à economia do país, mas também à qualidade de vida da população, tornando-se, portanto, um sério problema de saúde pública. Um modelo de previsão de qualidade do ar não resolve as questões supracitadas, contudo, previsões confiáveis auxiliam a tomada de decisão no que tange a gestão de qualidade do ar (AQI, 2001). Neste sentido, salienta-se o fato de que muitas cidades emitem alertas ao público quando os índices de determinados poluentes excedem os limites especificados. A partir de modelos preditivos confiáveis, agências governamentais podem realizar planejamentos mais detalhados no que diz respeito aos comunicados emitidos diretamente à população e também nas tomadas de decisão para controle dos eventos críticos de poluição.

A Tabela 2.4, elaborada por Bai *et al.* (2018), lista os diferentes modelos e técnicas, bem como as aplicações, de previsão da poluição do ar utilizados por diferentes autores. Conforme se observa, são diversas as metodologias aplicadas na previsão da poluição atmosférica. Em geral, pode-se classificar estas metodologias em quatro categorias: métodos de climatologia, estatísticos, três dimensões e híbridos.

Tabela 2.4 – *Diferentes modelos, técnicas e aplicações de previsão da poluição do ar encontrados em literatura (Bai et al., 2018).*

Métodos	Modelos	Comentários	Autores
Estatísticos	Filtro de Kalman (KF) e Previsão Híbrida (HF)	Usaram duas técnicas de ajuste, HF e KF, para melhorar a precisão das previsões fornecidas por um sistema de previsão de qualidade do ar	Silibello <i>et al.</i> (2015)

	Modelos lineares generalizados com <i>log-link</i> e distribuição gama	Aplicam variáveis meteorológicas previstas para uma boa previsão, por meio da análise comparativa dos dois modelos	Huebnerova <i>et al.</i> (2014)
	Rede neural artificial (ANN) e Média móvel integrada auto-regressiva (ARIMA)	Realizaram a análise de concentrações extremas por meio da integração de dois modelos em um conjunto (ensemble)	Catalano <i>et al.</i> (2016)
	Máquina de vetores de suporte / Algoritmo genético / Redes neurais de propagação reversa (SVM-GBPNN)	Avaliaram um modelo híbrido que utiliza SVM para classificar os dados, aplicando GA para otimizar o modelo BPNN	Feng <i>et al.</i> (2011)
Inteligência artificial	<i>Wavelet</i> / Redes neurais de propagação reversa (W-BPNN)	Usam a transformada <i>wavelet</i> para realizar extração de características e caracterização de poluentes atmosféricos	Bai <i>et al.</i> (2016)
	Transformada <i>wavelet</i> , <i>perceptron</i> multicamada, função de base radial, rede de Elman, SVM e modelo autoregressivo com entradas exógenas (ARX) linear	Realizaram a decomposição dos dados nos coeficientes <i>wavelet</i> e usaram diferentes NN para predição individual, combinando, assim, os poucos preditores no conjunto. Essa abordagem não requer informações exaustivas sobre os poluentes do ar e tem a capacidade de avaliar as relações não-lineares entre variáveis preditivas diferentes.	Siwek <i>et al.</i> (2012)
Híbridos	ANN Híbrido	Aplicaram parâmetros geográficos baseados em uma trajetória como entrada extra para o modelo de RNA; utilizaram a transformada <i>wavelet</i> na decomposição de séries originais em sub-séries com menor variabilidade	Feng <i>et al.</i> (2015)
	Rede neural <i>feed-forward</i> com mecanismo de rolamento e modelo cinza (RM-GM-FFNN)	Avaliaram o modelo FFNN aprimorado com RM e GM para analisar a possível correlação entre diferentes variáveis de entrada para melhorar a precisão das previsões	Fu <i>et al.</i> (2015)
	<i>Neuro-fuzzy</i> adaptativo (ANF), funções de	Aplicaram um método de predição de intervalo e ANF para	Song <i>et al.</i> (2015)

	distribuição	associar a incerteza de PMs de acordo com a distribuição de emissão de poluentes	
Três dimensões	Modelo-3/ Comunitário Multiescalar de Qualidade do Ar	Aplicaram um método para analisar a mudança da concentração de poluentes em situações de falta de dados práticos de poluição	Luo <i>et al.</i> (2012)
	Modelo de pesquisa e previsão do tempo acoplado a modelos químicos (WRF/Chem)	Utilizaram uma metodologia em que são avaliados diferentes módulos meteorológicos e químicos sob diferentes condições de separação e acoplamento para verificar a precisão da previsão	Grell <i>et al.</i> (2005)
Outros métodos	Redes neurais baseadas em modelos de previsão geográfica	São modelos que consideraram o fator geográfico e apresentaram melhor desempenho de que aqueles que não consideram	Kurt <i>et al.</i> (2010)
	GM Análise relacional cinza	Selecionaram 30 índices de 5 categorias, e encontraram os principais fatores de impacto usando análise relacional cinza. Depois usaram o modelo GM (1, 1) para prever a concentração de poluentes	Pan <i>et al.</i> (2011)

Os métodos climatológicos são razoavelmente simples, pouco custosos e de fácil implementação. Eles partem do pressuposto de que, uma vez conhecido o passado, é perfeitamente possível prever o futuro (AQI, 2001) por meio da associação de altos níveis de poluição com as respectivas condições meteorológicas específicas. A aplicação desta abordagem objetiva a previsão de níveis excedentes e não de concentrações ambientes.

No que se refere às metodologias estatísticas, estas realizam a associação bem como a quantificação entre parâmetros meteorológicos e um dado poluente atmosférico por meio de diversas técnicas (AQI, 2001). Aqui, salienta-se a CART, sigla para “*Classification and Regression Tree*”, análise de regressão e redes neurais. As metodologias CART são usadas para prever os níveis de determinado poluente utilizando *softwares* para correlacionar e identificar as variáveis meteorológicas ou de qualidade do ar aos níveis de poluição atmosférica (AQI, 2001). Outra abordagem estatística consagrada é a análise de regressão.

Trata-se da obtenção da equação de regressão linear multivariada para prever níveis de poluição. São utilizados pacotes de análise estatística padrão para associar os níveis de poluentes às variáveis meteorológicas e, por fim, obter a referida equação (AQI, 2001). Outro conjunto de técnicas dentro das metodologias estatísticas são as redes neurais. Redes neurais foram desenvolvidas de tal maneira a imitar o cérebro humano no reconhecimento de padrões, assim, estas abordagens utilizam técnicas de aprendizagem adaptativa e reconhecimento de padrões para identificar e quantificar a associação de parâmetros atmosféricos que influenciam a qualidade do ar (AQI, 2001). No Brasil, Borges, Andrade e Guardani (2012) desenvolveram um modelo de rede neural para prever a concentração de ozônio na Região Metropolitana de São Paulo utilizando como base os valores médios das variáveis meteorológicas em determinados períodos ao longo do dia. Os autores concluem o trabalho afirmando que para períodos de previsão de 1 a 5 dias foi atingido um nível de confiança de 95% para as previsões obtidas.

Outros métodos difundidos e com excelentes resultados são os determinísticos. Este tipo de modelos tem por objetivo determinar as concentrações dos poluentes atmosféricos emitidos por diversas fontes e se baseiam em equações da física, levando em conta, também, as reações químicas dos poluentes presentes. De maneira genérica, o estudo da poluição atmosférica por meio de modelos determinísticos passa pela solução das equações de Navier-Stokes e a equação de advecção-difusão para a concentração dos poluentes (Sportisse, 2008). São também consideradas as diversas reações químicas que um dado poluente pode sofrer na atmosfera. Devido à complexidade das equações envolvidas, as soluções destes sistemas são possíveis apenas por meio de cálculo numérico (Sportisse, 2008). Os modelos determinísticos podem ser divididos em duas categorias, modelos *on-line* acoplados (resolvem os modelos acoplados Navier-Stokes/evolução de concentrações) e *off-line* acoplados (dinâmica dos fluídos não está associada à evolução das concentrações sendo necessária a utilização de um modelo meteorológico como entrada para a equação de advecção-difusão) (Sportisse, 2008). Um exemplo de modelo determinístico é o CHIMERE, que simula os processos físicos e químicos que acontecem na atmosfera utilizando como base as previsões meteorológicas e a emissão de poluentes (Sportisse, 2008).

As técnicas mencionadas anteriormente são consagradas e apresentam pontos fortes, contudo todas elas admitem que a evolução temporal do fenômeno estudado exibe certa estabilidade, fazendo com que quaisquer mudanças no padrão de emissão do poluente e/ou no clima promovam uma redução na capacidade preditiva destes métodos. Como alternativa, as abordagens deterministas aperfeiçoam essas “falhas”. Os modelos determinísticos 3-D são classificados como Lagrangeanos ou Eulerianos e objetivam representar todos os principais processos que estão relacionados aos níveis de poluição atmosférica, utilizando, para tal, variados submodelos, a saber: modelos de emissão – simulam emissões dependentes do tempo ou seus percursores, tanto de fontes antropogênicas quanto naturais (AQI, 2001); modelos meteorológicos – permitem obter a evolução espaço-temporal das variáveis que descrevem a fluidodinâmica atmosférica e, por este motivo, exige, frequentemente, um conhecimento mais aprofundado da estrutura meteorológica da área em análise (Moreira, 2008); modelos químicos – usam propriedades como taxa cinética, espectroscopia, termodinâmica para simular a transformação da poluição primária em poluição secundária (AQI, 2001).

Já os sistemas híbridos, como o próprio nome sugere, são a combinação de dois ou mais métodos com o objetivo de se aproveitar as vantagens de cada método aumentando, portanto, a qualidade das previsões (Bai *et al*, 2018)

O aprendizado de máquina é um aprendizado por experiência que se aprimora à medida que a tarefa é executada, realizando, portanto, uma estruturação do conhecimento existente que conduz a um entendimento do aprendizado. Assim, o método k-NN, classificado como de regressão não paramétrica, parte da premissa de que a série reproduz seu próprio comportamento ao longo do tempo, asseverando que a observação futura apresenta relação com as anteriores. Esta abordagem é um algoritmo de aprendizado supervisionado que se baseia em exemplos (Parmezan e Batista, 2016) e objetiva encontrar padrões entre os pontos de um determinado conjunto de dados, sem se preocupar com sua localização no tempo, capturando, portanto, a dinâmica não linear de auto-similaridade na série temporal, característica semelhante à dinâmica fractal de uma evolução caótica (Perlin

e Ceretta, 2007). A metodologia k-NN pode ser aplicada em casos multivariados e univariados, sendo que apenas o último será descrito aqui, pois foi este utilizado nesta Tese.

O trabalho de Perlin e Ceretta (2007) foi adotado como base para a elaboração da descrição a seguir. É importante deixar claro que o k-NN já é um consagrado método (Imandoust e Bolandraftar, 2013) de predição de séries temporais a curto prazo (Parmezan e Batista, 2016). A Tabela 2.5 lista algumas das áreas em que a metodologia dos vizinhos próximos foi aplicada, demonstrando, portanto, sua ampla aceitação pela comunidade científica.

Tabela 2.5 – Áreas de aplicação da metodologia k-NN (o autor, 2018).

Autor(es)	Aplicação(ões)
Trstenjak <i>et al.</i> (2014)	Análise textual
Moldagulova e Sulaiman (2018)	
Krishnan <i>et al.</i> (2017)	Agricultura
Perlin e Ceretta (2007)	Finanças/Economia/Mercado
Imandoust e Bolandraftar (2013)	
Nie e Song (2018)	
Isa <i>et al.</i> (2017)	Medicina
Gupta e Mittal (2018)	
Chen <i>et al.</i> (2018)	Engenharia
Dragomir (2010)	Meio ambiente
Divya e Vijayalakshmi (2015)	
Chen <i>et al.</i> (2018)	Reconhecimento de imagem

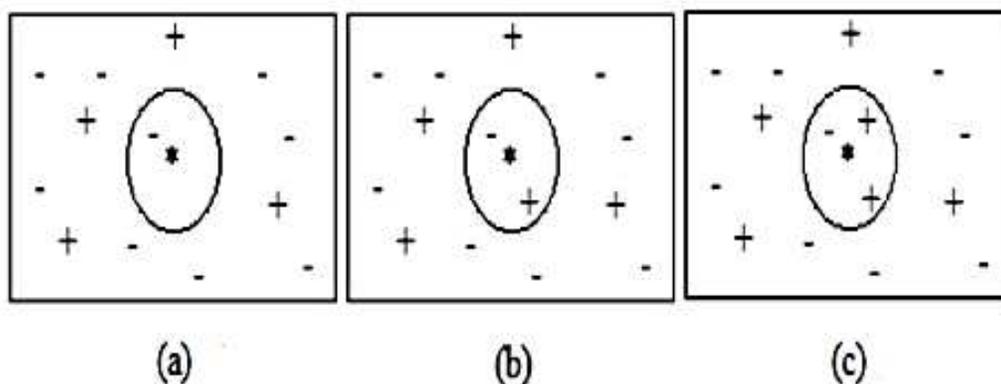
A abordagem k-NN parte, inicialmente, da definição do intervalo de treinamento e de sua divisão em vetores y_t^I de tamanho I , em que $t = 1, \dots, T$. Aqui, T é o número de observações no intervalo de treinamento e I é a dimensão de imersão que na, na prática, para um sistema caótico, é a mesma dimensão do atrator estranho, sendo possível determiná-la por meio da Teoria do Caos. O último vetor disponível antes dos pontos simulados é chamado de y_t^I . A seguir, realiza-se a seleção dos pedaços k mais parecidos com y_t^I por meio da avaliação na menor soma das distâncias entre os vetores, tal qual é apresentado na Equação 2.26, em que D é distância absoluta entre os vetores x e y . Por fim, constrói-se a previsão a partir da média dos k vizinhos escolhidos. Todas as etapas anteriores são realizadas em *loop* até que toda a previsão desejada seja criada. Por motivos

de acurácia, não é recomendada a geração de pontos preditivos *out of sample* que ultrapassem 10 – 20% do tamanho da série de treino (Hyndman e Athanasopoulos, 2014), em outras palavras, do número de pontos disponíveis para a análise da série em estudo.

$$D(x, y) = \sum_i^N |x_i - y_i| \quad (2.26)$$

Exemplificando o funcionamento do método, na Figura 2.12, são observadas classes positivas e negativas, denotando os vizinhos 1, 2 e 3 mais próximos de um conjunto de dados que se localiza no centro do círculo. No caso da Figura 2.12 (a), o vizinho mais próximo é negativo, logo essa classe foi atribuída ao conjunto de dados. Para o caso de ocorrer empate (Figura 2.12 (b)), a classe é escolhida aleatoriamente. No caso da Figura 2.12 (c) a maioria é tida como base para a escolha da classe, sendo assim, classe positiva (Moura *et al.*, 2017).

Figura 2.12 – Método dos vizinhos mais próximos (Mulak e Talhar, 2015).



Em geral, uma vez adequadamente determinados os valores de k e I , no que diz respeito às vantagens apresentadas pelo código k -NN, verifica-se que é uma técnica simples, bastante flexível e proporciona excelentes resultados. Já suas desvantagens residem no fato de, dependendo da série de treino, sua solução pode demandar um gasto computacional elevado e sua precisão ser comprometida pela presença de ruídos (Perlin and Ceretta, 2007).

2.6.2.1.1 – O método k-NN e a Teoria do Caos

No que diz respeito ao desempenho do método k-NN, duas decisões são bastante importantes. A primeira delas é a definição do valor mais adequado para k . De um modo geral, quando o conjunto de treino possui muitos ruídos é preferível utilizar o método k-NN com $k = 1$, caso contrário, com $k > 1$. Contudo, a determinação do valor de k é realizada por meio de experimentos (tentativa e erro) (Santos, 2009). A segunda decisão importante diz respeito à obtenção da dimensão de imersão (I).

Define-se dimensão de imersão como sendo a dimensão na qual um objeto (por exemplo, um atrator) está imerso, devendo ser suficientemente grande de modo a abrigar completamente a dinâmica do fenômeno estudado (Castilho, 2011). A escolha do método de determinação de I é um aspecto importante na análise da evolução temporal de um sistema, especialmente em se tratando de dados reais, uma vez que esta informação, a priori, não é conhecida (Harikrishnan *et al.*, 2017). Quando o valor de I obtido é menor que a dimensão real do objeto em estudo, todos os cálculos realizados, a partir de então, serão imprecisos, pois a dinâmica do sistema não estará completamente desenvolvida (Harikrishnan *et al.*, 2017). No caso de um valor de I maior do que o real, o custo computacional será excessivo (Harikrishnan *et al.*, 2017). Três metodologias são consagradas para este fim, o método da saturação de algum invariante do sistema (Grassberger e Procaccia, 1983), a decomposição em valores singulares (SVD) (Broomhead e King, 1986) e o método dos falsos vizinhos próximos (FNN) (Kennel *et al.*, 1992).

O método SVD realiza a decomposição da matriz de covariância do sistema, obtendo-se então, uma nova matriz diagonal. Uma vez diagonalizada, realiza-se a identificação dos autovalores e autovetores. Na prática, o número de autovalores não nulos é uma estimativa da dimensão de imersão mínima (Savi, 2017). Esta é, basicamente, uma abordagem linear com base teórica sólida (Mirmomeni e Lucas, 2008), contudo, ao usar o SVD com uma ferramenta não linear são observadas algumas dificuldades, especialmente devido aos problemas críticos na seleção dos valores singulares significativos e, devido a

este fato, muitos pesquisadores colocam em cheque a capacidade do método na avaliação de séries temporais não lineares (Paluš e Dvořák, 1992; Mirmomeni e Lucas, 2008).

A abordagem FNN baseia-se na busca de uma dimensão de imersão mínima de modo que não ocorram cruzamentos de órbitas relativas à evolução temporal (Savi, 2017). A ideia central desta metodologia é a de aumentar progressivamente a dimensão I e avaliar os vetores vizinhos próximos, verificando se a proximidade da vizinhança é verdadeira ou apenas uma projeção de dimensão superior (Lima e Nogueira, 2013). Este procedimento é realizado até o momento em que o aumento de I não resultar mais na redução do número de falsos vizinhos próximos (Lima e Nogueira, 2013). Esta metodologia, embora bastante eficaz, pode conduzir a resultados diferentes devido à presença de vários valores de vizinhos (Mirmomeni e Lucas, 2008).

O método da saturação de algum invariante baseia-se na variação da dimensão de imersão até o momento em que a propriedade estudada se mantenha constante e, portanto, saturada. A aplicação da Teoria do Caos na obtenção de I , por meio da curva de saturação da dimensão de correlação (D_2) utilizando, por exemplo, o algoritmo de Grassberger e Procaccia (1983) (discutido na seção 2.1.2.), é uma alternativa tanto viável quanto largamente utilizada (Di *et al.*, 2017) no estudo da imersão de séries temporais de poluentes atmosféricas, sendo especialmente interessante, pois permite, também, a determinação do número de graus de liberdade que regem a dinâmica do sistema (dimensão de correlação), informação importantíssima no estudo da evolução temporal de condições e poluentes atmosféricos (Sivakumar *et al.*, 2002; She e Yang, 2010; Dhanya e Kumar, 2011; Kyoung *et al.*, 2011; Lana *et al.*, 2016). Dentre os métodos destacados anteriormente, é o mais apropriado para o estudo de fenômenos inerentemente caóticos, característica consolidada da dinâmica das emissões atmosféricas (Raga e Le Moyne, 1996, Lin e Lee, 2008, Yu *et al.*, 2011, Yu *et al.*, 2013, Saeed *et al.*, 2017). Outras vantagens deste método são a otimização do tempo computacional quando utilizada uma séries de treinamento adequada, sua alta sensibilidade à heterogeneidade da distribuição de pontos (Souza, 2002), simplicidade, eficácia, intuição, desempenho de classificação competitiva em diversos domínios, robusto para utilização em dados com ruídos e bastante eficaz para séries de

treinamento grandes (Imandoust e Bolandraftar, 2013). Como desvantagens, verifica-se a baixa acurácia quando aplicados à séries curtas, sensibilidade à ruídos (Harikrishnan *et al.*, 2017), em caso de uma série de treino muito grande o tempo computacional pode ser alto, é muito sensível à características irrelevantes ou redundantes presentes nos dados e no aprendizado baseado em distância não há clareza quanto a qual tipo de distância usar (Imandoust e Bolandraftar, 2013). Sendo assim, a junção da Teoria do Caos, por meio do algoritmo de Grassberger e Procaccia (1983), ao método de predição k-NN é uma opção claramente oportuna no estudo preditivo a que se intenta nesta Tese.

CAPÍTULO III

MATERIAIS E MÉTODOS

Neste capítulo são descritos os materiais e métodos utilizados nesta Tese de Doutorado. Inicialmente são listadas quais estações de monitoramento da CETESB foram escolhidas bem como os poluentes que foram analisados. Na sequência os métodos de análise do Caos e de previsão são expostos. Esta Tese de Doutorado foi desenvolvida no Laboratório de Processos em Meio Porosos (LPMP), Departamento de Engenharia de Processos (DEPro), Faculdade de Engenharia Química (FEQ), Universidade Estadual de Campinas (UNICAMP).

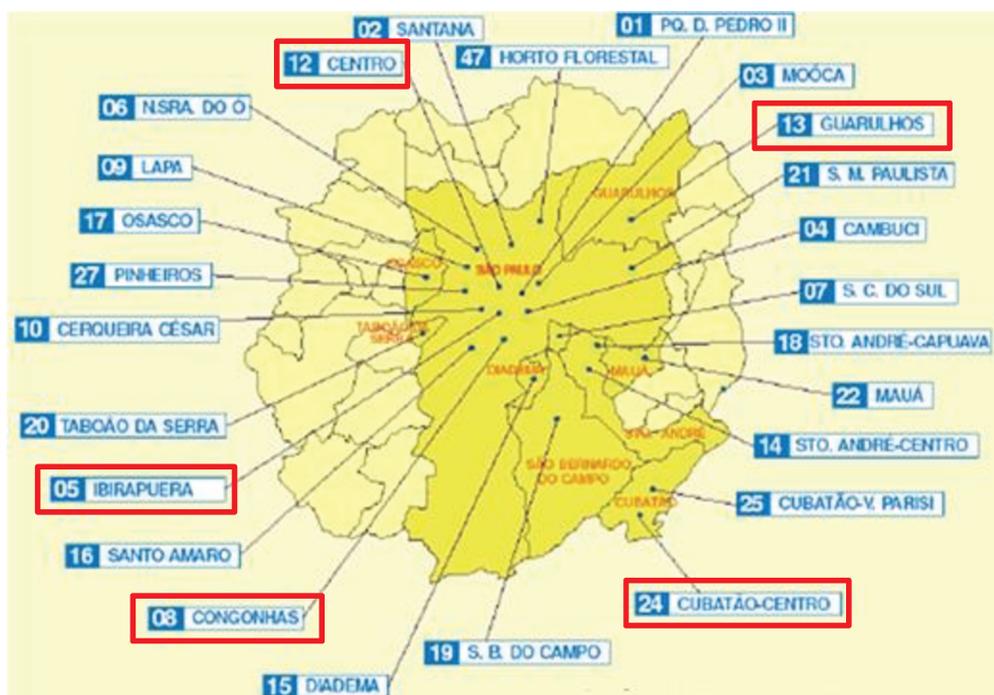
Os valores das séries temporais dos poluentes avaliados foram diretamente exportadas da plataforma eletrônica da Companhia de Tecnologia de Saneamento Ambiental do Estado de SP (CETESB). Todas as análises foram realizadas no período de 1998 a 2017. A Figura 3.1 apresenta a localização geográfica das estações que foram analisadas. Assim, os dados avaliados foram exportados das seguintes estações:

- Congonhas
- Guarulhos – Paço Municipal
- Ibirapuera
- Centro (São Paulo)
- Cubatão – Centro

As estações Centro, Congonhas e Guarulhos foram escolhidas por serem importantes centros comerciais e de alto tráfego automotivo, sendo as duas últimas sedes de dois aeroportos. A estação do Ibirapuera foi selecionada, pois apresenta comportamento atípico por estar em uma região arborizada, exclusivamente residencial, de baixo tráfego

veicular e ausência de indústrias. Já a estação Cubatão-Centro está inserida em uma localidade montanhosa, próxima ao nível do mar, altamente industrializada, onde o tráfego de veículos pesados exerce uma grande influência nos níveis de emissão de poluentes.

Figura 3.1 – Estações de monitoramento escolhidas para o estudo (CETESB, 2014).



Dois poluentes foram analisados, a concentração de óxidos de nitrogênio (NO_x) e os particulados inaláveis (MP₁₀). Particulados inaláveis são todas as partículas com menos de 10 micra de diâmetro e podem causar ou agravar diversos problemas de saúde. Os efeitos de seus níveis atmosféricos podem ocasionar desde tosse seca, cansaço, ardor nos olhos, nariz e garganta até sérios riscos de manifestação de doenças respiratórias e cardiovasculares e aumento das mortes prematuras em pessoas de grupos sensíveis (crianças, idosos e acometidos de doenças do aparelho respiratório e cardíaco) (CETESB, 2014; Vormittag e Saldiva, 2015; Dapper, 2016). Pessoas acometidas de doenças cardíacas, quando expostas à inalação de materiais particulados, podem apresentar dor torácica, palpitações, falta de ar e fadiga, ao que àqueles que apresentam doenças pulmonares podem

não ser capazes de respirar com a intensidade que normalmente fariam (Vormittag e Saldiva, 2015).

Em relação aos óxidos de nitrogênio (NOx), é uma forma genérica de chamar as misturas de ácido nítrico (NO) e dióxido de nitrogênio (NO₂). As fontes naturais de NOx são os incêndios florestais e relâmpagos e as fontes artificiais principais são a queima de combustíveis fósseis e de biomassa. Devido sua solubilidade em gordura, a exposição aos NOx causa edemas pulmonares causando, em situações extremas, a morte por asfixia.

A partir do descrito anteriormente, torna-se evidente a necessidade do estudo dos poluentes destacados e se justifica sua escolha como enfoque deste trabalho. A metodologia de obtenção das séries temporais está evidenciada no fluxograma da Figura 3.2.

Figura 3.2 – Fluxograma de obtenção das séries temporais (o autor, 2018).



A plataforma CETESB está hospedada sob o site <https://cetesb.sp.gov.br/ar/qualar/>. A consulta é gratuita e aberta, contudo é necessário que seja realizado um cadastro para gerar *login* e senha. Em relação às informações disponibilizadas pelo QUALAR, os dados são fornecidos como médias horárias da concentração de cada poluente. Após entrar no *Sistema de Informação da Qualidade do Ar* (QUALAR), é necessário acessar a opção “Consultas” no menu lateral e selecionar a opção “Exportar Dados Avançado”. A Figura 3.3 apresenta a tela de *download* do QUALAR. As informações necessárias para a obtenção dos dados são o período desejado, a estação e o parâmetro/poluente a ser estudado.

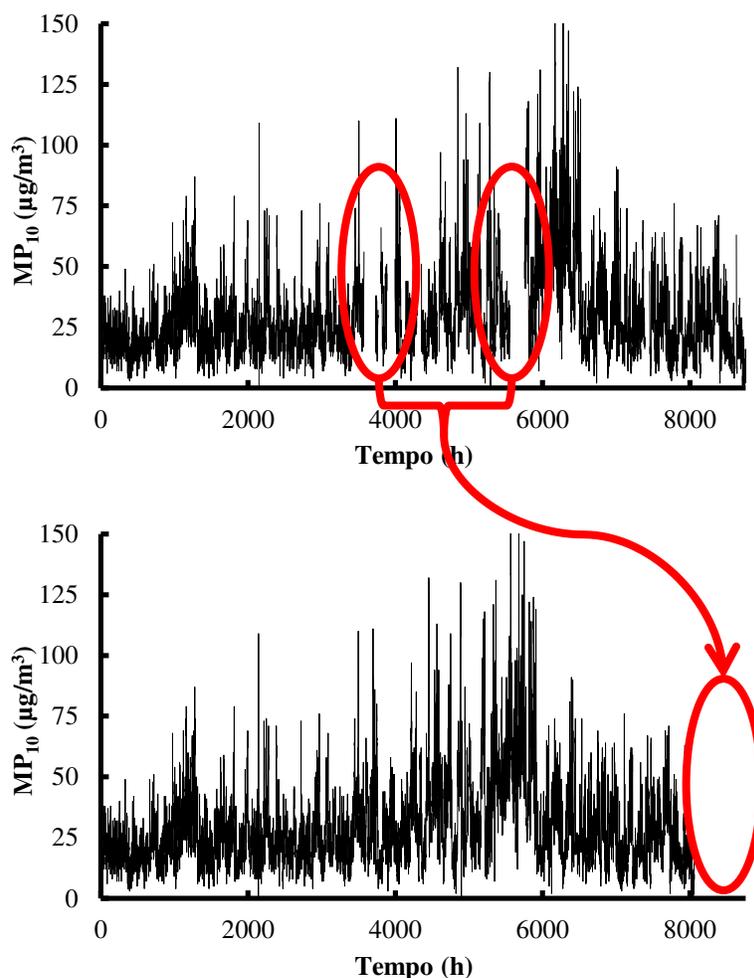
Figura 3.3 – Tela de aquisição de dados do sistema QUALAR (QUALAR CETESB: <http://qualar.cetesb.sp.gov.br>)

SISTEMA DE INFORMAÇÕES DA QUALIDADE DO AR		V 3.83
<ul style="list-style-type: none"> ESTAÇÃO PARÂMETRO ENTRADA DE DADOS CONSULTAS RELATÓRIOS SAIR 	Wesley Heleno Prieto ▾	
EXPORTAR DADOS AVANÇADO		
Tipo de Rede: Automática		
Data Inicial*:	<input type="text" value="01/01/2017"/>	Data Final*:
		<input type="text" value="31/12/2017"/>
Estação*:	<input style="border: 1px solid black;" type="text" value="Ibirapuera"/>	
PARÂMETROS (escolha no máximo 3)		
<input type="checkbox"/>	CO (Monóxido de Carbono)	
<input type="checkbox"/>	DV (Direção do Vento)	
<input type="checkbox"/>	DVG (Direção do Vento Global)	
<input type="checkbox"/>	MP10 (Partículas Inaláveis)	
<input type="checkbox"/>	MP2.5 (Partículas Inaláveis Finas)	
<input type="checkbox"/>	NO (Monóxido de Nitrogênio)	
<input type="checkbox"/>	NO2 (Dióxido de Nitrogênio)	
<input type="checkbox"/>	NOx (Óxidos de Nitrogênio)	
<input type="checkbox"/>	O3 (Ozônio)	
<input type="checkbox"/>	PRESS (Pressão Atmosférica)	
<input type="checkbox"/>	RADG (Radiação Solar Global)	
<input type="checkbox"/>	RADUV (Radiação Ultra-violeta)	
<input type="checkbox"/>	SO2 (Dióxido de Enxofre)	
<input type="checkbox"/>	TEMP (Temperatura do Ar)	
<input type="checkbox"/>	UR (Umidade Relativa do Ar)	
<input type="checkbox"/>	VV (Velocidade do Vento)	
<input type="button" value="Exportar"/>		

A plataforma exporta os dados em extensão *.csv* (que na prática pode ser aberto no *software Microsoft® Excel*) e, devido à imprecisões dos equipamentos de medição, problemas operacionais e instabilidades nas estações, é necessária a realização de um filtro de modo a eliminar células vazias oriundas de dados inválidos. Depois de filtrados, deve-se salvar a planilha em extensão *.exe* e agrupar os dados em um documento *.txt*. A extensão

.txt foi utilizada apenas no *upload* das séries no *software MATLAB R2016a*. A frequência de amostragem é de 1.0 hora, gerando, por ano, um total de 8.760 pontos. A filtragem dos dados é procedimento bastante simples. Depois de realizado o *download* da série, deve-se abrir o documento .csv no *software Microsoft® Excel* (ou equivalente) e na aba “*Dados*” aplicar a opção “*Filtro*” no rótulo da coluna retirando as células vazias e zeros. Esse procedimento tem algumas implicações. A Figura 3.4 exemplifica este processo, apresentado uma série temporal bruta e o mesmo conjunto de dados após a retirada de células vazias e zeradas.

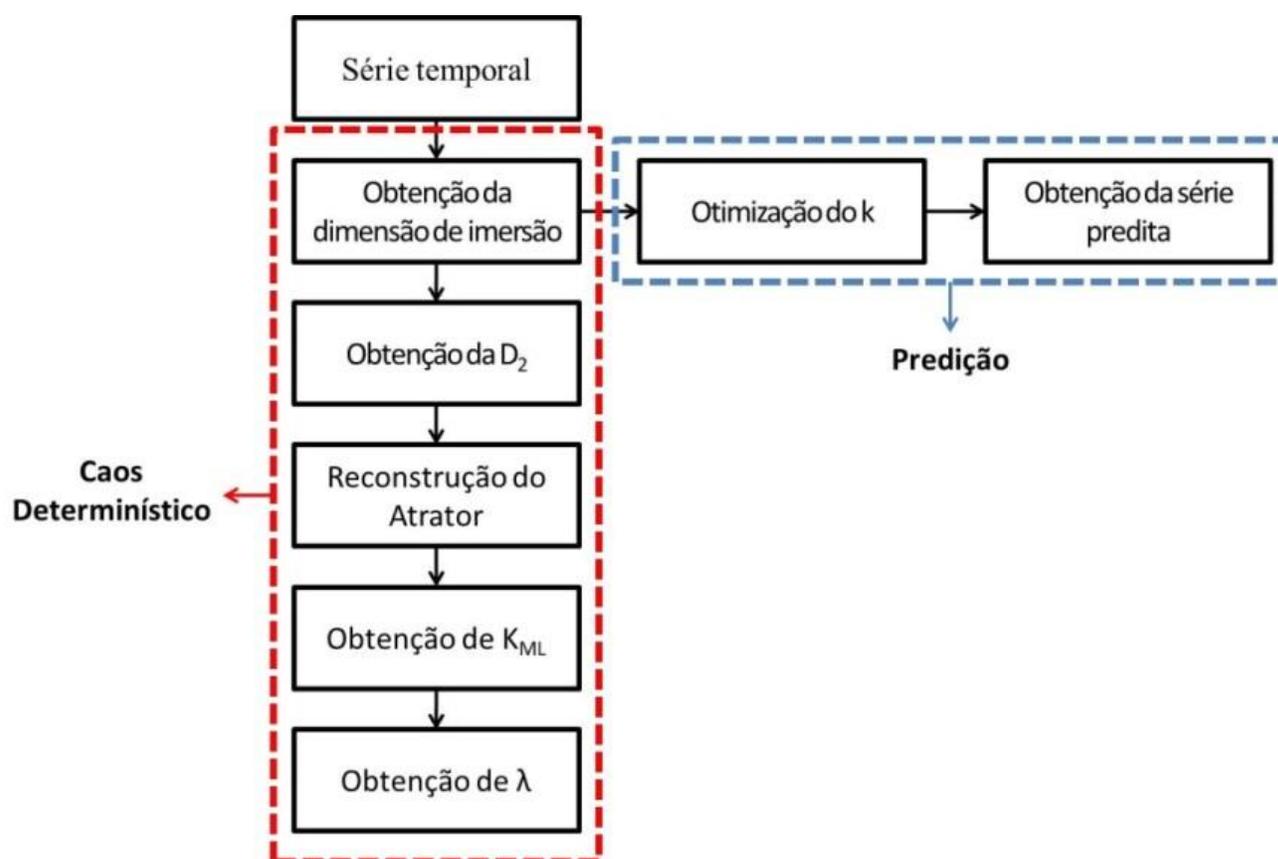
Figura 3.4 – Comparativo entre a série bruta e o mesmo conjunto de dados após a retirada de células vazias e zeradas (o autor, 2018).



Didaticamente, foram utilizados os dados da concentração horária de MP_{10} no ano de 2017 da estação de monitoramento de Congonhas. Na série bruta, verifica-se claramente a presença de lacunas de pontos que devem ser retiradas para tornar o conjunto contínuo. Desta forma, a falta de informações é lançada para o final da série, conforme procedimento explicitado na Figura 3.4 fazendo com que a série se torne contínua.

A Figura 3.5 apresenta o fluxo para a obtenção dos invariantes bem como da realização da predição.

Figura 3.5 – Fluxograma da análise do Caos determinístico e do método preditivo (o autor, 2018).

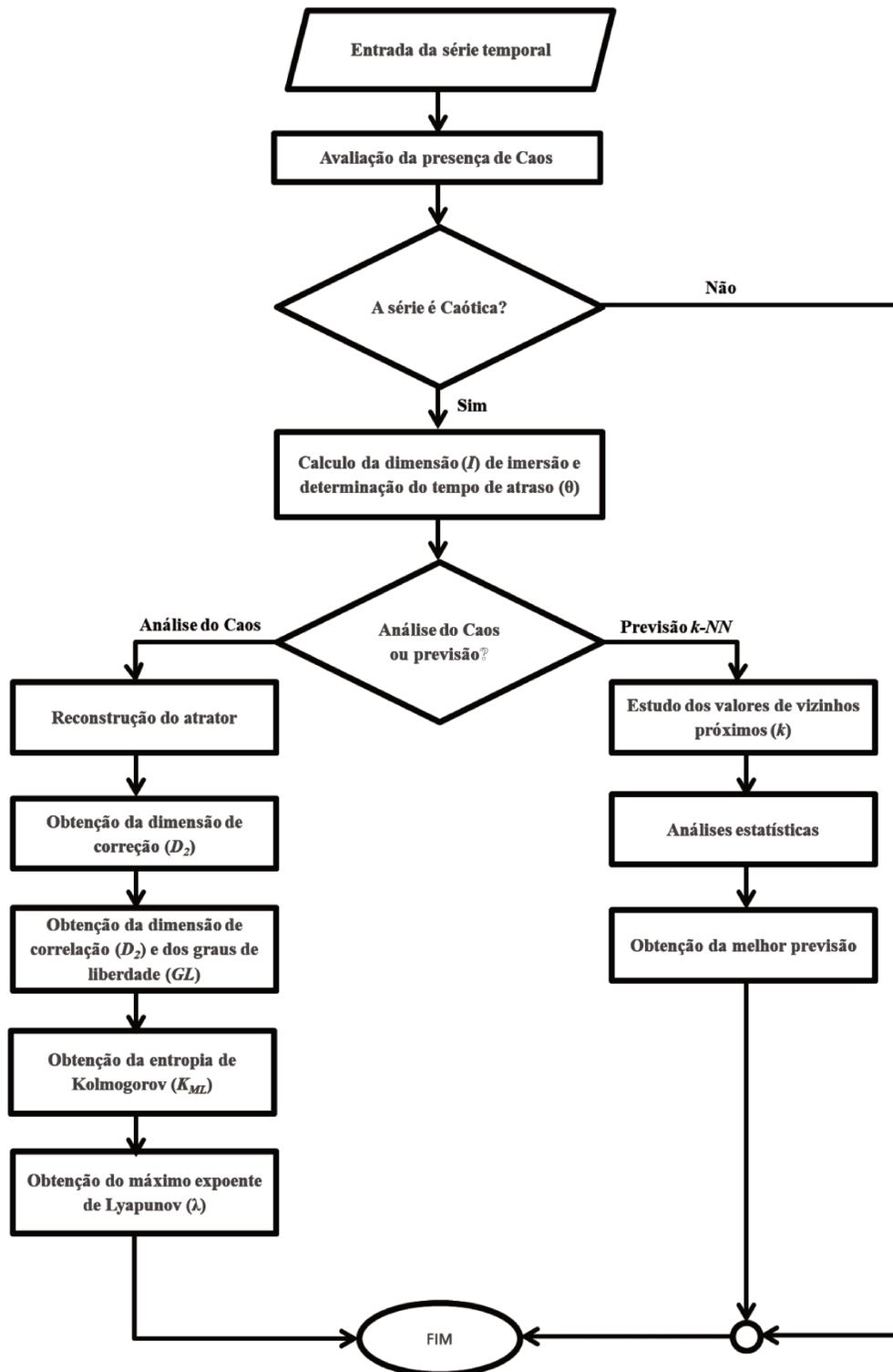


Depois de obtida a série em extensão *.txt*, passa-se à análise do Caos. O algoritmo que foi utilizado para a reconstrução dos atratores, cálculo da entropia de Kolmogorov,

dimensão de correlação e expoente de Lyapunov é o *Basic Resolution of Chaos (BRChaos)*, implementado por Moura (2013) e aprimorado pelo grupo de estudos do LPMP/DEPro/FEQ/UNICAMP (Prieto, 2014; Prieto *et al.*, 2015(a); Prieto *et al.*, 2015(b); Prieto *et al.*, 2016; Scatena *et al.*, 2016; Scatena *et al.*, 2017; Scatena, 2018) em um pacote computacional utilizando a linguagem da plataforma *Matlab*®. Os métodos encontrados no *BRChaos* são a decomposição em valores singulares (SVD) (Broomhead e King, 1986), máxima verossimilhança (Schouten *et al.*, 1994), Grassberger e Procaccia (1983) e de Rosenstein (1993), respectivamente para a obtenção dos atratores, entropia de Kolmogorov e dimensão de correlação e máximo expoente de Lyapunov. O apêndice A apresenta a comparação da eficiência do *BRChaos*, comprovando sua viabilidade. O software utilizado foi o *MATLAB R2016a* e o computador foi um *iMac Pro* modelo *Tela Retina 5K Mq2y2bz/A* processador *Intel Core 8, 3.2 Ghz, 32 Gb* de RAM e *HD* de *1 TB*. De modo a deixar explícita a metodologia adotada nesta Tese, o Fluxograma apresentado na Figura 3.6 especifica cada uma das etapas adotadas, bem como as tomadas de decisão. Por meio desta figura e da Figura 3.5, fica clara a proposta principal deste trabalho a de associar a Teoria do Caos, por meio da dimensão de imersão fornecida pelo algoritmo de Grassberger e Procaccia (1983), com o método *k-NN*. Esta fusão ocorre justamente no momento em que *I* é obtida por meio da metodologia do Caos e na sequência é utilizada como parâmetro para a geração das previsões.

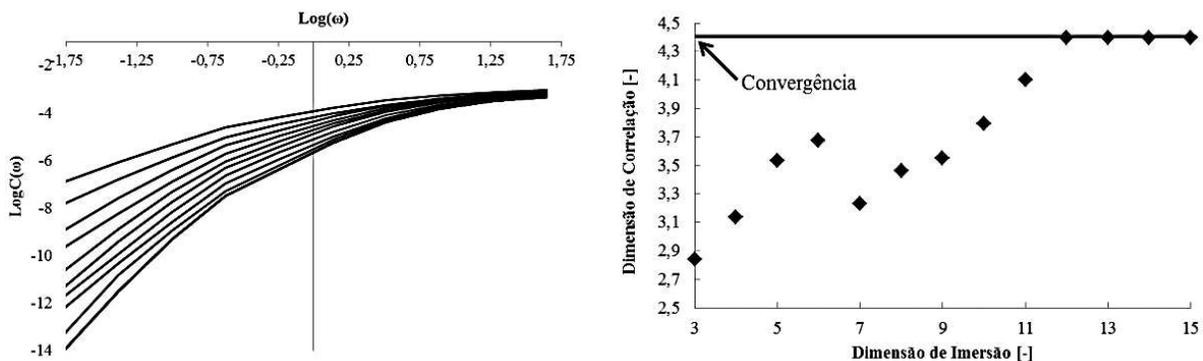
Como a estrutura de muitos atratores estranhos é heterogênea e algumas de suas regiões são mais visitadas que outras, torna-se necessário o cálculo de dimensões que considerem as não homogeneidades presentes no espaço de fases e, neste sentido, a dimensão de correlação é uma ferramenta que determina o número de graus de liberdade do sistema avaliado. Pelo método de Grassberger e Procaccia (1983) é imperativo o conhecimento do tempo de atraso (θ). Na prática, muitos autores utilizam a frequência de amostragem como o tempo de atraso (Glaser *et al.*, 2011; Yu *et al.*, 2013), sendo assim, no presente trabalho, utilizou-se $\theta = 1$ hora. A Figura 3.6 esboça um exemplo de como são obtidas a dimensão de imersão (*I*) e dimensão de correlação (D_2). Os valores de D_2 variam com aumento da dimensão de imersão, sendo indispensável a avaliação do valor de convergência deste invariante.

Figura 3.6 – Fluxograma detalhado das etapas de análise adotadas nesta Tese (o autor, 2018).



Exemplificando, a Figura 3.7 apresenta a relação da integral de correlação ($C(\omega)$) com o raio (ω) (lado esquerdo) e a convergência de D_2 com o aumento de I para dados genéricos. A saturação é o ponto de convergência e, portanto, os valores das dimensões de correlação de imersão.

Figura 3.7 – Convergência da dimensão de correlação e obtenção da dimensão de imersão (o autor, 2018).



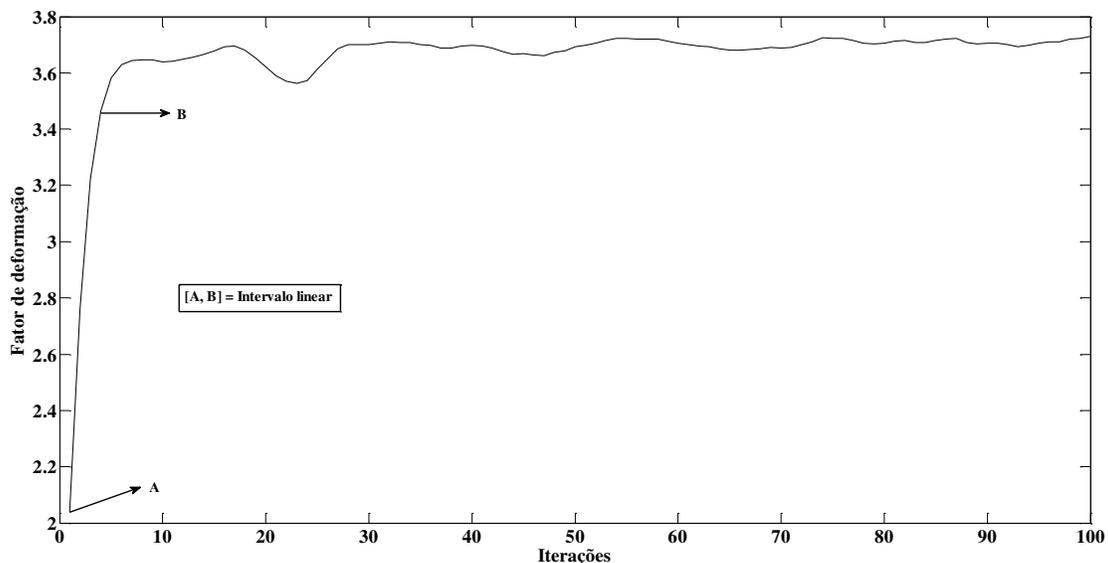
Para a avaliação da dinâmica caótica, utilizou-se, também, o máximo expoente de Lyapunov, método também encontrado no *BRChaos*. A abordagem para obtenção de λ pode ser observado na Figura 3.8.

O coeficiente angular da região linear do gráfico (fator de deformação pelo número de iterações) define o máximo expoente de Lyapunov. Aqui, novamente, utilizou-se o *software MATLAB R2016a* para o cálculo deste invariante.

Depois de determinados todos os invariantes caóticos, foram realizadas as previsões *out of sample*. Para tal, o comportamento das séries temporais relativas aos anos de 1998 à 2017 (série de treino) foram utilizadas para prever as concentrações horárias dos primeiros 60 dias de 2018 (1440 pontos), dado de que, conforme salientado na Revisão Bibliográfica, não é recomendada a geração de pontos preditivos *out of sample* que ultrapassem 10 – 20% do tamanho da série de treino (Hyndman e Athanasopoulos, 2014). Sendo assim, neste caso, o método do vizinho mais próximo realiza a varredura na série de treino, em busca de padrões, sem se preocupar com o tempo, gerando, por fim, a previsão de acordo com o

número de pontos escolhidos. O algoritmo utilizado nesta tese foi o k -NN adaptado por Perlin e Ceretta (2007). O valor de k (número de vetores que serão utilizados na previsão) é um parâmetro importante e deve ser determinado. Nesta Tese, variou-se o valor de k até que a melhor previsão fosse atingida. Outro parâmetro importante é a dimensão de imersão (I), determinada e otimizada previamente por meio da análise do Caos. A Figura 3.9 traz a integração do método de Grassberger e Procaccia (1983) (Teoria do Caos) com o algoritmo k -NN.

Figura 3.8 – Metodologia de obtenção do máximo expoente de Lyapunov (o autor, 2018).



O algoritmo k -NN depende da determinação da dimensão de imersão, nesta Tese, esta etapa foi realizada por meio da Teoria do Caos, conforme observado na Figura 3.9. A integração do caos determinístico com k -NN acontece no momento em que a dimensão de correlação (D_2) é determinada. De acordo com o procedimento descrito na Figura 3.7, ao se atingir a saturação, são determinados tanto I (dimensão de imersão) quando D_2 , o valor da dimensão de imersão é o mesmo utilizado no algoritmo preditivo. Uma vez obtida a série prevista, realiza-se a comparação com a série real.

No que tange às análises de médias e desvios padrão utilizadas nesta Tese, foram, para todos os casos, aplicadas as Equações 3.1 e 3.2. Todos os cálculos estatísticos foram realizados no *software Microsoft® Excel*.

Figura 3.9 – Integração da Teoria do Caos com a abordagem *k*-NN (o autor, 2018).

Código de D_2 (Grassberger e Procaccia, 1983)

```

1  %Dimensão de Correlação e Entropia de Kolmogorov
2  %pelo algoritmo Grassberger e Procaccia (1983a, 1983b).
3  %Entrada: A = série temporal
4  %      ts = tempo de amostragem
5  %      tau = tempo de atraso fornecido pela informação mútua
6  %      dim = dimensão de imersão
7  %Saída: Gráfico da Dimensão de correlação
8  %      Valor da Entropia de Kolmogorov e Dimensão de correlação
9  %*****
10 - clear all
11 - clc
12 - %*****
13 - %Definir os dados de entrada
14 - A=dlmread('TOTAL.txt');
15 - ts=1;
16 - tau=1;
17 - dim=13;
18 - %*****
19 - [N,M]=size(A);
20 - media=mean(A);
21 - desv=std(A);
22 - a=(A-media)/desv; % normalização da série
23 - lo=std(a)*0.66; % Comprimento de corte
24 - %*****
25 - %Definição do número de Bins

```

Código k-NN (Perlin e Ceretta, 2007)

```

1  % Example Script for NN
2
3  % This script will model a time series using the two different methods
4  % and then plot the forecasts versus real points
5
6  - addpath('m_Files');
7
8  - load Example_Data.mat;
9
10 - x=(TOTALMP10CONG); % Load data from .mat file
11 - d=15531; % Defines where to start the forecasts (and also the training data (1:d-1))
12 - m=13; % Size of histories (embedding dimension)
13 - k=2; % Number of nearest neighbors to use in the forecast's calculation
14
15 - %method_1='correlation';
16 - method_2='absolute_distance';
17
18 - %[OutSample_For_Corr,InSample_For_Corr,InSample_Res_Corr]=nn(x,d,m,k,method_1);
19
20 - [OutSample_For_Abs,InSample_For_Abs,InSample_Res_Abs]=nn(x,d,m,k,method_2,168);
21
22 - plot([x(d+1:end),InSample_For_Corr,InSample_For_Abs]);
23 - xlabel('Time');
24 - ylabel('Values');
25 - title(['Forecasts VS Real values (m=',num2str(m),' , k=',num2str(k),' d=',num2str(d),'')]);
26 - legend('Ibov Real Series','NN forecast with correlation Method','NN forecast with abs distance Method');
27
28 - rmpath('m_Files');

```

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad (3.1)$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{(n - 1)}} \quad (3.2)$$

Em que:

\bar{x} é o valor médio da amostra;

σ é o desvio padrão amostral;

x_i é cada um dos elementos da amostra;

n é o tamanho da amostra.

CAPÍTULO IV

RESULTADOS E DISCUSSÕES

O objetivo central desta tese é a aplicação da Teoria do Caos determinístico para avaliar a presença de não linearidades em séries temporais de poluentes atmosféricos e, em seguida, associar o caos, por meio da dimensão de imersão, à inteligência artificial objetivando realizar previsões de concentração horária da poluição do ar. O presente capítulo apresenta este estudo para cinco diferentes estações. Em cada subcapítulo são avaliadas as concentrações anuais e também as variações horárias de NO_x e MP₁₀. Ao final, o comportamento das séries temporais de 1998 a 2017 foi utilizado na previsão da concentração horária de cada poluente para os primeiros dois meses (total de 1440 pontos) do ano de 2018.

4.1 ESTAÇÃO DO IBIRAPUERA

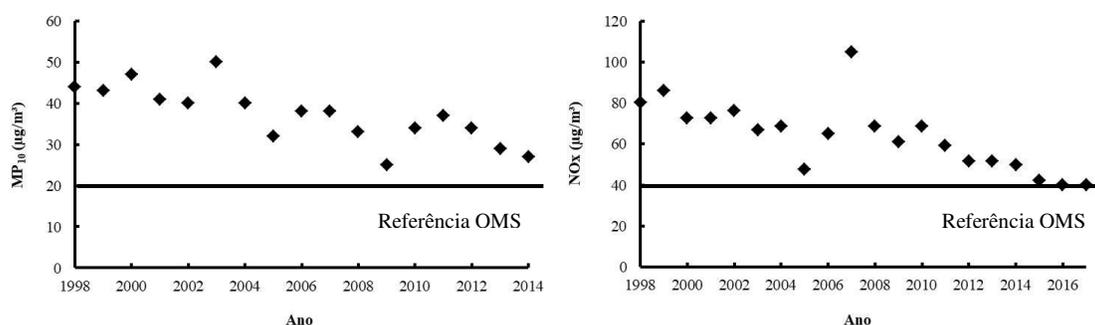
Inicialmente realizou-se a avaliação da variação da concentração anual média (Equação 3.1) dos poluentes NO_x e MP₁₀ na estação do Ibirapuera. A Figura 4.1 apresenta este panorama. Para auxiliar na interpretação dos gráficos da Figura 4.1, na Tabela 4.1 são apresentados os dados discretos das médias anuais de MP₁₀ e NO_x bem como os valores do desvio padrão de cada ponto. As casas decimais das médias dos dados foram ajustadas de acordo com a ordem de grandeza do desvio padrão (Montgomery, 2016). Tanto na Figura quanto na Tabela 4.1, comparando o perfil apresentado nos gráficos, é clara a redução dos níveis anuais de ambos os poluentes, chegando a 38,7% e 50% (quando comparado o primeiro com o último ano disponibilizado), respectivamente para o MP₁₀ e NO_x.

Diversos motivos influenciam essa redução, como a renovação da frota veicular paulistana, implementação de políticas públicas e etc, porém, nenhuma destas situações tem impacto direto na queda dos níveis de poluição atmosférica obtidos desta estação de

monitoramento da CETESB, dado fato de que nesta localização o fluxo de veículos (fontes móveis) é baixo, não existem indústrias (fontes fixas), sendo uma região puramente residencial e muito arborizada. Ainda assim, observa-se que os níveis de MP_{10} de todos os anos ultrapassaram a média anual de $20 \mu\text{g}/\text{m}^3$ recomendada pela OMS.

No que diz respeito aos NO_x , a OMS estipula o máximo de $40 \mu\text{g}/\text{m}^3$, neste sentido apenas em 2 anos (a saber, 2016 e 2017) não se observam ultrapassagens na estação do Ibirapuera.

Figura 4.1 – Variação da concentração anual de média de MP_{10} e NO_x na estação do Ibirapuera dos anos de 1998 à 2017 (o autor, 2018).



Sobre o desvio padrão, os valores obtidos são altos, da ordem da própria média e, no caso do NO_x , maiores que a μ , demonstrando o caráter disperso dos níveis horários que compõe o resultado médio. No caso do MP_{10} , a estação só realizou a aquisição de dados deste poluente até o ano de 2014.

Partindo da análise geral dos dados, intentou-se avaliar o caráter não linear da evolução temporal da concentração de MP_{10} e NO_x na estação do Ibirapuera. Devido à falta de dados em alguns anos, como por exemplo, o ano de 2014, os resultados bem como sua avaliação sofrem influência direta destas lacunas, pois tanto a abordagem do Caos, como a estatística, dependem diretamente do número de pontos da série. A Figura 4.2 apresenta as séries temporais e os atratores de concentração anual de MP_{10} dos anos de 1998 a 2014 da estação do Ibirapuera.

Tabela 4.1 – Médias anuais e desvios padrão da concentração de MP_{10} e NO_x na estação do Ibirapuera (o autor, 2018).

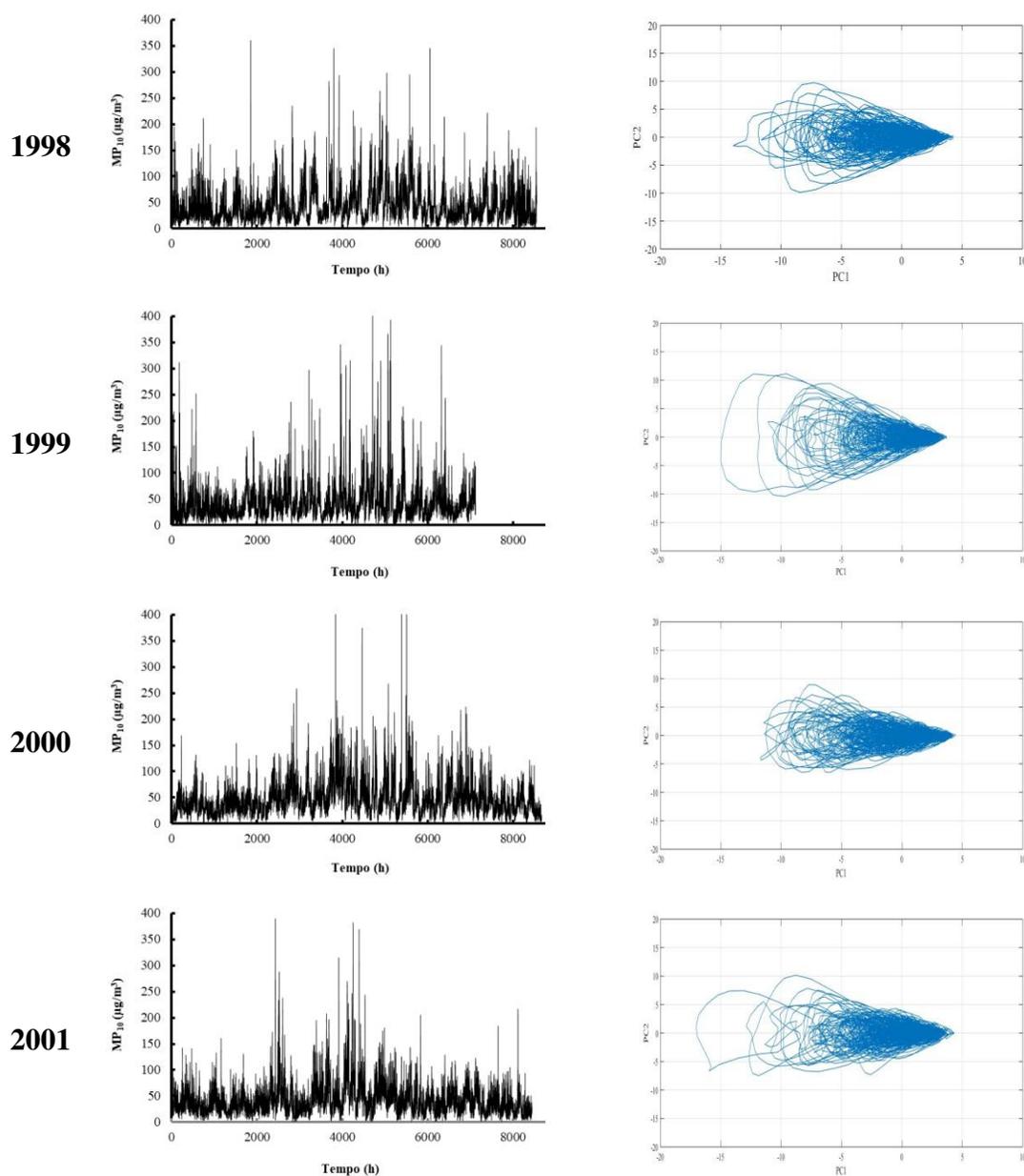
Ano	MP_{10}		NO_x	
	μ ($\mu\text{g}/\text{m}^3$)	σ ($\mu\text{g}/\text{m}^3$)	μ ($\mu\text{g}/\text{m}^3$)	σ ($\mu\text{g}/\text{m}^3$)
1998	44	33	80	97
1999	43	36	86	122
2000	47	34	73	90
2001	41	30	73	94
2002	40	30	76	105
2003	50	42	67	92
2004	40	31	69	88
2005	32	23	48	59
2006	38	29	65	69
2007	38	30	105	145
2008	33	25	69	92
2009	25	18	61	73
2010	38	34	69	90
2011	37	27	56	78
2012	34	26	52	55
2013	29	21	52	61
2014	27	17	50	61
2015	-	-	42	46
2016	-	-	40	46
2017	-	-	40	44

As séries temporais foram confeccionadas utilizando os dados tratados exportados diretamente da plataforma da CETESB. Devido à presença de dados inválidos em todos os anos, o número de pontos, em nenhum dos casos, chegou a 8760 (aquisição de 1 ponto por hora por 365 dias). Sendo assim, foi necessário um agrupamento dos dados de modo a não permitir descontinuidades no gráfico, pois para a aplicação do Caos determinístico as séries em estudo precisam ser contínuas. Desta forma, as lacunas foram lançadas no final da série.

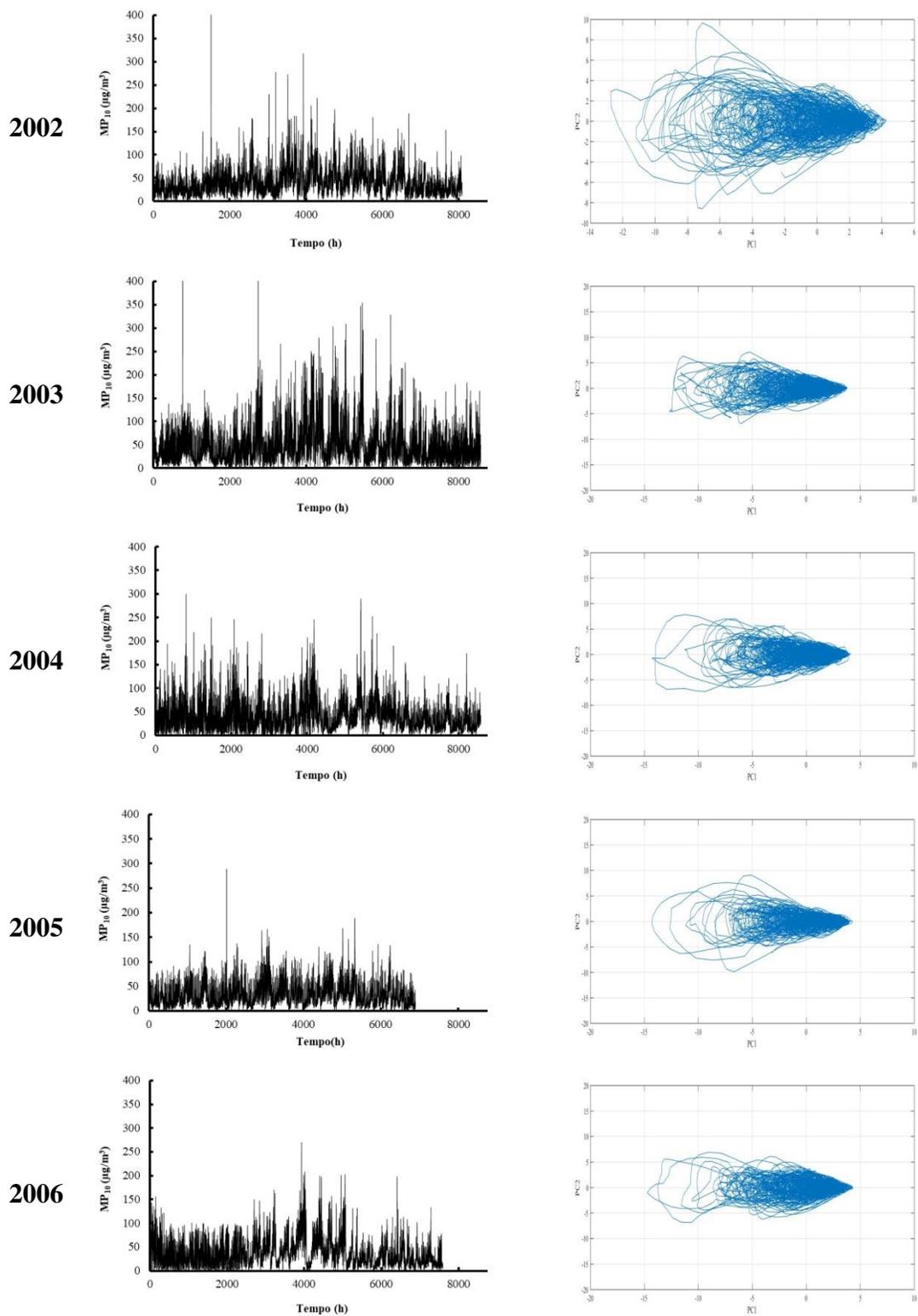
De modo geral, ao se observar a evolução temporal da $[MP_{10}]$ ($\mu\text{g}/\text{m}^3$) em todos os anos, nota-se grande variabilidade de caráter bastante dispersivo em relação à média. Isso ocorre, principalmente, pelas subjacentes características do fenômeno da poluição atmosférica, pois este apresenta sazonalidades que fazem com que, dependendo da época

do ano, os valores pontuais de concentração de MP_{10} fiquem mais ou menos concentrados. Neste sentido, parâmetros como volume e frequência das chuvas, velocidade e direção dos ventos, interferências locais e etc, influenciam diretamente os dados obtidos, fazendo com que cada ponto seja um sistema individual, passivo de estudo, e bastante sensível.

Figura 4.2 – Séries temporais e atratores de concentração anual MP_{10} obtidas na estação de monitoramento do Ibirapuera nos anos de 1998 a 2014 (o autor, 2018).

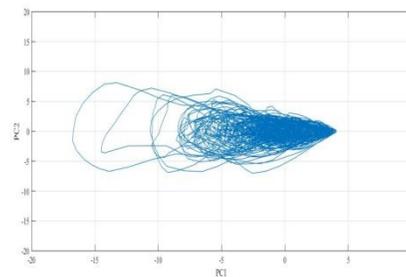
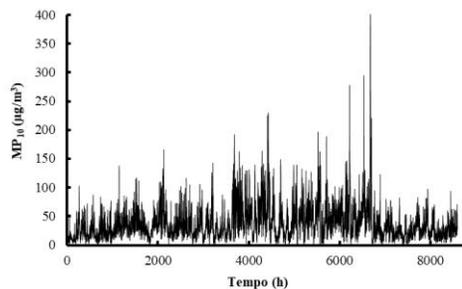


Capítulo IV – Resultados e Discussões

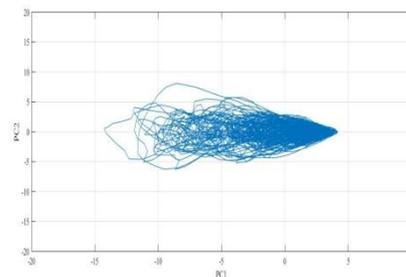
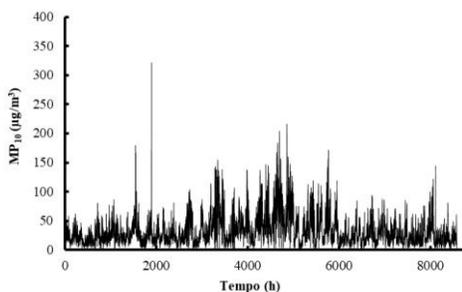


Capítulo IV – Resultados e Discussões

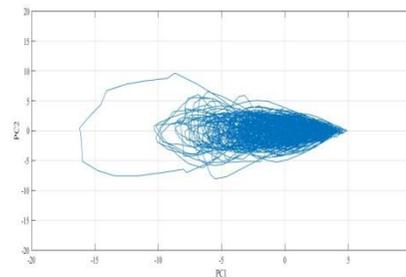
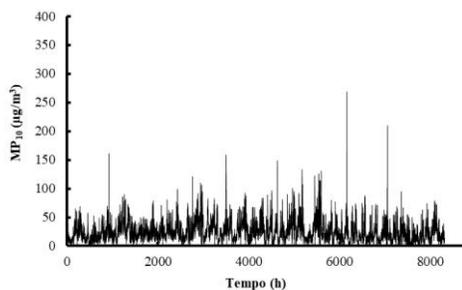
2007



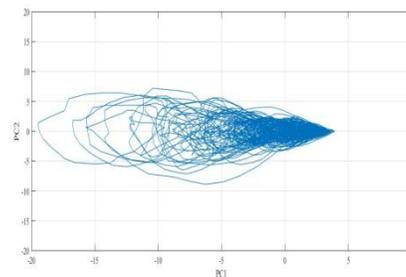
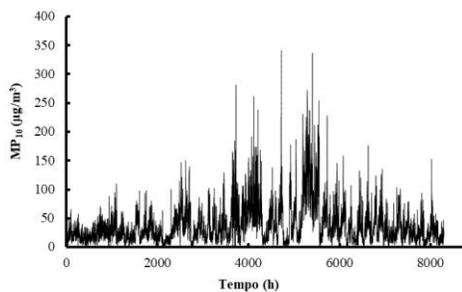
2008



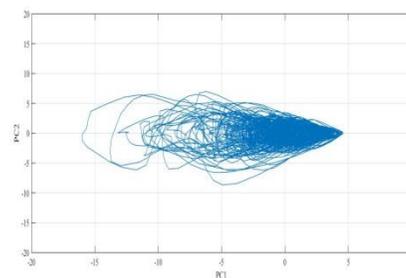
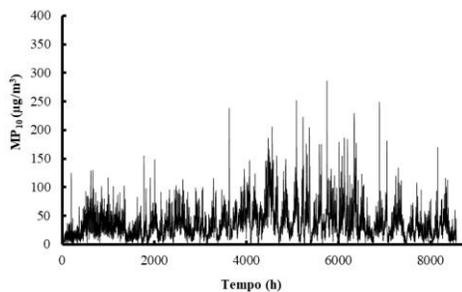
2009

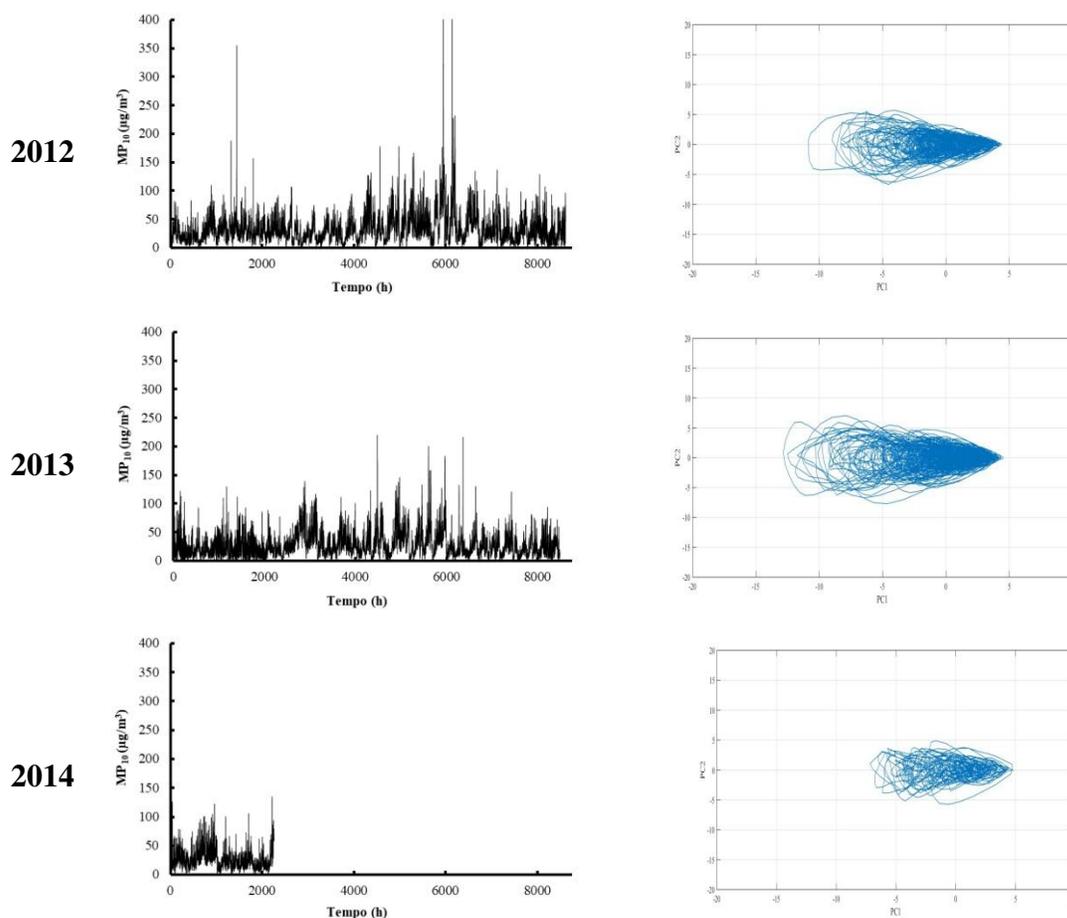


2010



2011

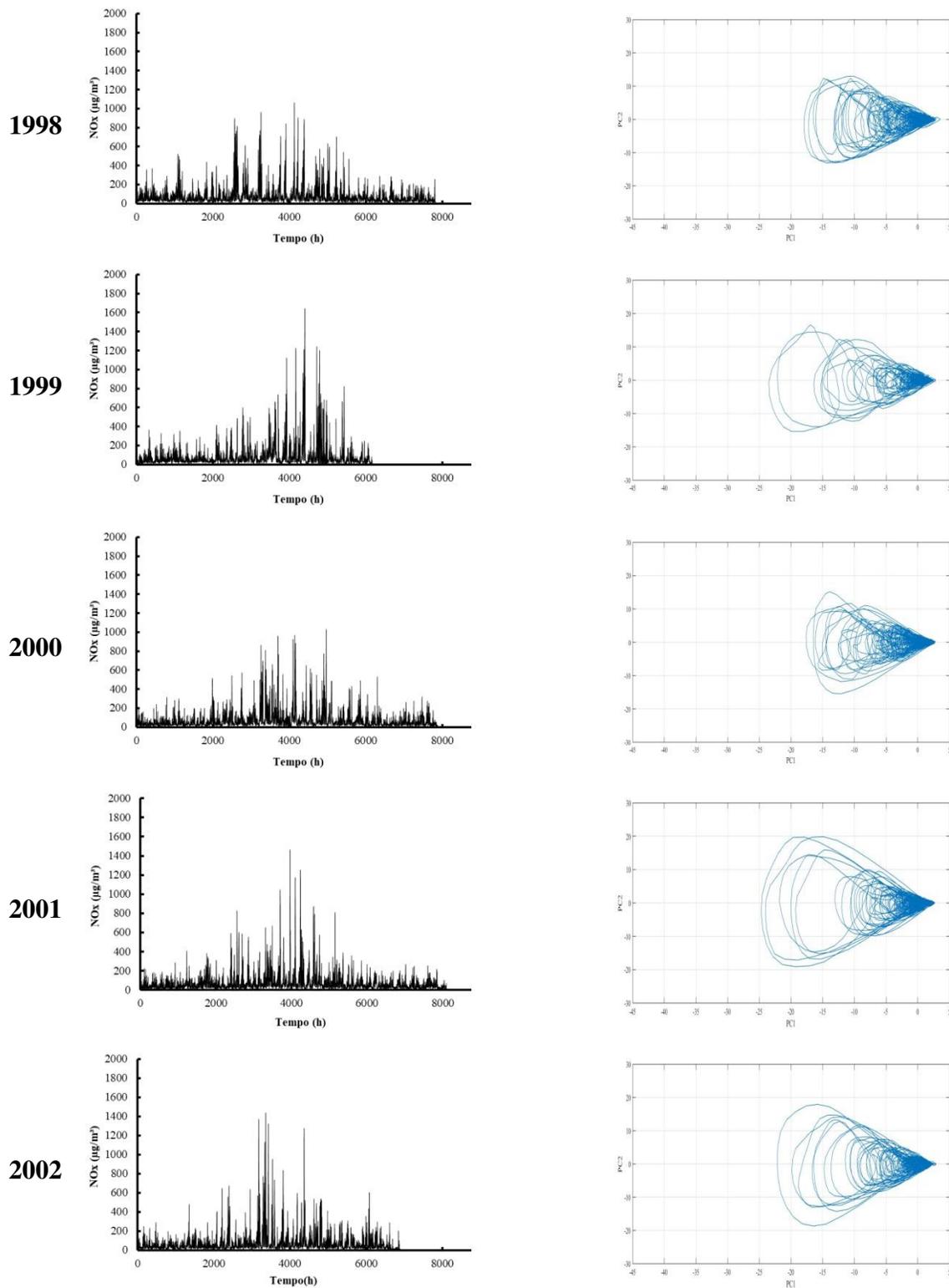




A reconstrução dos atratores no espaço de fases foi realizada a partir das respectivas séries temporais, contando com o mesmo número de pontos, dimensão de imersão (I) otimizada igual a 14 e tempo de atraso (θ) de 1 hora. Para todas as figuras reconstruídas, nota-se que as linhas de fluxo, para a concentração anual de MP_{10} , resultam em atratores estranhos, convergindo para uma região bem definida e estreita no espaço de fases (linha horizontal em PC2 (principal componente) = 0), sugerindo uma dinâmica caótica determinista.

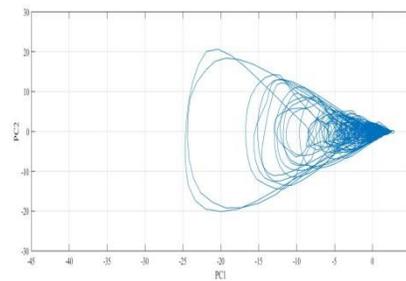
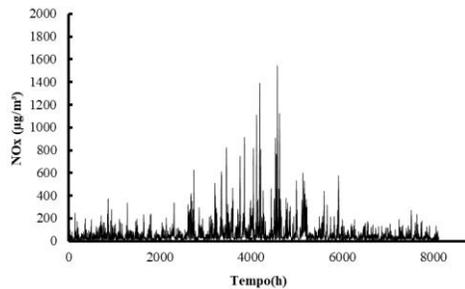
Em mãos da mesma metodologia, foram obtidas as séries temporais e reconstruídos os atratores da concentração anual de NO_x na estação de monitoramento do Ibirapuera. Para este poluente, a plataforma QUALAR disponibiliza os dados de 1998 até a presente data. A Figura 4.3 traz estas evoluções temporais e os respectivos atratores.

Figura 4.3 – Séries temporais e atratores de concentração anual NOx obtidas na estação de monitoramento do Ibirapuera nos anos de 1998 a 2017 (o autor, 2018).

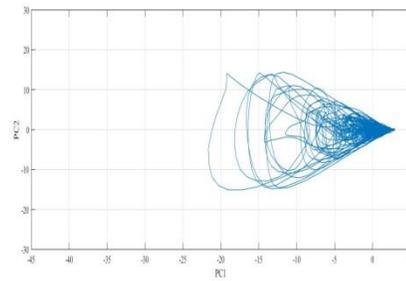
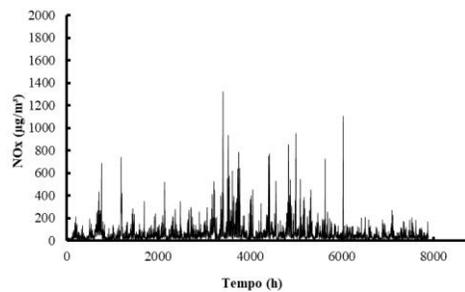


Capítulo IV – Resultados e Discussões

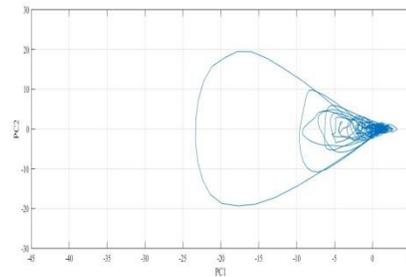
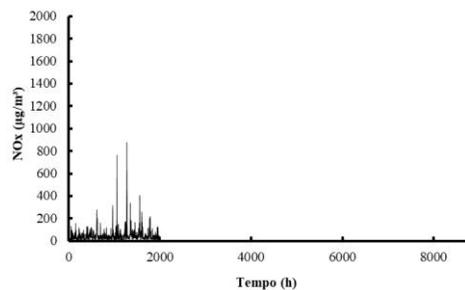
2003



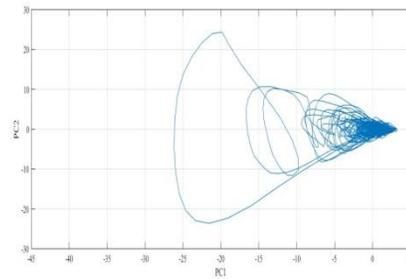
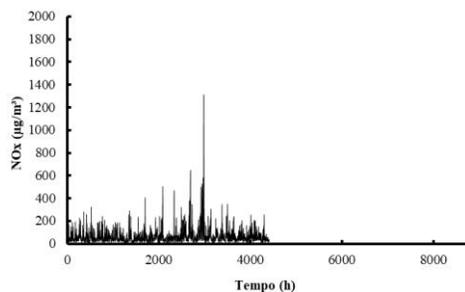
2004



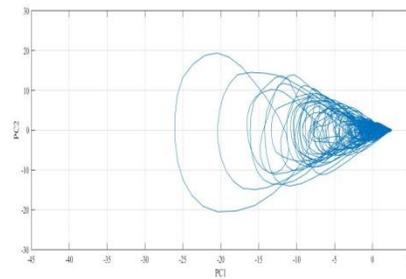
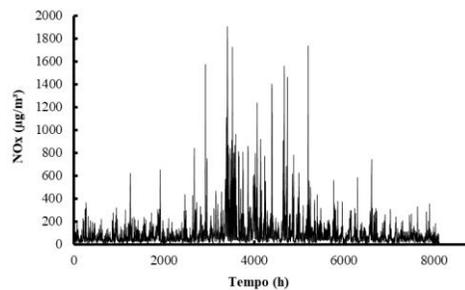
2005



2006

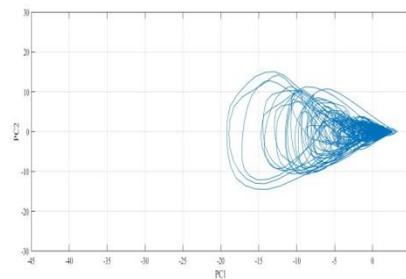
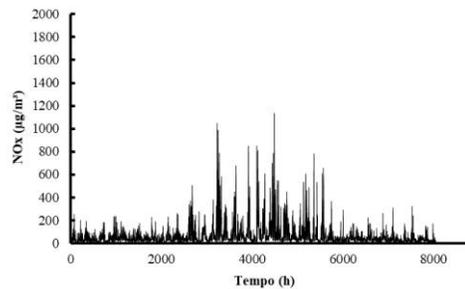


2007

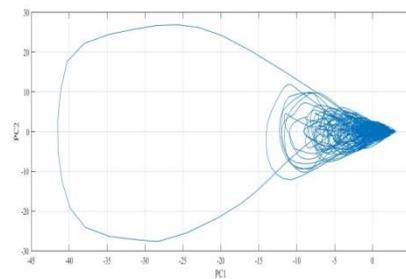
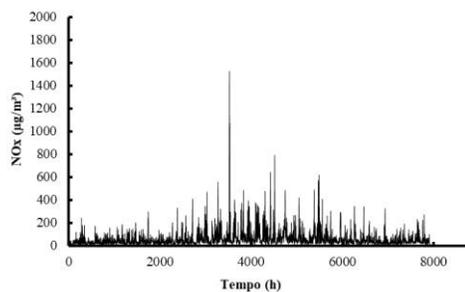


Capítulo IV – Resultados e Discussões

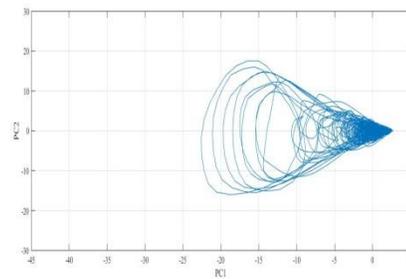
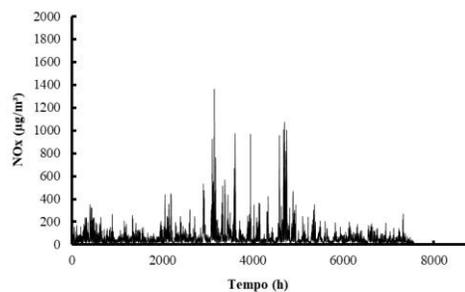
2008



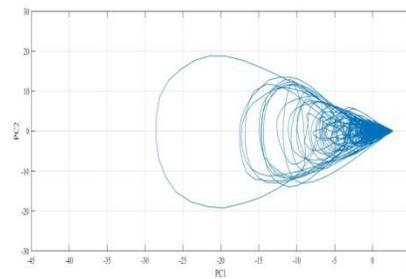
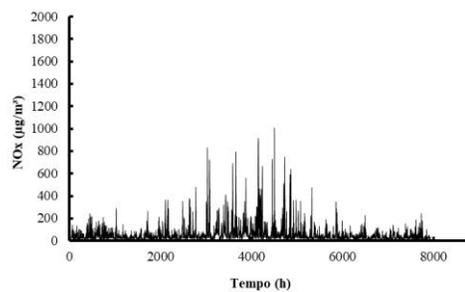
2009



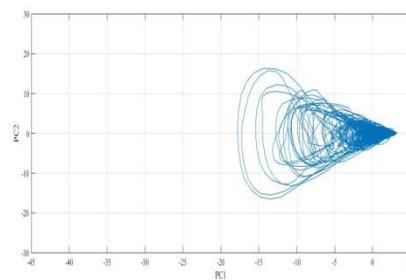
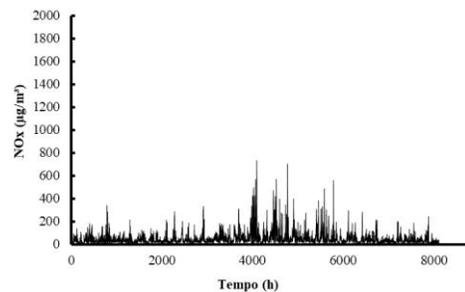
2010



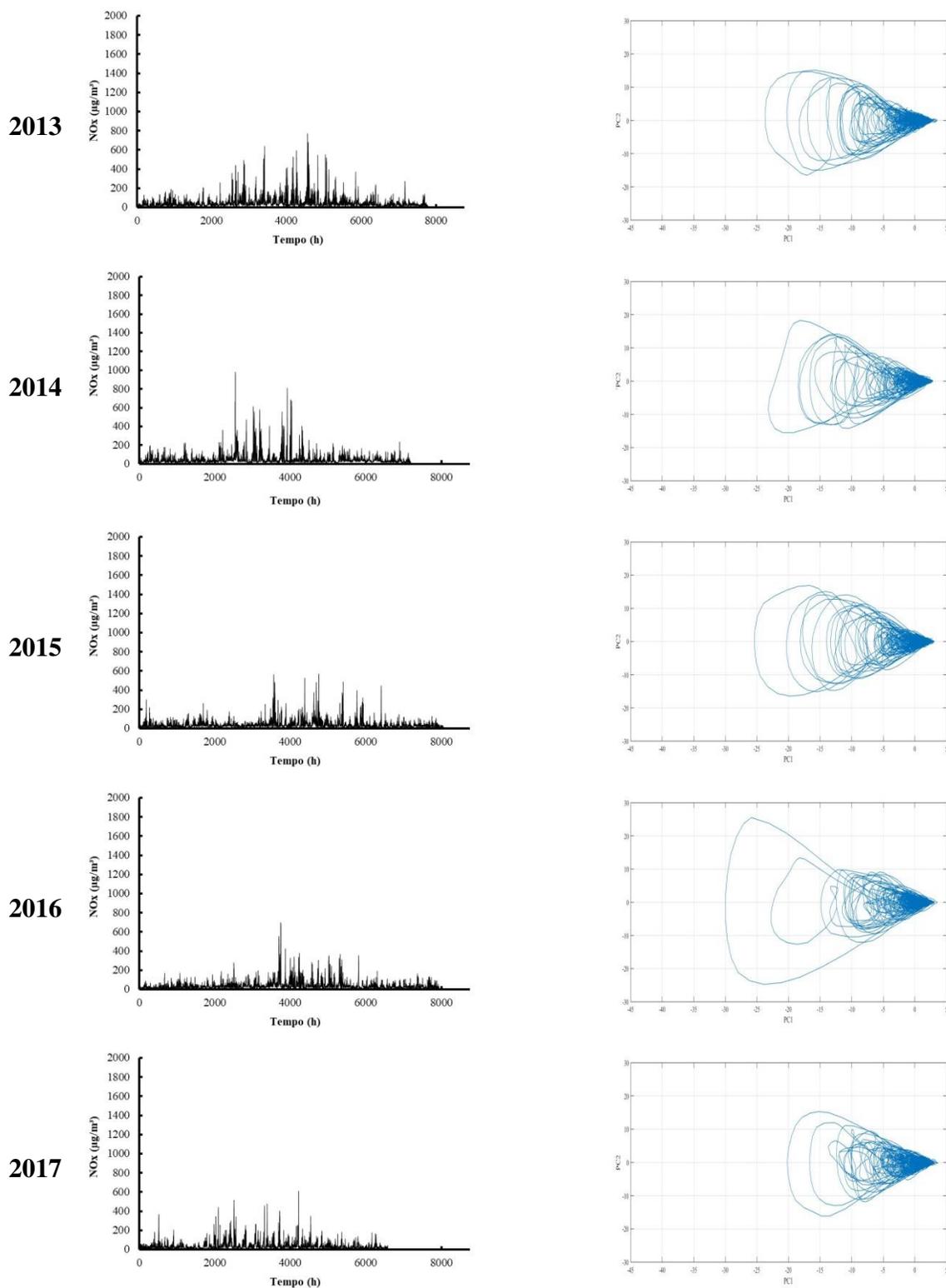
2011



2012



Capítulo IV – Resultados e Discussões



Pelos mesmos motivos supramencionados, observa-se a falta de pontos, relativos aos dados inválidos, em todos os anos avaliados, fato que obriga o agrupamento das

informações, mantendo a série contínua, por meio do deslocamento das lacunas para o fim da evolução temporal. Novamente, verifica-se um perfil bastante variável, dispersivo em relação ao valor médio e, para alguns casos, uma clara concentração de altos valores no meio da evolução. Este período coincide com o inverno, época em que, devido ao baixo volume das chuvas, ocorre a maior concentração dos poluentes atmosféricos.

Similar ao observado para o MP_{10} , os atratores reconstruídos para a $[NO_x]$ resultaram em dinâmica caótica, também convergindo para uma região bem definida e estreita no espaço de fases, asseverando uma possível dinâmica do Caos determinístico. Neste caso, a dimensão de imersão utilizada foi 19 e o tempo de atraso de 1 hora.

Quando se realiza a comparação entre os atratores reconstruídos para a $[MP_{10}]$ com os obtidos para a $[NO_x]$, verifica-se uma evolução mais clara e comportada para o NO_x , ao passo que para os MP_{10} a dinâmica se apresenta mais confusa, concentrada, sugerindo, portanto, uma evolução menos linear, logo, mais caótica para os particulados inaláveis. Dirimindo essa dúvida, a seguir são apresentados os invariantes do Caos.

A Tabela 4.2 traz os valores de dimensão de imersão (I), dimensão de correlação (D_2), entropia de Kolmogorov (K_{ML}) e o máximo expoente de Lyapunov para as concentrações anuais de MP_{10} e NO_x na estação de monitoramento da CETESB.

Devido às disformidades presentes no espaço de fases reconstruído, é importante que se avalie o grau de heterogeneidade do atrator para explicitar a complexidade global do sistema. Atualmente a grandeza mais utilizada para este fim é a dimensão de correlação (D_2) e justifica-se esta escolha por dois motivos: o primeiro se dá pela dificuldade para resolver numericamente as dimensões fractais, pois a metodologia que permite sua obtenção se baseia no algoritmo de contagem de caixas que, por sua vez, necessita de grande esforço computacional. Outro fator preponderante é que, uma vez calculada D_2 , pode-se obter o número de variáveis necessárias para descrever o comportamento caótico (Monteiro, 2011). Para a obtenção de D_2 , por meio da aplicação da metodologia de Grassberger e Procaccia (1983), conforme salientado anteriormente, é necessário o

conhecimento do tempo de atraso (θ). Com este fim, vários estudos versam sobre a melhor metodologia de determinação deste parâmetro. Alguns sugerem a utilização da informação mútua (Fraser e Swinney, 1986), função de auto correlação (Holzfuss e Mayer-Kress, 1985), integral de correlação (Yu *et al.*, 2011; Yu *et al.*, 2013). Na prática, tem-se utilizado o tempo de atraso igual à frequência de dados (Yu *et al.*, 2011; Yu *et al.*, 2013). Por via de consequência, no presente assume-se $\theta = 1$ hora.

Tabela 4.2 – Invariantes do Caos para os poluentes MP_{10} e NO_x avaliados na estação de monitoramento do Ibirapuera dos anos de 1998 a 2017 (o autor, 2018).

Ano	MP ₁₀					NO _x				
	I	D ₂ [-]	GL	K _{ML} (bits/s)	λ [-]	I	D ₂ [-]	GL	K _{ML} (bits/s)	λ [-]
1998	14	3,15	4	0,0864	0,11	19	3,78	4	0,0566	0,09
1999	12	4,40	5	0,0923	0,11	20	3,68	4	0,0457	0,09
2000	13	3,35	4	0,1217	0,11	18	3,27	4	0,0537	0,09
2001	12	3,95	4	0,1166	0,11	18	4,05	5	0,0536	0,10
2002	12	3,43	4	0,1278	0,11	19	3,10	4	0,0606	0,10
2003	9	4,21	5	0,1999	0,10	20	3,61	4	0,0511	0,09
2004	10	3,63	4	0,2912	0,10	18	3,87	4	0,0445	0,10
2005	11	3,93	4	0,0159	0,11	15	2,76	3	0,0838	0,07
2006	11	4,04	5	0,0882	0,10	15	3,51	4	0,0929	0,09
2007	16	3,48	4	0,0830	0,12	19	3,76	4	0,0411	0,10
2008	17	3,62	4	0,1246	0,12	21	3,48	4	0,0405	0,09
2009	14	3,85	4	0,1058	0,12	18	4,37	5	0,0481	0,10
2010	19	4,17	5	0,0796	0,11	18	3,96	4	0,0403	0,09
2011	16	3,57	4	0,1034	0,12	20	4,11	5	0,0384	0,09
2012	14	3,90	4	0,0939	0,12	18	3,17	4	0,0509	0,10
2013	12	4,04	5	0,0969	0,11	19	4,09	5	0,0440	0,09
2014	-	-	-	-	-	20	3,94	4	0,0495	0,09
2015	-	-	-	-	-	18	4,00	4	0,0578	0,09
2016	-	-	-	-	-	17	3,06	4	0,0515	0,09
2017	-	-	-	-	-	19	3,21	4	0,0559	0,09

Para a determinação da dimensão de correlação é importante o dimensionamento da dimensão de imersão, assim a Figura 3.5 esboçou um exemplo da metodologia admitida no cálculo de D_2 e I . Nas séries temporais obtidas, o valor da dimensão de correlação cresce até um ponto em que a taxa de aumento começa a diminuir e os valores de D_2 atingem a

saturação. Este ponto de saturação aponta, precisamente, a presença de uma dinâmica determinista (Yu *et al.*, 2013). O maior número inteiro mais próximo indica a quantidade de variáveis, ou graus de liberdade (GL) que regem a dinâmica do sistema (Yu *et al.*, 2013). Além disso, na evolução da dimensão de correlação, baixas dimensões são indicativos da formação de atratores do tipo ponto fixo no diagrama de fases. Após este ponto, ocorre formação de estruturas do tipo ciclo limite e o aumento de D_2 indica a presença de instabilidades causadoras de duplicação de período, resultando em dinâmica caótica para as dimensões maiores que 2 e 3.

No primeiro bloco de cada poluente da tabela (I , D_2 e GL), verifica-se o Método da Integral de Correlação em que consta a dimensão de correlação, I de convergência e número de variáveis para as concentrações anuais de MP_{10} e NO_x . Observa-se, para os dois poluentes, que o número de variáveis dinâmicas apresenta valores anuais diferentes, ou seja, não há unicidade no que diz respeito à evolução dos sistemas avaliados. Em termos de fenômeno, as séries temporais de concentração de poluentes atmosféricos são diretamente afetadas por fatores naturais e antropogênicos que variam constantemente, tanto em frequência quanto em magnitude, e impactam decisivamente na interação do poluente-ambiente e, portanto, sua mensuração, culminando em diferentes dinâmicas temporais, mais ou menos complexas, incrementando dificuldade na identificação destas forças motrizes e de suas contribuições presentes na série temporal (Yu *et al.*, 2013). Lee e Lin (2008) observaram o valor de 4,32 para D_2 quando avaliaram as concentrações de MP_{10} em Chung-Shan, Taipei (Taiwan), de Janeiro de 1998 à Dezembro de 1998. Para os dados de MP_{10} avaliados na estação de monitoramento do Ibirapuera dos anos de 1998 a 2017 os valores de D_2 variaram de 3,15 a 4,40, resultando, nos anos de 1999, 2003, 2006, 2010 e 2013 em 5 graus de liberdade, mesma informação reportada por Lee e Lin (2008).

Ainda discutindo a questão dos graus de liberdade, muito embora Yu *et al.* (2013) afirmem que os GL são expressos como o arredondamento do valor obtido para D_2 para número inteiro mais próximo, verifica-se que em diversos casos esse procedimento superestima a influência de um GL. Por exemplo, para a situação do MP_{10} no ano de 2006 o valor de $D_2 = 4,04$, logo, 5 graus de liberdade. Para este caso, assegura-se a presença de 4

GL, já o quinto está superestimado em 96%. Essa constatação ocorre para a maioria dos $GL = 5$. Ocorrência atípica é observada para o NOx no ano de 2005 em que $GL = 3$. Para este ano, nota-se que a série temporal apresenta menos de 25% dos pontos, sendo este o motivo indicado para o baixo valor de GL. Em geral, por meio dos dados apresentados na Tabela 4.2, é perfeitamente possível garantir a presença de 4 regentes da dinâmica dos sistemas, já o quinto pode ser entendido como uma possibilidade que exerce influência nas séries temporais, porém de forma menos incisiva. Similar ao realizado por Yu *et al.* (2011), a avaliação das características locais da estação permite destacar que os quatro graus de liberdades são:

- (i) *as condições meteorológicas;*
- (ii) *as emissões poluentes (predominantemente veiculares);*
- (iii) *as características do parque do Ibirapuera (presença de árvores e lagos inseridos em uma ampla área plana) e*
- (iv) *ausência de fontes fixas e pouquíssimas móveis.*

É importante mencionar que a série temporal de NOx para o ano de 2005 apresentou diversos dados inválidos, ou seja, informações não computadas neste estudo, desta forma impactando negativamente a avaliação do referido período. Em última análise, lança-se mão da entropia de Kolmogorov e máximo expoente de Lyapunov para certificar a natureza caótica dos sistemas em estudo.

Ainda na Tabela 4.2 estão relacionados os valores de entropia de Kolmogorov das séries temporais analisadas. Em todas as situações verificou-se a presença de Caos e, como já esperado, os valores anuais apresentam-se em mesma ordem de grandeza, porém diferentes. Esse resultado é corroborado pelo fato de que as séries temporais são definitivamente afetadas por fatores meteorológicos (frequência e volume de precipitações, velocidade do ar etc.), implementação de políticas públicas e alterações físicas e imprevisíveis na estação de monitoramento, conduzindo, assim, a situações anuais mais ou menos aleatórias. Comparando os valores de K_{ML} dos dois poluentes, nota-se que, conforme

já verificado na análise dos atratores, as séries temporais de MP_{10} apresentam um caráter inerentemente mais caótico.

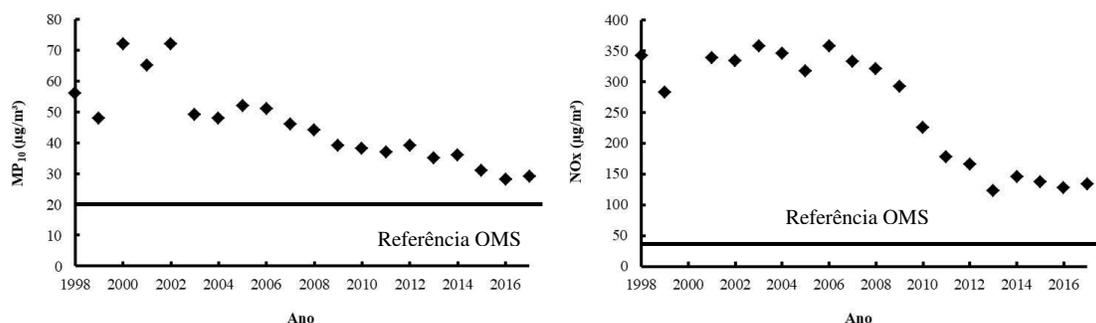
Em linhas gerais, os parâmetros meteorológicos influenciam diretamente a amplitude da evolução temporal de concentração de MP_{10} e NO_x , porém qualquer alteração na região onde está instalada a estação também afeta as séries temporais, podendo torná-las mais ou menos previsíveis, desta forma influenciando os invariantes caóticos. Algumas possíveis mudanças na região da estação que podem influenciar os resultados são: o aumento ou diminuição de fontes poluentes (veículos e indústrias); deslocamento de pessoas e veículos (causando suspensão de, por exemplo, material particulado depositado no solo); aumento ou diminuição do número de árvores e etc. Neste sentido, a estação do Ibirapuera situa-se em um ponto atípico da região metropolitana da cidade de São Paulo, pois, como já extensivamente salientado, trata-se de uma área residencial, bem arborizada, não apresentando fontes industriais em seu entorno, sendo as fontes veiculares as responsáveis pelas emissões locais.

Em derradeira análise, observa-se na última coluna de cada poluente da Tabela 4.2 os valores dos expoentes de Lyapunov. Resgatando o conceito apresentado na Tabela 2.1, todas as séries analisadas podem ser consideradas caóticas por este critério, pois em todos os casos $\lambda_{\text{Máximo}} > 0$. O método para a obtenção destes valores pode ser observado no exemplo presente na Figura 3.5. O coeficiente angular da região linear do gráfico (fator de deformação pelo número de iterações) define o máximo expoente de Lyapunov. No que diz respeito à análise de Caos, Yu *et al.* (2011) e Yu *et al.* (2013) concluem que esta metodologia é adequada para avaliar séries temporais de poluentes atmosféricos, pois por meio dela, mais especificamente do cálculo da dimensão de correlação, é possível obter resultados muito próximos daqueles postulados por análises clássicas como DFA, porém o Caos determinístico considera de forma ampla as relações não lineares associadas ao fenômeno com a vantagem de custo computacional muito menor daquele exigido por metodologias tradicionais.

4.2 ESTAÇÃO DE CONGONHAS

Similar ao realizado na estação do Ibirapuera, avaliou-se a variação da concentração anual média (Equação 3.1) de NO_x e MP₁₀ na estação de Congonhas. A Figura 4.4 apresenta as concentrações de ambos os poluentes de 1998 a 2017. Os dados discretos utilizados na confecção desta figura bem com os respectivos desvios padrão estão disponíveis na Tabela 4.3.

Figura 4.4 – Variação da concentração anual média de MP₁₀ e NO_x na estação de Congonhas dos anos de 1998 à 2017 (o autor, 2018).



Tanto para as concentrações anuais de NO_x quanto MP₁₀, verifica-se um claro perfil de redução dos níveis dos poluentes ao longo dos anos, chegando a 48% no caso dos particulados inaláveis e 61% para os óxidos de nitrogênio. Mesmo com uma tendência de queda, as médias anuais dos dois poluentes ultrapassaram, em todos os anos, os valores sugeridos pela OMS.

A proximidade com importantes avenidas, famosas por seu intenso tráfego automotivo, justificam os elevados níveis de poluição atmosférica. As políticas públicas, bem como a queda na idade da frota de veículos, auxiliaram muito a redução, contudo, a poucos metros desta estação, está localizado o aeroporto de Congonhas. O Inventário Nacional de Emissões Atmosféricas da Aviação Civil de 2014 aponta expressivo aumento das emissões de NO_x e MP₁₀ até 2014 (último ano avaliado), deixando claro que este incremento se deu devido ao intenso fluxo de aviões ter aumentado nos últimos anos. As

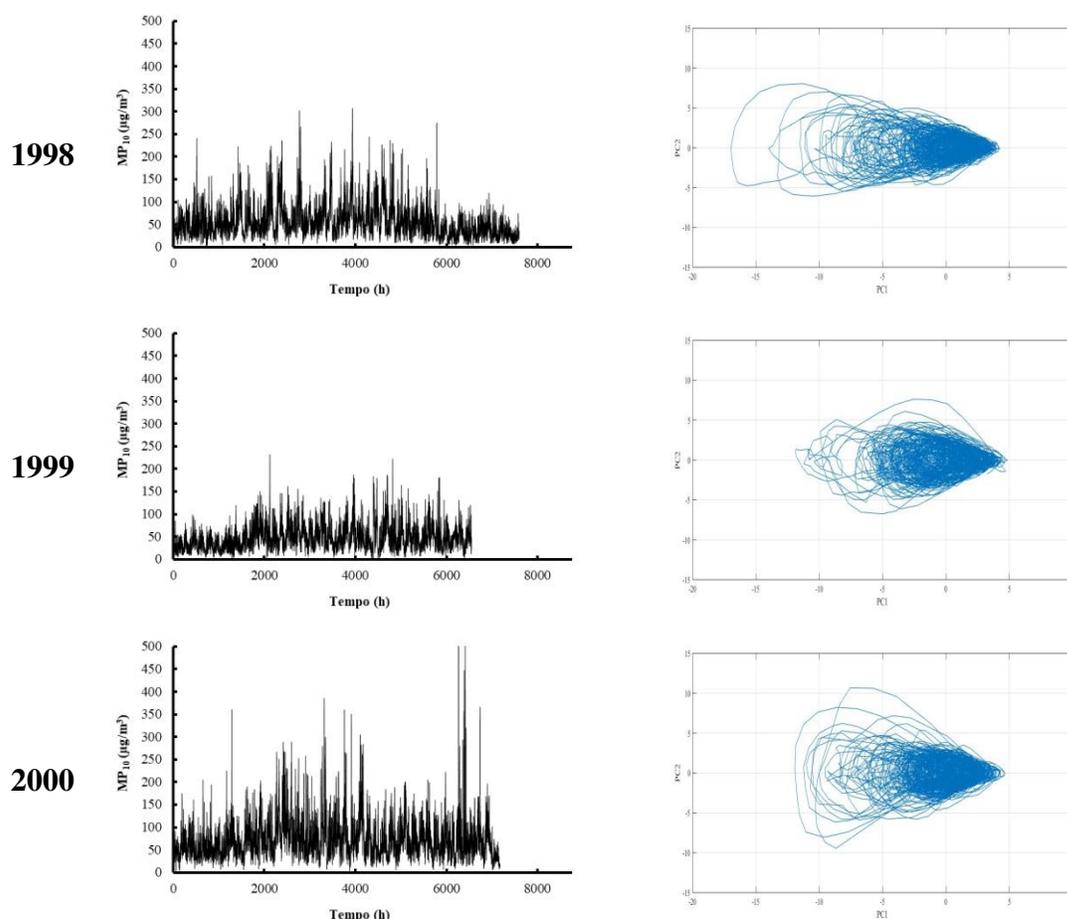
políticas públicas mais impactantes nas reduções observadas no período de 1998 as 2017 são os efeitos progressivos do Programa de Controle da Poluição do Ar por Veículos Automotores (PROCONVE), que desde a década de 1980 especificou mudanças tecnológicas nos veículos automotores e promoveu reduções significativas nas emissões de poluentes (Carvalho *et al.*, 2015). Desde o ano de 2003 pode-se salientar, também, a influência do Programa de Controle da Poluição do Ar por Motociclos e Veículos Similares (PROMOT). Outra iniciativa importante é a inspeção veicular que operou reduções expressivas de 2009 a 2014.

Tabela 4.3 – Médias anuais e desvios padrão da concentração de MP_{10} e NO_x na estação de Congonhas (o autor, 2018).

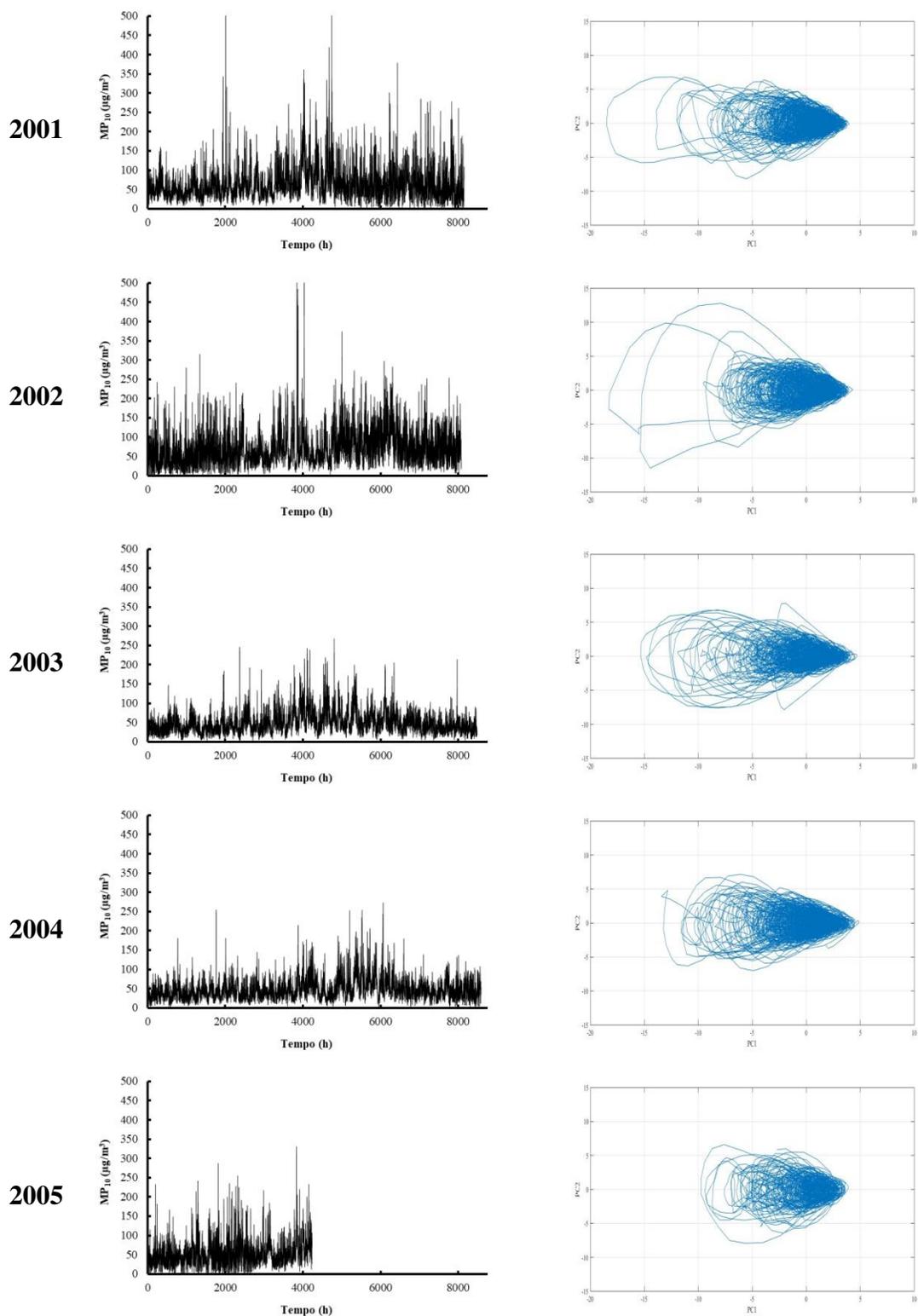
Ano	MP_{10}		NO_x	
	μ ($\mu\text{g}/\text{m}^3$)	σ ($\mu\text{g}/\text{m}^3$)	μ ($\mu\text{g}/\text{m}^3$)	σ ($\mu\text{g}/\text{m}^3$)
1998	56	37	342	244
1999	48	28	283	189
2000	72	44	-	-
2001	65	43	338	256
2002	72	50	334	248
2003	49	30	357	271
2004	48	28	346	241
2005	52	34	317	218
2006	51	31	357	269
2007	46	26	332	250
2008	44	27	321	235
2009	39	21	292	200
2010	38	25	225	176
2011	37	23	176	136
2012	39	22	166	109
2013	35	20	122	63
2014	36	23	145	107
2015	31	20	138	107
2016	28	17	128	97
2017	29	18	134	88

No que diz respeito aos desvios padrão, verificou-se, novamente, que seus valores assumem a amplitude da média, deixando claro o caráter disperso dos níveis horários que compõe o resultado final médio. Observa-se, também, que no ano de 2000, a plataforma QUALAR não disponibilizou os dados relativos à [NOx]. Depois de verificadas as características gerais da variação anual de cada poluente, realizou-se a análise das não linearidades presentes na evolução temporal da [MP₁₀] e da [NOx] nos anos de 1998 a 2017. Inicialmente foram obtidas as séries temporais e os atratores descritivos da dinâmica dos fenômenos. A Figura 4.5 apresenta as séries temporais e os atratores de concentração anual de MP₁₀ da estação de Congonhas.

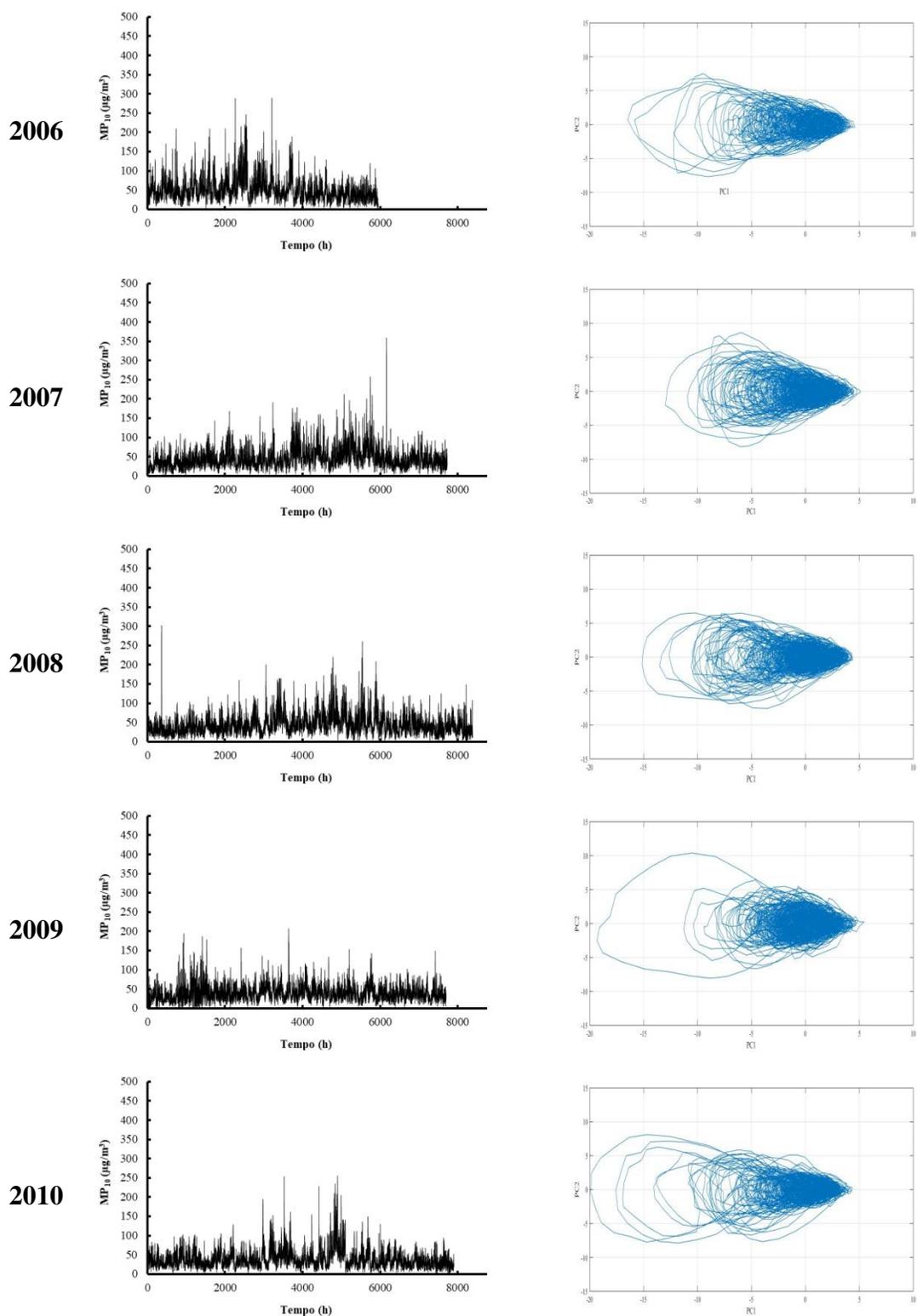
Figura 4.5 – Séries temporais e atratores de concentração anual de MP₁₀ obtidas na estação de monitoramento de Congonhas nos anos de 1998 a 2017 (o autor, 2018).



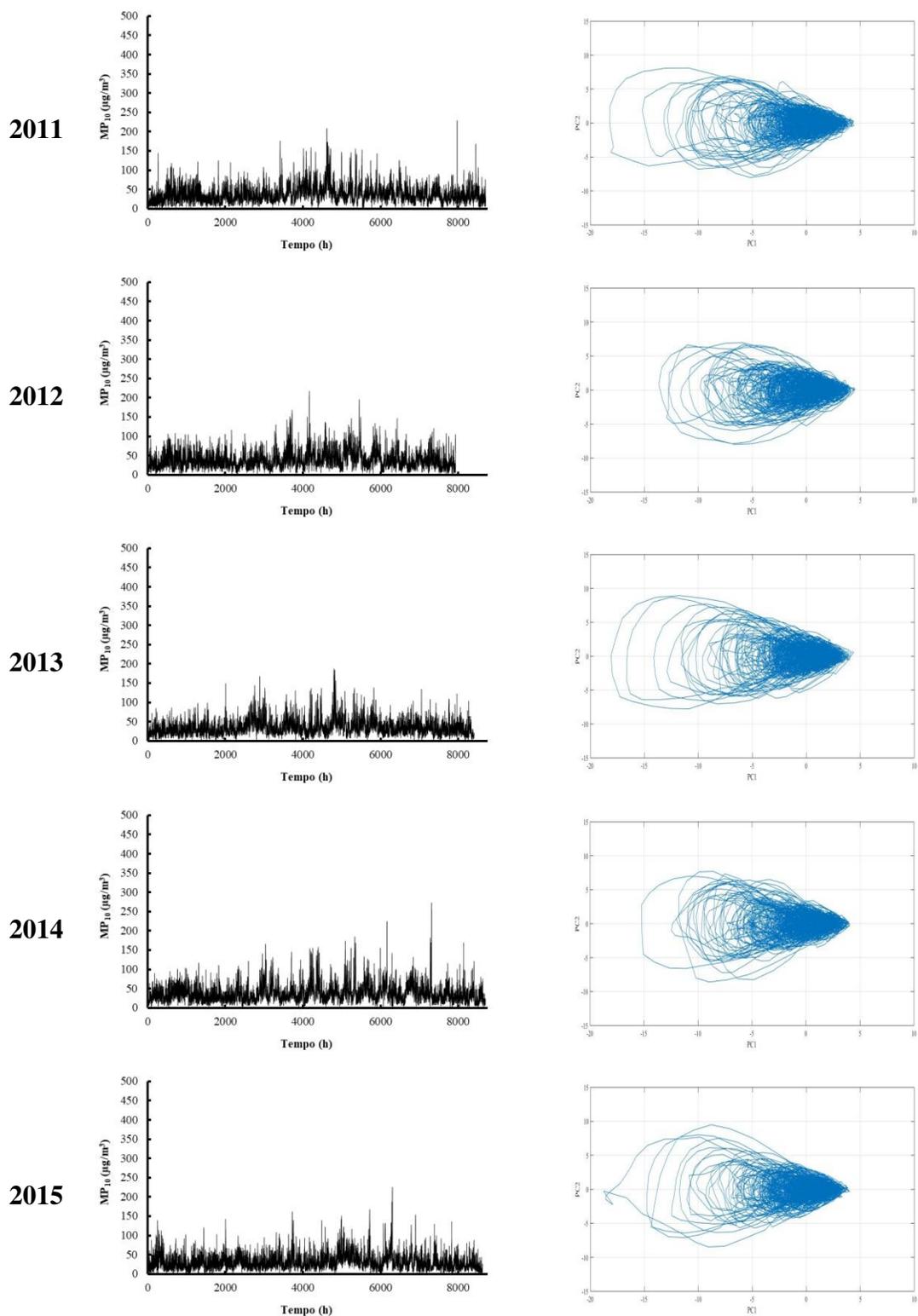
Capítulo IV – Resultados e Discussões

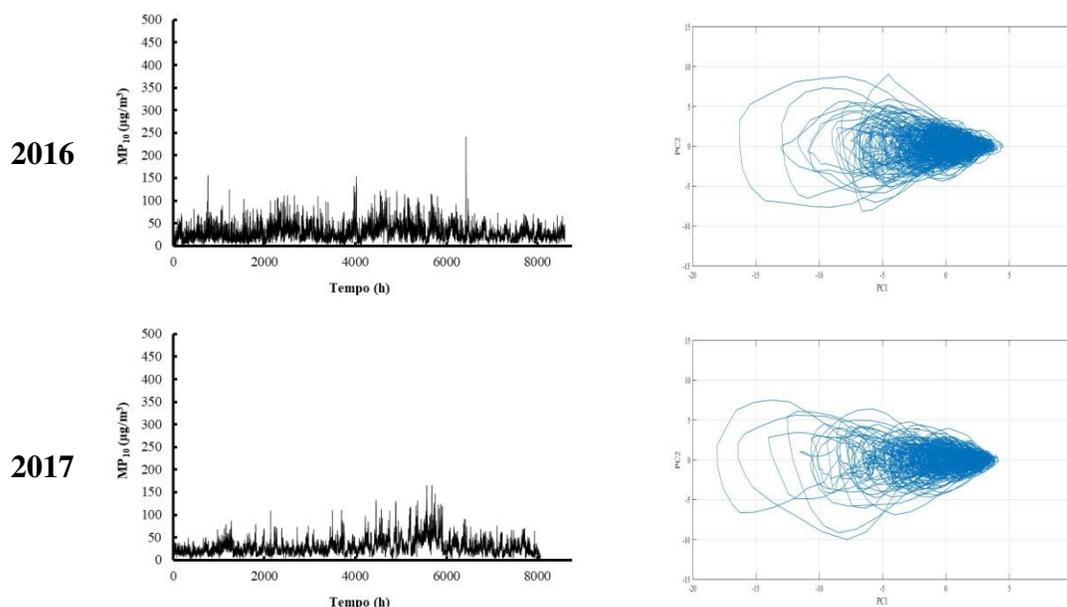


Capítulo IV – Resultados e Discussões



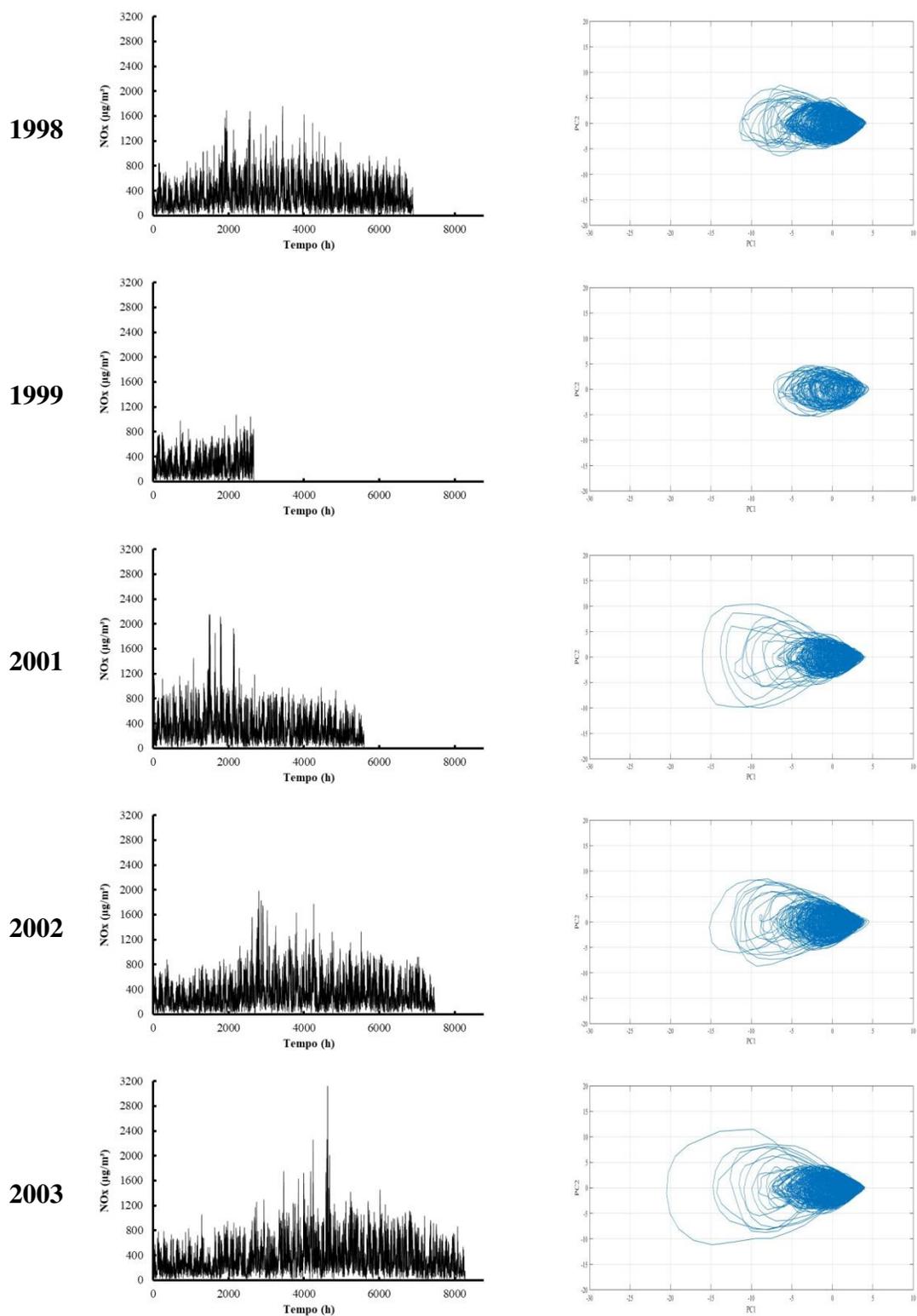
Capítulo IV – Resultados e Discussões



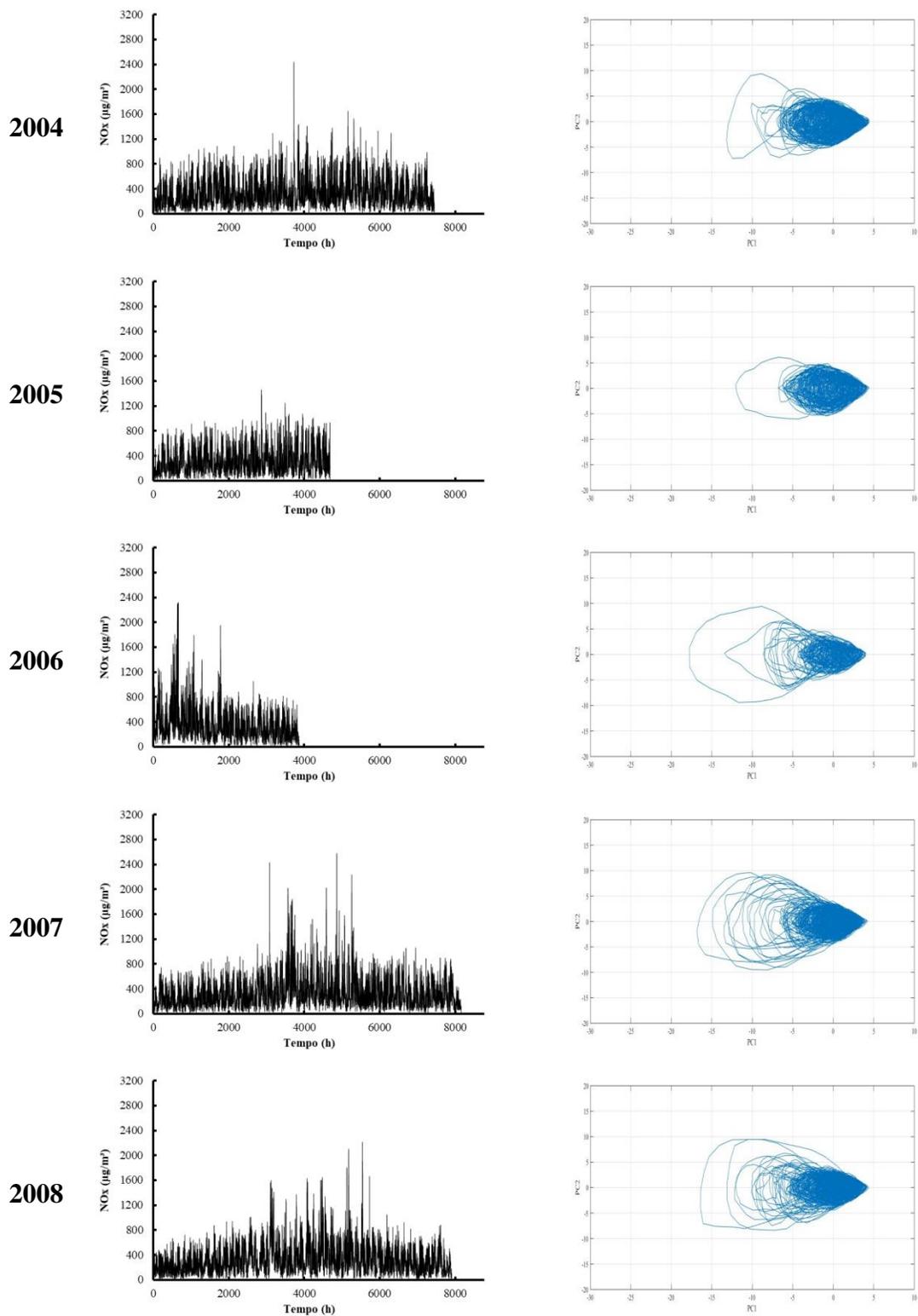


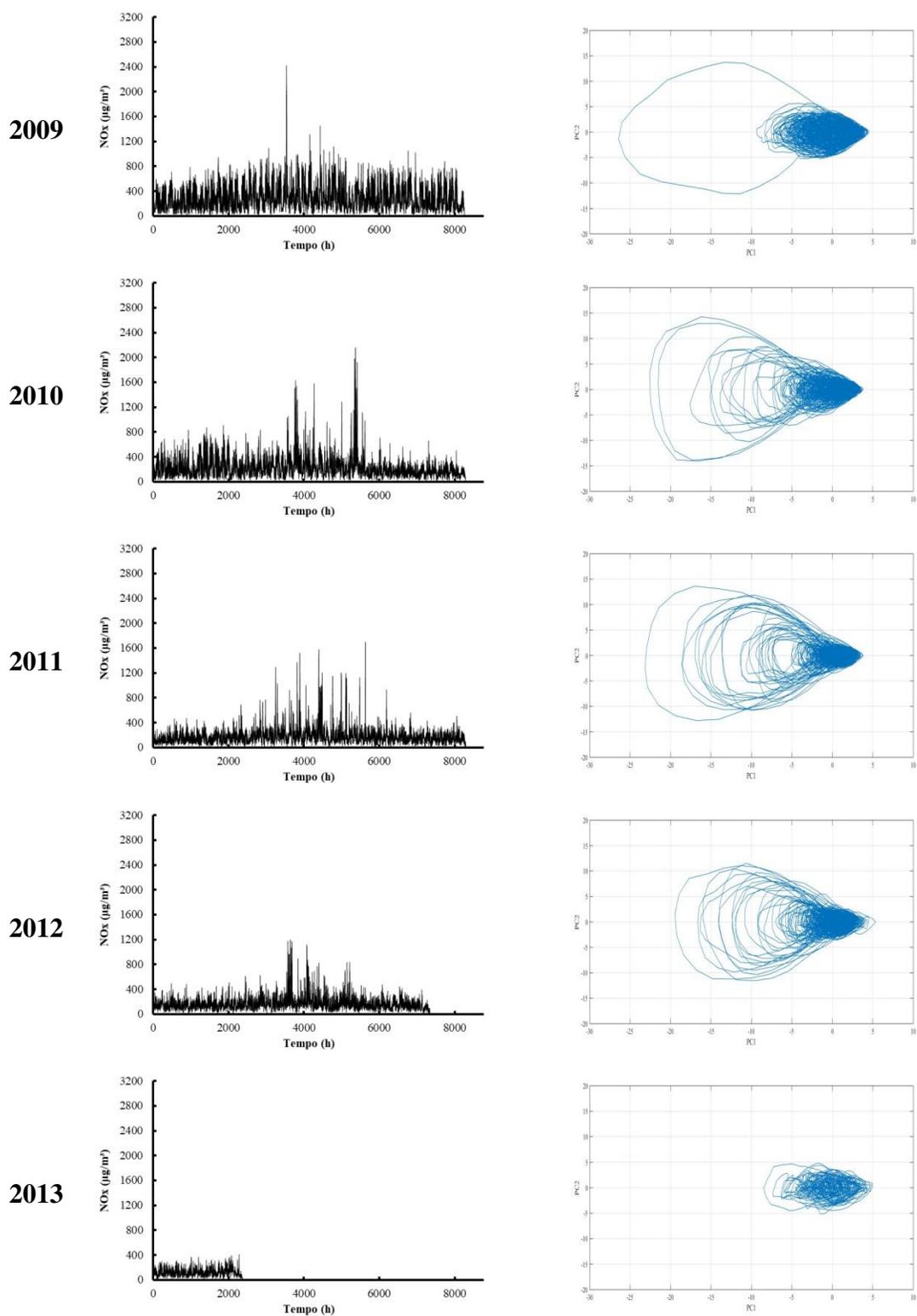
Assim como verificado nas séries obtidas na estação do Ibirapuera, aqui é possível notar a falta de dados ao final das séries. Pelos mesmos motivos já mencionados, foi necessário deslocar todos os pontos faltantes do meio da série para o final, tornando-a, assim, contínua. É claro o caráter variável bem dispersivo em relação à média presente na evolução temporal, ocasionado por características do fenômeno da poluição atmosférica, como a sazonalidade, que fazem com que os valores pontuais de concentração de MP₁₀ sejam maiores ou menores dependendo das condições climáticas. A reconstrução dos atratores no espaço de fases foi realizada com o mesmo número de pontos da série temporal, dimensão de imersão (I) otimizada igual a 12 e tempo de atraso (θ) de 1 hora. Para todos os atratores obtidos, verifica-se que as linhas de fluxo culminam em atratores estranhos, exibindo uma clara convergência para uma região bem definida e estreita no espaço de fases (linha horizontal em PC2 (principal componente) = 0), sugerindo uma dinâmica caótica determinista. Analogamente, as mesmas conclusões são extraídas da Figura 4.6 em que são apresentadas as séries temporais e a reconstrução dos atratores da concentração anual de NO_x na estação de monitoramento de Congonhas.

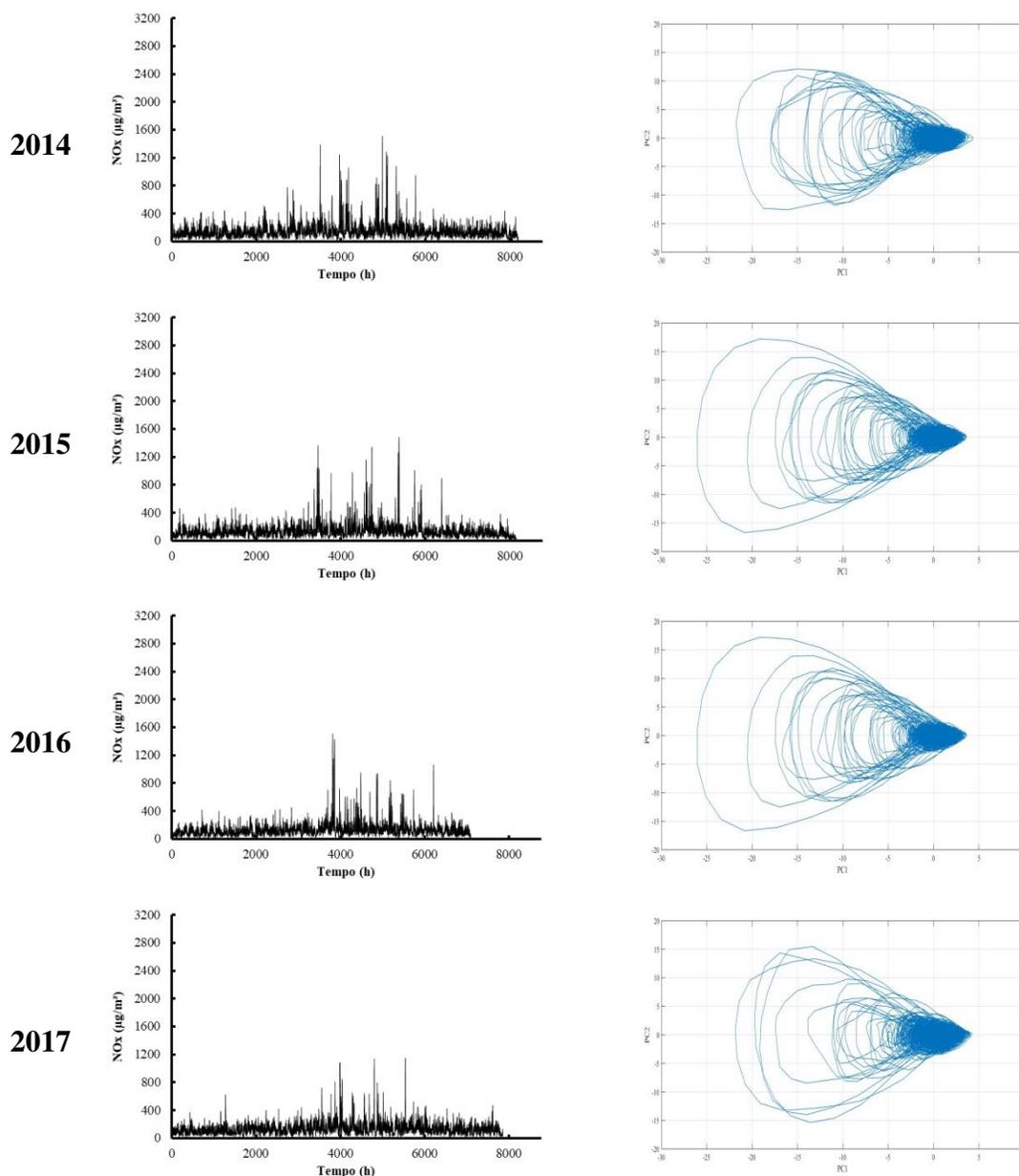
Figura 4.6 – Séries temporais e atratores de concentração anual de NO_x obtidas na estação de monitoramento de Congonhas nos anos de 1998 a 2017 (o autor, 2018).



Capítulo IV – Resultados e Discussões







Nestas séries, também, nota-se o caráter dispersivo e perfil variável das séries, características espelhadas na dinâmica do atrator, claramente caótico, obtido. No que tange a dinâmica do Caos determinístico, para as concentrações anuais de NOx da estação de Congonhas, utilizou-se tempo de atraso igual a 1 hora e dimensão de imersão otimizada igual 13. Nesta estação a avaliação foi bastante comprometida devido à falta de muitos pontos em diversos anos estudados.

De modo a asseverar as conclusões anteriores, no que se refere à dinâmica caótica, apresenta-se a Tabela 4.4. Nela é possível observar o estudo da dimensão de correlação (D_2), graus de liberdade (GL), entropia de Kolmogorov (K_{ML}) e máximo expoente de Lyapunov (λ).

Nos primeiros blocos da tabela (I , D_2 e GL) de cada poluente, verifica-se o I de convergência e número de variáveis para as concentrações anuais de MP_{10} e NO_x . Observa-se, para os dois poluentes, que o número de variáveis do sistema apresenta valores anuais diferentes, ou seja, demonstrando que não há unicidade no que diz respeito à dinâmica dos sistemas avaliados.

Ambos os poluentes apresentam de três a cinco graus de liberdade que regem a dinâmica do sistema. Estes GLs são difíceis de serem determinados, vez que a poluição atmosférica é afetada por diversos fatores que variam desde as condições meteorológicas (umidade, frequência e intensidade de precipitações, velocidade e direção do vento, raios UV, temperatura e etc), condições geográficas, fontes emissoras e etc. Novamente aqui, pode-se mencionar como comparativo, para o MP_{10} , o trabalho de Lee e Lin (2008) que obtiveram $D_2 = 4,32$ para análise realizada em Chung-Shan, Taipei (Taiwan), de Janeiro de 1998 à Dezembro de 1998. No caso desta Tese, para estação de Congonhas dos anos de 1998 a 2017 os valores de D_2 variaram de 3,46 a 4,64, resultando, nos anos de 1998, 1999, 2000, 2001, 2004, 2006, 2009, 2010, 2011, 2014 e 2016 em 5 graus de liberdade, similar ao reportado por Lee e Lin (2008).

Novamente, ao seguir o preconizado por Yu *et al.* (2013), para vários casos, ocorre a superestimação da influência de um GL. Um claro exemplo é o MP_{10} no ano de 2009, para o qual afirmar $GL = 5$ é superestimar a influência do quinto grau de liberdade em 95%. Sendo assim, por meio dos cálculos da Tabela 4.4, assegura-se a existência de 4 GL, sendo o quinto uma possibilidade que exerce menos influência na evolução temporal. Por meio das características do local onde está instalada a estação é possível destacar que os quatro graus de liberdades são:

- (i) as condições meteorológicas;
- (ii) as emissões poluentes;
- (iii) a proximidade com avenidas de alto tráfego (Avenida dos Bandeirantes e Alameda dos Tupiniquins) e
- (iv) a presença do aeroporto de Congonhas.

Tabela 4.4 – Invariantes do Caos para os poluentes MP_{10} e NO_x avaliados na estação de monitoramento de Congonhas dos anos de 1998 a 2017 (o autor, 2018).

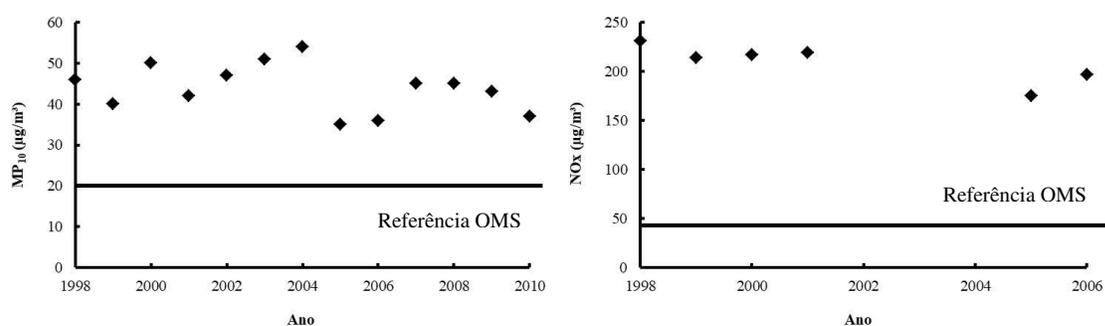
Ano	MP_{10}					NO_x				
	I	D_2 [-]	GL	K_{ML} (bits/s)	λ [-]	I	D_2 [-]	GL	K_{ML} (bits/s)	λ [-]
1998	14	4,24	5	0,1574	0,14	14	3,74	4	0,1369	0,13
1999	14	4,25	5	0,1889	0,14	13	3,56	4	0,1375	0,12
2000	13	4,29	5	0,2215	0,14	-	-	-	-	-
2001	11	4,14	5	0,1110	0,14	13	2,99	3	0,2008	0,13
2002	10	3,68	4	0,1602	0,13	14	3,63	4	0,1602	0,13
2003	12	3,81	4	0,1437	0,14	14	3,62	4	0,1438	0,13
2004	12	4,39	5	0,1919	0,13	13	3,84	4	0,1686	0,14
2005	12	3,99	4	0,2196	0,11	13	4,10	5	0,1839	0,13
2006	13	4,64	5	0,2913	0,13	13	3,34	4	0,1722	0,12
2007	11	3,65	4	0,1302	0,13	14	4,14	5	0,1776	0,14
2008	11	3,72	4	0,1849	0,13	13	3,47	4	0,1964	0,13
2009	10	4,02	5	0,2345	0,13	13	3,75	4	0,1678	0,13
2010	12	4,16	5	0,1721	0,13	13	3,36	4	0,1142	0,13
2011	11	4,19	5	0,2324	0,13	11	4,53	5	0,1574	0,13
2012	10	3,91	4	0,3370	0,13	10	4,52	5	0,2157	0,12
2013	12	3,78	4	0,3049	0,13	11	4,43	5	0,2730	0,10
2014	10	4,32	5	0,2553	0,13	11	4,15	5	0,1557	0,13
2015	9	3,46	4	0,2235	0,12	10	4,05	5	0,1260	0,13
2016	10	4,13	5	0,2379	0,12	10	4,16	5	0,1318	0,13
2017	15	3,93	4	0,1855	0,13	11	4,10	5	0,2082	0,12

Sobre os expoentes de Lyapunov e entropia de Kolmogorov, verifica-se que $0 < K_{ML} < \infty$ e $\lambda > 0$ para todas as séries analisadas, sendo assim está evidente que a evolução temporal dos poluentes avaliados é caótica.

4.3 ESTAÇÃO DO CENTRO DE SÃO PAULO

Na Figura 4.7, verifica-se a variação concentração anual média (Equação 3.1) de NO_x e MP₁₀ obtidas na estação do Centro de São Paulo nos anos disponibilizados pela plataforma QUALAR - CETESB. Os dados discretos que foram utilizados na elaboração desta figura e seus desvios padrão estão apresentados na Tabela 4.5.

Figura 4.7 – Variação da concentração anual média de MP₁₀ e NO_x na estação do Centro de São Paulo desde 1998 até o último ano disponibilizado pela plataforma (o autor, 2018).



A estação do Centro de São Paulo disponibiliza apenas dados históricos de 1998 a 2010 para o MP₁₀ e de 1998 a 2006 para NO_x, sendo que para este último não estão disponíveis os anos de 2002, 2003 e 2004. Por meio do perfil observado na Figura 4.6 e os dados apresentados na Tabela 4.5, verifica-se uma redução de quase 20% e 15% (comparando o primeiro e último ano analisados) dos níveis anuais de, respectivamente, MP₁₀ e NO_x. Pode-se salientar como os impulsionadores destas reduções, a renovação da frota veicular da capital paulista, políticas públicas e etc. Mesmo com as reduções destacadas, observa-se que as emissões de MP₁₀ de todos os anos ultrapassaram a média anual de 20 µg/m³ recomendada pela OMS. Para o NO_x, a OMS estipula o máximo de 40 µg/m³ e, sendo assim, são verificadas ultrapassagens em todos os anos avaliados.

No que diz respeito ao desvio padrão, como observado nas estações anteriores, os valores obtidos são altos, apresentando a ordem da própria média, assim, asseverando, o

caráter disperso dos valores horários que foram utilizados na composição do resultado médio.

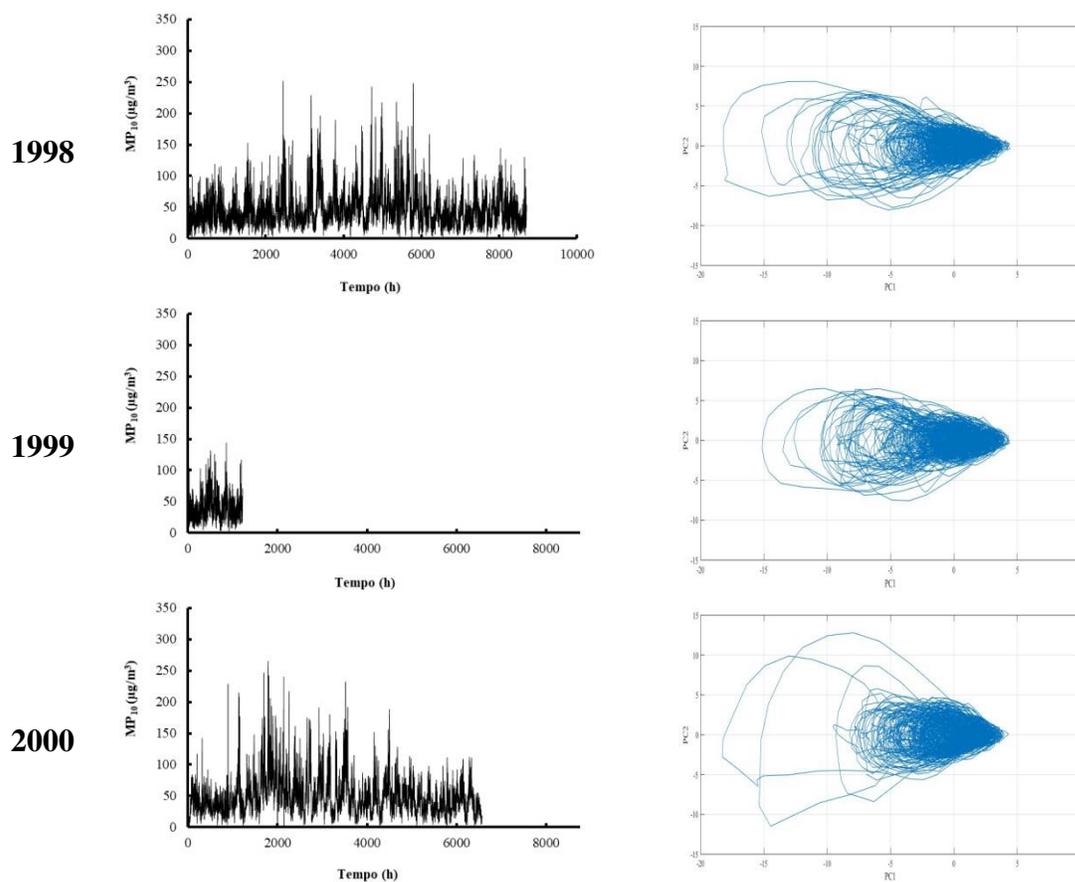
Tabela 4.5 – Médias anuais e desvios padrão da concentração de MP_{10} e NO_x na estação do Centro de São Paulo (o autor, 2018).

Ano	MP_{10}		NO_x	
	μ ($\mu\text{g}/\text{m}^3$)	σ ($\mu\text{g}/\text{m}^3$)	μ ($\mu\text{g}/\text{m}^3$)	σ ($\mu\text{g}/\text{m}^3$)
1998	46	28	231	135
1999	40	21	214	169
2000	50	31	217	144
2001	42	26	219	151
2002	47	29	-	-
2003	51	33	-	-
2004	54	38	-	-
2005	35	25	175	118
2006	36	20	197	137
2007	45	27	-	-
2008	45	32	-	-
2009	43	27	-	-
2010	37	27	-	-

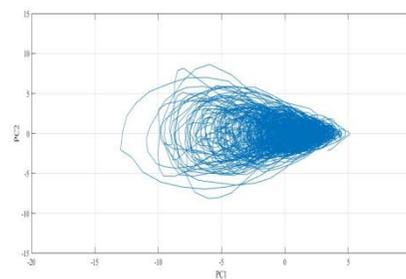
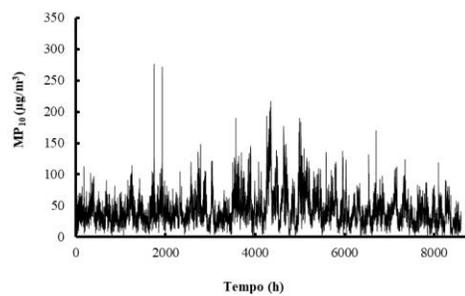
A presença de vias de alto tráfego automotivo de veículos leves como a Rua da Consolação, Avenida São Luis, Avenida Ipiranga e outras, justificam os elevados níveis de poluição atmosférica. As políticas públicas que possivelmente mais impactaram nas reduções observadas no período de 1998 a 2010 são os efeitos paulatinos do Programa de Controle da Poluição do Ar por Veículos Automotores (PROCONVE), política implementada na década de 1980 e que cujas exigências de alteração tecnológicas nos veículos automotores promoveram reduções significativas nas emissões de poluentes (Carvalho *et al.*, 2015). Partindo do ano de 2003, pode-se salientar, também, a influência do Programa de Controle da Poluição do Ar por Motociclos e Veículos Similares (PROMOT). Já em relação à inspeção veicular, esta não pode ser levada em conta no caso desta estação, vez que operou reduções expressivas na poluição de 2009 a 2014 e no caso aqui considerado, verifica-se que os dados históricos não ultrapassam 2010.

Uma vez extraídas e analisadas as características gerais da variação anual da concentração de cada poluente, passa-se avaliação das subjacentes não linearidades apresentadas na evolução temporal da $[MP_{10}]$ e da $[NO_x]$ nos anos de 1998 a 2010, para tal, inicialmente foram obtidas as séries temporais e os atratores representativos da dinâmica dos fenômenos. A Figura 4.8 e 4.9 apresentam as séries temporais e os atratores de concentração anual de MP_{10} e NO_x da estação do Centro de São Paulo do ano de 1998 a 2010.

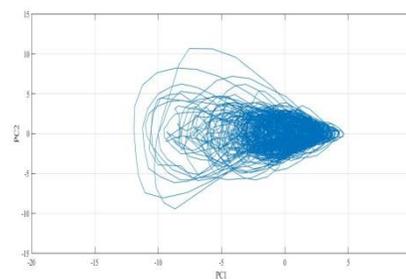
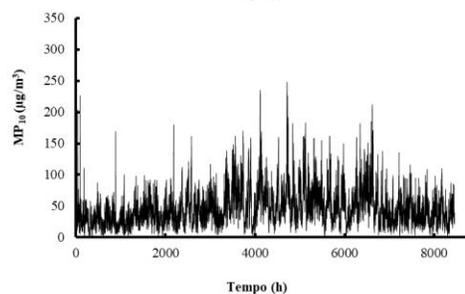
Figura 4.8 – Séries temporais e atratores de concentração anual de MP_{10} obtidas na estação de monitoramento do Centro de São Paulo nos anos de 1998 a 2010 (o autor, 2018).



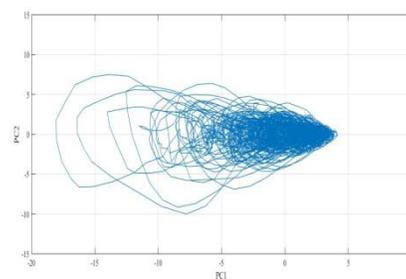
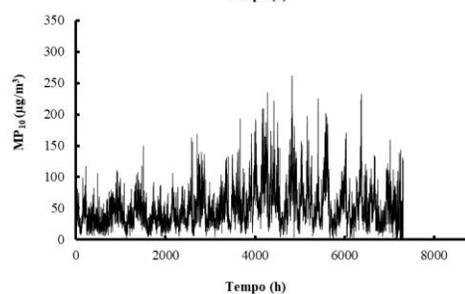
2001



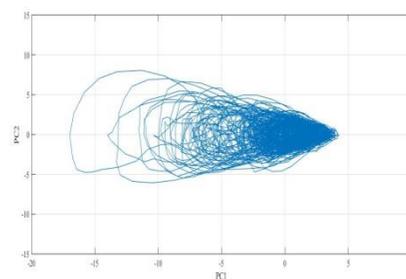
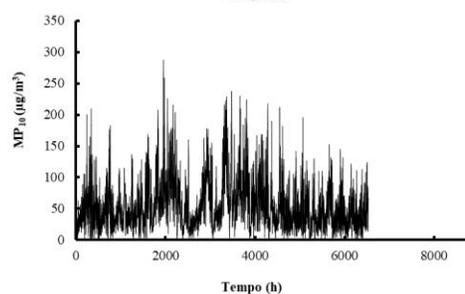
2002



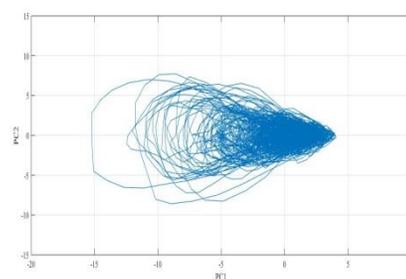
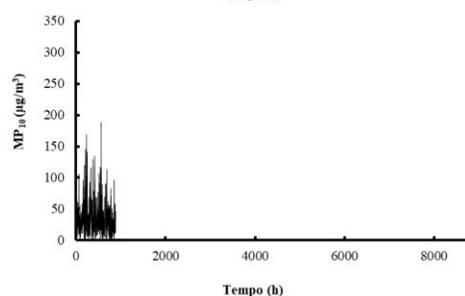
2003



2004

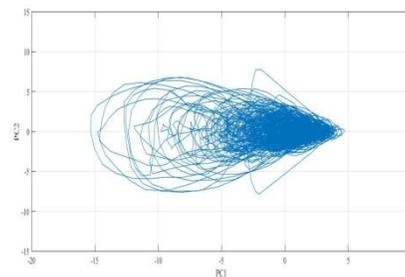
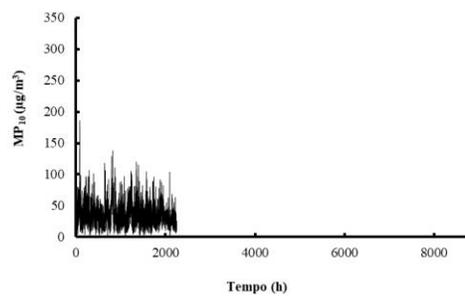


2005

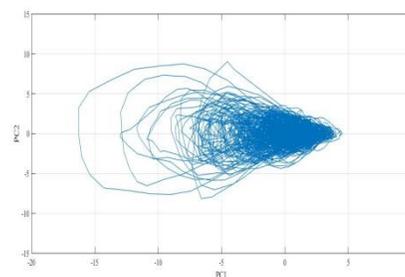
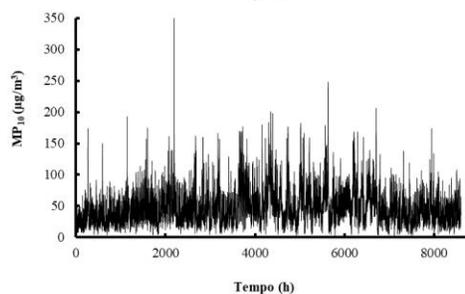


Capítulo IV – Resultados e Discussões

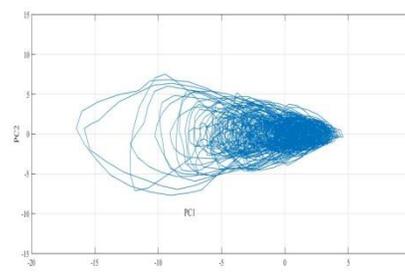
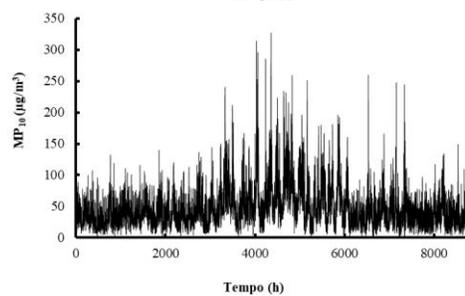
2006



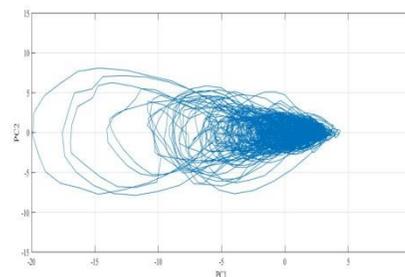
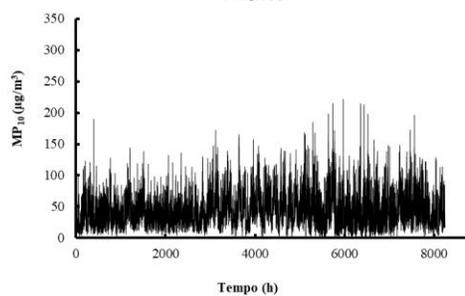
2007



2008



2009



2010

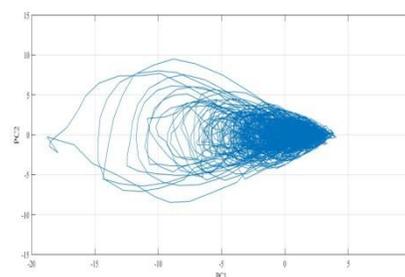
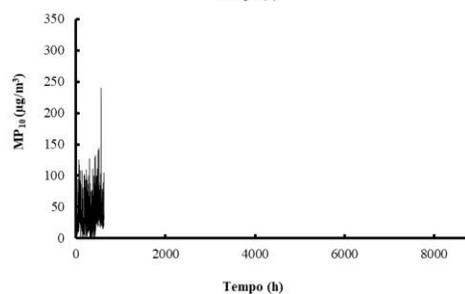
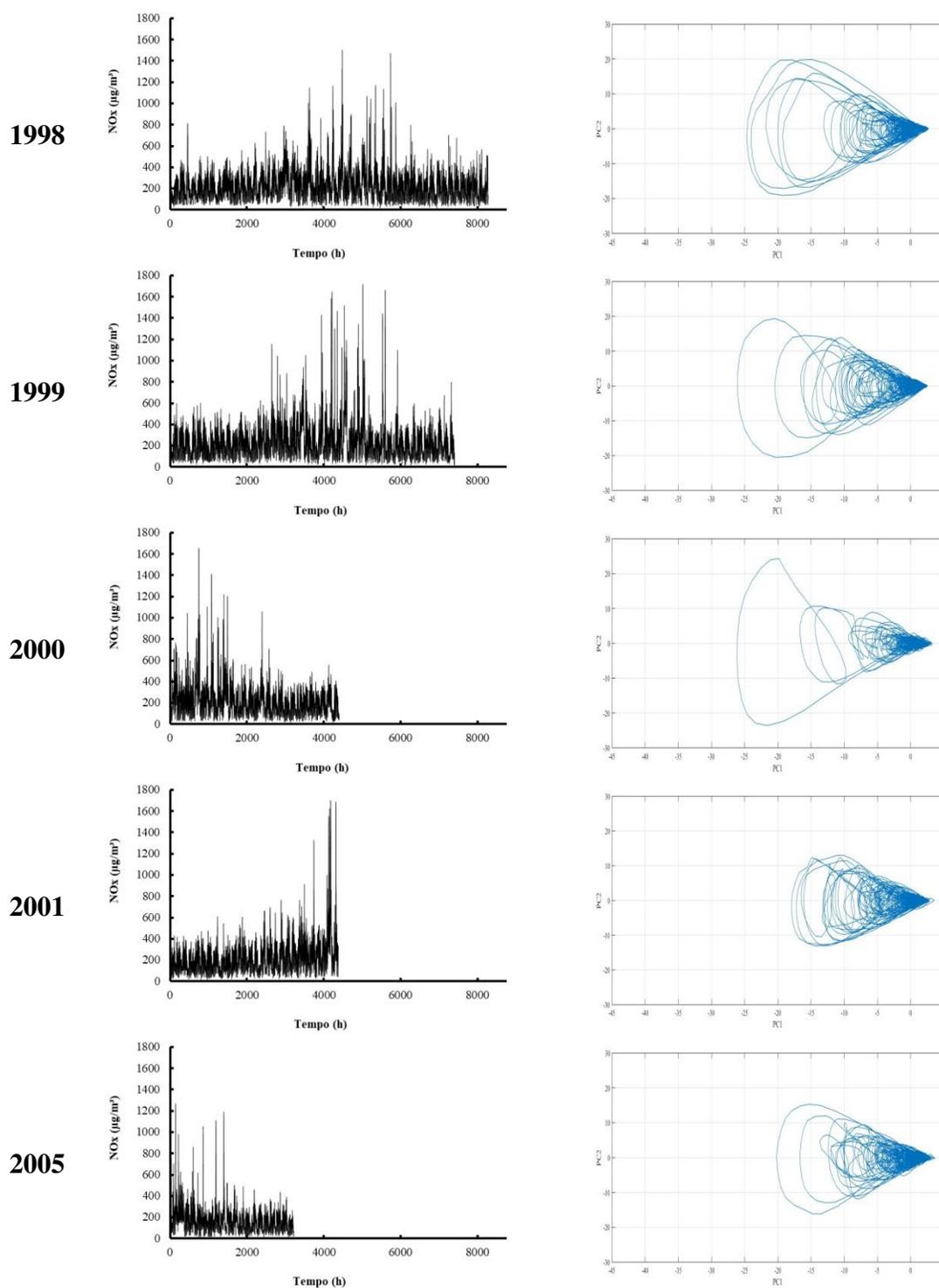


Figura 4.9 – Séries temporais e atratores de concentração anual de NOx obtidas na estação de monitoramento do Centro de São Paulo nos anos de 1998 a 2005 (o autor, 2018).



Como padrão adotado nesta Tese e constatado nas estações anteriores, é possível se observar, em todos os anos e para os dois poluentes, a falta de dados ao final das séries, não completando os 8760 pontos que deveria apresentar uma evolução temporal anual com aquisição horária. Pelos mesmos motivos salientados anteriormente, as lacunas relativas aos zeros e dados inválidos foram lançados para o final da evolução temporal, tornando-a, assim, contínua. Especialmente para esta estação, no caso do MP_{10} , três anos, a saber, 1999, 2005 e 2006, e para o NO_x (2000, 2001 e 2005), apresentaram muitas lacunas e ausência de dados, prejudicando, sobremaneira, a avaliação das informações extraídas destes anos. No caso do NO_x , outros três anos, 2002, 2003 e 2004, tiveram suas aquisições interrompidas. Ainda assim, é evidente o caráter variável dispersivo em relação à média apresentado pelas séries temporais. Estas constatações são resultados diretos das características do fenômeno da poluição atmosférica, como, por exemplo, a sazonalidade, que, dependendo da estação do ano, incidência de chuvas, saturação da atmosfera e outras, fazem com que os valores ponto a ponto de concentração do poluente sob análise sejam maiores ou menores dependendo das condições climáticas. A reconstrução dos atratores no espaço de fases foi realizada com o mesmo número de pontos da série temporal, dimensão de imersão (I) otimizada igual a 11 para o MP_{10} e 13 para o NO_x , com tempo de atraso (θ) de 1 hora. Novamente, verificou-se que em todos os atratores obtidos as linhas de fluxo resultam em atratores estranhos com convergência para uma região bem definida e estreita no espaço de fases (linha horizontal em PC_2 (principal componente) = 0), sugerindo uma dinâmica caótica determinista.

Confirmando as conclusões obtidas anteriormente, a Tabela 4.6 traz o estudo da dimensão de correlação (D_2), graus de liberdade (GL), entropia de Kolmogorov (K_{ML}) e máximo expoente de Lyapunov (λ). A primeira coluna de cada um dos poluentes traz o valor do I de convergência que, na prática é o valor da dimensão de imersão obtida por meio do algoritmo de Grassberger e Procaccia (1983), na sequência é apresentada a dimensão de correlação D_2 e também o número de graus de liberdade que regem a dinâmica do sistema. Nas séries temporais de concentração horária de MP_{10} e NO_x da estação de monitoramento do Centro de São Paulo, durante a otimização de I , o valor de D_2 cresceu até um momento em que a taxa de aumento começa a diminuir e os valores de D_2 atingem a

saturação, ponto em que se verifica de uma dinâmica determinista (Yu *et al.*, 2013). A partir do valor de D_2 , o número inteiro mais próximo indica a quantidade de variáveis, ou graus de liberdade (GL) que regem a dinâmica do sistema (Yu *et al.*, 2013). Nota-se que tanto para o NOx quanto MP₁₀, o número de variáveis do sistema apresenta valores anuais diferentes, ou seja, não há unicidade no que diz respeito à evolução dos sistemas avaliados. Ambos os poluentes apresentam entre três a cinco graus de liberdade. Similar ao comentando anteriormente, os GLs são parâmetros de difícil determinação, pois o fenômeno da poluição atmosférica é afetado por muitos fatores, como as condições meteorológicas (umidade, frequência e intensidade de precipitações, velocidade e direção do vento, raios UV, temperatura e etc), condições geográficas, fontes emissoras e etc. Realizando um comparativo, para o MP₁₀, com o trabalho de Lee e Lin (2008), que obtiveram $D_2 = 4,32$ para análise realizada em Chung-Shan, Taipei (Taiwan), de Janeiro à Dezembro de 1998, para estação Centro de São Paulo, dos anos de 1998 a 2010 os valores de D_2 variaram de 2,25 a 4,31, resultando, nos anos de 1998, 2001, 2007, 2008 e 2009 em 5 graus de liberdade, similar ao que foi publicado por Lee e Lin (2008).

Tabela 4.6 – *Invariantes do Caos para os poluentes MP₁₀ e NOx avaliados na estação de monitoramento do Centro de São Paulo dos anos de 1998 a 2010 (o autor, 2018).*

Ano	MP ₁₀					NOx				
	I	D ₂ [-]	GL	K _{ML} (bits/s)	λ [-]	I	D ₂ [-]	GL	K _{ML} (bits/s)	λ [-]
1998	13	4,18	5	0,1968	0,15	13	3,93	4	0,1314	0,12
1999	6	2,97	3	0,4150	0,11	14	3,83	4	0,1204	0,12
2000	15	3,55	4	0,1837	0,16	13	3,10	4	0,1289	0,12
2001	14	4,31	5	0,2110	0,15	13	3,90	4	0,1285	0,11
2002	15	3,91	4	0,1396	0,15	-	-	-	-	-
2003	15	3,35	4	0,1690	0,15	-	-	-	-	-
2004	14	3,31	4	0,1155	0,15	-	-	-	-	-
2005	8	2,49	3	0,3219	0,11	12	2,76	3	0,1591	0,11
2006	5	2,98	3	0,2996	0,11	13	3,48	4	0,1198	0,12
2007	11	4,12	5	0,7036	0,14	-	-	-	-	-
2008	11	4,25	5	0,3726	0,14	-	-	-	-	-
2009	7	4,21	5	0,3859	0,13	-	-	-	-	-
2010	4	2,25	3	0,3219	0,12	-	-	-	-	-

Continuando na discussão dos graus de liberdade, novamente se verifica a superestimação do quinto grau de liberdade. Essa constatação pode ser feita no caso do MP_{10} para o ano de 1998 em que 82% da influência do GL sob a evolução temporal é superestimada, sendo possível garantir a presença de 4 GL, sendo o quinto uma possibilidade. Similar ao observado nas estações anteriores, esta constatação é observada para a maioria dos $GL = 5$. Ocorrências atípicas são observadas para os anos de 1999, 2005, 2006 e 2010 (MP_{10}) em que $GL = 3$. Similar ao discutido anteriormente, nota-se que as séries temporais apresentam menos de 25% dos pontos, sendo este o provável motivo indicado para o baixo valor de GL. Via de regra, tem-se observado que é perfeitamente possível a garantia da presença de 4 GL, sendo que o quinto pode ser entendido como uma possibilidade menos influente. A avaliação das características locais da estação permite destacar que os quatro graus de liberdades são:

- (i) *as condições meteorológicas;*
- (ii) *as emissões poluentes (predominantemente veiculares);*
- (iii) *a presença de vias de alto tráfego e*
- (iv) *ausência de fontes fixas.*

Da análise dos expoentes de Lyapunov e entropia de Kolmogorov, obteve-se $0 < K_{ML} < \infty$ e $\lambda > 0$ para todas as séries analisadas, sendo assim é claro que a evolução temporal dos poluentes avaliados é caótica.

É importante salientar que para os dois poluentes, diversos dados inválidos precisaram ser retirados, especialmente no caso do MP_{10} nos anos de 1999, 2005 e 2006, e para o NO_x (2000, 2001 e 2005) fazendo com que muitas informações deixassem de ser computadas, desta forma causando um impacto negativo na avaliação dos referidos períodos.

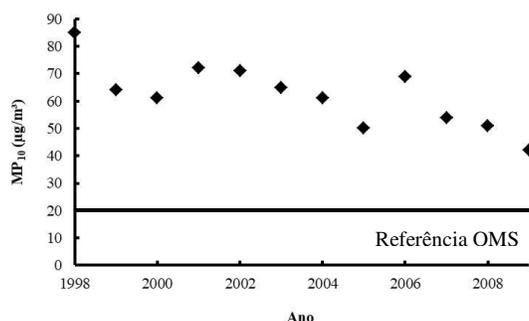
No caso de todos os invariantes, verificou-se a presença de Caos com valores anuais em mesma ordem de grandeza. Esta constatação é a asseveração de que as séries temporais são afetadas por fatores meteorológicos (frequência e volume de precipitações, velocidade

do ar etc.), implementação de políticas públicas e alterações físicas e imprevisíveis na estação de monitoramento, fazendo com que surjam situações anuais mais ou menos aleatórias. Similar ao observado para as demais estações, comparando os valores de K_{ML} do MP_{10} com o NO_x , verifica-se que as séries temporais de MP_{10} apresentaram um caráter inerentemente mais caótico.

4.4 ESTAÇÃO GUARULHOS

Seguindo o mesmo procedimento adotado para as estações anteriores, a Figura 4.10 apresenta a variação concentração anual média (Equação 3.1) de MP_{10} obtidas na estação de Guarulhos nos anos de 1998 a 2009. Todos os dados horários foram obtidos na plataforma QUALAR - CETESB. Os dados que deram origem a este gráfico bem como os respectivos desvios padrão são apresentados na Tabela 4.7

Figura 4.10 – Variação da concentração anual média de MP_{10} na estação de Guarulhos de 1998 a 2009 (o autor, 2018).



A estação de monitoramento de Guarulhos disponibiliza apenas dados históricos de 1998 a 2009, sendo assim, este foi o período considerado na análise. A estação não realizou medidas da concentração horária de NO_x . Tanto por meio do perfil da Figura 4.10 quanto pelos dados apresentados na Tabela 4.7, verifica-se uma clara redução da concentração anual de MP_{10} . Comparando os anos de 1998 e 2009, observa-se uma queda de quase 51% dos níveis de material particulado inalável. Conforme mencionado

anteriormente, os prováveis responsáveis por esta redução são a implementação de políticas públicas e a renovação da frota de veículos do estado de São Paulo. Mesmo com as reduções destacadas, observa-se que as emissões de MP_{10} de todos os anos ultrapassaram a média anual de $20 \mu\text{g}/\text{m}^3$ recomendada pela OMS. No caso do CONAMA N° 491 e Decreto Paulista 2013, que estipulam a média anual máxima de $40 \mu\text{g}/\text{m}^3$, verifica-se que dos doze anos analisados nenhum apresentou valores em conformidade. Evidentemente que aqui tanto o CONAMA N° 491 e Decreto Paulista 2013 servem apenas como base, pois estes foram promulgados apenas, respectivamente, em 2018 e 2013 e as aquisições realizadas na estação de Guarulhos se encerram em 2009.

Tabela 4.7 – Médias anuais e desvios padrão da concentração de MP_{10} na estação de Guarulhos (o autor, 2018).

Ano	MP_{10}	
	μ ($\mu\text{g}/\text{m}^3$)	σ ($\mu\text{g}/\text{m}^3$)
1998	85	48
1999	64	51
2000	61	49
2001	72	54
2002	71	55
2003	65	51
2004	61	49
2005	50	36
2006	69	52
2007	54	40
2008	51	46
2009	42	31

Os altos índices de MP_{10} apresentados na estação de monitoramento de Guarulhos estão fortemente relacionados ao local onde ela está instalada. A poucos metros do trevo entre duas importantes rodovias de acesso à capital paulista, Rodovia Presidente Dutra e Rodovia Hélio Smidt, esta estação capta a emissão de particulados inaláveis advindas dos escapamentos de milhares de veículos de carga e automóveis leves que passam diariamente pelo local. Outro fator a ser destacado é a presença do Aeroporto Internacional de São Paulo (Governador André Franco Montoro) que, segundo o Inventário Nacional de

Emissões Atmosféricas da Aviação Civil, publicado em 2014 pela Agência Nacional de Aviação Civil (ANAC), constatou o aumento bastante acentuado das emissões de material particulado inalável pelas dezenas de aeronaves que diariamente aterrissam e decolam no aeroporto. Este aumento é observado devido ao crescimento no setor da aviação civil.

Quanto às políticas públicas que, possivelmente, tenham influenciado as reduções mencionadas, destaca-se o PROCONVE, com início na década de 80, e o PROMOT (2003).

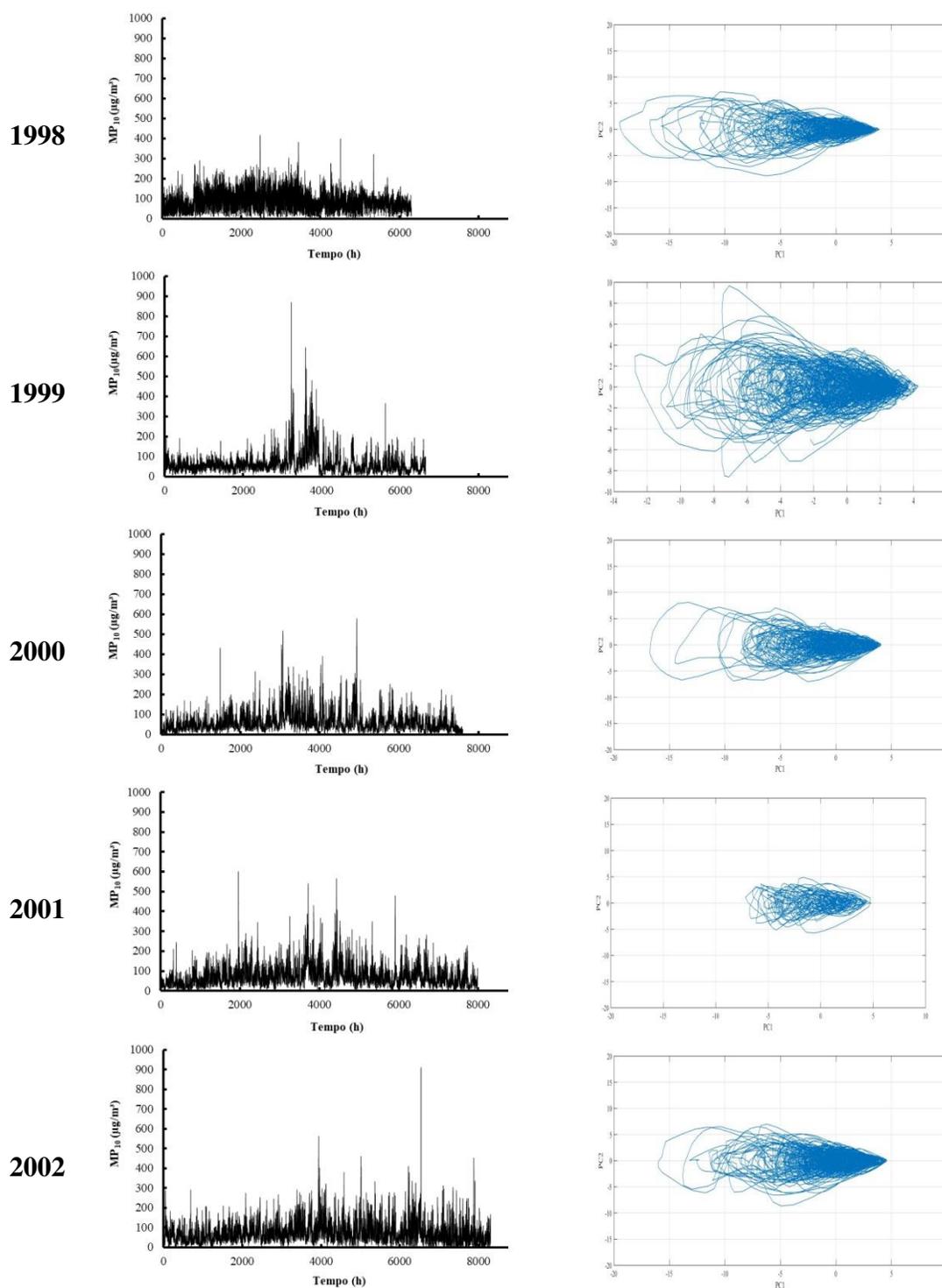
Sobre o desvio padrão, verifica-se, novamente, o caráter dispersivo e bastante variável dos dados sob análise, vez que seus valores estão na amplitude da própria média.

Depois de avaliadas as características gerais anuais dos dados, passa-se a avaliação das séries temporais que deram origem à esses dados. Estas séries são compostas de dados horários anuais e sua evolução temporal foi avaliada à luz da Teoria do Caos, cujos objetivos são avaliar a presença de subjacentes características caóticas nos dados.

Desta feita, inicialmente reconstrói-se, a partir da série temporal, a dinâmica do sistema no espaço fases. A figura obtida recebe o nome de atrator e sua reconstrução, com características fractais, está imersa em um espaço I dimensional chamada dimensão de imersão. Das características topográficas do atrator são abstraídos o máximo expoente de Lyapunov, entropia de Kolmogorov e dimensão de correlação, invariantes que auxiliam no estudo do caos determinístico.

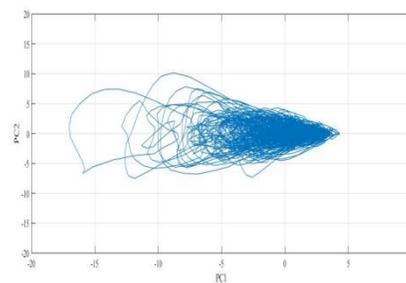
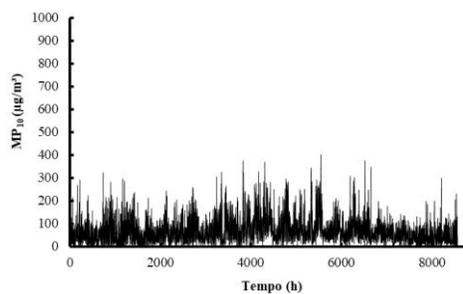
Neste sentido, a Figura 4.11 apresenta as séries temporais da concentração horária de MP_{10} e os respectivos atratores reconstruídos para cada ano analisado (1998 a 2009).

Figura 4.11 – Séries temporais e atratores de concentração anual de MP_{10} obtidas na estação de monitoramento de Guarulhos nos anos de 1998 a 2009 (o autor, 2018).

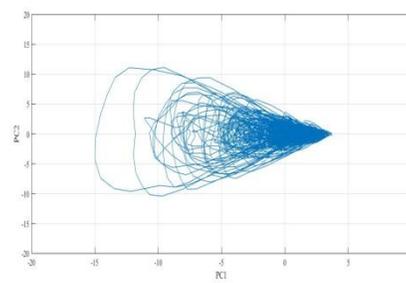
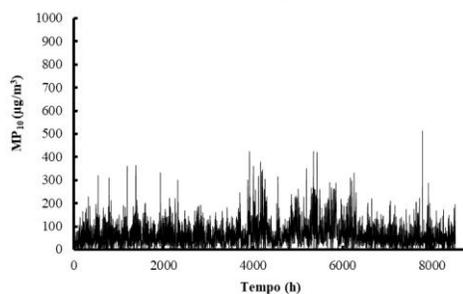


Capítulo IV – Resultados e Discussões

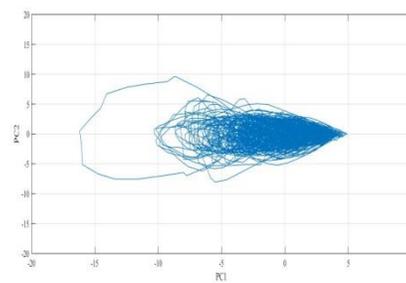
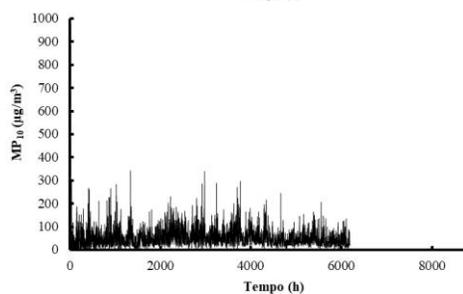
2003



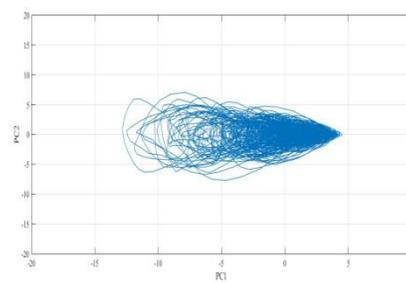
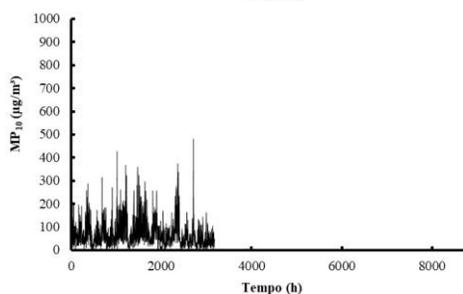
2004



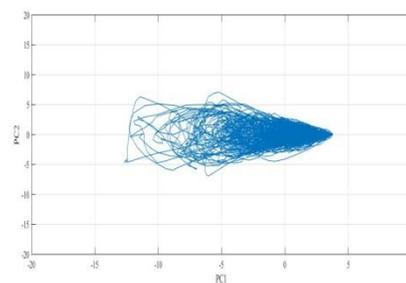
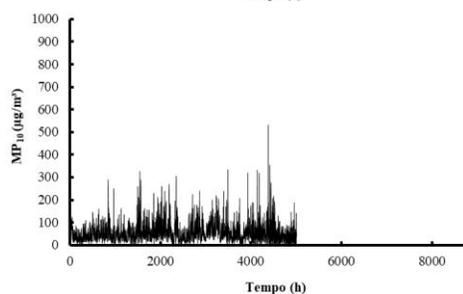
2005

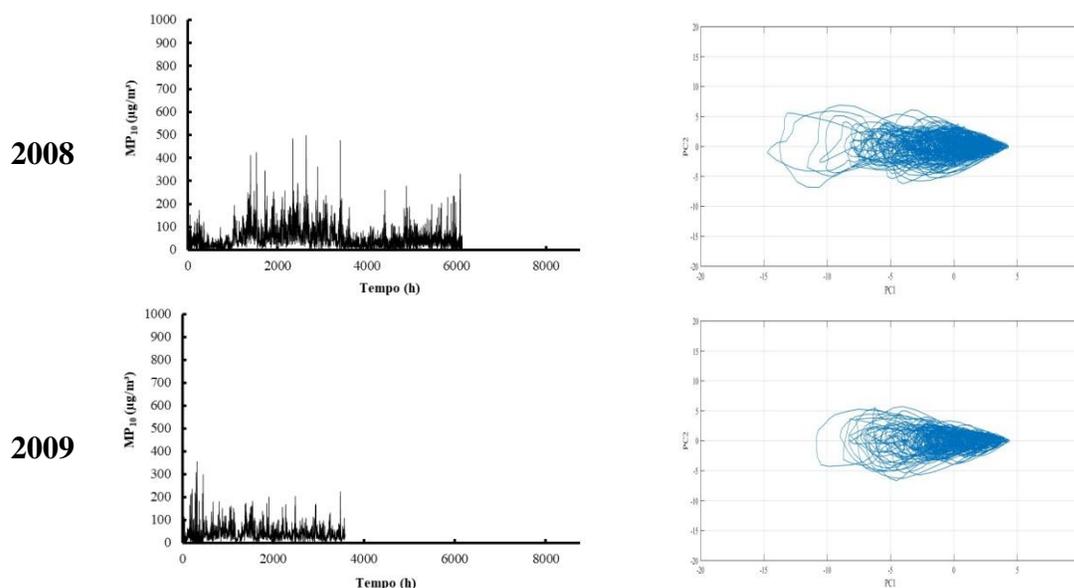


2006



2007





Fazendo uma análise das séries, verifica-se, assim como nas estações anteriores, a falta de dados ao final da evolução temporal. A existência de diversos dados inválidos ao longo do ano obriga o lançamento destas lacunas ao final da série de modo a torná-la contínua para que, assim, seja possível aplicar a técnica do Caos. Este procedimento faz com que algumas características importantes do fenômeno sejam perdidas. A poluição atmosférica está intrinsicamente relacionada às características meteorológicas e apresenta sazonalidade devido ao ciclo das estações do ano (primavera, verão outono e inverno), por isso, quando um grande volume de dados é agrupado devido à falta de informações precedentes, características cíclicas associadas ao período são perdidas, acarretando em erros à análise. Os anos de 2005, 2006, 2007, 2008 e 2009 foram afetados por esta ausência de informações. De mais a mais, corroboram-se as constatações feitas anteriormente por meio da análise do desvio padrão, são verificadas séries bastante variáveis com valores dispersos. No que se refere à reconstrução dos atratores, esta foi realizada contando com o mesmo número de pontos da série temporal, dimensão de imersão (I) igual a 11 e tempo de atraso (θ) de 1 hora. Similar ao notado para as demais estações, as linhas de fluxo resultaram em atratores estranhos, convergindo para uma região bem definida e estreita no espaço de fases (linha horizontal em PC2 (principal componente) = 0), asseverando uma dinâmica caótica determinista.

Continuando com a análise do caos determinístico, a Tabela 4.8 apresenta a dimensão de imersão (I), dimensão de correlação (D_2), número de graus de liberdade (GL), entropia de Kolmogorov (K_{ML}) e máximo expoente de Lyapunov obtidos a partir da variação da concentração horária de MP_{10} na estação de monitoramento de Guarulhos nos anos de 1998 a 2009.

Tabela 4.8 – *Invariantes do Caos para o MP_{10} avaliados na estação de monitoramento de Guarulhos dos anos de 1998 a 2009 (o autor, 2018).*

MP_{10}					
Ano	I	D_2 [-]	GL	K_{ML} (bits/s)	λ [-]
1998	6	4,00	4	0,3149	0,11
1999	10	3,07	4	0,0880	0,14
2000	13	4,29	5	0,1472	0,15
2001	10	3,70	4	0,1396	0,10
2002	10	3,68	4	0,2023	0,15
2003	10	4,27	5	0,2380	0,15
2004	9	4,13	5	0,2215	0,14
2005	11	3,96	4	0,3160	0,16
2006	12	3,21	4	0,1244	0,14
2007	10	4,08	5	0,3219	0,14
2008	12	3,93	4	0,0780	0,14
2009	11	3,77	4	0,1635	0,14

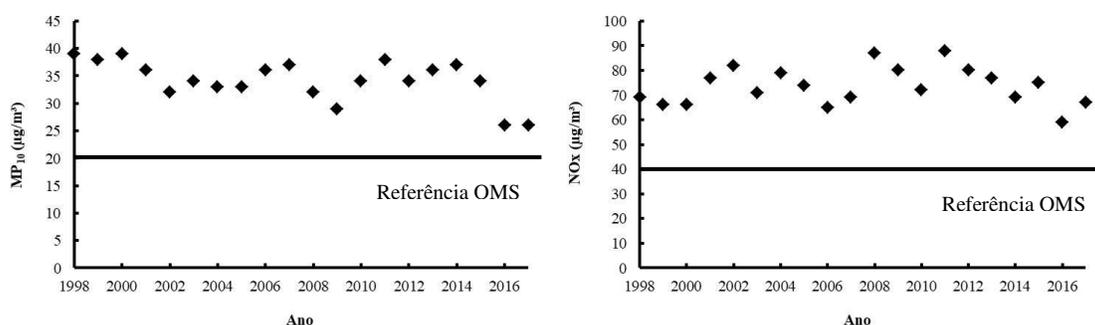
A Tabela 4.8 confirma exatamente as constatações feitas por meio dos atratores, todas as evoluções apresentam características claramente caóticas pois $0 < K_{ML} < \infty$ e $\lambda > 0$ para todas as séries analisadas. Análogo ao realizado para as estações anteriores, quando são comparados os valores de D_2 da Tabela 4.8 com os obtidos por Lee e Lin (2008), que obtiveram $D_2 = 4,32$, verifica-se que em diversos anos os valores são bem próximos ao encontrado pelos autores, inclusive no que tange o número de graus de liberdade sugeridos por Lee e Lin (2008), $GL = 5$, na estação de Guarulhos, nos anos de 2000, 2003 e 2007, o mesmo valor foi obtido. Extensivamente discutido nas estações anteriores, o quinto grau de liberdade é uma superestimação de um grau de liberdade, sendo categórica a presença de 4 GL. No caso da estação de Guarulhos, os quatro graus de liberdades podem ser:

- (i) as condições meteorológicas;
- (ii) as emissões poluentes;
- (iii) a proximidade com pontos de alto tráfego de veículos leves e pesados (Rodovia Presidente Dutra, interligação com a Rodovia Hélio Smidt e a Avenida Presidente Tancredo Neves) e
- (iv) a presença do aeroporto internacional de Guarulhos.

4.5 ESTAÇÃO DE CUBATÃO

Partindo para a última estação analisada, a Figura 4.12 apresenta a variação anual média da concentração de MP_{10} e NO_x dos anos de 1998 a 2017 em Cubatão. Cada um dos pontos representa uma média (Equação 3.1) a partir da aquisição horário (frequência de 1 hora) durante um ano inteiro. Os dados discretos bem como os respectivos desvios padrão de cada ano estão elencados na Tabela 4.9.

Figura 4.12 – Variação da concentração anual média de MP_{10} e NO_x na estação de Cubatão de 1998 a 2017 (o autor, 2018).



De imediato, verifica-se pela Figura 4.12 um discreto perfil de redução nos materiais particulados inaláveis. Por meio da Tabela 4.9 é possível especificar uma redução de 33% nos níveis de MP_{10} quando comparados os anos de 1998 e 2017, contudo, 2016 e 2017 foram os anos em que esta redução se acentuou, sendo que até este período as emissões variaram pouco. No caso do NO_x , quando comparados os anos de 1998 e 2017, verifica-se uma redução de menos de 3%, sendo que os valores de concentração anual de

óxidos de nitrogênio pouco variaram. Esta foi a única estação analisada nesta Tese em que foram observados percentuais de redução quase nulos com perfis de concentração de poluentes quase constantes. A explicação para este reside nas características da localidade onde a estação está instalada.

Tabela 4.9 – Médias anuais e desvios padrão da concentração de MP_{10} e NO_x na estação de Cubatão (o autor, 2018).

Ano	MP_{10}		NO_x	
	μ ($\mu\text{g}/\text{m}^3$)	σ ($\mu\text{g}/\text{m}^3$)	μ ($\mu\text{g}/\text{m}^3$)	σ ($\mu\text{g}/\text{m}^3$)
1998	39	32	69	64
1999	38	32	66	57
2000	39	35	66	58
2001	36	35	77	68
2002	32	21	82	70
2003	34	26	71	69
2004	33	21	79	73
2005	33	27	74	67
2006	36	28	65	56
2007	37	28	69	69
2008	32	23	87	80
2009	29	20	80	69
2010	34	26	72	71
2011	38	26	88	78
2012	34	27	80	74
2013	36	27	77	71
2014	37	28	69	60
2015	34	26	75	62
2016	26	21	59	50
2017	26	18	67	57

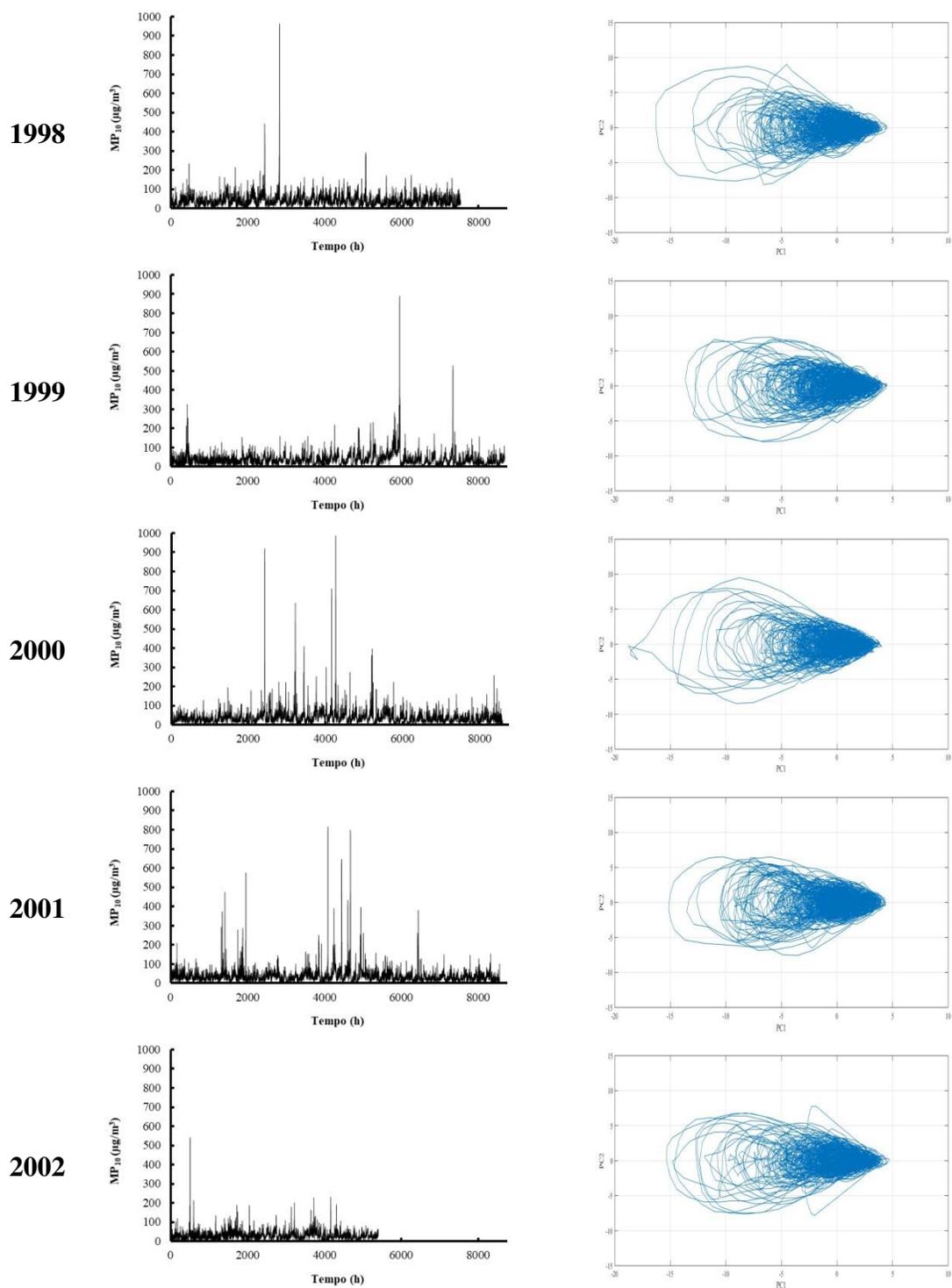
A região que abriga a estação de Cubatão apresenta uma geografia bastante acidentada, com montes, vales e uma região verde bastante acentuada onde há também a presença do Rio Cubatão. São observadas a existência de ruas de alto fluxo de automóveis pesados e leves (Avenida 9 de Abril, Avenida Tiradentes, Rodovia Cônego Domenico Rangoni e outras) que servem de escoamento das pesadas cargas transportadas pelos caminhões que trafegam do e para o imenso parque industrial que a cidade de Cubatão

possui. Mesmo com a implementação de diversas políticas públicas no Estado de São Paulo e no Brasil, exaustivamente salientadas anteriormente, os níveis dos poluentes aqui estudados mantem-se praticamente constante, demonstrando uma saturação da atmosfera local, resultante da contínua poluição fotoquímica, com produção de material particulado secundário e ozônio durante a movimentação atmosférica (Machado, 2008). Diversos danos aos ecossistemas e na saúde da população local são conhecidos (Pereira, 2004; Moraes, 2003) e estão diretamente relacionados às emissões do parque industrial de Cubatão. Com uma topografia que dificulta a dispersão dos poluentes, pelo fato de as indústrias se localizarem na base da Serra do Mar, o transporte dos poluentes sofre resistência física dos montes e da vegetação que compõe a serra, transformando a região em uma atmosfera a parte do restante do Estado de São Paulo. A movimentação atmosférica também auxilia a concentração dos poluentes, pois durante o dia a brisa do mar predomina já ao anoitecer, via de regra, a direção dos ventos sofre uma inversão, transportando os poluentes na direção do mar, já o ar de São Paulo descende passando sobre a Serra do Mar.

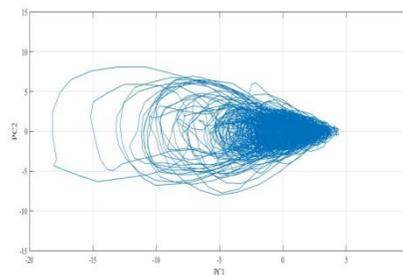
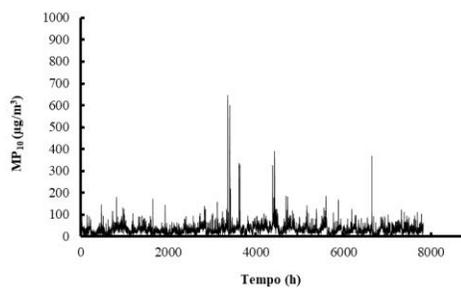
Muito embora localizada em área de difícil dispersão, os níveis de MP_{10} e NO_x encontrados na estação de Cubatão não são os mais elevados quando comparados às demais estações, contudo observa-se que as emissões de MP_{10} de todos os anos ultrapassaram a média anual de $20 \mu\text{g}/\text{m}^3$ recomendada pela OMS. Já em relação ao estipulado pelo CONAMA N° 491 e pelo Decreto Paulista 2013, ambos máxima de $40 \mu\text{g}/\text{m}^3$, nenhum dos anos apresentou níveis superiores aos dois limites estipulados. Para o NO_x , a OMS estipula o máximo de $40 \mu\text{g}/\text{m}^3$ e, sendo assim, são verificadas ultrapassagens em todos os anos avaliados.

Passando para a análise do Caos, a Figura 4.13 e 4.14 apresentam as séries temporais bem como os atratores reconstruídos para, respectivamente, a concentração anual de MP_{10} e NO_x dos anos de 1998 a 2017.

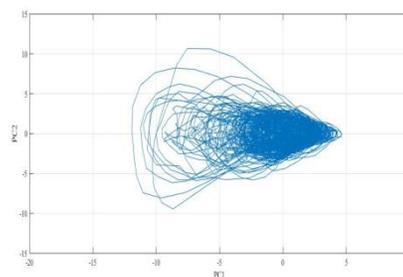
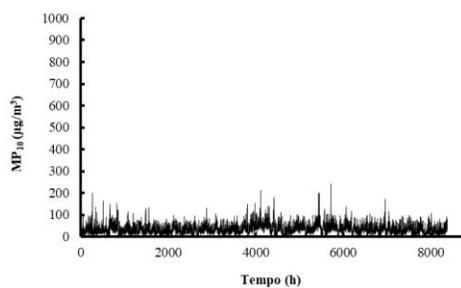
Figura 4.13 – Séries temporais e atratores de concentração anual de MP_{10} obtidas na estação de monitoramento de Cubatão nos anos de 1998 a 2017 (o autor, 2018).



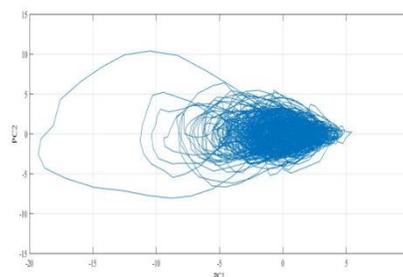
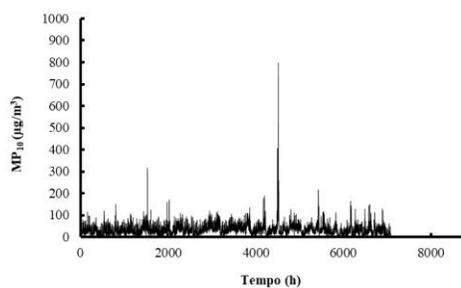
2003



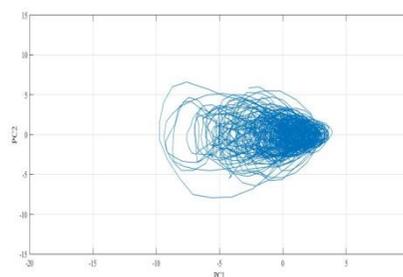
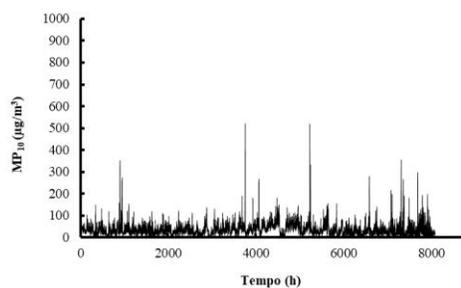
2004



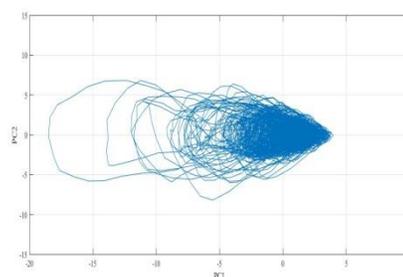
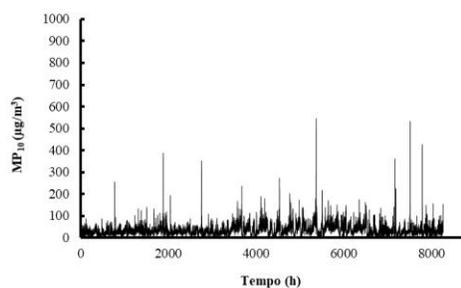
2005



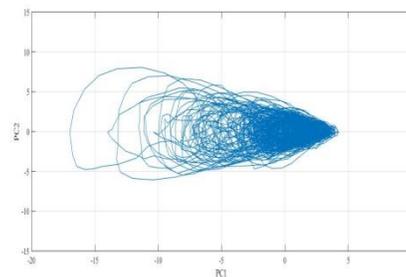
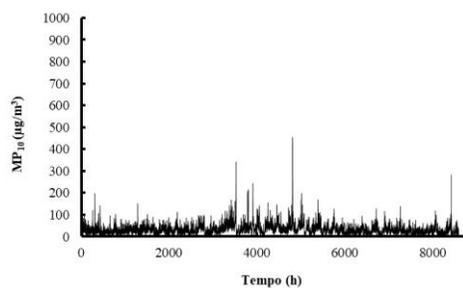
2006



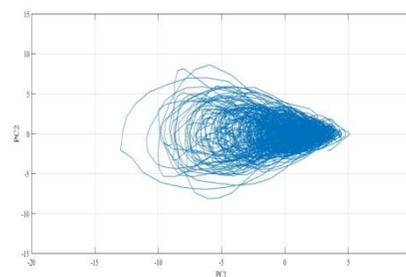
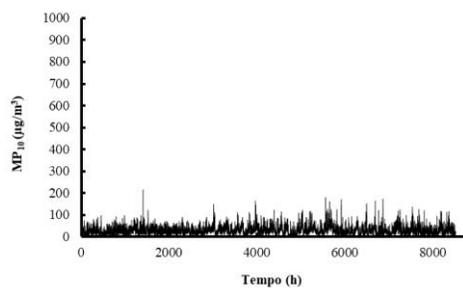
2007



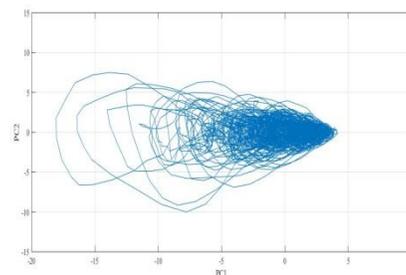
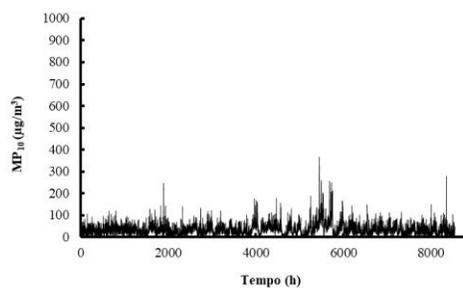
2008



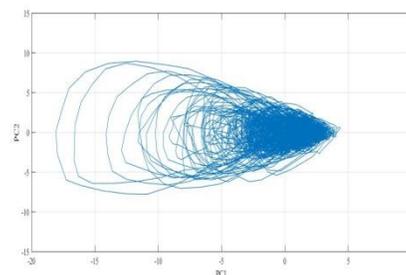
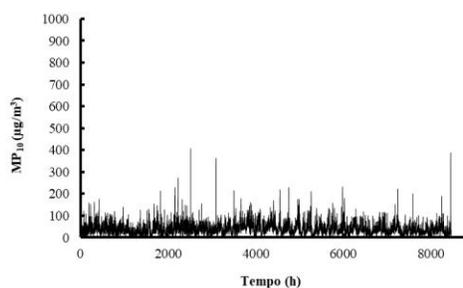
2009



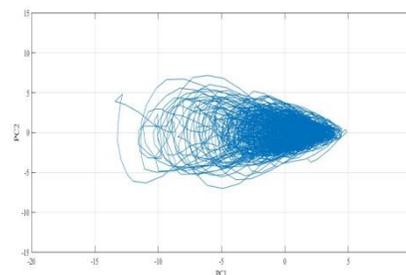
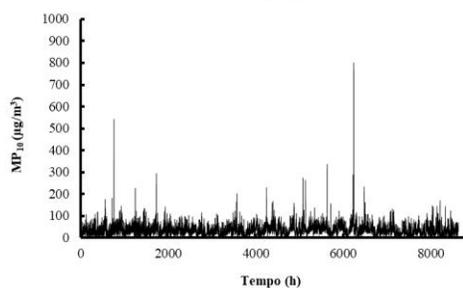
2010



2011

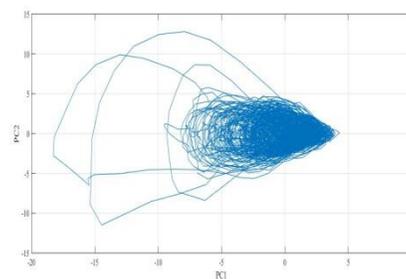
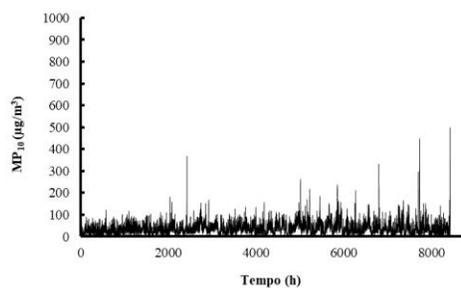


2012

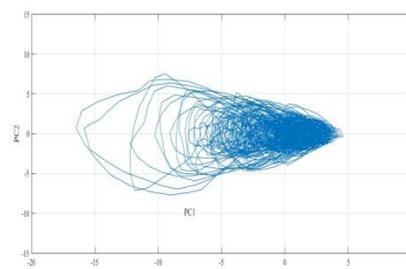
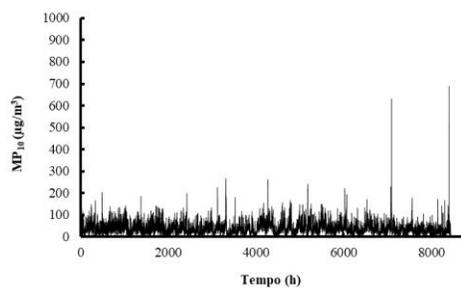


Capítulo IV – Resultados e Discussões

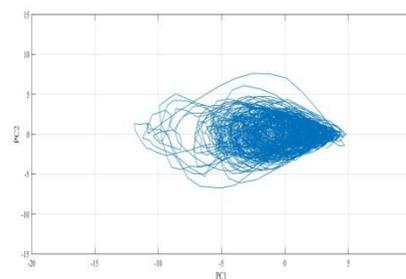
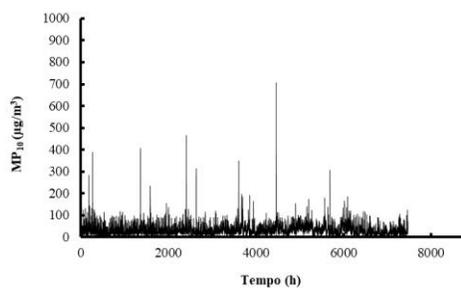
2013



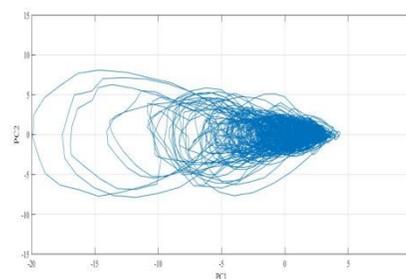
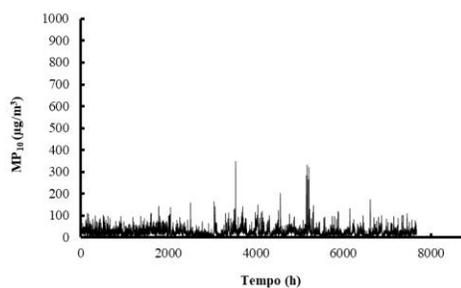
2014



2015



2016



2017

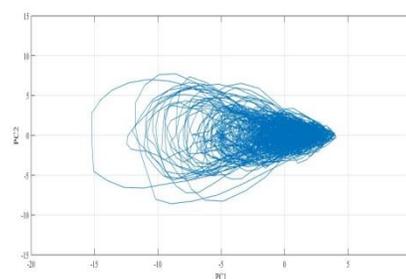
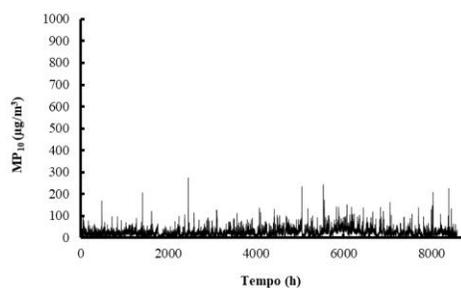
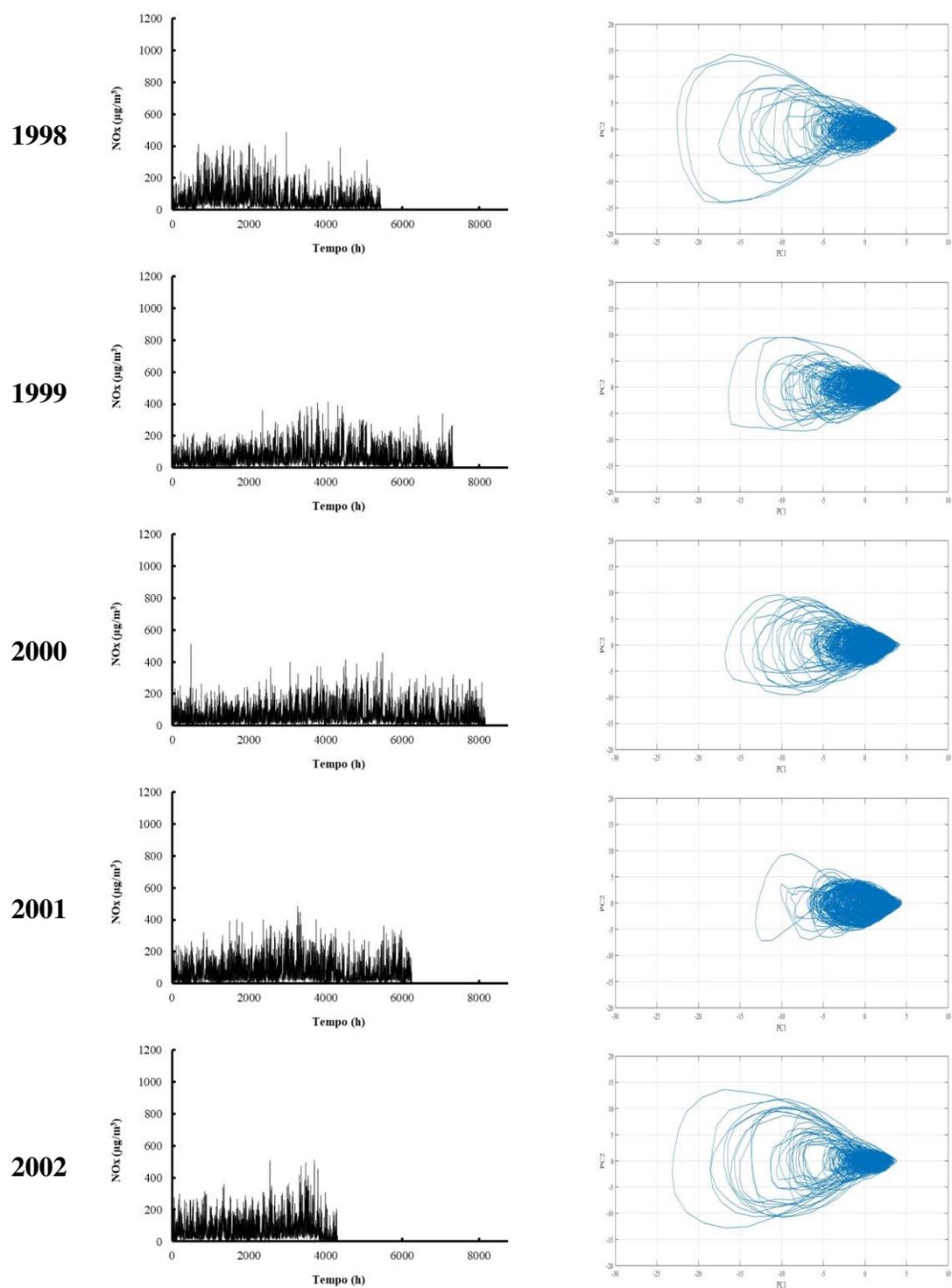
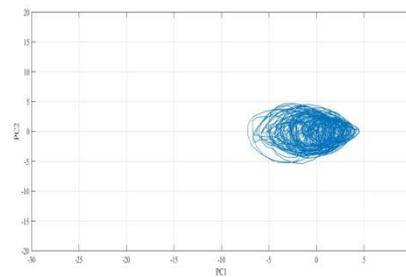
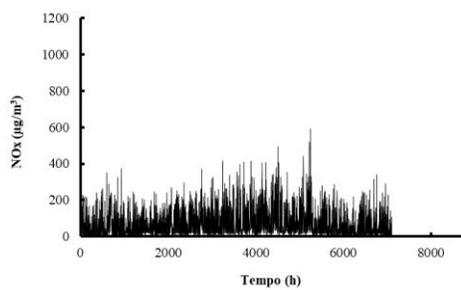


Figura 4.14 – Séries temporais e atratores de concentração anual de NOx obtidas na estação de monitoramento de Cubatão nos anos de 1998 a 2017 (o autor, 2018).

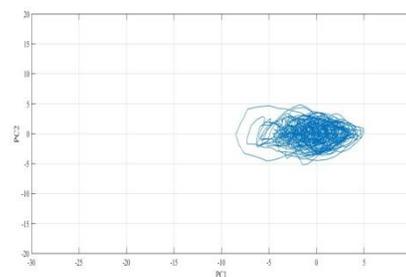
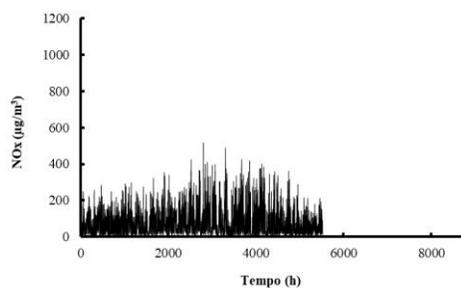


Capítulo IV – Resultados e Discussões

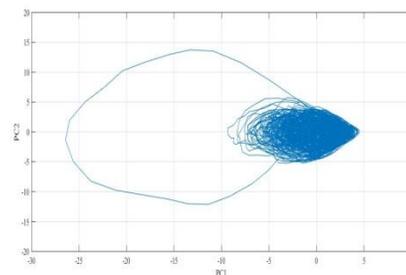
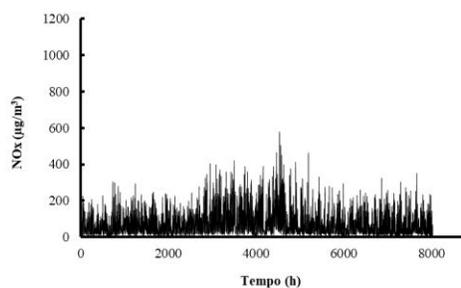
2003



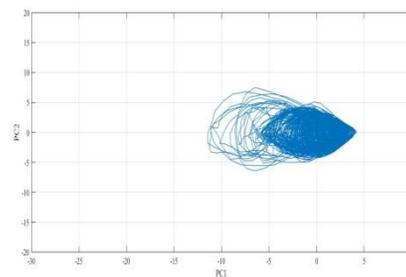
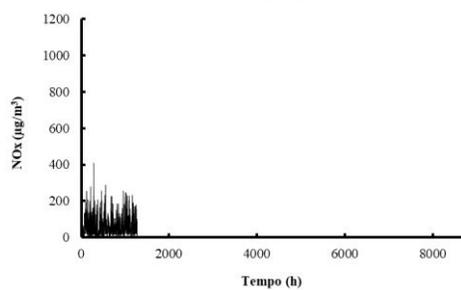
2004



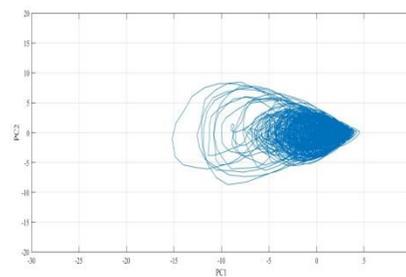
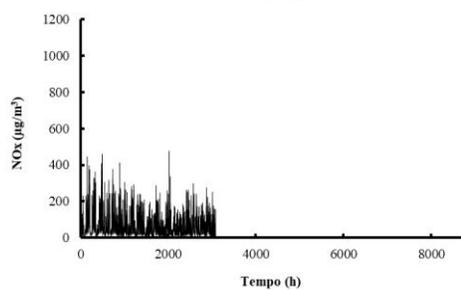
2005



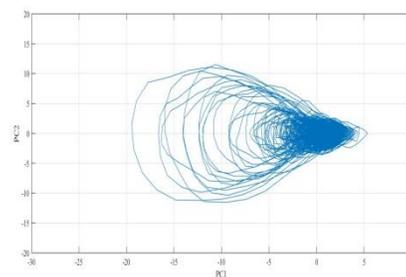
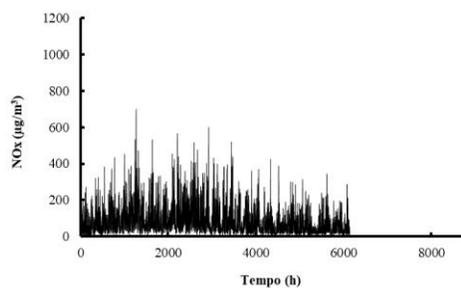
2006



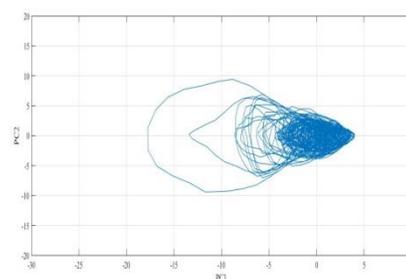
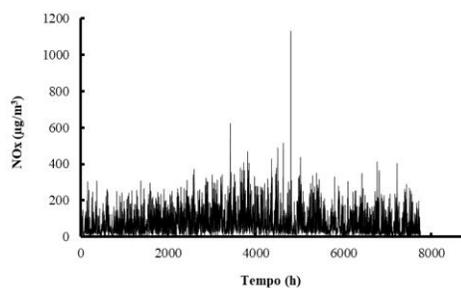
2007



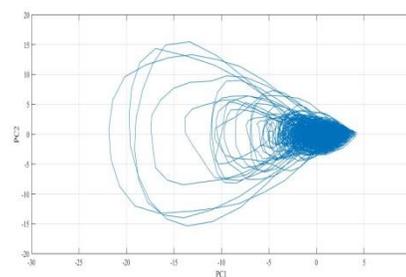
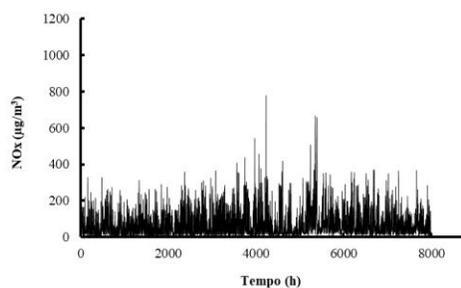
2008



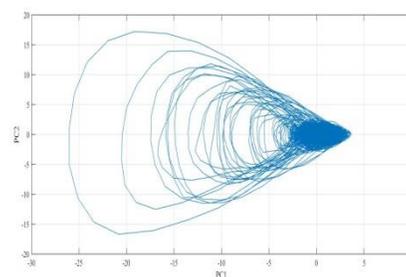
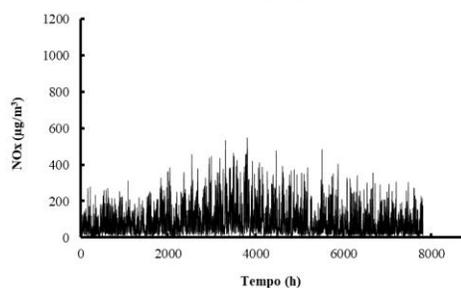
2009



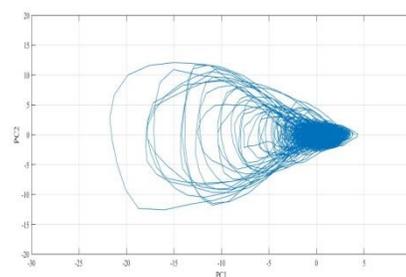
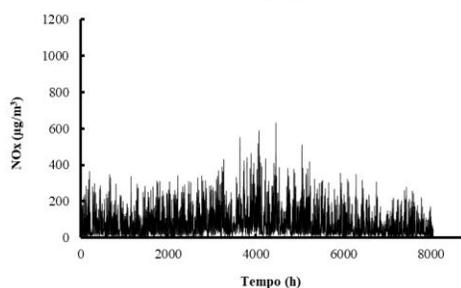
2010

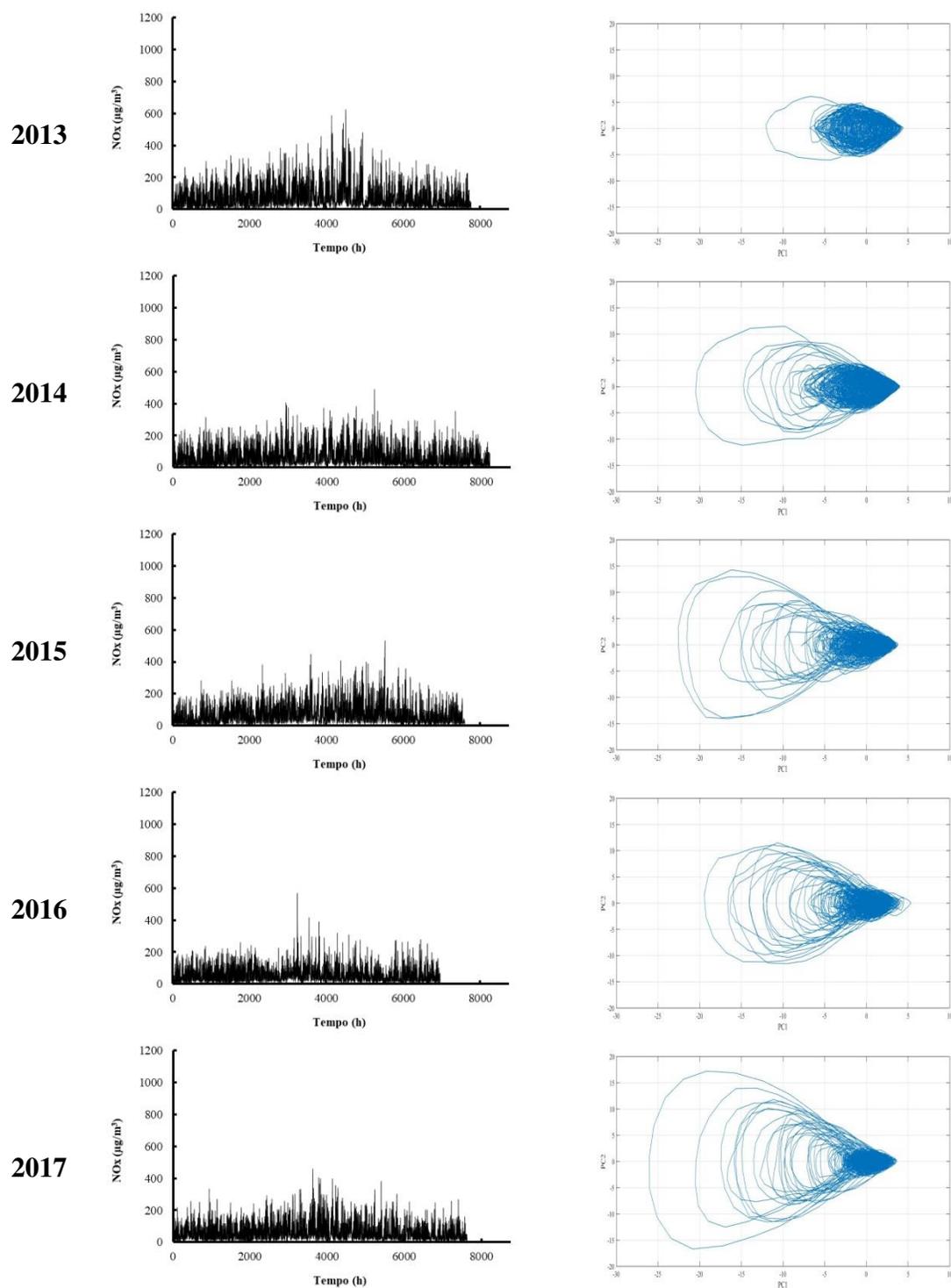


2011



2012





Análogo ao notado nas séries obtidas para todas as demais estações, verifica-se a inexistência de dados ao término da evolução, consequência direta do deslocamento das lacunas oriundas de dados inválidos. No caso da estação da Cubatão este problema é

bastante recorrente nas séries de concentração anual de NOx e o prejuízo acarretado pela falta destes dados se espelha diretamente na análise do Caos, especialmente quando da avaliação dos invariantes. Ainda sobre as séries temporais, é notável o caráter variável dispersivo em relação à média presente no conjunto de dados, características essas ocasionadas pela própria natureza do fenômeno da poluição atmosférica que já foram discutidas anteriormente. Os atratores apresentados nas Figuras 4.13 e 4.14 foram reconstruídos, respectivamente para o MP₁₀ e NOx, com $I = 10$ e $I = 14$ e em ambos os casos, o tempo de atraso (θ) foi de 1 hora. Em todas as situações avaliadas, observa-se que as linhas de fluxo culminam em atratores estranhos, exibindo uma clara convergência para uma região bem definida e estreita no espaço de fases (linha horizontal em PC2 (principal componente) = 0), sugerindo uma dinâmica caótica determinista.

De modo a asseverar as conclusões supracitadas, a Tabela 4.10 lista os invariantes do caos determinístico obtidos para a concentração anual de MP₁₀ e NOx na estação de Cubatão dos anos de 1998 a 2017.

Não diferente das demais estações, realizou-se, inicialmente, a otimização da dimensão de imersão por meio do algoritmo de Grassberger e Procaccia (1983) e na sequência D_2 , K_{ML} e λ foram calculados adotando a média dos valores de I . Esse procedimento garante que todos os invariantes representem a mesma dimensão e, portanto, sejam elegíveis à comparação. Em linhas gerais, as conclusões aqui não são diferentes das abstraídas para as demais estações. O número de GL variou entre 4 e 5 para o MP₁₀ e 3 e 5 para o NOx. Aqui também é notada a superestimação do quinto grau de liberdade, sendo possível assegurar a presença de 4 GL exercendo influência na evolução temporal. Conforme realizado para as demais estações, nota-se ser possível, a partir das características locais da aquisição, determinar os GL. Sendo assim, para a Cubatão:

- (i) *as condições meteorológicas;*
- (ii) *as emissões poluentes;*
- (iii) *a presença de vias de alto tráfego e do parque indústria e*
- (iv) *geografia do local onde está instalada a estação.*

Tabela 4.10 – Invariantes do Caos para os poluentes MP_{10} e NO_x avaliados na estação de monitoramento de Cubatão dos anos de 1998 a 2017 (o autor, 2018).

Ano	MP ₁₀					NO _x				
	I	D ₂ [-]	GL	K _{ML} (bits/s)	λ [-]	I	D ₂ [-]	GL	K _{ML} (bits/s)	λ [-]
1998	11	4,08	5	0,2340	0,16	14	3,65	4	0,1177	0,13
1999	12	4,17	5	0,1300	0,15	13	3,73	4	0,1772	0,13
2000	12	4,63	5	0,1458	0,15	13	4,00	4	0,1218	0,13
2001	10	4,48	5	0,1081	0,15	14	3,33	4	0,1148	0,13
2002	9	4,00	4	0,1993	0,14	13	3,44	4	0,1481	0,12
2003	11	3,79	4	0,1459	0,15	14	3,77	4	0,1680	0,13
2004	9	4,16	5	0,3959	0,15	14	3,13	4	0,1504	0,13
2005	10	4,21	5	0,2248	0,15	14	3,82	4	0,1087	0,13
2006	10	4,25	5	0,1601	0,14	12	2,66	3	0,2345	0,10
2007	10	4,06	5	0,1687	0,14	15	3,14	4	0,1043	0,13
2008	9	4,65	5	0,2442	0,14	15	3,68	4	0,1103	0,13
2009	8	3,58	4	0,2816	0,14	14	3,78	4	0,1265	0,12
2010	9	4,11	5	0,2053	0,14	13	3,44	4	0,0893	0,13
2011	8	3,67	4	0,3041	0,15	13	3,10	4	0,1456	0,13
2012	8	4,09	5	0,3023	0,14	14	3,75	4	0,1543	0,13
2013	8	3,97	4	0,2607	0,14	14	4,02	5	0,1083	0,13
2014	8	3,39	4	0,2756	0,14	15	4,07	5	0,1964	0,13
2015	8	4,38	5	0,2758	0,14	15	3,81	4	0,1360	0,12
2016	7	3,32	4	0,1609	0,14	13	3,71	4	0,1175	0,13
2017	8	4,23	5	0,1550	0,14	15	3,60	4	0,1800	0,13

Já em relação a dinâmica das séries analisadas, é evidente a presença de caos uma vez que $0 < K_{ML} < \infty$ e $\lambda > 0$ para todas as situações.

Fazendo uma verificação geral de todas as estações analisadas, no que tange os índices de poluição, nota-se que a estação de Guarulhos foi a que apresentou os maiores valores de concentração horária média de MP_{10} dentro dos períodos analisados. Já a estação de Congonhas foi a com maior concentração de NO_x . Destacando os perfis de redução, a estação de Guarulhos foi a que apresentou o maior índice de redução de material particulado inalável, ao passo que em Congonhas a redução da concentração horária de NO_x foi a mais acentuada. Em geral, exceto raros casos, as estações analisadas se encontram na mesma faixa de concentração quando comparados os dois poluentes. É muito

difícil precisar os motivos particulares que levam a determinadas reduções ou estabilizações, uma vez que as movimentações atmosféricas estão interligadas por características meteorológicas que, por sua vez, exercem influência direta na concentração e dispersão dos poluentes do ar. Pode-se citar a influência da estabilidade atmosférica como um dos fatores mais importantes, pois ela afeta o movimento vertical do ar, influenciando na taxa com o qual os compostos poluentes são misturados ao ar limpo, causando uma inibição do transporte dos poluentes no ar. A altura de mistura é outro fator bastante importante, pois está relacionada à distância vertical entre a superfície da Terra e a altitude das correntes de convecção, sendo assim, quanto maior a altura de mistura, verifica-se uma maior quantidade de ar limpo em mistura com uma pequena quantidade de poluentes.

4.6 PREVISÕES

O método ou algoritmo do vizinho mais próximo é um instrumento popular para realizar previsões sobre, por exemplo, o comportamento do mercado de ações em curto prazo. Trata-se de uma ferramenta de abordagem não paramétrica baseada na análise de indicadores quantitativos e também em padrões gráficos, a fim de identificar o tempo de entrada (compra) ou saída do mercado (venda) produzindo, portanto, previsões da série temporal em $t + 1$ (Perlin e Ceretta, 2007). O principal objetivo desta seção é o de estudar o desempenho de previsão do método do vizinho mais próximo para os dados horários de concentração dos poluentes atmosféricos (MP₁₀ e NO_x) para as estações do Ibirapuera, Congonhas, Centro de São Paulo, Guarulhos e Cubatão. Para tal, o método inicialmente faz o reconhecimento de padrões em um conjunto de dados conhecidos (série de treino) e depois “extrapola” (*out of sample*) uma nova série temporal baseada nos padrões encontrados.

O procedimento para todas as estações se iniciou com o estudo de k (número de vizinhos) e a avaliação da qualidade das previsões em cada uma das localidades mantendo como dimensão de imersão o valor gerado pelo algoritmo de Grassberger e Procaccia (1983) (I_{GP}) para as séries de treino. Na sequência realizou-se a proposta de generalização

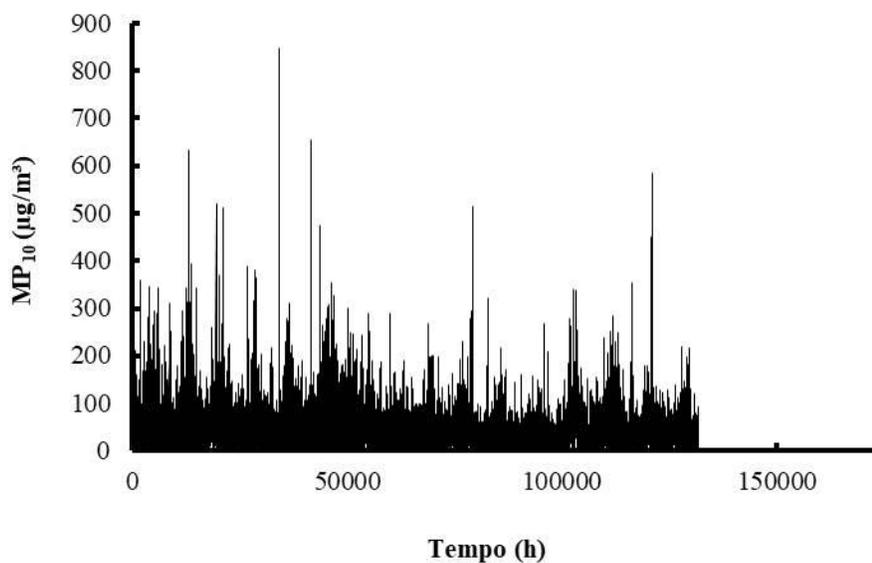
de k (com $I = I_{GP}$) e, por último, avaliou-se a qualidade das previsões com k e I generalizados.

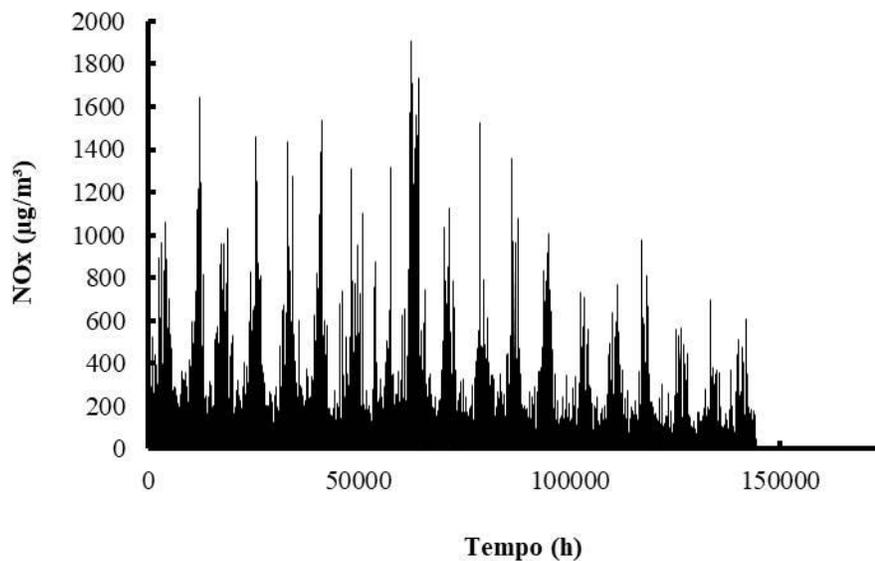
4.6.1 Estação de Ibirapuera

A Figura 4.15 apresenta a série de treino utilizada na previsão da concentração horária de MP_{10} e NO_x na estação do Ibirapuera. Nesta figura são encontrados os dados horários de MP_{10} e NO_x , obtidos na estação do Ibirapuera, dos anos de 1998 a 2017. Conforme observado nas séries individuais, os dados agrupados apresentam um perfil dispersivo, pouco comportado e com vários picos.

O valor de concentração médio encontrado por meio da Equação 3.1 para o MP_{10} foi de 38 ± 30 ($\mu\text{g}/\text{m}^3$) e para o NO_x de 64 ± 87 ($\mu\text{g}/\text{m}^3$). Novamente, para os dois casos, verifica-se um desvio padrão médio da ordem da própria média, assegurando, portanto, o caráter bem variável e dispersivo dos dados.

Figura 4.15 – Série de treino da concentração horária de MP_{10} e NO_x dos anos de 1998 a 2017 na estação do Ibirapuera (o autor, 2018).





No que diz respeito à dinâmica temporal destes conjuntos, a Figura 4.16 e 4.17 apresentam o atratores reconstruídos para as séries de treino da concentração horária de, respectivamente, MP₁₀ e NOx de 1998 a 2017.

Figura 4.16 – Atrator reconstruído a partir da série de treino da concentração horária de MP₁₀ para os anos de 1998 a 2017 (Ibirapuera) (o autor, 2018).

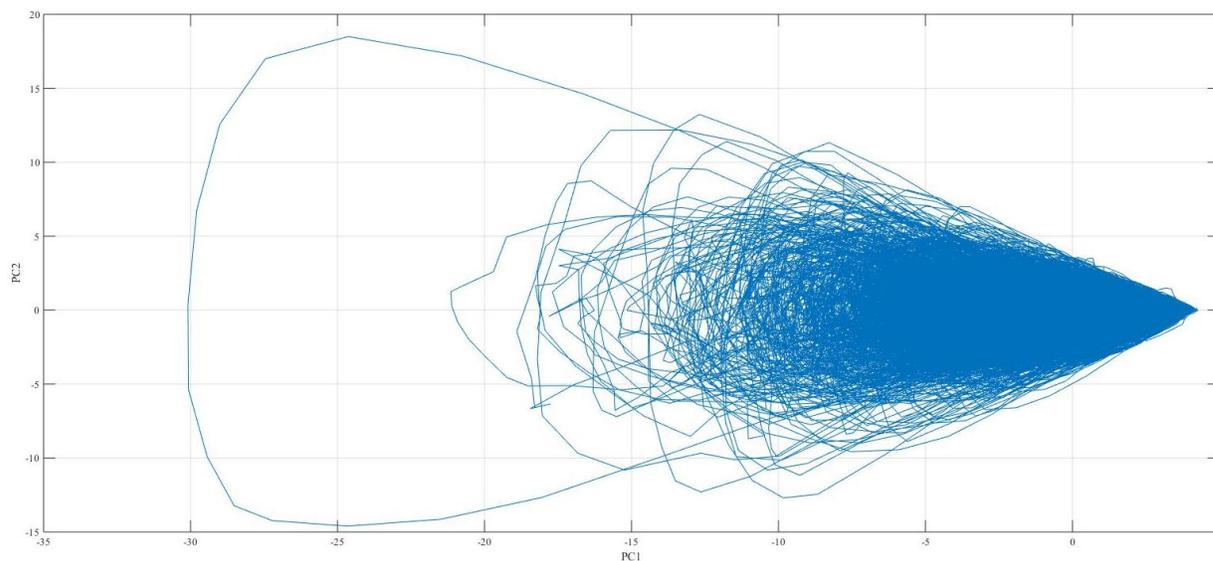
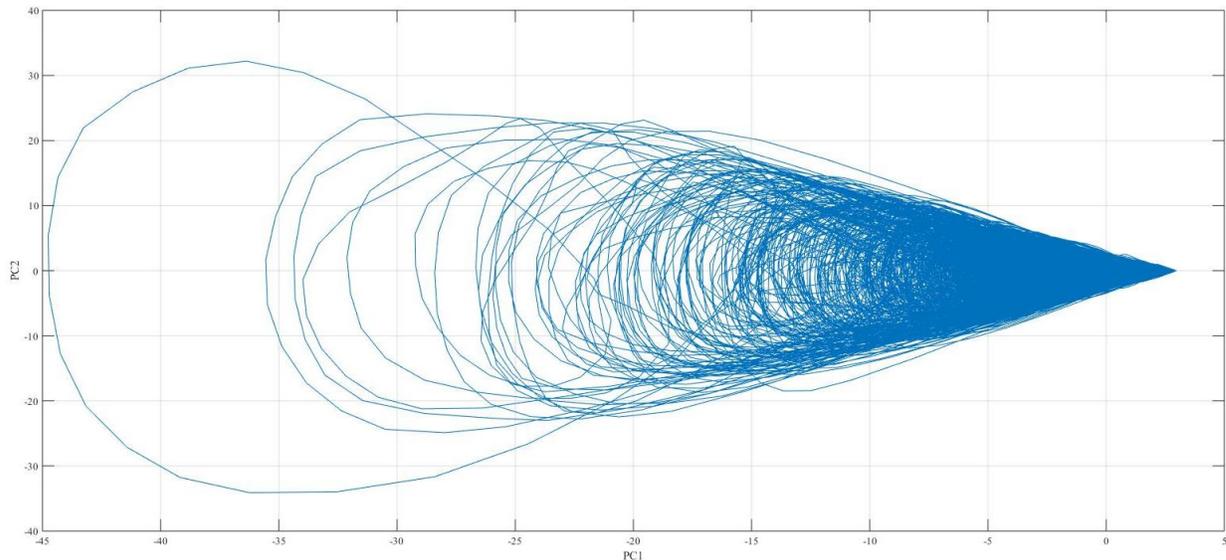


Figura 4.17 – Atrator reconstruído a partir da série de treino da concentração horária de NOx para os anos de 1998 a 2017 (Ibirapuera) (o autor, 2018).

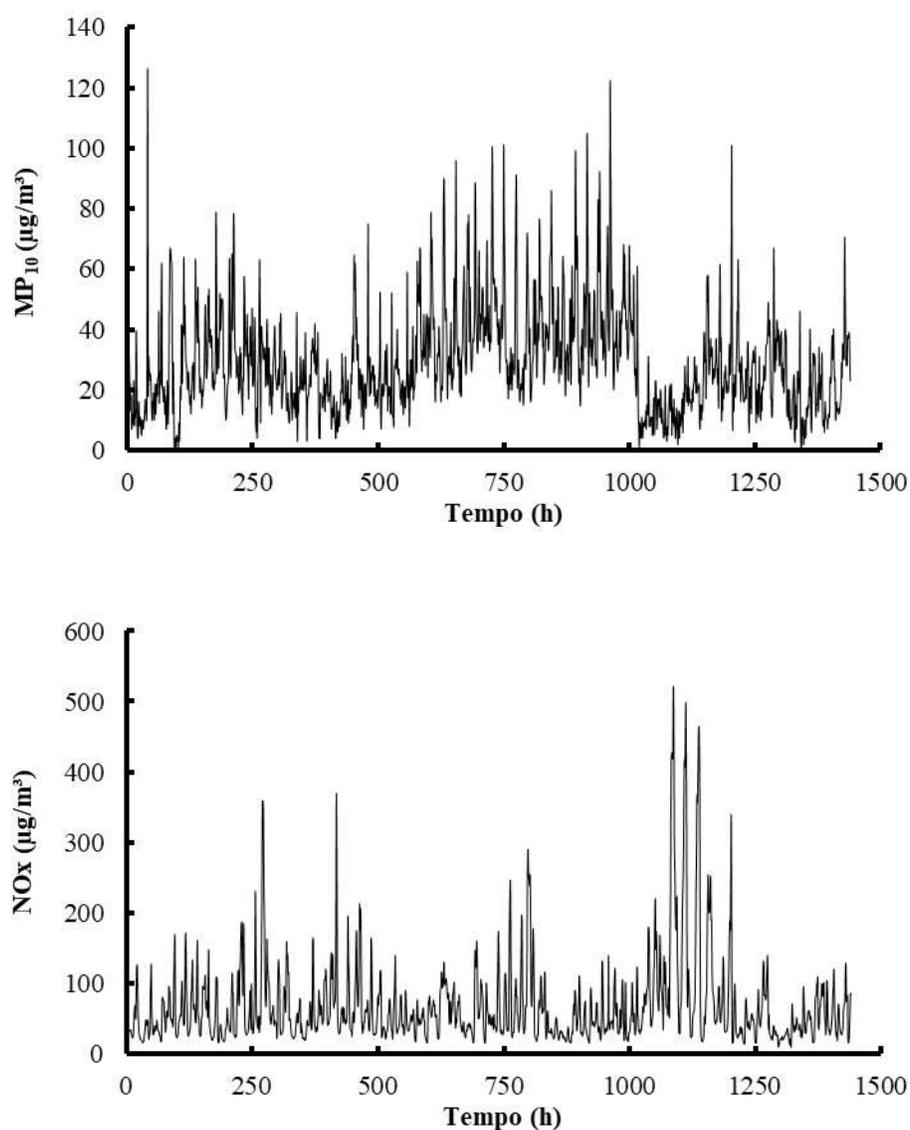


É evidente o fato de que as evoluções das linhas dos dois atratores culminam em dinâmica caótica, contudo, para certificar, os valores de D_2 , K_{ML} e λ foram determinados, resultando em $D_2 = 4,7026$, $K_{ML} = 0,0831$ bits/s e $\lambda = 0,15$ para o MP₁₀ e $D_2 = 4,7640$, $K_{ML} = 0,0373$ bits/s e $\lambda = 0,12$ para o NOx. O tempo de atraso utilizado foi igual à frequência de amostragem (1 hora) e a dimensão de imersão otimizada por meio da metodologia de Grassberger e Procaccia (1983) resultou em $I = 13$ para o MP₁₀ e $I = 19$ para o NOx.

Por meio dos invariantes do Caos determinístico, fica absolutamente evidente a evolução caótica dos dados que apresentaram 5 graus de liberdade regentes da dinâmica dos conjuntos.

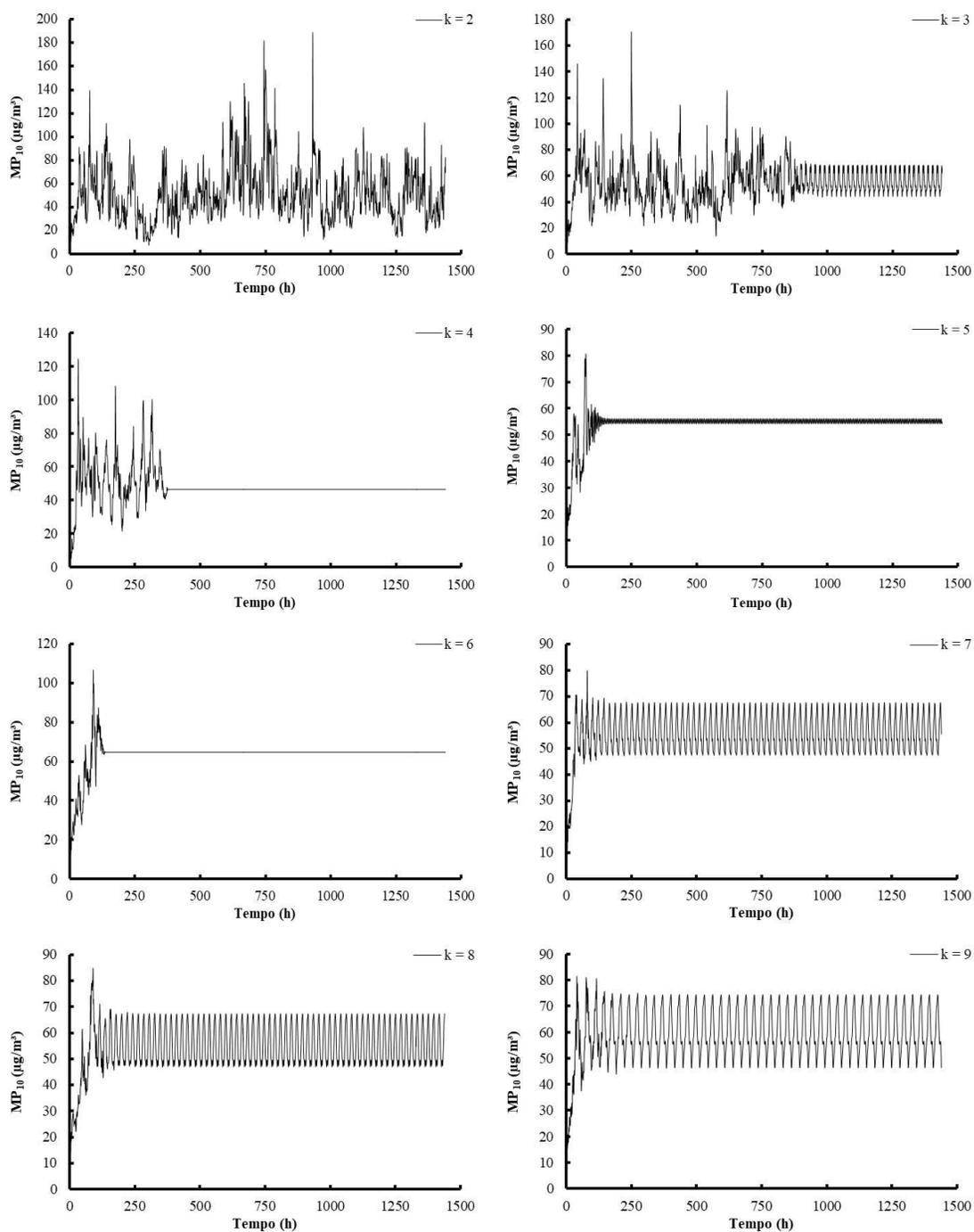
Partindo agora para a simulação, a Figura 4.18 apresenta a série original da variação de MP₁₀ e NOx na estação do Ibirapuera para os primeiros dois meses de 2018. Os dados oriundos dessas séries temporais foram utilizados para a comparação da eficiência da previsão e não foram inseridos na série de treino (*out of sample*).

Figura 4.18 – Séries temporais da concentração de MP_{10} e NO_x obtidas na estação de monitoramento do Ibirapuera para os meses de janeiro e fevereiro de 2018 (o autor, 2018).



Por fim, a Figura 4.19 e 4.20 são os resultados das simulações *out of sample* para o MP_{10} e NO_x utilizando o algoritmo *k-NN* e variando o número de vizinhos (*k*) de 2 a 10. Os gráficos comparativos das séries previstas com as originais podem ser encontrados no Apêndice D, Figuras D1 e D2.

Figura 4.19 – Previsão *out of sample* da concentração horária de MP_{10} nos primeiros dois meses de 2018 na estação do Ibirapuera utilizando o algoritmo k - NN (o autor, 2018).



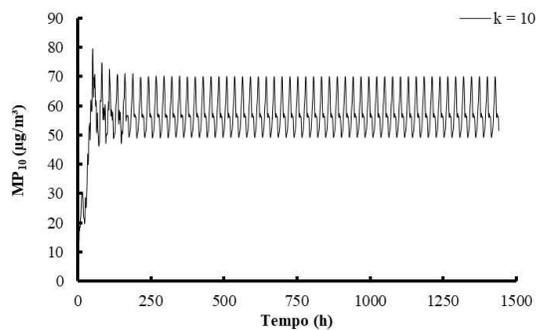
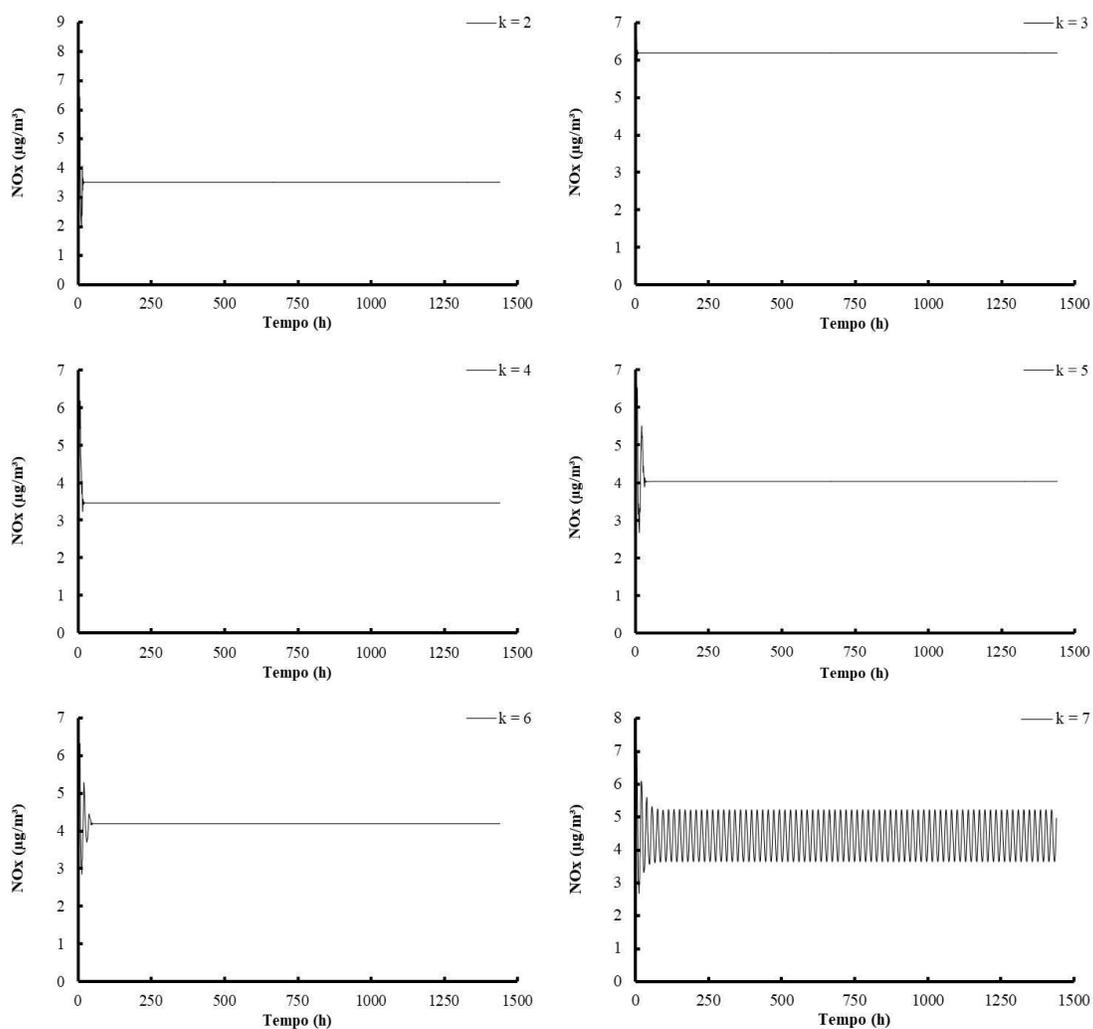
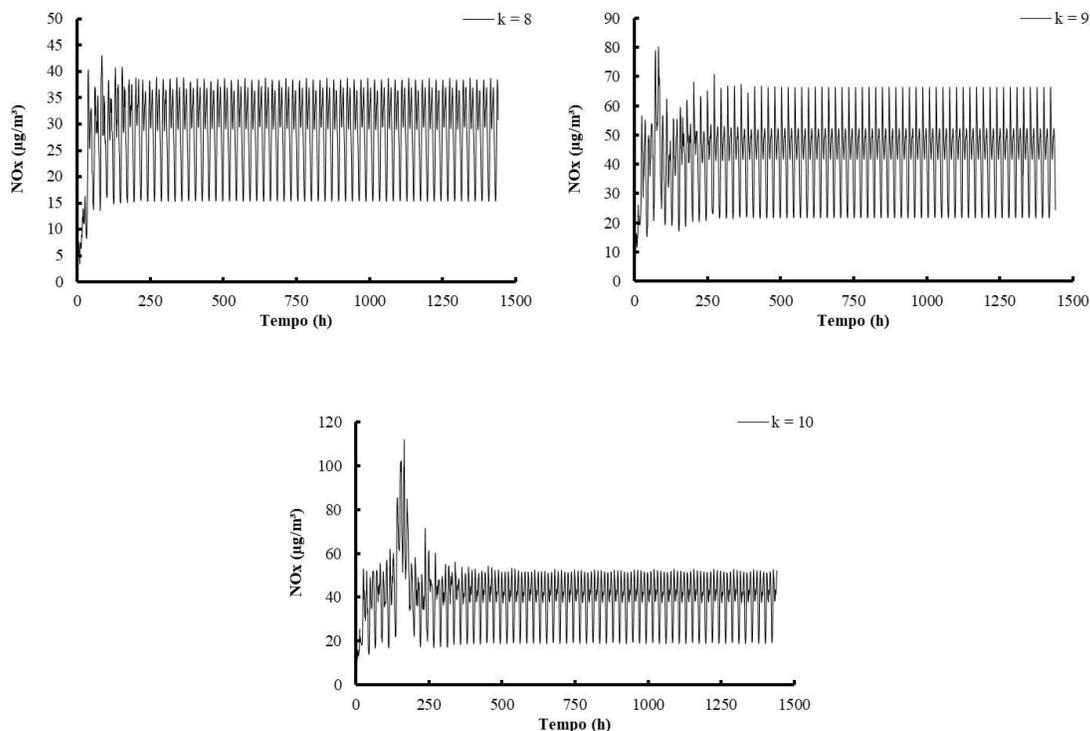


Figura 4.20 – Previsão *out of sample* da concentração horária de NO_x nos primeiros dois meses de 2018 na estação do Ibirapuera utilizando o algoritmo k -NN (o autor, 2018).





Por meio do exposto nas Figuras 4.19 e 4.20, verifica-se que para o MP₁₀ o valor de $k = 2$ apresentou uma série temporal completa até o último ponto gerado ao passo que para o NO_x nenhum dos k proporcionou uma previsão adequada para os 1440 pontos intentados, contudo, se observado atentamente, nota-se para os primeiros pontos obtidos existe uma variação similar a uma série temporal. Neste sentido, as análises sugerem que é possível a previsão, porém para um intervalo de tempo menor do que aquele testado. Em algumas situações é possível uma convergência para um valor médio ou ainda uma variação harmônica que não descreve a real dinâmica do fenômeno. Desta forma, as Tabela 4.11 (MP₁₀) e 4.12 (NO_x) trazem a comparação dos dados reais com as previsões realizadas para 60 dias (1440 pontos), 30 dias (720 pontos), 15 dias (360 pontos), 7 dias (168 pontos), 2 dias (48 pontos) e 1 dia (24 pontos). A qualidade das previsões foi avaliada por meio do desvio da média real em relação à média do conjunto simulado e por meio do método da Raiz do Erro Quadrático Médio (RMSE) e o Erro Absoluto Médio (MAE). O RMSE mede a magnitude média do erro por meio da raiz quadrada da média das diferenças entre a previsão e os dados reais, sendo usado como uma métrica estatística padrão para medir o desempenho de modelos nas áreas de meteorologia, qualidade do ar, climatologia e geociências (Chai e Draxler, 2014; Montgomery *et al.*, 2015). Já o erro absoluto médio

(MAE), também largamente utilizado para este fim, verifica a magnitude média dos erros em um conjunto de previsões, desconsiderando as direções, tratando-se de uma média das diferenças absolutas entre a informação prevista e a real, sendo que todas as diferenças individuais apresentam o mesmo peso (Chai e Draxler, 2014; Montgomery *et al.*, 2015). As formalizações matemáticas de ambos os métodos são apresentadas nas Equações 4.1 e 4.2.

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}| \quad (4.1)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (4.2)$$

Em que:

n é o número de amostras;

x_i é o valor fornecido pelo classificador para a i -ésima amostra;

\bar{x} é a média dos valores de todas as amostras.

Tabela 4.11 - Comparação entre as séries simuladas e reais para a concentração de MP_{10} – Ibirapuera (o autor, 2018).

	Média (60 dias) ($\mu\text{g}/\text{m}^3$)	Desvio	Média (30 dias) ($\mu\text{g}/\text{m}^3$)	Desvio	Média (15 dias) ($\mu\text{g}/\text{m}^3$)	Desvio	Média (1 semana) ($\mu\text{g}/\text{m}^3$)	Desvio	Média (2 dias) ($\mu\text{g}/\text{m}^3$)	Desvio	Média (1 dia) ($\mu\text{g}/\text{m}^3$)	Desvio
Real	29,00	-	28,30	-	26,50	-	24,50	-	19,40	-	18,00	-
<i>k</i>												
2	51,25	77%	50,01	77%	44,79	69%	53,34	118%	37,44	93%	22,43	24%
3	54,52*	88%	52,54	86%	54,16	104%	53,73	120%	43,05	122%	21,85	21%
4	47,76*	65%	49,06*	73%	51,78	95%	48,62	99%	38,92	100%	15,97	12%
5	54,28*	87%	53,41*	89%	51,64*	95%	47,61*	95%	34,66	79%	22,98	27%
6	63,48*	119%	62,37*	120%	60,14*	127%	55,06*	125%	31,05	60%	23,95	33%
7	55,41*	91%	54,73*	93%	53,57*	102%	50,89*	108%	37,78*	94%	21,64	20%
8	54,65*	88%	53,70*	90%	53,68*	102%	47,82*	95%	29,57	52%	22,92	27%
9	59,19*	104%	58,42*	106%	53,80*	103%	52,83*	116%	40,28*	107%	23,93	32%
10	56,83*	96%	56,04*	98%	53,92*	103%	51,36*	110%	32,85*	69%	21,99	21%
	60 dias		30 dias		15 dias		1 semana		2 dias		1 dia	
<i>k</i>	RMSE ($\mu\text{g}/\text{m}^3$)	MAE ($\mu\text{g}/\text{m}^3$)										
2	35,3	22,3	33,9	21,7	32,9	18,3	40,1	28,9	31,7	18,0	16,8	4,3
3	33,5	25,5	34,0	24,3	36,6	27,7	38,7	29,3	38,4	23,6	18,8	3,7
4	27,2	18,8	29,4	20,8	34,8	25,3	36,1	24,2	38,2	19,5	16,5	2,1
5	30,8	25,3	30,2	25,1	30,6	25,1	31,3	23,1	30,1	15,2	17,7	4,9
6	38,8	34,5	38,3	34,1	38,4	33,6	38,1	30,6	24,9	11,6	17,8	5,9
7	32,4	26,4	32,1	26,5	32,5	27,1	34,1	26,4	30,7	18,3	16,4	3,6
8	31,8	25,7	31,0	25,4	30,8	25,4	30,3	23,4	22,9	10,1	16,0	4,8
9	35,7	30,2	35,0	30,2	34,6	30,2	34,6	28,4	30,8	20,8	17,0	5,8
10	33,1	27,8	32,6	27,8	32,8	28,1	33,5	26,9	24,9	13,4	15,3	3,9

*valores desconsiderados na análise devido à presença de convergência harmônica ou para um valor médio.

Tabela 4.12 - Comparação entre as séries simuladas e reais para a concentração de NOx – Ibirapuera (o autor, 2018).

	Média (60 dias) ($\mu\text{g}/\text{m}^3$)		Média (30 dias) ($\mu\text{g}/\text{m}^3$)		Média (15 dias) ($\mu\text{g}/\text{m}^3$)		Média (1 semana) ($\mu\text{g}/\text{m}^3$)		Média (2 dias) ($\mu\text{g}/\text{m}^3$)		Média (1 dia) ($\mu\text{g}/\text{m}^3$)	
	Desvio		Desvio		Desvio		Desvio		Desvio		Desvio	
Real	24,22	-	22,27	-	21,71	-	16,51	-	10,40	-	9,30	-
k												
2	3,53*	85%	3,53*	84%	3,54*	84%	3,57*	78%	3,7*	64%	3,9	58%
3	6,19*	74%	6,19*	72%	6,18*	71%	6,18*	63%	6,1*	41%	6,1	34%
4	3,47*	86%	3,49*	84%	3,51*	84%	3,57*	78%	3,8*	63%	4,2	55%
5	4,04*	83%	4,05*	82%	4,08*	81%	4,13*	75%	4,4*	58%	4,7	50%
6	4,20*	83%	4,20*	81%	4,21*	81%	4,23*	74%	4,3*	58%	4,5	51%
7	4,41*	82%	4,42*	80%	4,43*	80%	4,45*	73%	4,6*	56%	4,9	47%
8	28,43*	17%	27,95*	26%	4,43*	80%	24,78*	50%	15,6*	51%	8,4	10%
9	43,00*	78%	42,50*	92%	4,42*	80%	39,26*	137%	28,5*	174%	18,9	103%
10	40,26*	66%	40,91*	84%	4,42*	80%	42,74*	159%	25,9*	149%	17,1	84%
	60 dias		30 dias		15 dias		1 semana		2 dias		1 dia	
k	RMSE ($\mu\text{g}/\text{m}^3$)	MAE ($\mu\text{g}/\text{m}^3$)	RMSE ($\mu\text{g}/\text{m}^3$)	MAE ($\mu\text{g}/\text{m}^3$)	RMSE ($\mu\text{g}/\text{m}^3$)	MAE ($\mu\text{g}/\text{m}^3$)	RMSE ($\mu\text{g}/\text{m}^3$)	MAE ($\mu\text{g}/\text{m}^3$)	RMSE ($\mu\text{g}/\text{m}^3$)	MAE ($\mu\text{g}/\text{m}^3$)	RMSE ($\mu\text{g}/\text{m}^3$)	MAE ($\mu\text{g}/\text{m}^3$)
2	27,8	20,7	25,3	18,7	23,7	18,1	17,5	13,0	10,3	6,7	8,8	5,4
3	25,9	18,0	23,3	16,0	21,7	15,5	15,5	10,4	8,5	4,2	6,8	3,2
4	27,9	20,7	25,3	18,7	253,7	18,2	17,5	13,0	10,1	6,6	8,4	5,1
5	27,4	20,2	24,9	18,1	23,3	17,6	17,1	12,4	9,9	6,0	8,4	4,6
6	27,3	20,0	24,8	18,0	23,1	17,5	17,0	12,3	9,8	6,1	8,3	4,8
7	27,2	19,8	24,6	17,8	23,1	17,2	17,0	12,1	9,7	5,8	8,3	4,4
8	20,4	4,2	18,6	5,8	17,1	5,2	15,8	8,2	13,2	5,2	6,5	0,9
9	27,3	18,8	27,5	20,4	26,4	19,7	28,9	22,7	24,4	18,1	14,2	9,6
10	27,8	16,1	29,1	18,8	29,3	20,5	34,6	26,2	21,2	15,5	10,7	7,8

*valores desconsiderados na análise devido à presença de convergência harmônica ou para um valor médio.

Para interpretar os resultados das Tabelas 4.11 e 4.12 é importante fazê-lo sempre observando as Figuras 4.19 e 4.20, vez que os valores de desvio, RMSE e MAE só tem sentido quando estão associados a uma previsão resultante de uma série temporal e, conforme já observado, em diversos casos, dependendo do número de pontos considerados, a média pode estar englobando resultados constantes ou com variação harmônica, o que, de forma alguma, representa a fenomenologia aqui estudada.

Observando as previsões para os dois poluentes na estação do Ibirapuera, verifica-se que a única situação em que foi possível prever uma série completa (1440 pontos) foi para o caso do MP_{10} com $k = 2$. Nesta situação a média prevista foi de $51,25 \mu\text{g}/\text{m}^3$ ao passo que a real foi de $29,00 \mu\text{g}/\text{m}^3$, um desvio de 77% com $RMSE = 35,3 \mu\text{g}/\text{m}^3$ e $MAE = 22,3 \mu\text{g}/\text{m}^3$. Por estes valores é bastante evidente que a previsão não foi satisfatória. Sobre o RMSE, este é análogo de várias maneiras ao desvio padrão (Longley *et al.*, 2013) de uma dada amostra. Muito embora no caso do RMSE o processo envolva a extração da raiz quadrada do erro quadrado médio, é interessante considerar seu valor como sendo um erro médio de cada observação, não importando se negativo ou positivo (Longley *et al.*, 2013). Neste caso, um RMSE maior que a própria média indica alta variabilidade dos dados e baixa confiança. De modo geral, se forem encontrados grandes valores de RMSE a conclusão é de que existem grandes erros na variável ajustada e se esses valores forem próximos de zero então há o indicativo de um ajuste quase perfeito (Camelo *et al.*, 2017). Observando o MAE, é possível verificar o valor médio do erro entre a série observada e ajustada (Camelo *et al.*, 2017). Tanto o MAE quanto RMSE apresentam as mesmas unidades de medida da variável sob estudo (Camelo *et al.*, 2017). Para o caso em discussão anteriormente, MP_{10} ($k = 2$), na estação do Ibirapuera o valor de MAE também indica uma previsão insatisfatória para todas as situações (60, 30, 15, 7, 2 e 1 dias). Para $k = 3$ (MP_{10}) ainda é possível analisar a situação de uma previsão de até 30 dias, pois acima disso a série fica constante, porém os valores tanto de desvio quanto RMSE e MAE se mantêm inadequados. Neste caso, verifica-se que as situações ótimas (independentemente do k) ocorrem para as previsões de 24 horas e a melhor de todas as previsões da concentração de MP_{10} na estação do Ibirapuera é verificada para $k = 4$ (média de 24 horas), com desvio de 12%, $RMSE = 16,5 \mu\text{g}/\text{m}^3$ e $MAE = 2,1 \mu\text{g}/\text{m}^3$.

Partindo para a análise da Tabela 4.12 (NO_x – Ibirapuera) é possível notar que a situação preditiva é bem pior que a apresentada pelo MP₁₀, pois nenhum dos casos permite a avaliação de previsões acima de 24 horas, uma vez que, pela Figura 4.20, para todos os valores de k ocorreram convergências e/ou variação harmônica para um dado médio. Sendo assim, a previsão de um dia é a mais indicada, sendo que a situação ótima ocorreu para $k = 8$ com desvio de 10%, $RMSE = 6,5 \mu\text{g}/\text{m}^3$ e $MAE = 0,9 \mu\text{g}/\text{m}^3$.

Contextualizando estes dados, as características do próprio fenômeno são importantes para o entendimento destes valores. Tratando-se de um evento de caráter altamente dispersivo, influenciado por, no mínimo 3 graus de liberdade, é esperado que a previsão não resultasse em dados horários tão fiéis e, por isso, os resultados para RMSE e MAE foram razoavelmente elevados. Ainda assim, tendo em vista a importância de previsões com esta, pode-se, sem dúvida, considerar que tais valores são muito bons, especialmente quando as variações do poluente são avaliadas em termos de dias ou semanas (quando a previsão assim permite), períodos em que os erros médios são amortecidos pela amplitude dos dados. Ao avaliar a raiz quadrada dos erros quadrados médios, verificam-se algumas informações interessantes. Como o método RMSE eleva os erros ao quadrado antes da média, isso acarreta em uma valorização considerável do peso dos grandes erros, ou seja, o valor de RMSE é útil quando erros grandes são especialmente indesejáveis. Via de regra, conforme observado, MAE tende a ser menor que RMSE, justamente pelos motivos salientados anteriormente, enquanto o primeiro tende a atribuir pesos iguais aos erros o segundo atribui um peso relativamente alto a erros maiores (Chai e Draxler, 2014).

4.6.2 Estação de Congonhas

Seguindo o procedimento realizado para a estação do Ibirapuera, a Figura 4.21 apresenta a série de treino utilizada na previsão da concentração horária de MP₁₀ e NO_x obtidos na estação de Congonhas dos anos de 1998 a 2017. Similar ao constatado para as séries anuais, os dados de treino apresentam um perfil bastante variável e dispersivo. O valor de concentração médio encontrado por meio da Equação 3.1 para o MP₁₀ foi de $45 \pm$

32 ($\mu\text{g}/\text{m}^3$) e para o NO_x de 254 ± 216 ($\mu\text{g}/\text{m}^3$). Novamente, para os dois poluentes, observa-se um desvio padrão alto da ordem da própria média, deixando claro o caráter variável e dispersivo dos dados. De modo a assegurar que as séries utilizadas como treino apresentam, assim com as anuais, características puramente caóticas, as Figuras 4.22 e 4.23 trazem o atratores reconstruídos para as séries de treino da concentração horária de MP_{10} e NO_x na estação de monitoramento de Congonhas dos anos de 1998 a 2017.

Figura 4.21 – Série de treino da concentração horária de MP_{10} e NO_x dos anos de 1998 a 2017 na estação de Congonhas (o autor, 2018).

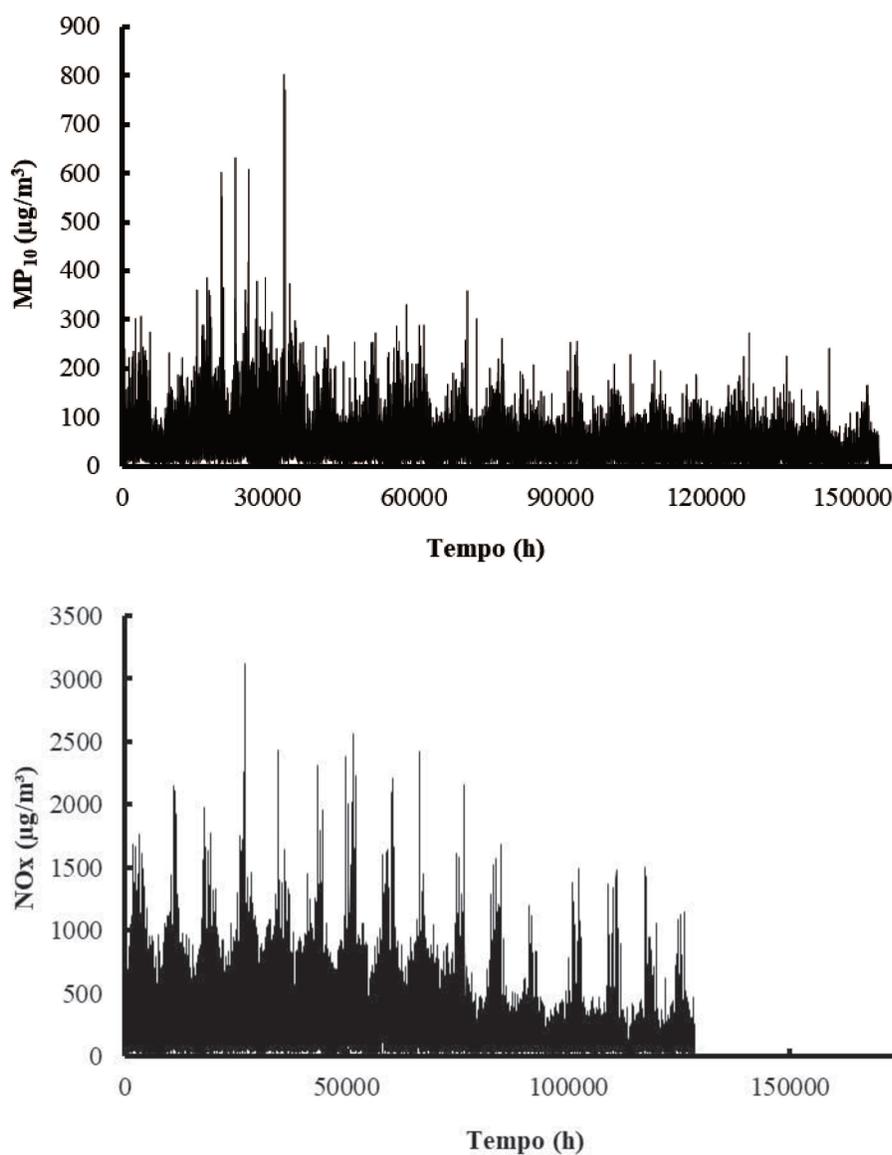


Figura 4.22 – Atrator reconstruído a partir da série de treino da concentração horária de MP_{10} para os anos de 1998 a 2017 (Congonhas) (o autor, 2018).

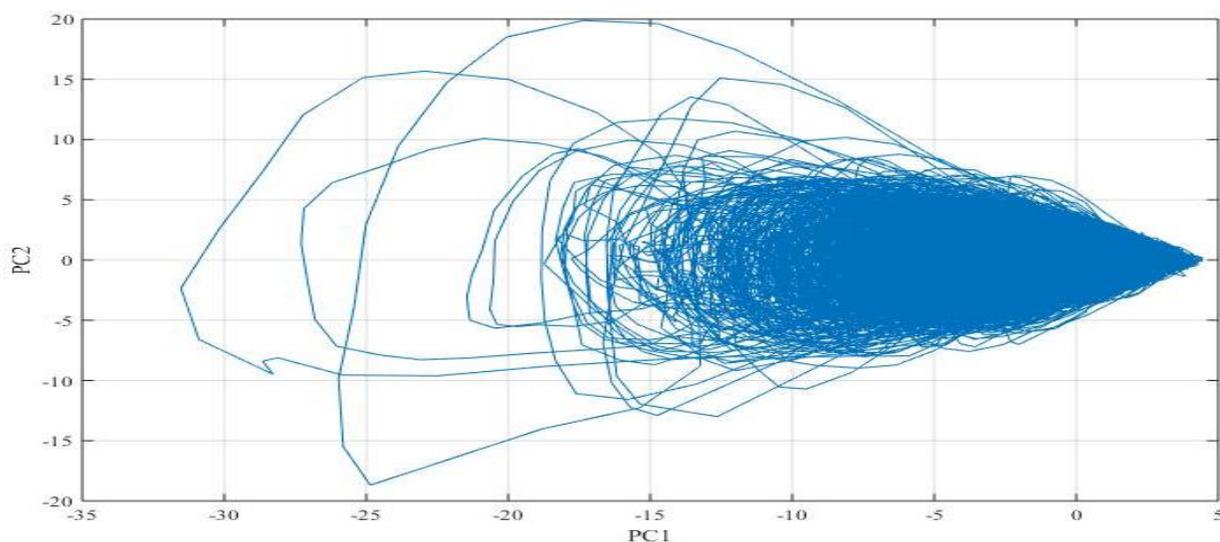
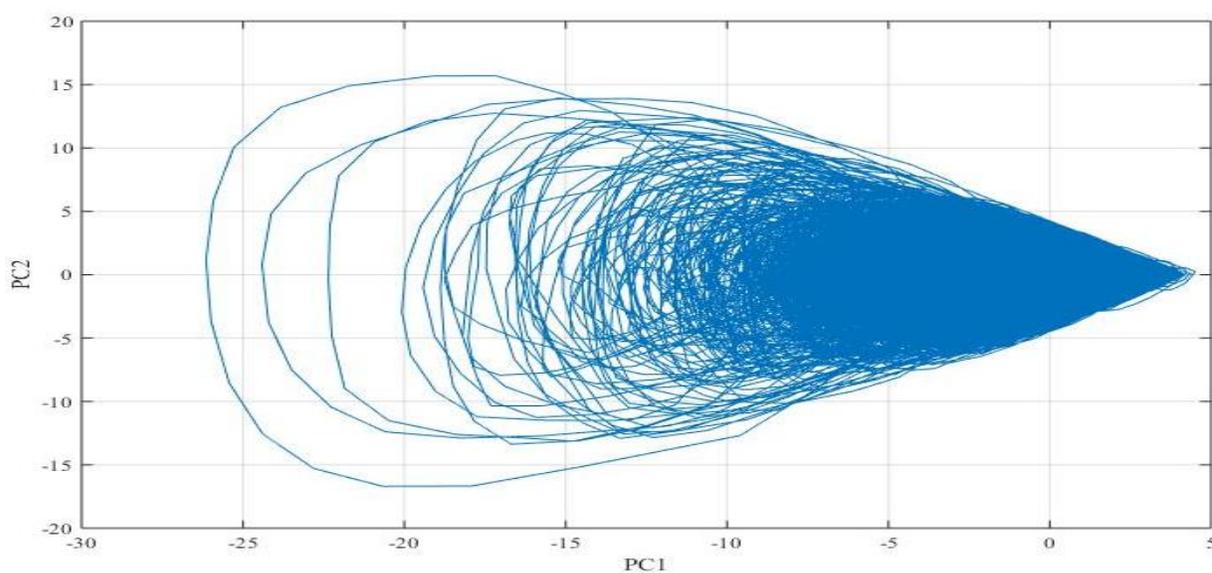


Figura 4.23 – Atrator reconstruído a partir da série de treino da concentração horária de NO_x para os anos de 1998 a 2017 (Congonhas) (o autor, 2018).



Os atratores foram reconstruídos no espaço de fases com dimensão de imersão, otimizada pelo algoritmo de Grassberger e Procaccia (1983) com tempo de atraso de 1 hora, igual a 13 e 16 para, respectivamente a concentração de MP_{10} e NO_x . A evolução dos

sistemas no espaço de fases descreve claramente um atrator estranho com soluções que convergem para uma região bem definida e estreita no espaço de fases (linha horizontal em PC2 (principal componente) = 0), sugerindo uma dinâmica caótica determinista, contudo, intentando asseverar essa conclusão, D_2 , K_{ML} foram calculados, resultando em $D_2 = 4,9692$, $K_{ML} = 0,1083$ bits/s e $\lambda = 0,15$ para o MP₁₀ e $D_2 = 4,7999$, $K_{ML} = 0,0480$ bits/s e $\lambda = 0,13$ para o NOx. Em mão dos invariantes caóticos, é possível assegurar que a evolução dos dados é caótica e que apresentam 5 graus de liberdade que regem a dinâmica dos conjuntos.

Uma vez confirmada a subjacente dinâmica do caos presente nos dados, passa-se a realização das previsões de fato. Neste ponto é importante destacar que o algoritmo k -NN necessita de duas informações básicas, o número de vizinhos próximos (k) e a dimensão de imersão (I).

Nesta Tese a proposta é justamente utilizar nas previsões o valor da dimensão de imersão gerado pelo algoritmo de Grassberger e Procaccia (1983), pois conforme demonstrado diversas vezes ao longo deste capítulo, os dados apresentam características caóticas sendo, portanto, intuitiva a aplicação de uma metodologia que respeite essas características ao calcular o valor de I imputado ao k -NN.

Na Figura 4.24 é possível observar a série original da variação de MP₁₀ e NOx na estação do Congonhas para os primeiros dois meses de 2018. Os dados advindos dessas séries temporais foram utilizados para a comparação da eficiência da previsão e não foram inseridos na série de treino (*out of sample*).

Intentando reproduzir o mesmo comportamento do conjunto de dados presentes na Figura 4.24, as Figuras 4.25 e 4.26 trazem os resultados das simulações *out of sample* para o MP₁₀ e NOx utilizando o algoritmo k -NN, variando o número de vizinhos próximos (k) de 2 a 10 e com dimensão de imersão igual ao valor gerado pelo algoritmo de Grassberger e Procaccia (1983). Os gráficos comparativos das séries previstas com as originais podem ser encontrados no Apêndice D, Figuras D3 e D4.

Figura 4.24 – Séries temporais da concentração de MP_{10} e NOx obtidas na estação de monitoramento de Congonhas para os meses de janeiro e fevereiro de 2018 (o autor, 2018).

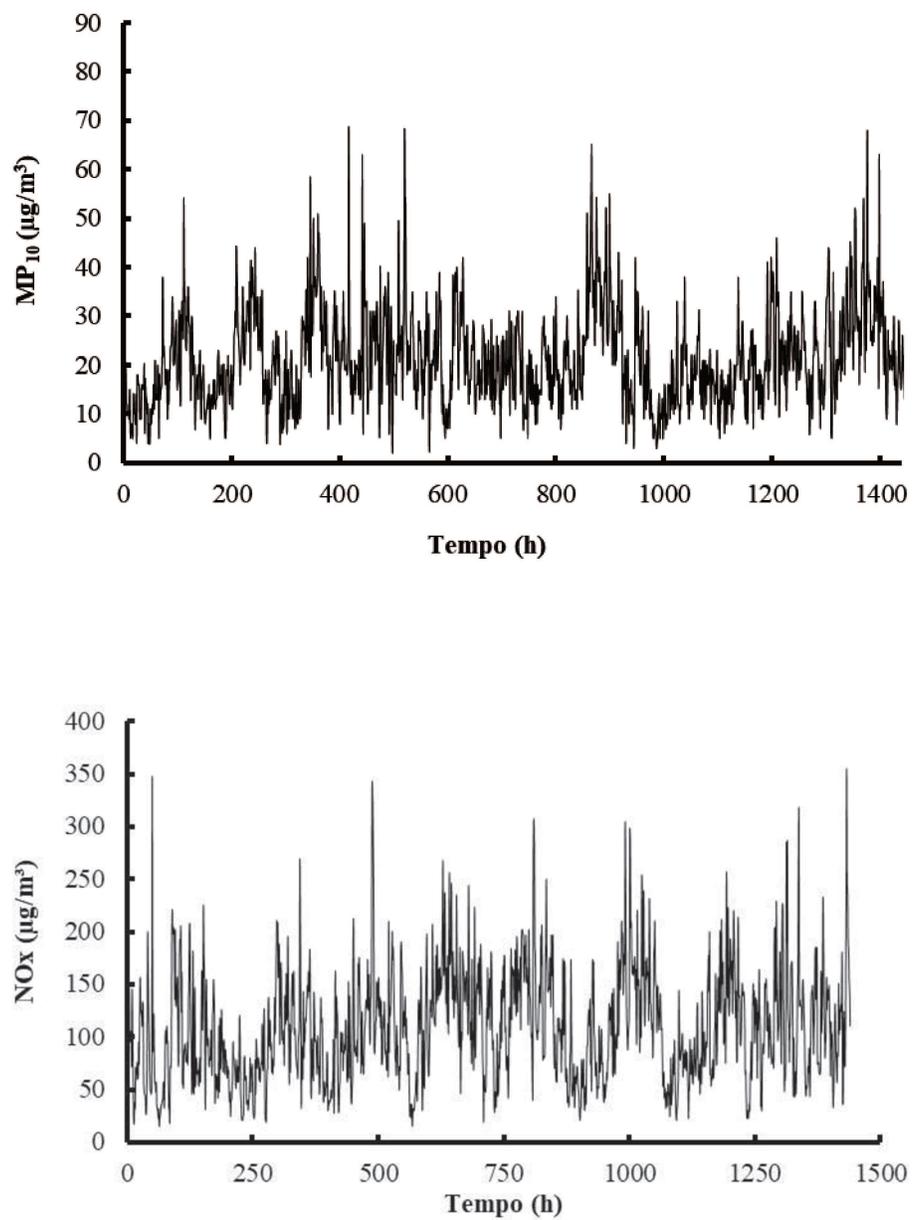
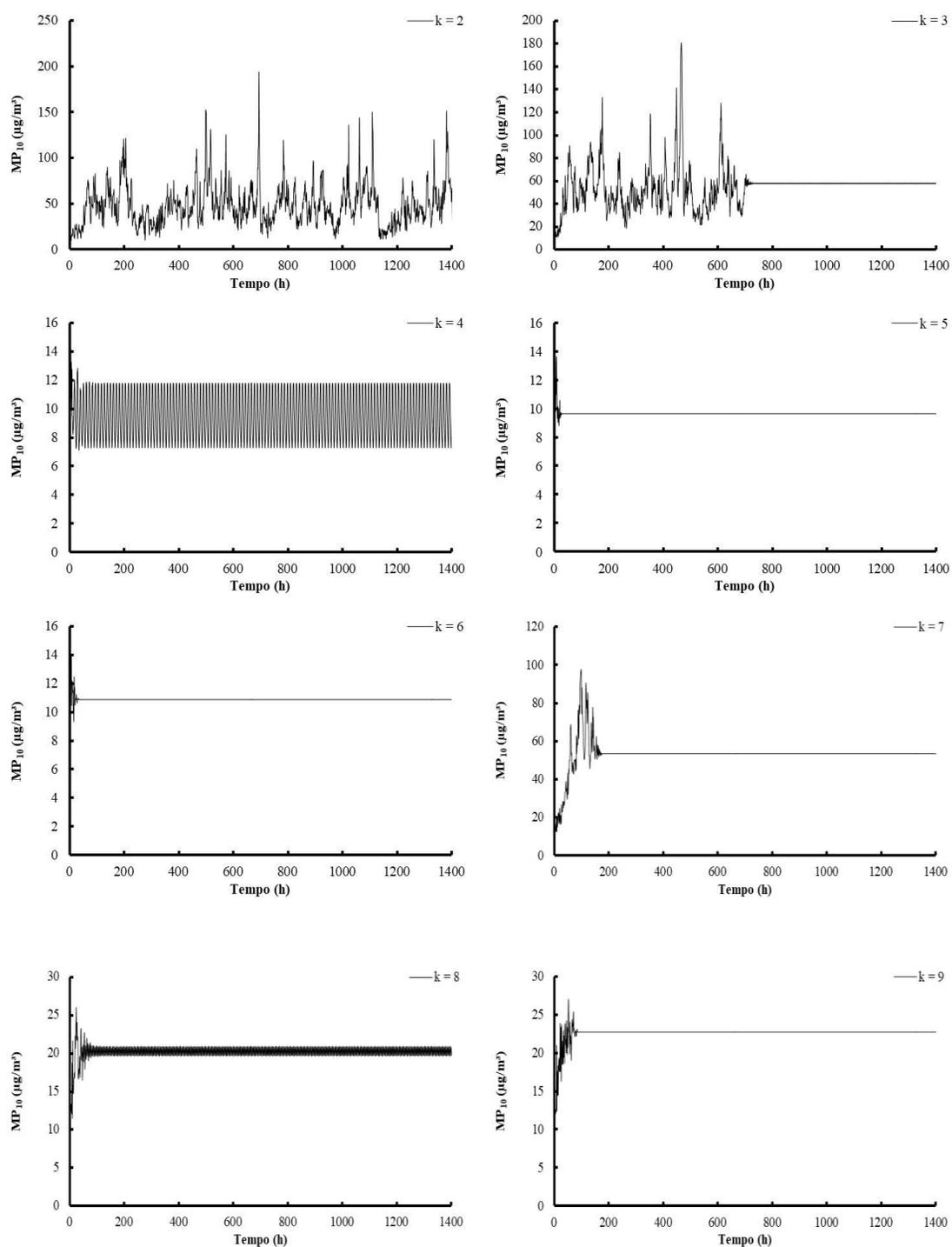


Figura 4.25 – Previsão *out of sample* da concentração horária de MP_{10} nos primeiros dois meses de 2018 na estação de Congonhas utilizando o algoritmo k -NN (o autor, 2018).



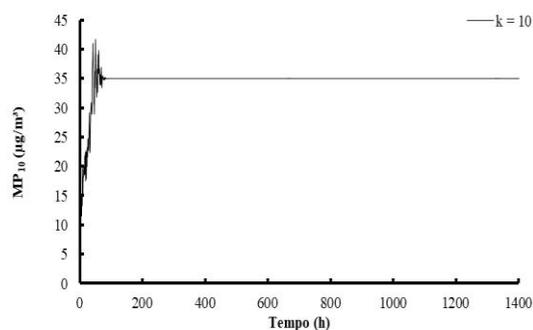
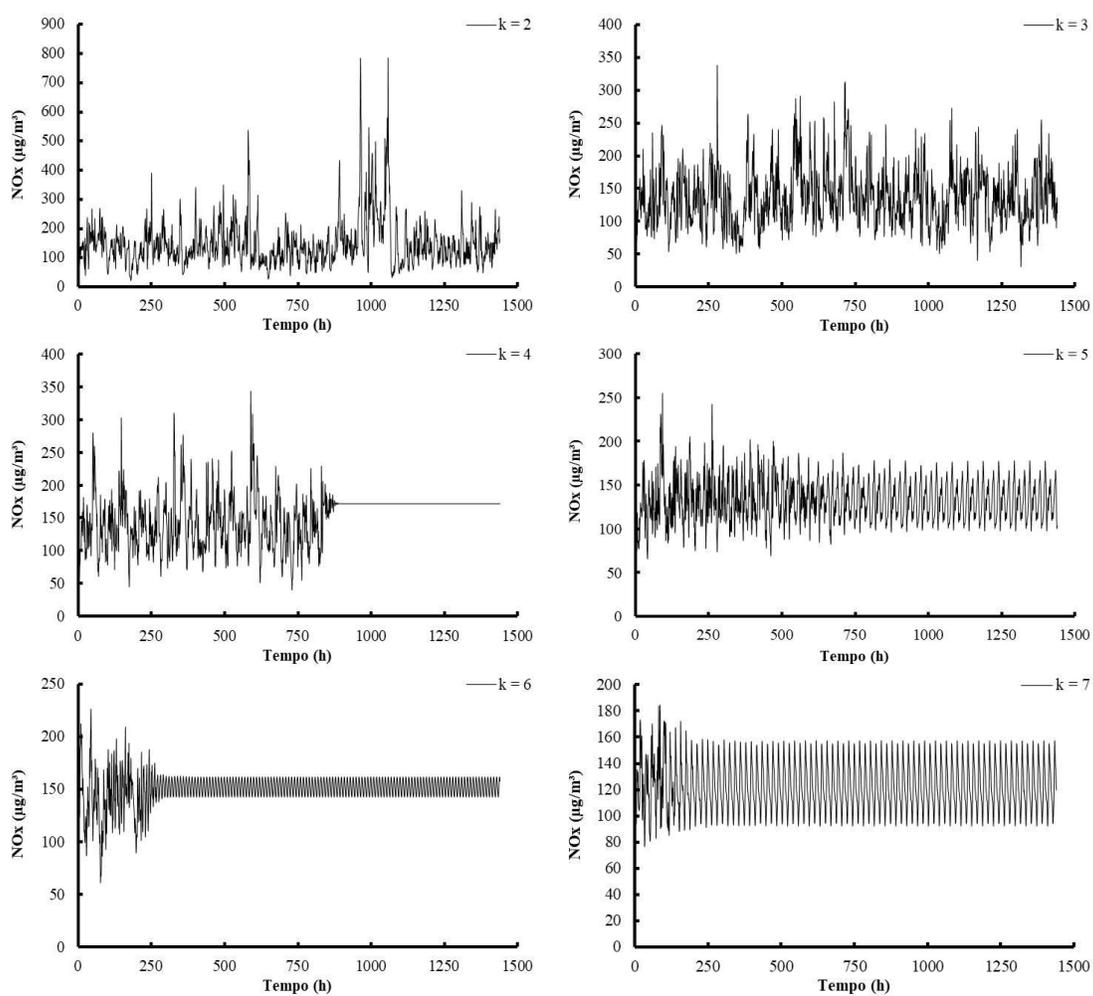
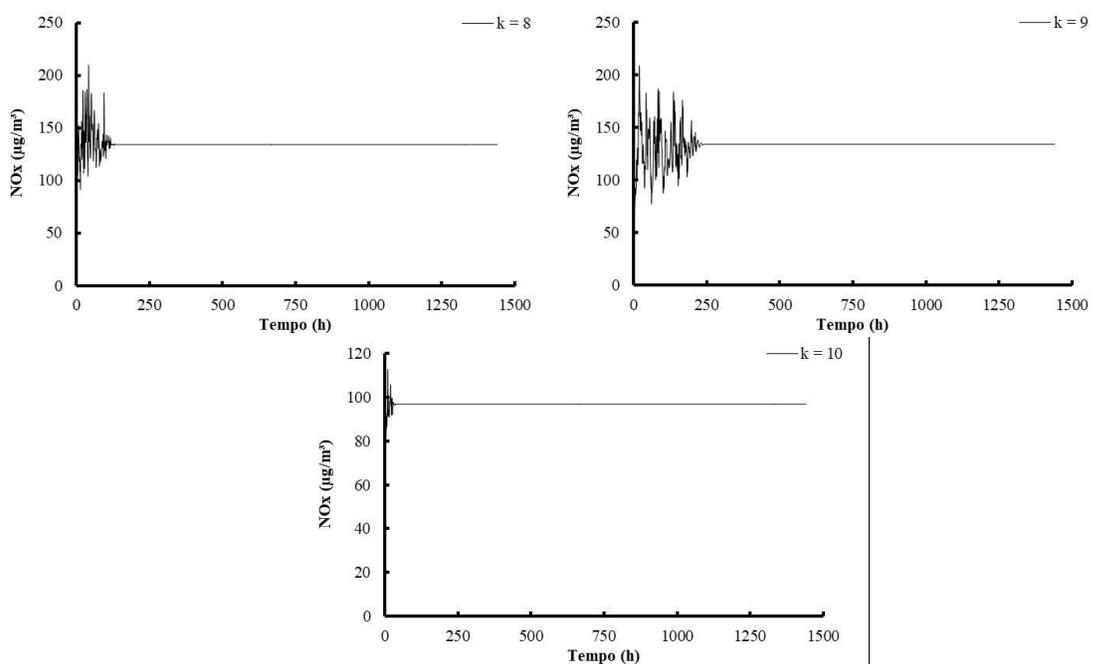


Figura 4.26 – Previsão *out of sample* da concentração horária de NO_x nos primeiros dois meses de 2018 na estação de Congonhas utilizando o algoritmo k -NN (o autor, 2018).





Nas Figuras 4.25 e 4.26 é possível notar que para diversos valores de k ocorrem, tanto para o MP_{10} quanto NO_x , convergências para um valor médio ou ainda variação ao redor de um único valor. Isso faz com que a série, a partir desses pontos, não demonstre mais representatividade em relação aos dados originais. Em muitos casos fica evidente a necessidade de se considerar apenas os primeiros previstos. Assim, no caso do MP_{10} , observa-se que a única situação para a qual o algoritmo foi capaz de gerar uma série de 1440 pontos foi para $k = 2$ e para $k = 3$ se verifica uma série de aproximadamente 720 pontos. No caso do NO_x , duas séries completas foram geradas ($k = 2$ e $k = 3$) sendo que as demais apresentaram problemas de convergência. Para ambos os poluentes, mesmo para as séries com problemas de convergência, nota-se que para os primeiros pontos obtidos existe uma variação similar a uma série temporal. Novamente, similar ao constatado na estação anterior, as análises sugerem que é possível a previsão, porém para um intervalo de tempo menor do que aquele testado.

Buscando avaliar a qualidade das previsões, as Tabelas 4.13 (MP_{10}) e 4.14 (NO_x) listam a comparação dos dados reais com as previsões realizadas para 60 dias (1440 pontos), 30 dias (720 pontos), 15 dias (360 pontos), 7 dias (168 pontos), 2 dias (48 pontos) e 1 dia (24 pontos). As séries previstas foram submetidas a verificação por meio do desvio

da média real em relação à média do conjunto simulado e por meio o método da Raiz do Erro Quadrático Médio (RMSE) e o Erro Absoluto Médio (MAE).

Da mesma forma que o comentando anteriormente, a interpretação dos dados das Tabelas 4.13 e 4.14 precisa ser realizada à luz das Figuras 4.25 e 4.26. No caso do MP₁₀, só tem sentido a avaliação das previsões de 1440 pontos para $k = 2$, pois as demais apresentam erros de convergência. Sobre esta previsão, verifica-se um desvio de 124% com $RMSE = 35,8$ e $MAE = 26,4 \mu\text{g}/\text{m}^3$. Considerando que a média real foi $21,27 \mu\text{g}/\text{m}^3$ versus $47,71 \mu\text{g}/\text{m}^3$ da previsão, torna-se claro que os dados gerados não representam adequadamente a concentração de MP₁₀ no período avaliado. Os melhores casos para $k = 2$ ocorreram para a previsão de 1 dia (desvio de 39%, $RMSE = 11,2$ e $MAE = 4,2 \mu\text{g}/\text{m}^3$) e 2 dias (desvio de 48%, $RMSE = 10,8$ e $MAE = 5,8 \mu\text{g}/\text{m}^3$). Observando ainda $k = 3$ (MP₁₀), o erro de convergência ocorre a partir do ponto 720, logo é possível analisar a qualidade da previsão anterior que, para 30 dias apresentou desvio de 150% ($RMSE = 39,3$ e $MAE = 31,3 \mu\text{g}/\text{m}^3$), 15 dias com desvio de 156% ($RMSE = 37,4$ e $MAE = 30,9 \mu\text{g}/\text{m}^3$), 7 dias desvio de 196% ($RMSE = 40,1$ e $MAE = 33,2 \mu\text{g}/\text{m}^3$), 2 dias desvio de 118% ($RMSE = 22,0$ e $MAE = 14,1 \mu\text{g}/\text{m}^3$) e a melhor previsão foi para 1 dia com desvio de 23% ($RMSE = 10,7$ e $MAE = 2,4 \mu\text{g}/\text{m}^3$). A melhor previsão para o MP₁₀ em Congonhas se deu para $k = 6$ (1 dia), com desvio de 11%, $RMSE = 7,8$ e $MAE = 1,8 \mu\text{g}/\text{m}^3$.

Observando as previsões para o NO_x na estação de Congonhas, verifica-se que para $k = 2$ e $k = 3$ foi possível obter séries temporais completas (1440 pontos). Para $k = 2$ (1440 pontos) foi possível obter um desvio de 36% $RMSE = 102,1$ e $MAE = 39,3 \mu\text{g}/\text{m}^3$, já para $k = 3$ o desvio foi de 25% com $RMSE = 74,5$ e $MAE = 27,6 \mu\text{g}/\text{m}^3$. Considerando as características do fenômeno, para uma previsão longa como esta, a previsão para $k = 3$ pode ser considerada um bom resultado. Ainda assim, nota-se que a melhor de todas as previsões é notada para $k = 10$ (1 dia).

De modo geral, salvo algumas situações, verifica-se que os melhores valores de desvio, RMSE e MAE são encontrados para as previsões de 24 horas, sugerindo, novamente que esta seja a situação mais confiável a ser explorada.

Tabela 4.13 - Comparação entre as séries simuladas e reais para a concentração de MP_{10} – Congonhas (o autor, 2018).

	Média (60 dias) ($\mu\text{g}/\text{m}^3$)		Desvio		Média (30 dias) ($\mu\text{g}/\text{m}^3$)		Desvio		Média (15 dias) ($\mu\text{g}/\text{m}^3$)		Desvio		Média (1 semana) ($\mu\text{g}/\text{m}^3$)		Desvio		Média (2 dias) ($\mu\text{g}/\text{m}^3$)		Desvio		Média (1 dia) ($\mu\text{g}/\text{m}^3$)		Desvio		
Real	21,27	-	21,06	-	19,71	-	17,15	-	12,7	-	12,8	-													
k																									
2	47,71	124%	47,87	127%	42,93	118%	42,89	150%	18,84	48%	17,73	39%													
3	55,27*	160%	52,59	150%	50,53	156%	50,70	196%	27,82	118%	15,76	23%													
4	9,55*	55%	9,57*	55%	9,59*	51%	9,60*	44%	9,83*	23%	10,05	22%													
5	9,65*	55%	9,66*	54%	9,69*	51%	9,75*	43%	10,04*	21%	10,47	19%													
6	10,89*	49%	10,90*	48%	10,91*	45%	10,95*	36%	11,18*	13%	11,44	11%													
7	52,91*	149%	52,45*	149%	51,52*	161%	49,42	188%	23,19	82%	17,35	36%													
8	20,21*	5%	20,15*	4%	51,63*	162%	19,71*	15%	18,37	43%	16,86	31%													
9	22,58*	6%	22,44*	7%	51,75*	163%	21,56*	26%	18,65	46%	16,39	28%													
10	34,64*	63%	34,27*	63%	51,87*	163%	31,83*	86%	23,76	86%	17,68	38%													
	60 dias				30 dias				15 dias				1 semana				2 dias				1 dia				
k	RMSE ($\mu\text{g}/\text{m}^3$)	MAE ($\mu\text{g}/\text{m}^3$)	RMSE ($\mu\text{g}/\text{m}^3$)	MAE ($\mu\text{g}/\text{m}^3$)	RMSE ($\mu\text{g}/\text{m}^3$)	MAE ($\mu\text{g}/\text{m}^3$)	RMSE ($\mu\text{g}/\text{m}^3$)	MAE ($\mu\text{g}/\text{m}^3$)	RMSE ($\mu\text{g}/\text{m}^3$)	MAE ($\mu\text{g}/\text{m}^3$)	RMSE ($\mu\text{g}/\text{m}^3$)	MAE ($\mu\text{g}/\text{m}^3$)	RMSE ($\mu\text{g}/\text{m}^3$)	MAE ($\mu\text{g}/\text{m}^3$)	RMSE ($\mu\text{g}/\text{m}^3$)	MAE ($\mu\text{g}/\text{m}^3$)	RMSE ($\mu\text{g}/\text{m}^3$)	MAE ($\mu\text{g}/\text{m}^3$)	RMSE ($\mu\text{g}/\text{m}^3$)	MAE ($\mu\text{g}/\text{m}^3$)	RMSE ($\mu\text{g}/\text{m}^3$)	MAE ($\mu\text{g}/\text{m}^3$)	RMSE ($\mu\text{g}/\text{m}^3$)	MAE ($\mu\text{g}/\text{m}^3$)	
2	35,8	26,4	36,8	26,8	33,2	23,3	32,5	25,8	10,8	5,8	11,2	4,2													
3	38,6	34,0	39,3	31,6	37,4	30,9	40,1	33,2	22,0	14,1	10,7	2,4													
4	15,4	11,7	15,1	11,5	14,0	10,1	11,3	7,6	6,9	3,2	8,2	3,1													
5	15,2	11,6	14,9	11,4	13,8	10,0	11,1	7,4	6,9	2,9	8,0	2,7													
6	14,2	10,4	14,0	10,2	12,9	8,7	10,3	6,2	6,6	1,8	7,8	1,8													
7	33,6	31,6	33,8	31,4	35,2	31,9	37,5	32,2	14,8	10,0	11,3	4,1													
8	9,8	1,1	9,6	0,9	9,6	0,4	8,9	2,6	9,8	5,3	10,9	3,4													
9	9,8	1,3	9,6	1,4	9,8	2,5	9,5	4,4	10,2	5,6	10,4	2,9													
10	16,6	13,4	16,4	13,2	17,0	13,9	17,5	14,7	15,2	10,5	11,4	4,3													

*valores desconsiderados na análise devido à presença de convergência harmônica ou para um valor médio.

Tabela 4.14 - Comparação entre as séries simuladas e reais para a concentração de NO_x – Congonhas (o autor, 2018).

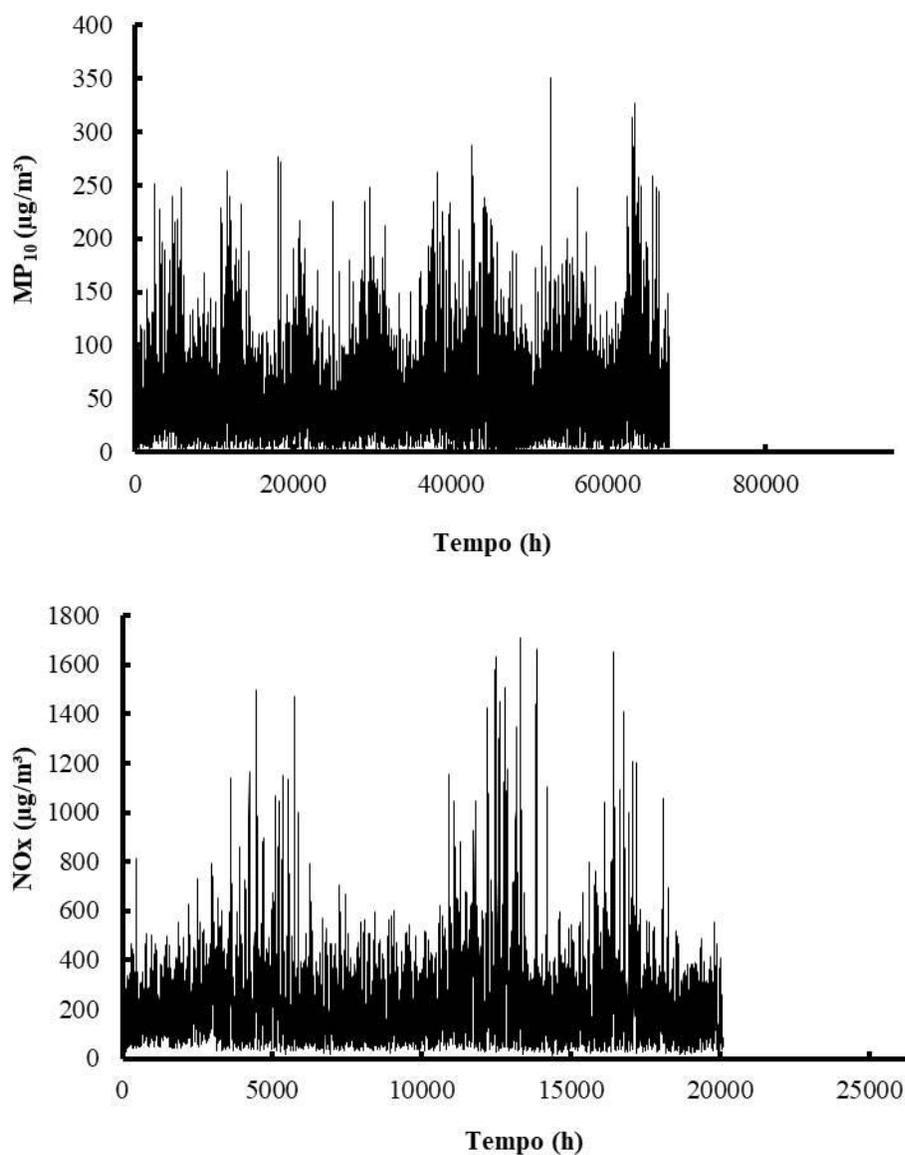
	Média (60 dias) (µg/m³)	Desvio	Média (30 dias) (µg/m³)	Desvio	Média (15 dias) (µg/m³)	Desvio	Média (1 semana) (µg/m³)	Desvio	Média (2 dias) (µg/m³)	Desvio	Média (1 dia) (µg/m³)	Desvio
Real	108,70	-	102,80	-	91,80	-	95,40	-	86,20	-	80,50	-
k												
2	147,95	36%	139,17	35%	135,91	48%	141,24	48%	134,72	56%	109,25	36%
3	136,31	25%	136,92	33%	127,53	39%	130,66	37%	117,31	36%	117,08	45%
4	152,27*	40%	140,00	36%	138,37	51%	140,83	48%	118,83	38%	111,42	38%
5	132,65*	22%	133,25*	30%	131,84	44%	128,32	34%	112,88	31%	100,54	25%
6	150,15*	38%	148,30*	44%	144,60*	58%	140,91	48%	145,97	69%	145,07	80%
7	122,26*	13%	122,42*	19%	122,53*	33%	123,44*	29%	120,39	40%	123,21	53%
8	134,55*	24%	134,71*	31%	122,68*	34%	135,76*	42%	134,55	56%	124,56	55%
9	133,27*	23%	132,50*	29%	122,79*	34%	127,95	34%	126,34	47%	123,29	53%
10	96,67*	11%	96,60*	6%	122,91*	34%	96,12*	1%	94,66*	10%	92,43	15%
	60 dias		30 dias		15 dias		1 semana		2 dias		1 dia	
k	RMSE (µg/m³)	MAE (µg/m³)	RMSE (µg/m³)	MAE (µg/m³)	RMSE (µg/m³)	MAE (µg/m³)	RMSE (µg/m³)	MAE (µg/m³)	RMSE (µg/m³)	MAE (µg/m³)	RMSE (µg/m³)	MAE (µg/m³)
2	102,1	39,3	91,0	36,4	85,8	44,1	86,7	45,8	83,5	48,5	64,1	28,8
3	74,5	27,6	75,8	34,2	72,7	35,7	67,8	35,2	61,1	31,1	70,0	36,6
4	76,6	43,6	75,2	37,2	77,5	46,6	82,2	45,4	67,0	32,6	70,5	30,9
5	62,8	24,0	65,7	30,5	68,7	40,0	70,4	32,9	59,4	26,7	51,8	20,1
6	68,3	41,5	70,2	45,6	75,1	52,8	76,9	45,5	80,5	59,7	82,4	64,5
7	58,0	13,6	59,0	19,6	62,7	30,7	67,1	28,0	60,2	34,2	62,6	42,8
8	59,4	25,9	61,1	32,0	65,6	43,2	68,5	40,3	68,8	48,3	63,3	44,1
9	59,2	24,6	60,9	29,7	65,3	39,2	68,2	32,5	66,7	40,1	74,7	42,7
10	54,5	12,0	52,0	6,2	48,7	4,7	53,7	0,7	42,7	8,4	37,9	11,9

*valores desconsiderados na análise devido à presença de convergência harmônica ou para um valor médio.

4.6.3 Estação do Centro de São Paulo

Seguindo com a estação do Centro de São Paulo, a Figura 4.27 apresenta a série de treino utilizada na previsão da concentração horária de MP_{10} e NO_x .

Figura 4.27 – Série de treino da concentração horária de MP_{10} (de 1998 a 2009) e NO_x (de 1998 a 2000) na estação do Centro de São Paulo (o autor, 2018).



Para o MP_{10} a estação realizou a aquisição até o ano de 2010, sendo assim, a série de treino foi montada de 1998 a 2009 e a série comparativa abrangeu os meses de janeiro e fevereiro de 2010. Já para o NOx a aquisição de iniciou em 1998 até 2001, reiniciou em 2005 finalizando em 2006. Neste caso, a série de treino foi elaborada de 1998 a 2000 e a comparativa englobou os meses de janeiro de fevereiro de 2001. Novamente, os dados de treino apresentam um perfil bastante variável e dispersivo. O valor de concentração médio encontrado por meio da Equação 3.1 para o MP_{10} foi de 46 ± 30 ($\mu g/m^3$) e para o NOx de 231 ± 147 ($\mu g/m^3$). Para os dois poluentes, observa-se um desvio padrão alto da ordem da própria média, deixando claro o caráter variável e dispersivo dos dados. Outra constatação importante, também notada em todas as estações, é a ausência de dados ao final da evolução das séries de treino. Essa ausência se dá pelo fato de que os dados inválidos presentes nas tabelas exportados diretamente da plataforma CETESB foram lançados ao final de modo a tornar as séries contínuas, permitindo sua aplicação no algoritmo k -NN e avaliação do caos.

Similar ao realizado para as estações anteriores, garantindo que as séries de treino utilizadas, similar às anuais, guardam características puramente caóticas, as Figuras 4.28 e 4.28 trazem o atratores reconstruídos para as séries de treino da concentração horária de MP_{10} e NOx na estação de monitoramento do Centro de São Paulo.

Ambos os atratores, assim como todos os apresentados nesta Tese, foram obtidos por meio do método da decomposição em valores singulares (SVD) (Broomhead e King, 1986) e reconstruídos no espaço de fases com dimensão de imersão (Grassberger e Procaccia, 1983) igual a 13 e tempo de atraso de 1 hora tanto para o MP_{10} quanto NOx . A solução dos dois sistemas no espaço de fases descreve um atrator estranho, característico de uma dinâmica caótica determinista, pois as soluções convergem para uma região bem definida e estreita no espaço de fases (linha horizontal em PC2 (principal componente) = 0). Ainda assim, realizou-se o cálculo dos invariantes entropia de Kolmogorov e dimensão de correlação. Para o MP_{10} , $D_2 = 4,4478$, $K_{ML} = 0,1504$ bits/s e $\lambda = 0,14$ e para o NOx , $D_2 = 4,5691$, $K_{ML} = 0,1254$ bits/s e $\lambda = 0,13$. A partir deste invariantes, é possível

assegurar, de forma quantitativa, que a evolução dos dados é caótica (pois $0 < K_{ML} < \infty$) apresentando 5 graus de liberdade regentes da dinâmica de ambos os sistemas.

Figura 4.28 – Atrator reconstruído a partir da série de treino da concentração horária de MP_{10} para os anos de 1998 a 2009 (Centro de São Paulo) (o autor, 2018).

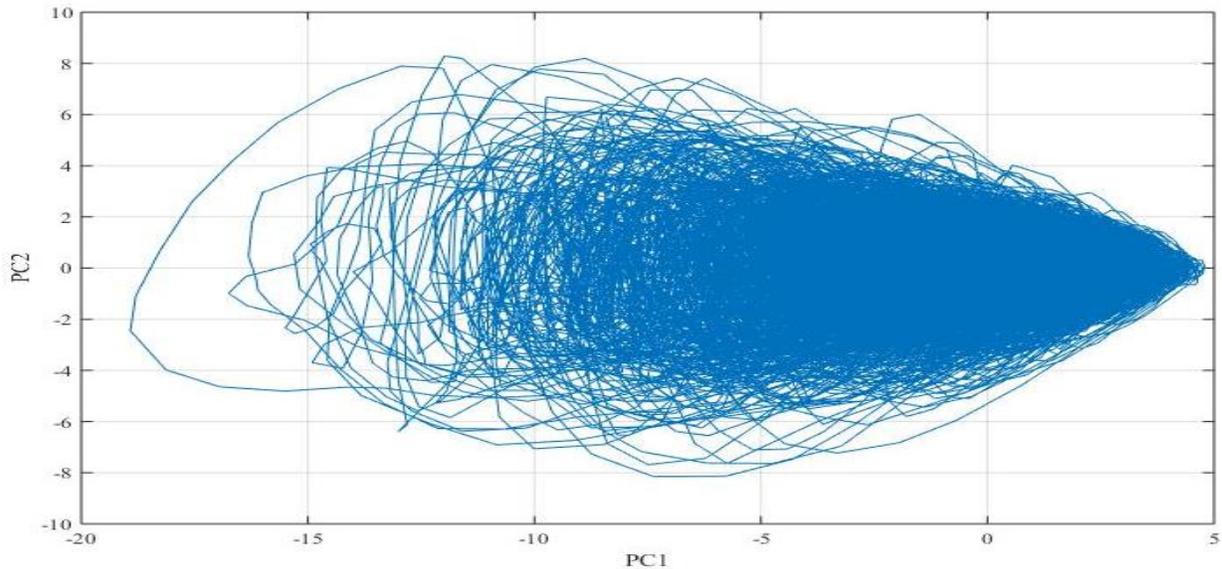
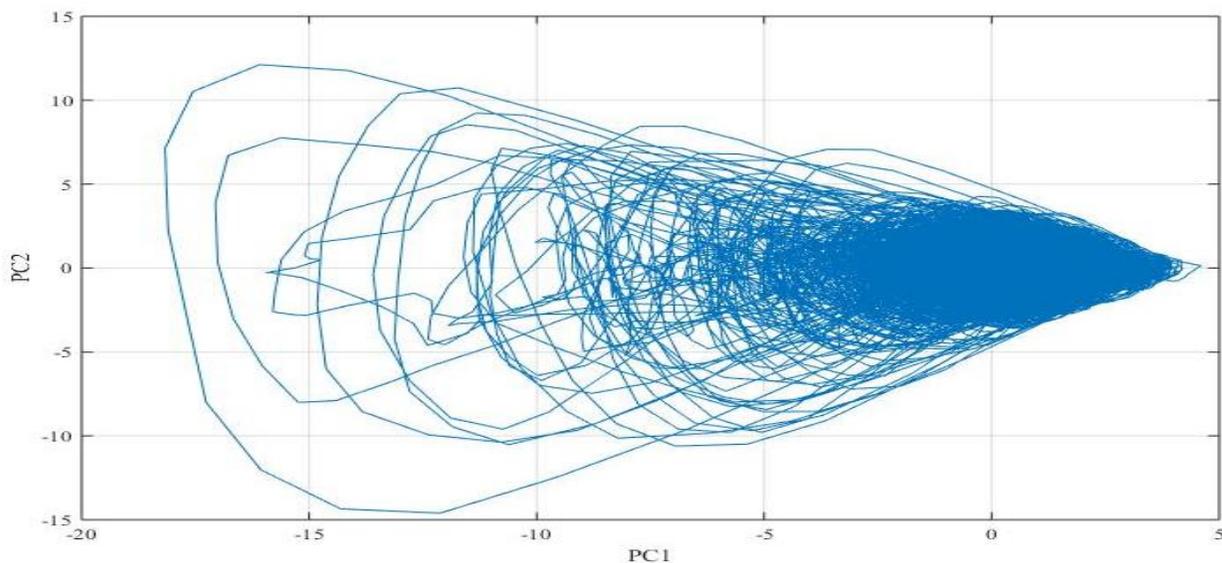


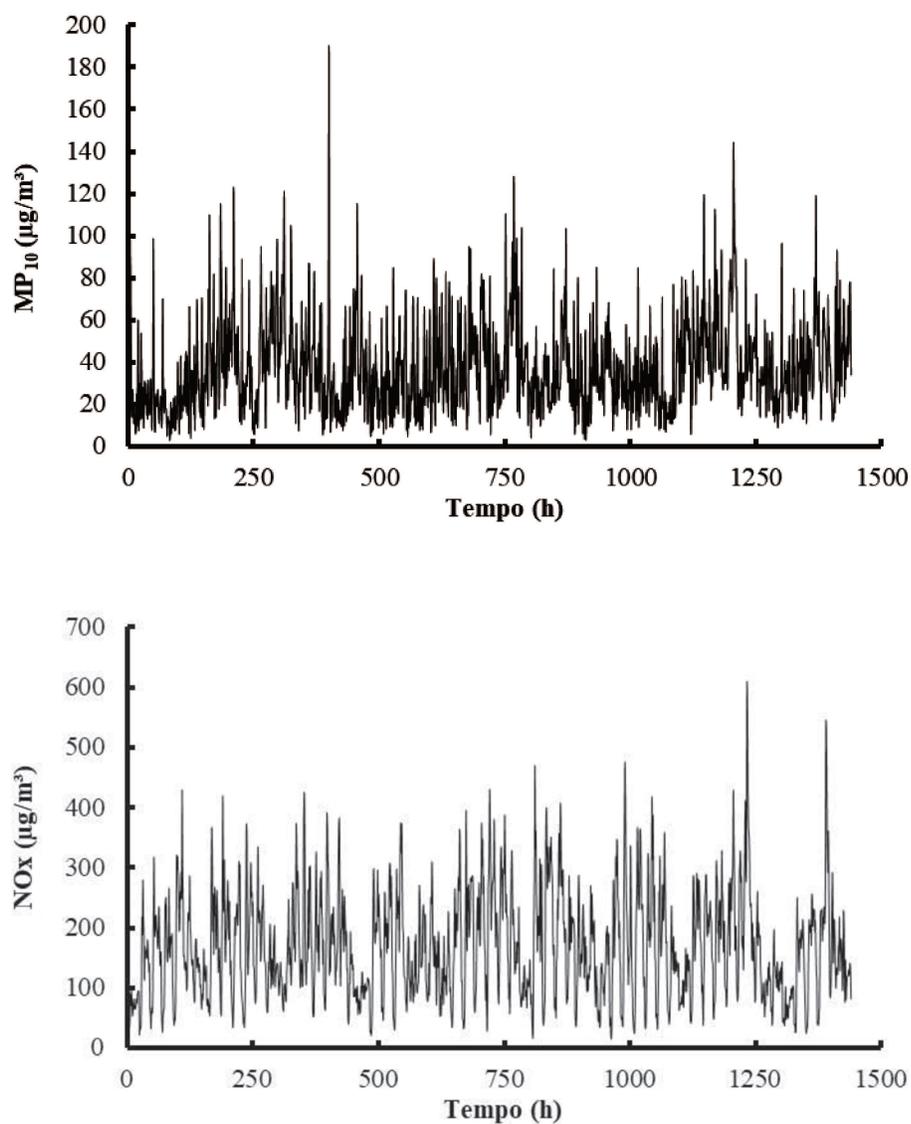
Figura 4.29 – Atrator reconstruído a partir da série de treino da concentração horária de NO_x para os anos de 1998 a 2000 (Centro de São Paulo) (o autor, 2018).



Garantida a dinâmica caótica determinista por meio dos atratores, entropia de Kolmogorov e dimensão de correlação, realizou-se as previsões por meio do algoritmo k - NN utilizando como valor da dimensão de imersão o dado gerado pelo algoritmo de

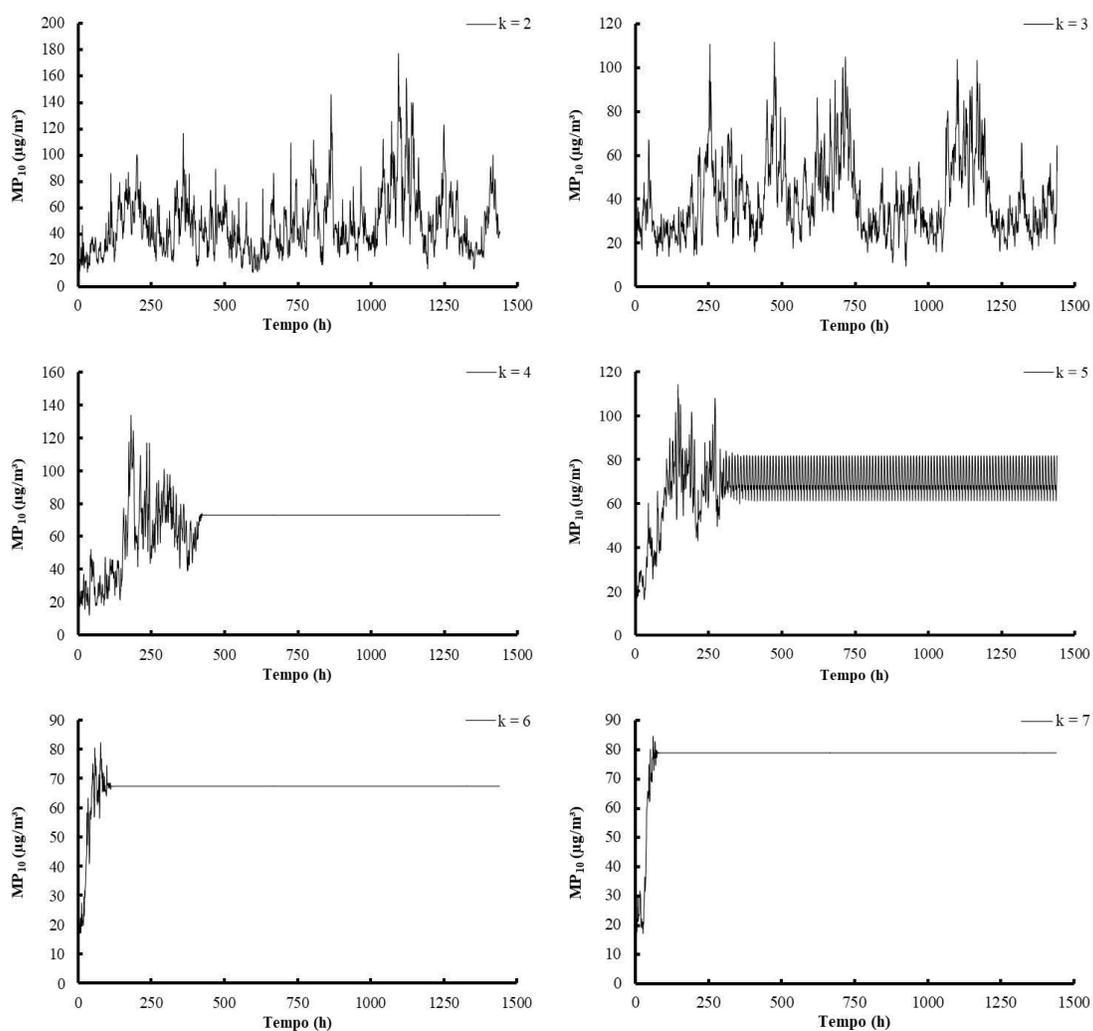
Grassberger e Procaccia (1983). A primeira etapa foi a solução das previsões variando o valor de vizinhos próximos (k). De modo a comparar qualitativamente as previsões realizadas, são apresentadas, na Figura 4.30, as séries relativas aos primeiros dois meses de aquisição do ano de 2010 para o MP_{10} e do mesmo período do ano de 2001 para o NO_x .

Figura 4.30 – Séries temporais da concentração de MP_{10} (ano de 2010) e NO_x (ano de 2001) obtidas na estação de monitoramento do Centro de São Paulo para os meses de janeiro e fevereiro dos anos relativos as últimas aquisições de cada um dos poluentes (o autor, 2018).



Os dados obtidos por meio dessas séries temporais foram utilizados como objeto para comparação da eficiência da previsão e não foram inseridos na série de treino (*out of sample*). As Figuras 4.31 e 4.32 trazem as previsões para, respectivamente, o MP_{10} e NO_x , variando k de 2 a 10. Os gráficos comparativos das séries previstas com as originais podem ser encontrados no Apêndice D, Figuras D5 e D6.

Figura 4.31 – Previsão *out of sample* da concentração horária de MP_{10} nos primeiros dois meses de 2010 na estação do Centro de São Paulo utilizando o algoritmo k -NN (o autor, 2018).



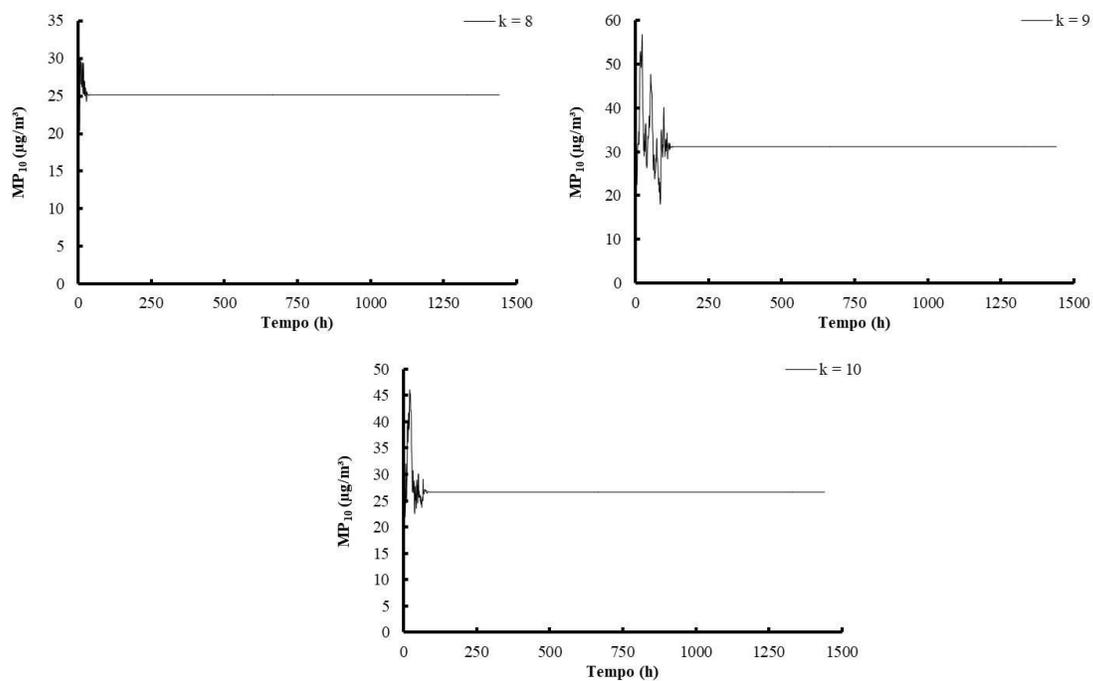
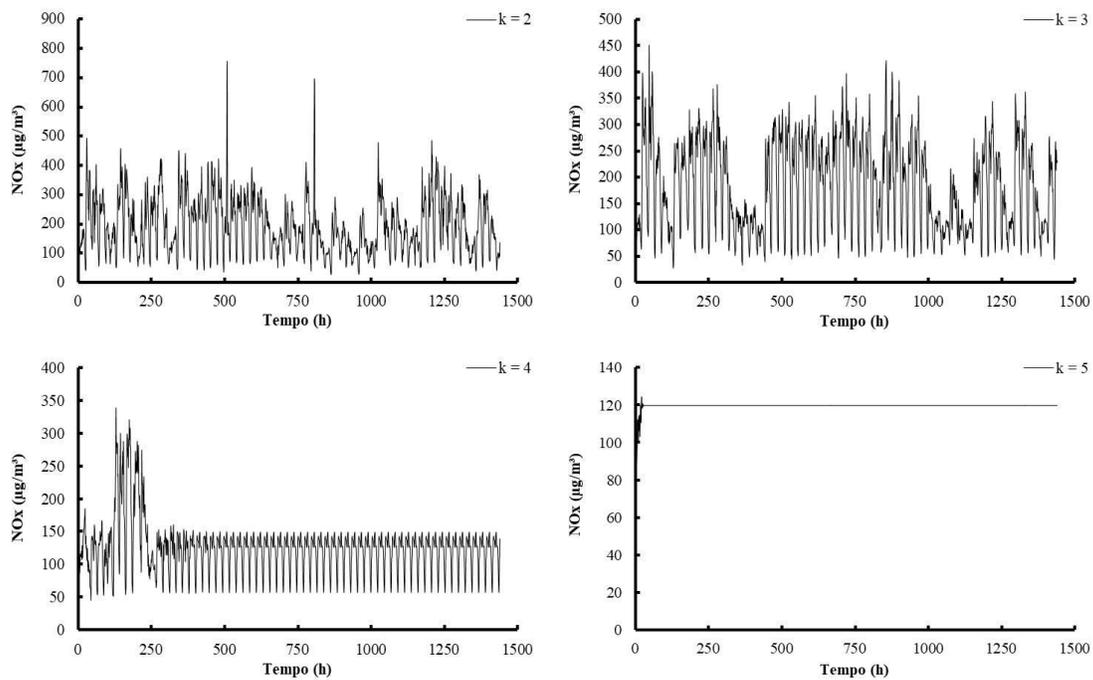
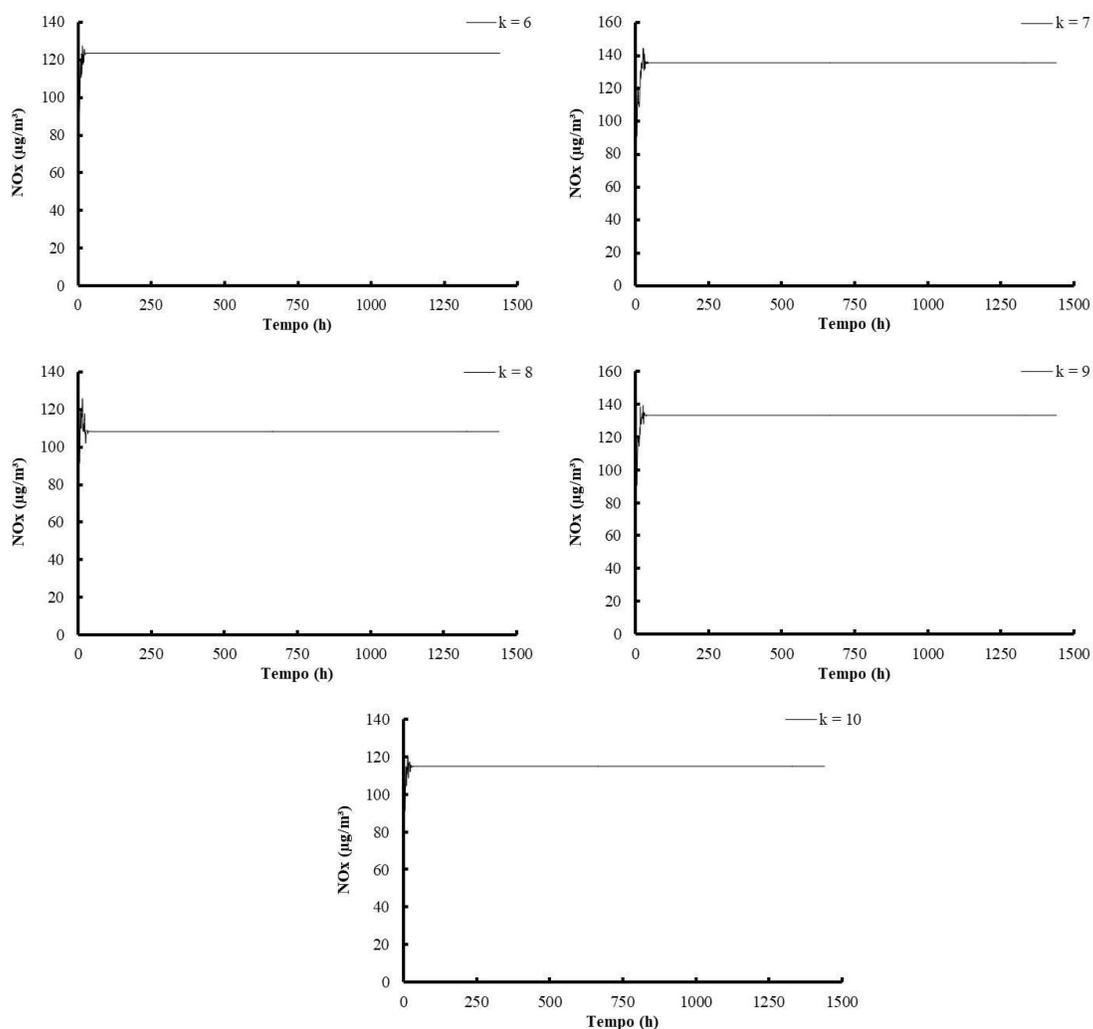


Figura 4.32 – Previsão *out of sample* da concentração horária de NO_x nos primeiros dois meses de 2000 na estação do Centro de São Paulo utilizando o algoritmo $k\text{-NN}$ (o autor, 2018).





Partindo da análise das previsões do MP_{10} para os primeiros dois meses de 2010, verifica-se que o valor de k , assim como constatado nas demais estações, exerce influência direta na série gerada. Os valores de $k = 2$ e $k = 3$ foram os únicos que permitiram a geração de uma série completa de 1440 pontos, já para $k = 4$ e $k = 5$ se verifica uma série com problemas de convergência a partir de, aproximadamente, 300 pontos (ou 300 horas). Para os demais valores de vizinhos próximos, a saber, 6, 7, 8, 9 e 10, são observadas convergências para um ponto médio ou variação harmônica a partir do dado relativo à 24^a hora.

Já em relação ao NO_x (Figura 4.32), séries completas (1440 pontos) foram obtidas para $k = 2$ e 3 , uma série de 250 pontos para $k = 4$ e nas demais situações foram verificados problemas de convergência para simulações com mais de 24 pontos.

Seguindo agora com a avaliação quantitativa das previsões, as Tabelas 4.15 e 4.16 apresentam a comparação dos dados reais com as previsões realizadas para 60 dias (1440 pontos), 30 dias (720 pontos), 15 dias (360 pontos), 7 dias (168 pontos), 2 dias (48 pontos) e 1 dia (24 pontos). Os critérios adotados para a análise das séries previstas foram o desvio da média real em relação à média do conjunto simulado e também os métodos do Erro Quadrático Médio (RMSE) e o Erro Absoluto Médio (MAE). Novamente, é importante destacar, que não é possível considerar os valores das Tabelas 4.15 e 4.16 sem antes fazer uma verificação nas Figuras 4.31 e 4.32. Para exemplificar o motivo desta afirmação, toma-se com base o caso para $k = 9$, previsão de 60 dias (1440 pontos) do MP₁₀. Nesta configuração, nota-se uma média prevista de 31,27 $\mu\text{g}/\text{m}^3$ versus 37,43 $\mu\text{g}/\text{m}^3$ da real, com desvio de 16%, $RMSE = 23,0 \mu\text{g}/\text{m}^3$ e $MAE = 6,2 \mu\text{g}/\text{m}^3$. Em análise qualitativa, é possível considerar essa, uma vez conhecendo as características do fenômeno estudado, uma boa previsão, contudo ao recorrer à Figura 4.31 esta conclusão não se sustenta, haja vista que para simulações com mais de 24 pontos ocorreu convergência para um valor médio, invalidando qualquer previsão que ultrapasse o limite de 24 horas. Partindo deste pressuposto, diversos valores apresentados nas Tabelas 4.15 e 4.16 estão na mesma situação e, por este motivo, serão discutidos, assim como em todas as estações desta Tese, apenas os melhores casos fundamentados nas conclusões abstraídas do estudo da variação de k .

Sendo assim, para o MP₁₀, verifica-se ser possível uma previsão de 60 dias muito interessante quando $k = 3$, com desvio de 9%, $RMSE = 27,0$ e $MAE = 3,2 \mu\text{g}/\text{m}^3$. Para uma previsão de 15 dias, o erro baixa para 0%, $RMSE = 27,8$ e $MAE = 0,1 \mu\text{g}/\text{m}^3$. A melhor simulação de 24 horas ocorreu para $k = 10$, com desvio de 4%, $RMSE = 39,4$ e $MAE = 1,5 \mu\text{g}/\text{m}^3$.

Observando o NO_x (Tabela 4.16), observa-se que para $k = 3$ em uma previsão de 60 dias o desvio da média foi de 8%, $RMSE = 135,3$ e $MAE = 25,0 \mu g/m^3$. Para o mesmo valor de k , agora considerando uma média de 30 dias, o desvio foi de 16%, $RMSE = 135,3$ e $MAE = 25,0 \mu g/m^3$, para 15 dias o desvio foi 19%, $RMSE = 126,9$ e $MAE = 30,8 \mu g/m^3$, 1 semana (desvio = 30%, $RMSE = 136,1$ e $MAE = 72,2 \mu g/m^3$), 2 dias (desvio = 91%, $RMSE = 148,1$ e $MAE = 91,4 \mu g/m^3$) e, por fim, para 1 dia, desvio = 85%, $RMSE = 65,0$ e $MAE = 56,3 \mu g/m^3$. Em geral, verifica-se que o a diminuição do número de dias simulados implica em aumento do desvio, isso ocorre pelo fato de que quanto mais dias previstos maior será o número de valores para compor a média e, portanto, ocorre uma suavização do valor da média acarretando em menores erros. Ainda no NO_x, a melhor situação para uma simulação de 24 horas ocorreu para $k = 5$, com desvio de 55%, $RMSE = 41,7$ e $MAE = 38,5 \mu g/m^3$. Novamente, contextualizando as características do fenômeno e conhecendo as subjacentes particularidades caóticas, pode-se afirmar que algumas dessas previsões apresentam bom desempenho.

A qualidade das previsões ao se utilizar a metodologia k -NN está intrinsicamente relacionada à série de treino. Por se tratar de um método de aprendizado de máquina, ele parte da premissa de que a série reproduz seu próprio comportamento ao longo do tempo, sendo assim, a observação futura está associada às anteriores (Parmezan e Batista, 2016). Logo, fica clara a sua importante relação com a série de treino e é direta a conclusão de que problemas no treino gerarão erros nas previsões. As séries utilizadas neste trabalho, por diversos motivos já salientados, apresentam lacunas oriundas de dados inválidos. Tanto para a aplicação da metodologia do Caos quanto do próprio k -NN é necessário que as séries sejam contínuas, logo se realizou a retirada das referidas brechas, lançando-as ao final da evolução. Esse procedimento implica em juntar dados muitas vezes sem conexão fenomenológica, pois a poluição atmosférica, sazonal e cíclica, assim como as estações do ano, apresentam características bem definidas pelo período considerado e, ao agrupar um dado advindo, por exemplo, do inverno com a primavera, muita informação é perdida e o próprio algoritmo k -NN “aprende de forma errada”, gerando por fim uma previsão equivocada. Este é destacado nesta Tese, como sendo o principal motivo do insucesso de algumas previsões e se repete para todas as estações aqui consideradas.

Tabela 4.15 - Comparação entre as séries simuladas e reais para a concentração de MP_{10} – Centro de São Paulo (o autor, 2018).

	Média (60 dias) ($\mu\text{g}/\text{m}^3$)	Desvio	Média (30 dias) ($\mu\text{g}/\text{m}^3$)	Desvio	Média (15 dias) ($\mu\text{g}/\text{m}^3$)	Desvio	Média (1 semana) ($\mu\text{g}/\text{m}^3$)	Desvio	Média (2 dias) ($\mu\text{g}/\text{m}^3$)	Desvio	Média (1 dia) ($\mu\text{g}/\text{m}^3$)	Desvio
Real	37,43	-	35,88	-	36,55	-	27,79	-	28,48	-	34,33	-
k												
2	47,81	28%	40,91	14%	43,00	18%	35,35	27%	22,65	21%	21,31	38%
3	40,65	9%	42,46	18%	36,49	0%	27,71	0%	33,18	16%	25,65	25%
4	68,14*	82%	63,26*	76%	55,70	52%	33,55	21%	26,82	6%	23,69	31%
5	68,93*	84%	66,66*	86%	62,08*	70%	52,89	90%	27,99	2%	23,87	31%
6	66,39*	77%	65,45*	82%	63,57*	74%	59,27*	113%	37,76	32%	22,65	34%
7	77,45*	107%	75,91*	112%	72,83*	99%	65,79*	137%	34,43*	21%	23,03	33%
8	25,16*	33%	25,18*	30%	72,98*	100%	25,24*	9%	25,57*	11%	25,76	25%
9	31,27*	16%	31,41*	12%	73,12*	100%	32,32*	16%	35,44	24%	38,05	11%
10	26,77*	28%	26,90*	25%	73,28*	101%	27,77*	0%	30,81*	8%	32,94	4%
	60 dias		30 dias		15 dias		1 semana		2 dias		1 dia	
k	RMSE ($\mu\text{g}/\text{m}^3$)	MAE ($\mu\text{g}/\text{m}^3$)	RMSE ($\mu\text{g}/\text{m}^3$)	MAE ($\mu\text{g}/\text{m}^3$)	RMSE ($\mu\text{g}/\text{m}^3$)	MAE ($\mu\text{g}/\text{m}^3$)	RMSE ($\mu\text{g}/\text{m}^3$)	MAE ($\mu\text{g}/\text{m}^3$)	RMSE ($\mu\text{g}/\text{m}^3$)	MAE ($\mu\text{g}/\text{m}^3$)	RMSE ($\mu\text{g}/\text{m}^3$)	MAE ($\mu\text{g}/\text{m}^3$)
2	32,7	10,4	27,6	5,0	28,3	6,5	25,4	7,6	29,2	5,9	39,5	13,1
3	27,0	3,2	28,3	6,6	27,8	0,1	23,0	0,1	31,3	4,6	37,6	8,7
4	39,0	30,7	38,1	27,4	33,5	1931	22,4	5,8	28,4	1,6	36,7	10,8
5	39,6	31,5	40,3	30,8	38,0	25,5	39,3	25,1	29,7	0,5	37,4	10,5
6	36,7	29,0	38,2	29,6	37,5	27,0	42,3	31,5	35,4	9,2	36,9	11,8
7	46,2	40,0	47,4	40,1	45,6	36,3	49,7	38,0	35,3	6,0	37,7	11,3
8	25,2	12,3	25,3	10,7	26,4	11,4	21,7	2,5	27,4	3,0	37,4	8,6
9	23,0	6,2	23,7	4,5	25,0	4,9	23,1	4,5	31,5	6,9	42,1	3,6
10	24,5	10,7	24,8	9,0	26,0	9,4	22,1	0,0	29,0	2,3	39,4	1,5

*valores desconsiderados na análise devido à presença de convergência harmônica ou para um valor médio.

Tabela 4.16 - Comparação entre as séries simuladas e reais para a concentração de NO_x – Centro de São Paulo (o autor, 2018).

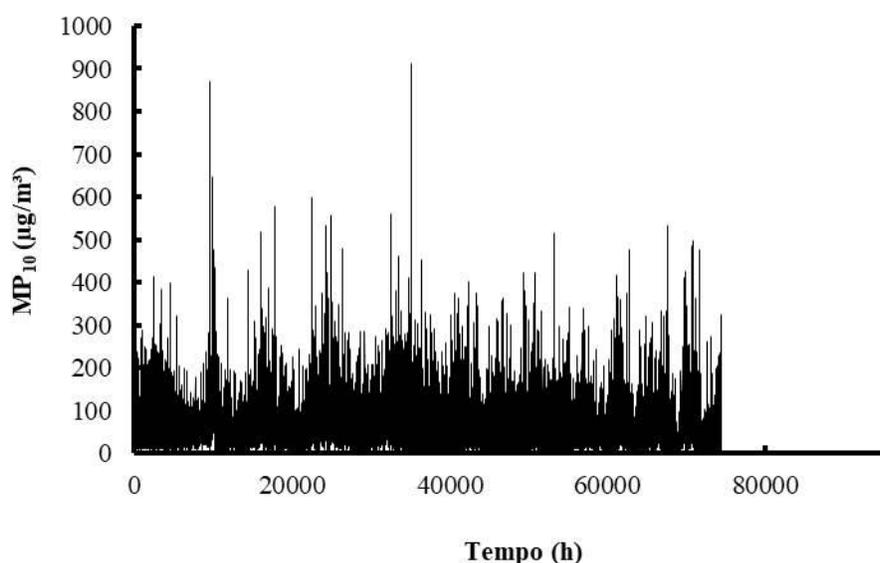
	Média (60 dias) (µg/m ³)	Desvio	Média (30 dias) (µg/m ³)	Desvio	Média (15 dias) (µg/m ³)	Desvio	Média (1 semana) (µg/m ³)	Desvio	Média (2 dias) (µg/m ³)	Desvio	Média (1 dia) (µg/m ³)	Desvio
Real	168,94	-	163,69	-	160,08	-	142,55	-	100,59	-	70,19	-
k												
2	193,91	15%	211,34	29%	206,05	29%	214,78	51%	194,38	93%	126,54	80%
3	182,36	8%	189,40	16%	190,87	19%	185,07	30%	192,09	91%	129,89	85%
4	125,93*	25%	131,63*	20%	142,59*	11%	140,81	1%	114,71	14%	125,07	78%
5	119,49*	29%	119,31*	27%	118,94*	26%	118,11*	17%	114,22*	14%	108,73	55%
6	123,30*	27%	123,15*	25%	122,86*	23%	122,20*	14%	119,16*	18%	114,86	64%
7	135,36*	20%	135,06*	17%	134,43*	16%	133,02*	7%	126,44*	26%	116,34	66%
8	108,29*	36%	108,32*	34%	134,54*	16%	108,48*	24%	109,05*	8%	110,25	57%
9	132,86*	21%	132,60*	19%	134,62*	16%	130,90*	8%	125,38*	25%	117,62	68%
10	114,78*	32%	114,69*	30%	134,72*	16%	114,10*	20%	112,27*	12%	109,58	56%
	60 dias		30 dias		15 dias		1 semana		2 dias		1 dia	
k	RMSE (µg/m³)	MAE (µg/m³)	RMSE (µg/m³)	MAE (µg/m³)	RMSE (µg/m³)	MAE (µg/m³)	RMSE (µg/m³)	MAE (µg/m³)	RMSE (µg/m³)	MAE (µg/m³)	RMSE (µg/m³)	MAE (µg/m³)
2	135,3	25,0	142,8	47,7	133,7	46,0	136,1	72,2	119,7	93,7	65,0	56,3
3	121,0	13,4	113,7	25,7	126,9	30,8	127,2	42,5	148,1	91,4	102,5	59,6
4	109,1	43,0	100,4	32,1	97,7	17,5	99,7	1,7	75,5	14,1	60,2	54,8
5	109,1	49,4	93,1	44,4	89,5	41,1	79,6	24,4	56,3	13,6	41,7	38,5
6	99,6	45,6	91,4	40,6	87,9	37,2	78,7	20,3	58,3	18,5	47,6	44,6
7	94,5	33,6	86,4	28,7	83,1	25,6	75,6	9,5	59,6	25,8	49,5	46,1
8	107,4	60,6	99,1	55,4	95,6	51,7	84,5	34,1	58,2	8,4	43,2	40,0
9	95,5	36,1	87,4	31,1	84,0	28,0	76,1	11,6	59,2	24,8	50,8	47,4
10	103,8	54,2	95,5	49,1	92,0	45,6	81,5	28,4	57,2	11,6	42,7	39,3

*valores desconsiderados na análise devido à presença de convergência harmônica ou para um valor médio.

4.6.4 Estação de Guarulhos

Para a estação de Guarulhos, a Figura 4.33 apresenta a série de treino utilizada na previsão da concentração horária de MP_{10} . A estação de Guarulhos não realizou aquisição de NO_x e para o MP_{10} o período disponibilizado na plataforma QUALAR foi de 1998 a 2009, sendo assim, a série de treino foi elaborada de 1998 a 2008 e a série comparativa abrangeu os meses de janeiro e fevereiro de 2009. Análogo ao que foi observado nas estações anteriores, os dados de treino apresentam um perfil bastante variável e dispersivo. O valor de concentração médio encontrado por meio da Equação 3.1 para o MP_{10} foi de 64 ± 50 ($\mu\text{g}/\text{m}^3$). O caráter variável e dispersivo dos dados é confirmado ao se observar um desvio padrão alto da ordem da própria média. Novamente, a ausência de dados ao final da evolução da série de treino é constatada, sendo esta observação resultado direto dos dados inválidos presentes nas tabelas exportados da plataforma CETESB.

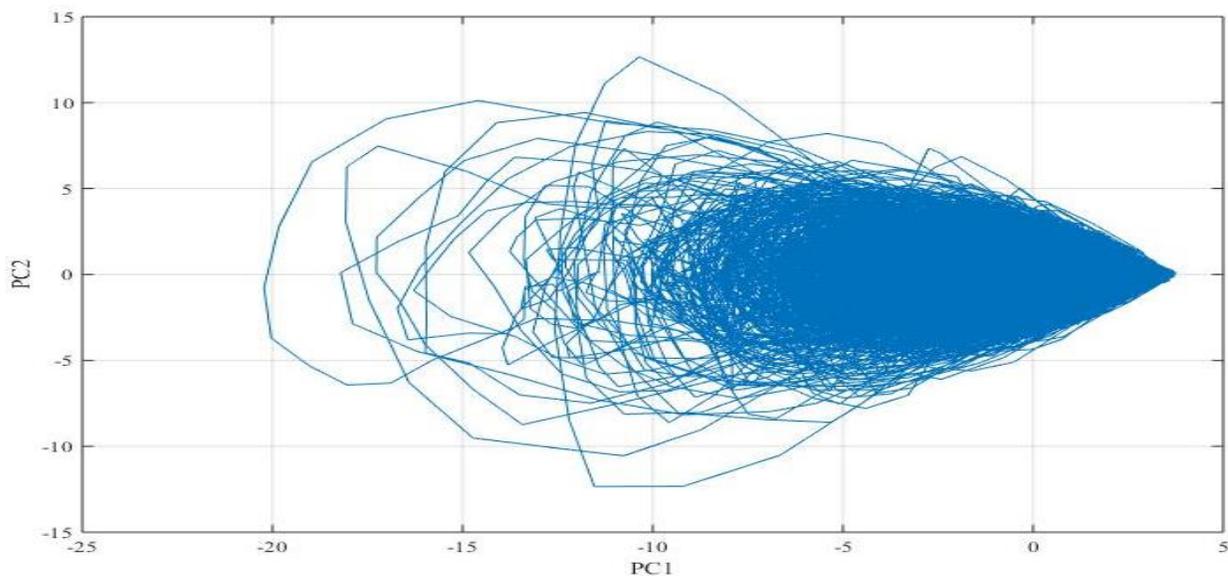
Figura 4.33 - Série de treino da concentração horária de MP_{10} dos anos de 1998 a 2009 na estação de Guarulhos (o autor, 2018).



Conforme procedimento adotado nesta Tese, de modo a garantir que a série de treino utilizada presente, durante sua evolução temporal, características caóticas, a Figuras

4.34 traz o atrator reconstruído para as séries de treino da concentração horária de MP_{10} na estação de monitoramento do Centro de São Paulo. A solução no espaço de fases foi obtida por meio do método da decomposição em valores singulares (SVD) (Broomhead e King, 1986), com dimensão de imersão otimizada (Grassberger e Procaccia, 1983) $I = 10$ e tempo de atraso de 1 hora. Por meio da evolução das linhas que descrevem o atrator é possível verificar que se trata de um atrator do tipo estranho (característico do caos) com soluções convergentes para uma região bem definida e estreita no espaço de fases (linha horizontal em PC2 (principal componente) = 0). Para assegurar essas conclusões, realizou-se o cálculo dos invariantes entropia de Kolmogorov e dimensão de correlação. Sendo assim, $D_2 = 4,4793$, $K_{ML} = 0,1346$ bits/s e $\lambda = 0,18$. Dos invariantes do caos determinístico, pode-se asseverar, quantitativamente, que a evolução dos dados é caótica (pois $0 < K_{ML} < \infty$) apresentando 5 graus de liberdade regentes da dinâmica de ambos os sistemas.

Figura 4.34 – Atrator reconstruído a partir da série de treino da concentração horária de MP_{10} para os anos de 1998 a 2008 (Guarulhos) (o autor, 2018).



Uma vez discutidas as características da dinâmica caótica por meio do atrator, entropia de Kolmogorov e dimensão de correlação, passa-se à previsão que foi realizada utilizando o algoritmo k -NN e o valor de dimensão de imersão obtido aplicando o método

de Grassberger e Procaccia (1983). Para comparar de forma qualitativa as previsões, apresenta-se na Figura 4.35 a série real da concentração de MP_{10} na estação de Guarulhos para os primeiros dois meses de 2009. Os dados advindos dessa série temporal foram utilizados na comparação da eficiência da previsão e não foi inseridos na série de treino (*out of sample*). Ao todo nove previsões são apresentadas para valores de vizinhos próximos (k) variando de 2 a 10. A Figura 4.36 traz as previsões para o MP_{10} , variando k de 2 a 10. Os gráficos comparativos das séries previstas com a original podem ser encontrados no Apêndice D, Figura D7.

Analisando as previsões para o MP_{10} para os primeiros dois meses de 2009 é possível notar, novamente, que o número de vizinhos próximos é parâmetro fundamental para a geração de uma boa previsão. A única situação para a qual foi possível obter uma série completa (1440 pontos) é observada para $k = 2$. Para $k = 3$ uma série bem desenvolvida é observada até aproximadamente o ponto 625, já para os demais valores de k , são verificadas convergências para um ponto médio ou variação harmônica a partir do dado relativo à 24^a hora.

Figura 4.35 – Série temporal da concentração de MP_{10} obtida na estação de monitoramento de Guarulhos para os meses de janeiro e fevereiro de 2009 (o autor, 2018).

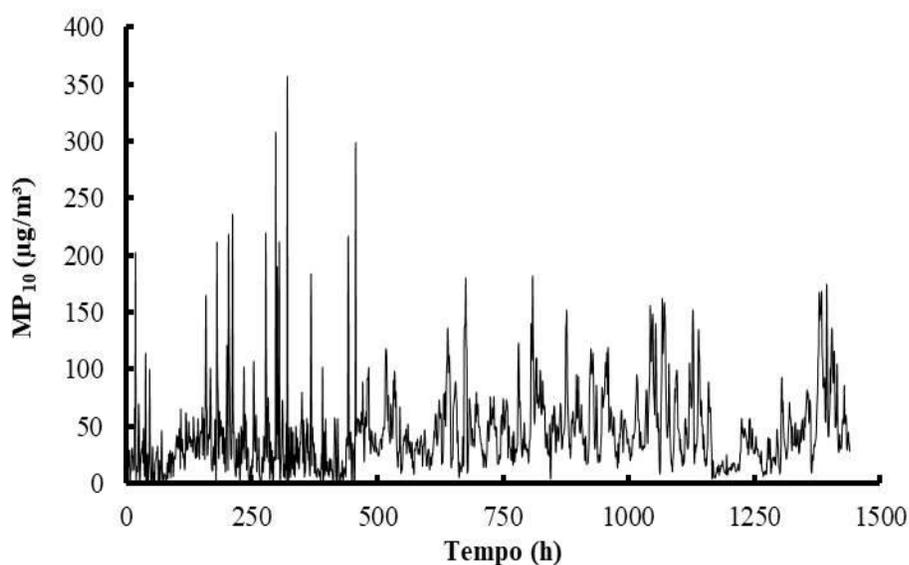
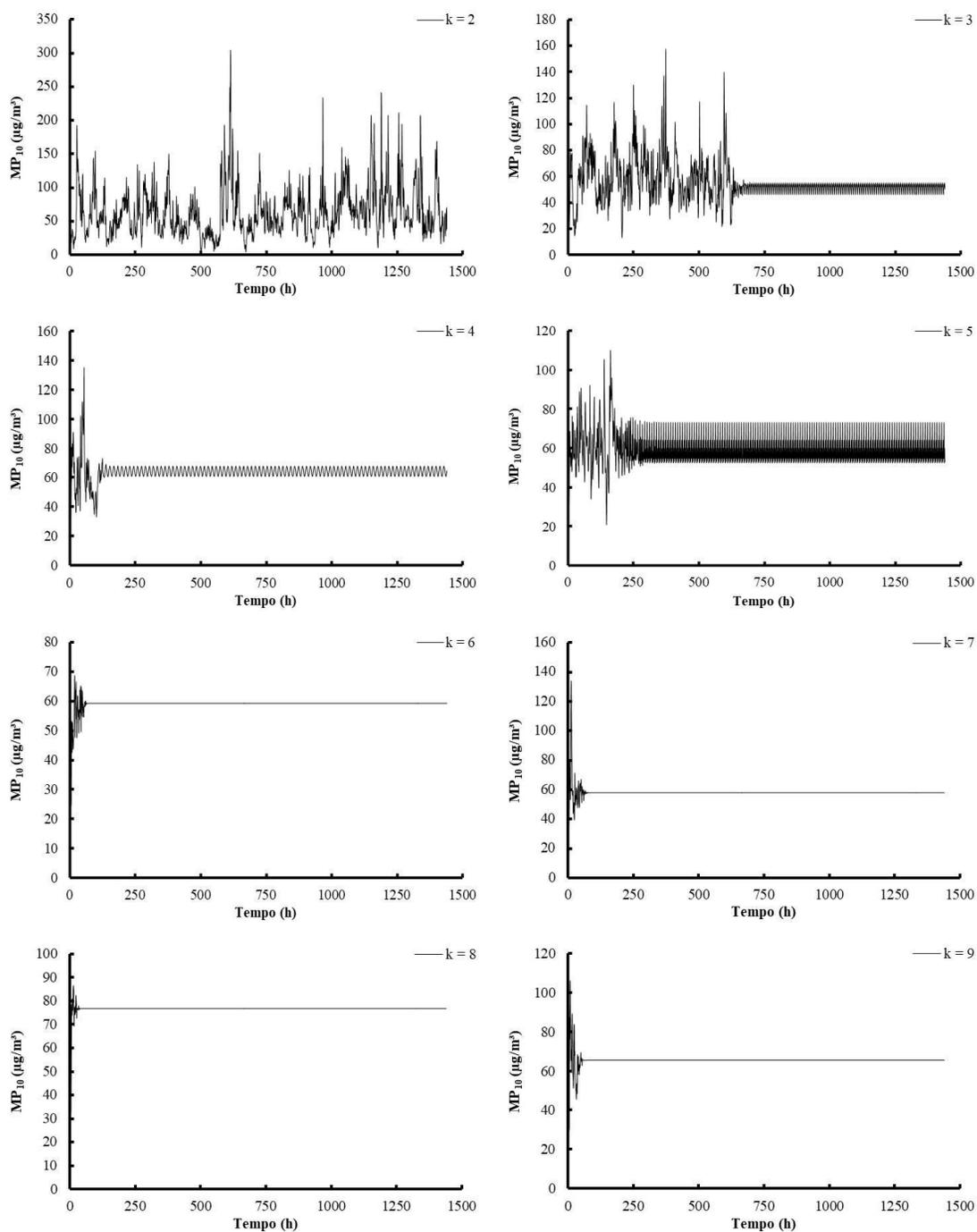
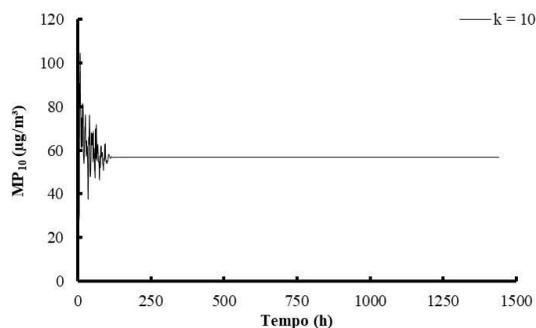


Figura 4.36 – Previsão *out of sample* da concentração horária de MP_{10} nos primeiros dois meses de 2009 na estação de Guarulhos utilizando o algoritmo k - NN (o autor, 2018).





Realizando a análise quantitativa das previsões, a Tabela 4.17 apresenta a comparação dos dados reais com as previsões realizadas para 60 dias (1440 pontos), 30 dias (720 pontos), 15 dias (360 pontos), 7 dias (168 pontos), 2 dias (48 pontos) e 1 dia (24 pontos). Os critérios adotados para a análise das séries previstas foram os mesmo utilizados para as outras estações, o desvio da média real em relação à média do conjunto simulado e também os métodos do Erro Quadrático Médio (RMSE) e o Erro Absoluto Médio (MAE). Aqui também não é possível considerar os valores da Tabela 4.17 apartados da Figura 4.36, pelos motivos e exemplos salientados ao final da seção 4.3.3.

Avaliando $k = 2$ (única série inteira de 1440 pontos), para a média de 60 dias o desvio foi de 42%, $RMSE = 54,2$ e $MAE = 18,8 \mu g/m^3$, no caso da previsão de 30 dias o desvio foi de 47%, $RMSE = 55,4$ e $MAE = 18,4 \mu g/m^3$, para 15 dias o desvio foi 61%, $RMSE = 55,4$ e $MAE = 22,4 \mu g/m^3$, 1 semana (desvio = 83%, $RMSE = 53,3$ e $MAE = 25,4 \mu g/m^3$), 2 dias (desvio = 97%, $RMSE = 69,6$ e $MAE = 30,8 \mu g/m^3$) e, por fim, para 1 dia, desvio = 24%, $RMSE = 47,3$ e $MAE = 8,2 \mu g/m^3$. De modo geral, um aumento do erro com a diminuição do número de dias simulados. Esta é uma conclusão direta do fato de que quanto mais dias previstos maior será o número de valores para compor a média e, portanto, ocorre uma suavização do valor da média acarretando em menores erros quando comparado à média real. Ainda sobre o MP_{10} , a melhor situação para uma simulação de 24 horas ocorreu para $k = 2$, com desvio de 24%, $RMSE = 47,3$ e $MAE = 8,2 \mu g/m^3$. De forma análoga ao discutido na estação anterior, as características caóticas e fenomenológicas apresentadas pelo fenômeno permitem dizer que as melhores previsões obtidas são boas, não sendo melhores devido à massiva falta de dados na série de treino.

Tabela 4.17 - Comparação entre as séries simuladas e reais para a concentração de MP_{10} – Guarulhos (o autor, 2018).

	Média (60 dias) ($\mu\text{g}/\text{m}^3$)		Desvio		Média (30 dias) ($\mu\text{g}/\text{m}^3$)		Desvio		Média (15 dias) ($\mu\text{g}/\text{m}^3$)		Desvio		Média (1 semana) ($\mu\text{g}/\text{m}^3$)		Desvio		Média (2 dias) ($\mu\text{g}/\text{m}^3$)		Desvio		Média (1 dia) ($\mu\text{g}/\text{m}^3$)		Desvio		
Real	44,79	-	39,17	-	36,58	-	30,61	-	31,71	-	33,63	-													
k																									
2	63,59	42%	57,49	47%	58,95	61%	56,02	83%	62,55	97%	25,41	24%													
3	54,65*	22%	57,95*	48%	59,62	63%	54,90	79%	43,72	38%	44,62	33%													
4	63,83*	43%	63,47*	62%	62,75*	72%	61,17	100%	62,00	96%	57,81	72%													
5	59,22*	32%	59,66*	52%	60,57*	66%	61,37	101%	58,63	85%	53,91	60%													
6	58,98*	32%	58,77*	50%	58,33*	59%	57,33*	87%	53,05	67%	48,42	44%													
7	58,03*	30%	58,09*	48%	58,20*	59%	58,45*	91%	59,54	88%	62,41	86%													
8	76,65*	71%	76,51*	95%	58,32*	59%	75,56*	147%	72,52*	129%	68,32	103%													
9	65,26*	46%	65,14*	66%	58,40*	60%	64,37*	110%	61,85*	95%	64,35	91%													
10	56,97*	27%	57,16*	46%	58,47*	60%	58,40*	91%	61,31*	93%	62,14	85%													
	60 dias				30 dias				15 dias				1 semana				2 dias				1 dia				
k	RMSE ($\mu\text{g}/\text{m}^3$)	MAE ($\mu\text{g}/\text{m}^3$)	RMSE ($\mu\text{g}/\text{m}^3$)	MAE ($\mu\text{g}/\text{m}^3$)	RMSE ($\mu\text{g}/\text{m}^3$)	MAE ($\mu\text{g}/\text{m}^3$)	RMSE ($\mu\text{g}/\text{m}^3$)	MAE ($\mu\text{g}/\text{m}^3$)	RMSE ($\mu\text{g}/\text{m}^3$)	MAE ($\mu\text{g}/\text{m}^3$)	RMSE ($\mu\text{g}/\text{m}^3$)	MAE ($\mu\text{g}/\text{m}^3$)	RMSE ($\mu\text{g}/\text{m}^3$)	MAE ($\mu\text{g}/\text{m}^3$)	RMSE ($\mu\text{g}/\text{m}^3$)	MAE ($\mu\text{g}/\text{m}^3$)	RMSE ($\mu\text{g}/\text{m}^3$)	MAE ($\mu\text{g}/\text{m}^3$)	RMSE ($\mu\text{g}/\text{m}^3$)	MAE ($\mu\text{g}/\text{m}^3$)	RMSE ($\mu\text{g}/\text{m}^3$)	MAE ($\mu\text{g}/\text{m}^3$)	RMSE ($\mu\text{g}/\text{m}^3$)	MAE ($\mu\text{g}/\text{m}^3$)	
2	54,2	18,8	55,4	18,4	55,4	22,4	53,3	25,4	69,6	30,8	47,3	8,2													
3	40,9	9,9	47,3	18,8	51,9	23,0	44,2	24,3	45,8	12,0	53,0	11,0													
4	41,2	19,0	46,0	24,3	50,3	26,2	44,6	30,6	53,7	30,3	57,0	24,2													
5	39,8	14,4	44,4	20,5	49,3	24,0	43,3	30,8	48,0	26,9	48,7	20,3													
6	38,7	14,2	42,8	19,6	46,9	21,7	38,6	26,7	43,1	21,3	46,9	14,7													
7	38,6	13,2	43,0	18,9	47,5	21,6	40,9	27,8	51,2	27,8	61,0	28,8													
8	48,2	31,9	53,6	37,4	57,8	39,6	53,7	45,0	58,0	40,8	60,9	34,6													
9	41,6	20,5	46,5	26,0	50,9	28,3	45,2	33,8	52,5	30,1	61,9	30,7													
10	38,3	12,2	42,6	18,0	47,3	21,0	40,7	27,8	51,3	29,6	59,6	28,5													

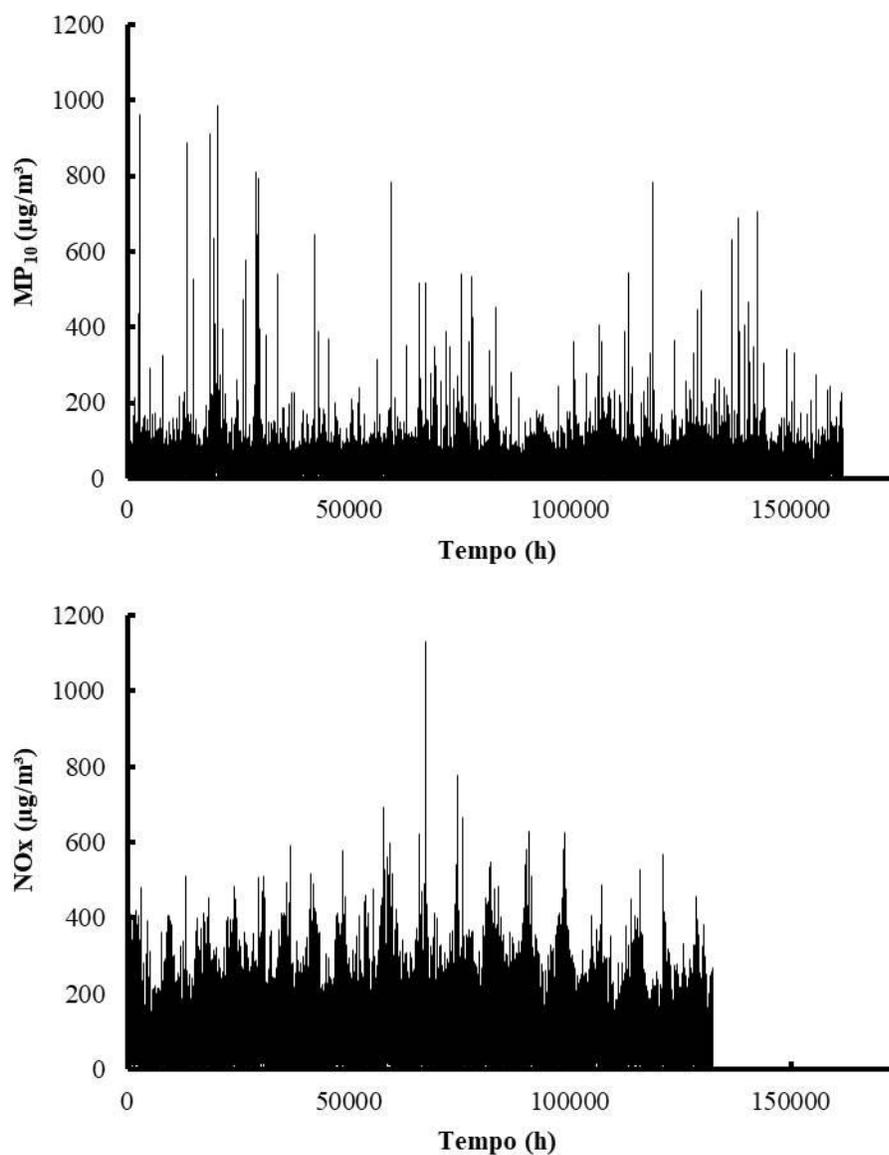
*valores desconsiderados na análise devido à presença de convergência harmônica ou para um valor médio.

4.6.5 Estação de Cubatão - Centro

A Figura 4.37 apresenta a série de treino utilizada na previsão da concentração horária de MP_{10} e NOx na estação de monitoramento de Cubatão - Centro. As séries de treino de ambos os poluentes foram elaboradas com dados de 1998 a 2017 e a série comparativa abrangeu os meses de janeiro e fevereiro de 2018. Conforme notado em todas as estações analisadas nesta Tese, os dados de treino apresentam um perfil bastante variável e dispersivo. O valor de concentração médio encontrado por meio da Equação 3.1 para o MP_{10} foi de $34 \pm 27 \mu\text{g}/\text{m}^3$ e para o NOx foi de $74 \pm 67 \mu\text{g}/\text{m}^3$. O desvio padrão da ordem da própria média confirma o caráter variável e dispersivo dos dados. A falta de informações, relativa aos dados inválidos presentes nas tabelas exportados da plataforma CETESB, também é verificada para os poluentes da estação de Cubatão – Centro.

Para garantir que as séries utilizadas como treino, similar ao observado nas séries anuais individuais, representam uma dinâmica caótica, as Figuras 4.38 e 4.39 ilustram os atratores reconstruídos para as séries de treino da concentração horária de MP_{10} e NOx na estação de monitoramento de Cubatão - Centro dos anos de 1998 a 2017. Ambos os atratores foram reconstruídos no espaço de fases utilizando o método da decomposição em valores singulares (SVD) (Broomhead e King, 1986), dimensão de imersão otimizada pelo algoritmo de Grassberger e Procaccia (1983) igual a 9 para o MP_{10} e 14 para o NOx . O tempo de atraso utilizado foi igual à frequência de amostragem (Yu *et al.*, 2011; Yu *et al.*, 2013), sendo assim, $\theta = 1 h$. Os dois atratores, bem como todos os apresentados nesta Tese, foram reconstruídos com o mesmo número de pontos da série temporal. Por meio da inspeção dos atratores é possível verificar que a evolução dos sistemas no espaço de fases descreve claramente um atrator estranho com soluções que convergem para uma região bem definida e estreita no espaço de fases (linha horizontal em PC2 (principal componente) = 0), sugerindo uma dinâmica caótica determinista.

Figura 4.37 – Séries de treino da concentração horária de MP_{10} e NO_x de 1998 a 2017 na estação de Cubatão – Centro (o autor, 2018).



Para confirmar essa constatação, lança-se mão do cálculo dos invariantes do caos determinístico, a saber, entropia de Kolmogorov (K_{ML}) e dimensão de correlação (D_2). Para o MP_{10} , $D_2 = 4,7556$, $K_{ML} = 0,1757$ bits/s e $\lambda = 0,19$ e para o NO_x , $D_2 = 4,9203$, $K_{ML} = 0,1201$ bits/s e $\lambda = 0,16$. A partir dos valores obtidos para os invariantes é possível garantir que a evolução dos dados apresentam características caóticas cuja dinâmica é representada por 5 graus de liberdade.

Figura 4.38 – Atrator reconstruído a partir da série de treino da concentração horária de MP_{10} para os anos de 1998 a 2017 (Cubatão - Centro) (o autor, 2018).

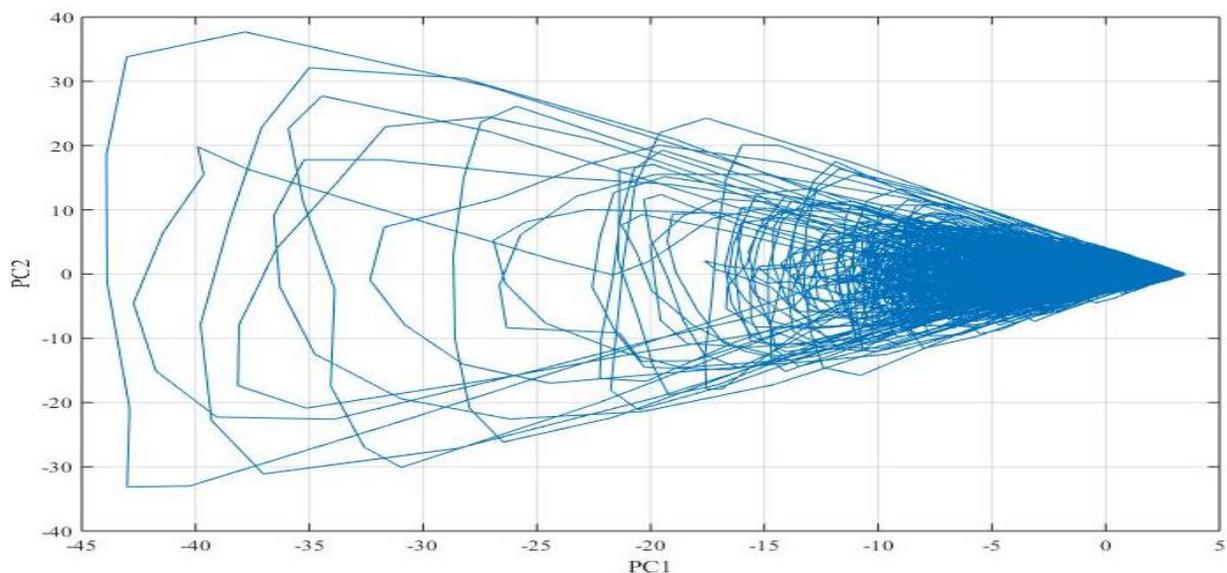
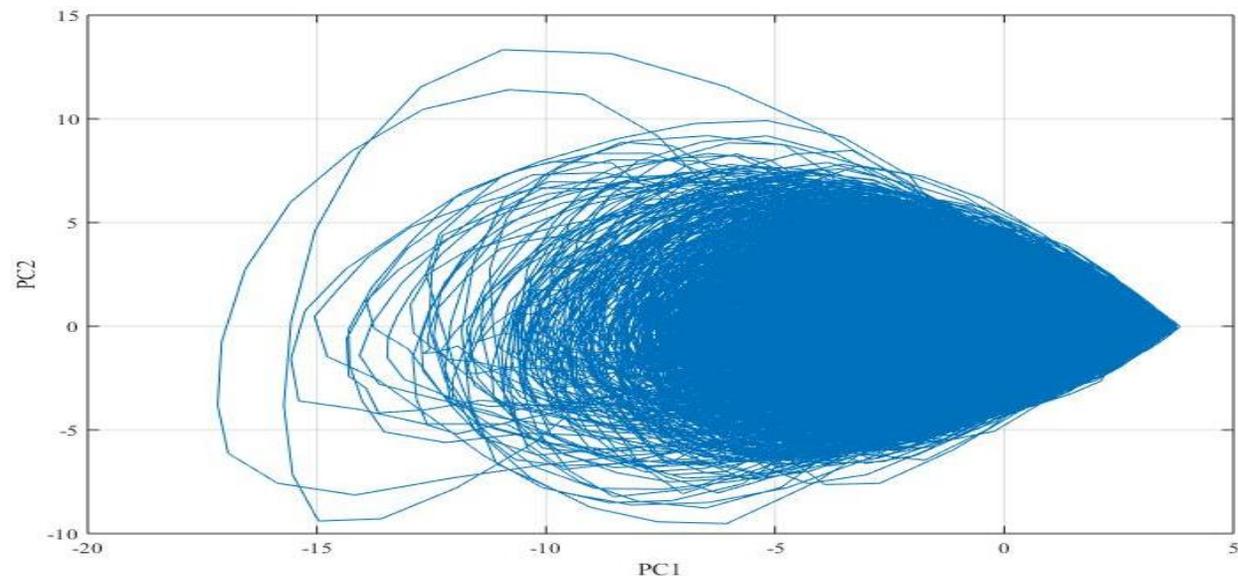


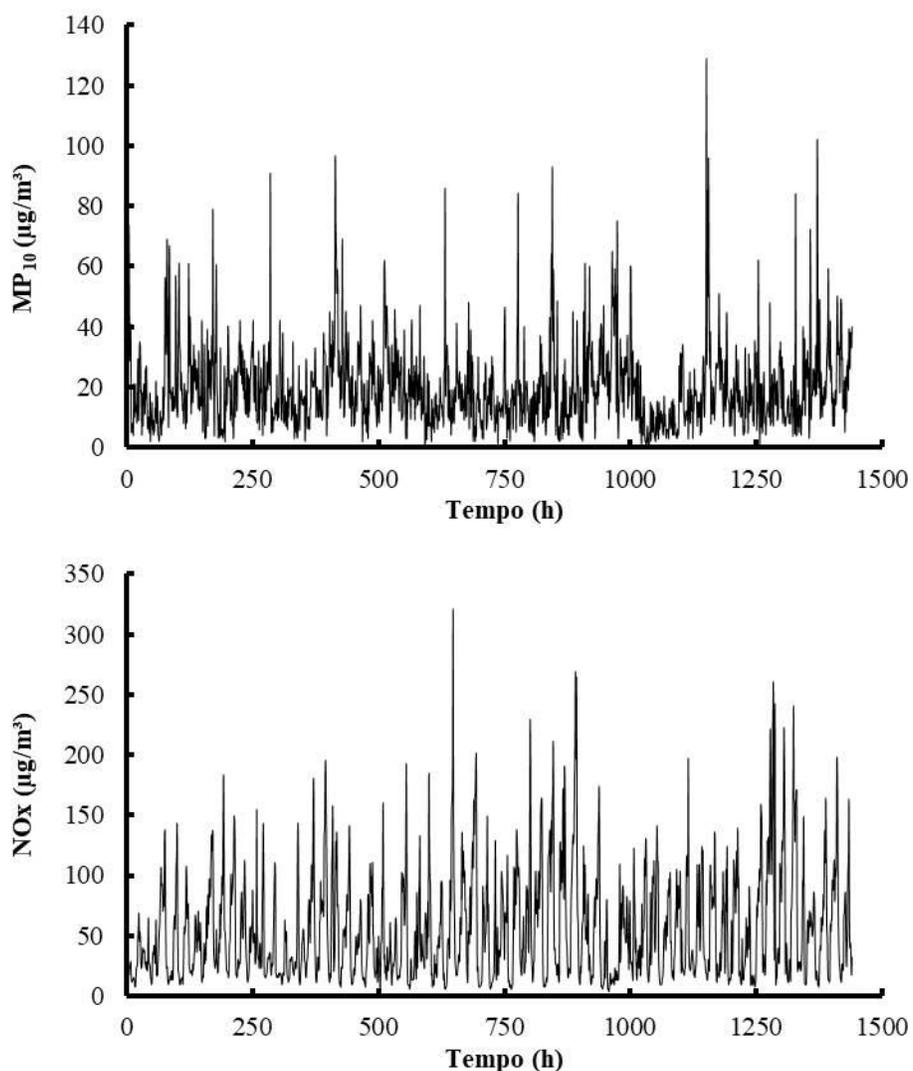
Figura 4.39 – Atrator reconstruído a partir da série de treino da concentração horária de NO_x para os anos de 1998 a 2017 (Cubatão - Centro) (o autor, 2018).



A Figura 4.40 apresenta a série original da variação de MP_{10} e NO_x na estação de Cubatão – Centro para os primeiros dois meses de 2018. Os dados obtidos por meio destas séries temporais foram utilizados para a comparação da eficiência da previsão e não foram inseridos na série de treino (*out of sample*). Objetivando a reprodução do mesmo

comportamento do conjunto de dados presentes na Figura 4.40, as Figura 4.41 e 4.42 apresentam os resultados das simulações *out of sample* para o MP_{10} e NOx utilizando o algoritmo *k-NN*, variando o número de vizinhos próximos (k) de 2 a 10 e com dimensão de imersão igual ao valor gerado pelo algoritmo de Grassberger e Procaccia (1983). Os gráficos comparativos das séries previstas com as originais podem ser encontrados no Apêndice D, Figuras D8 e D9.

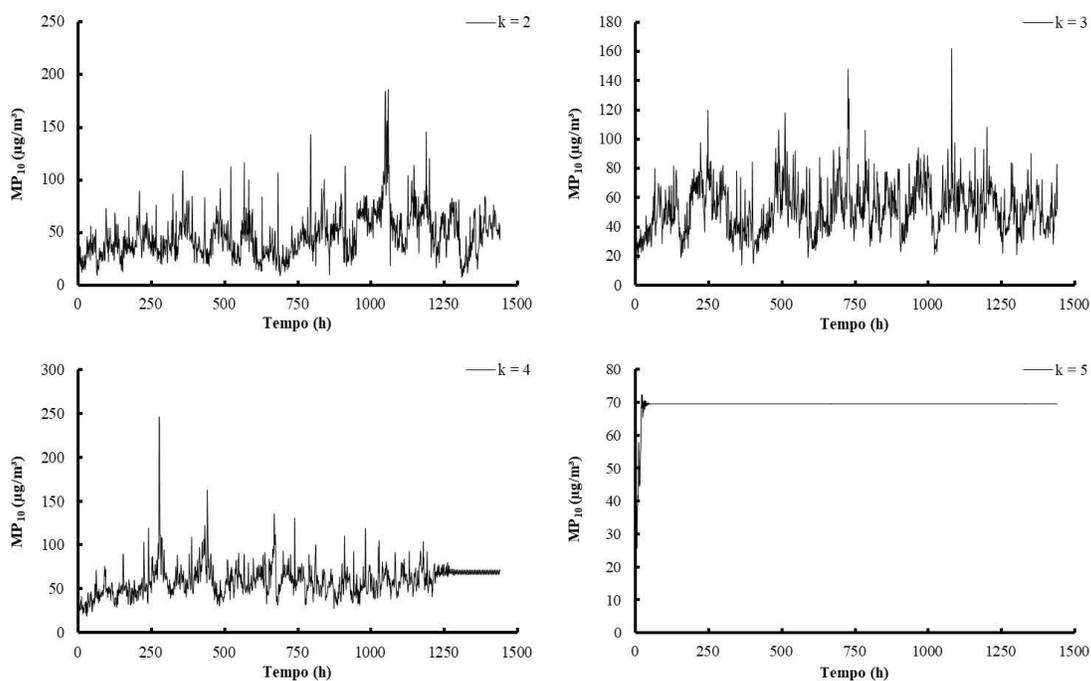
Figura 4.40 – Séries temporais da concentração de MP_{10} e NOx obtidas na estação de monitoramento de Cubatão para os meses de janeiro e fevereiro de 2018 (o autor, 2018).



Iniciando pela análise das previsões do MP_{10} para os primeiros dois meses de 2018, verifica-se que o valor de k é um importante parâmetro na qualidade das previsões, pois direta, tanto quali quanto quantitativamente na série gerada. Os valores de $k = 2$ e $k = 3$ foram os únicos que permitiram a geração de uma série completa de 1440 pontos, já para $k = 4$, $k = 6$ e $k = 7$ são verificadas séries com problemas de convergência a partir de, aproximadamente, respectivamente, 1100, 500 e 450 pontos. Para os demais valores de vizinhos próximos, a saber, 5, 8, 9 e 10, são observadas convergências para um ponto médio ou variação harmônica a partir do dado relativo à 24^a hora.

Considerando as previsões realizada para o NO_x (Figura 4.42), séries completas (1440 pontos) foram obtidas para $k = 2$ e 3. Para $k = 4$ são observados 1200 pontos viáveis, para $k = 5$, 900 pontos, $k = 7$ observa-se o erro de convergência para previsões com mais de 250 pontos e para $k = 8$ verifica-se uma variação harmônica a partir do ponto 325. Nas demais situações foram verificados problemas de convergência para simulações com mais de 24 pontos.

Figura 4.41 – Previsão *out of sample* da concentração horária de MP_{10} nos primeiros dois meses de 2018 na estação de Cubatão - Centro utilizando o algoritmo k -NN (o autor, 2018).



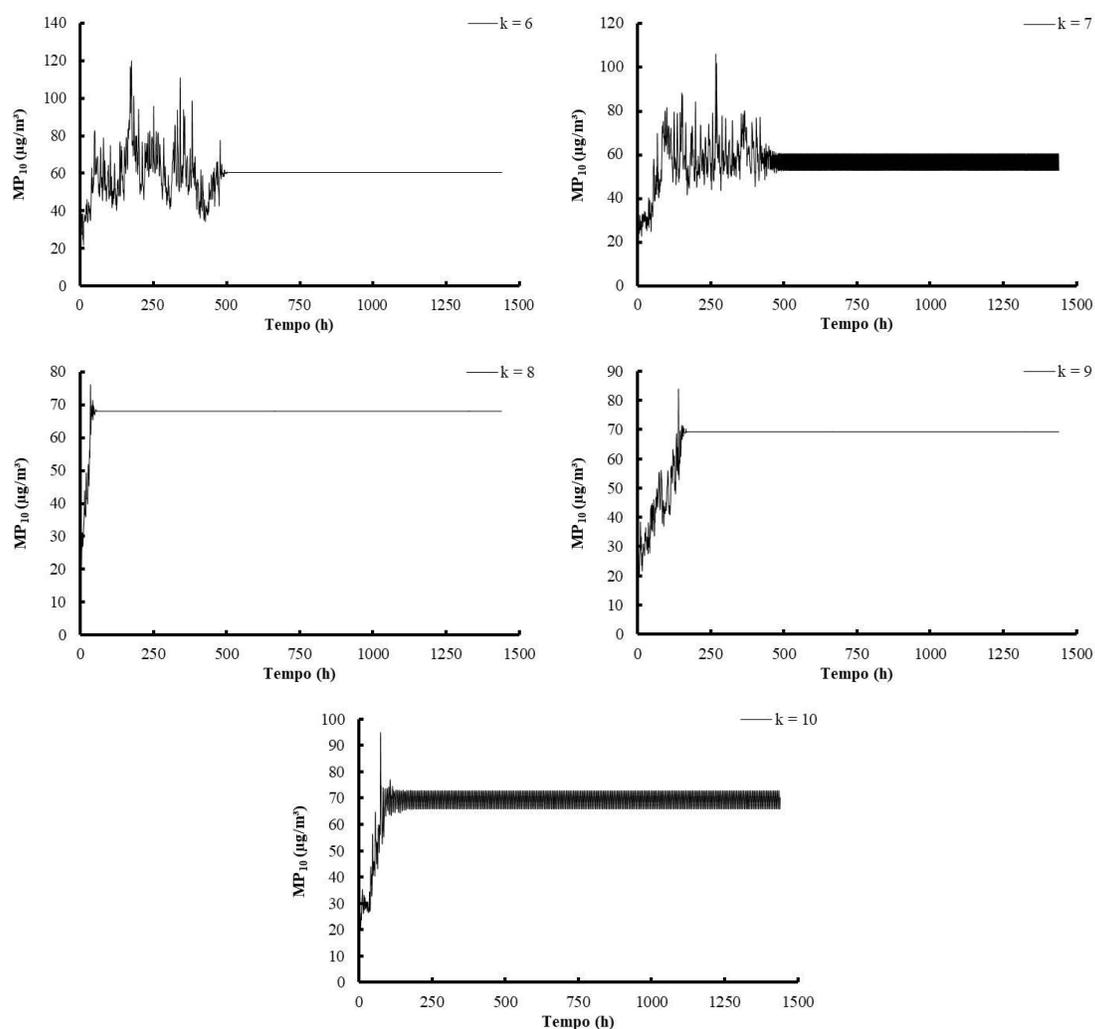
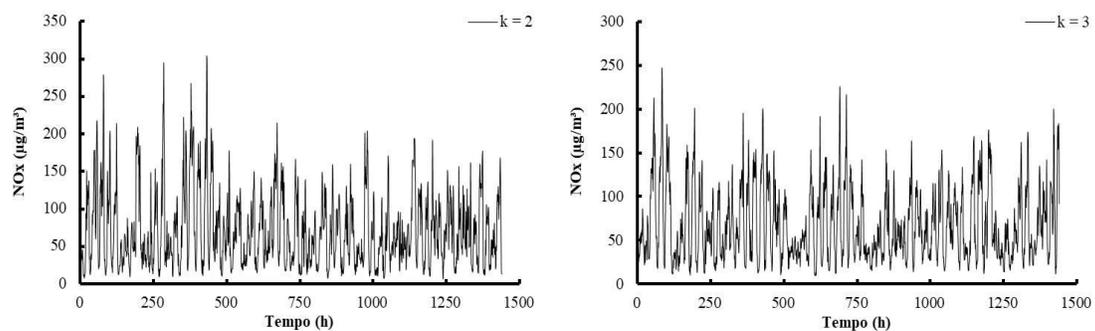
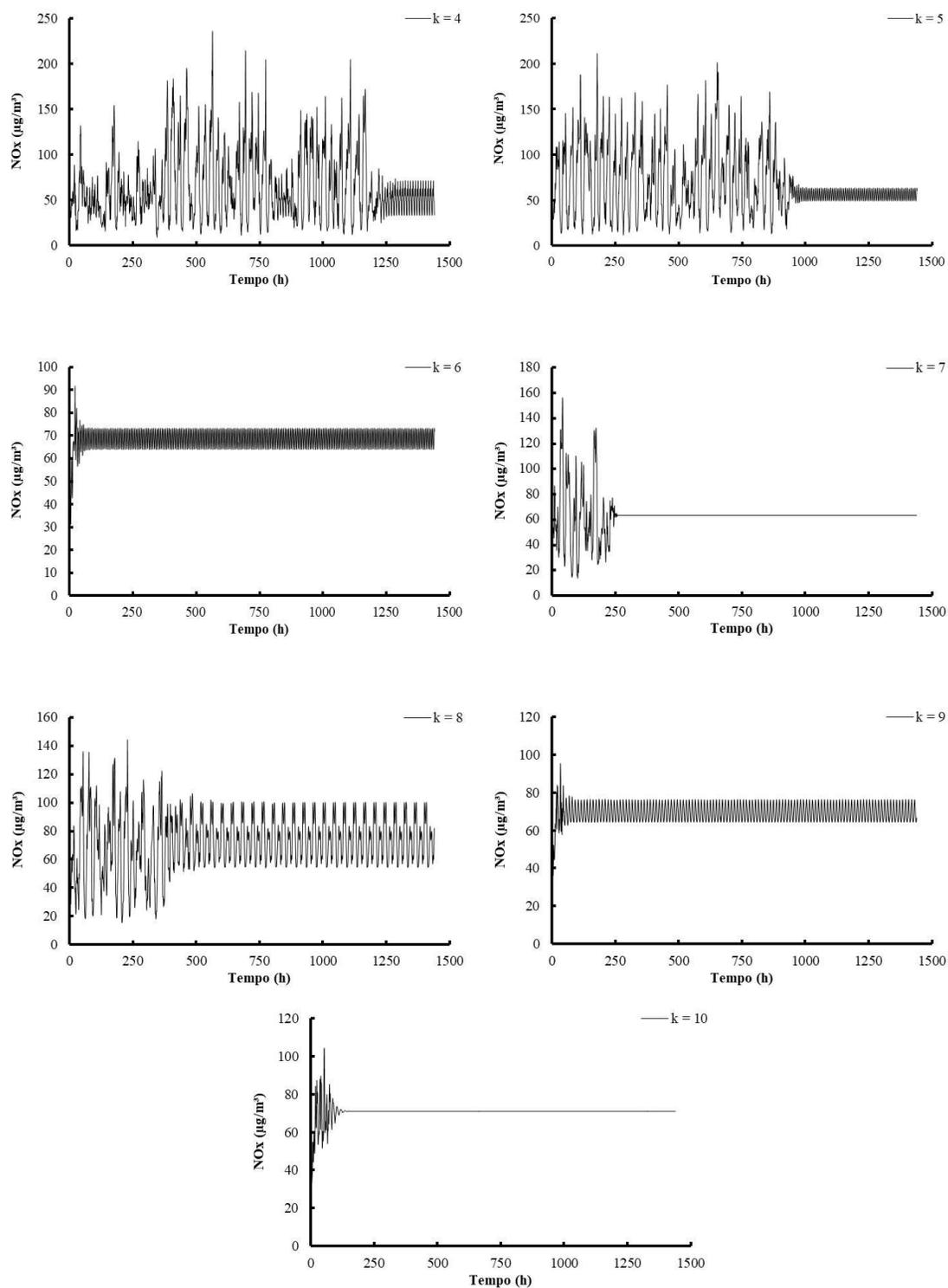


Figura 4.42 – Previsão *out of sample* da concentração horária de NO_x nos primeiros dois meses de 2018 na estação de Cubatão - Centro utilizando o algoritmo k -NN (o autor, 2018).





A última etapa desta seção é a avaliação quantitativa das previsões. Neste sentido, as Tabelas 4.18 e 4.19 apresentam a comparação dos dados reais com as previsões realizadas para 60 dias (1440 pontos), 30 dias (720 pontos), 15 dias (360 pontos), 7 dias

(168 pontos), 2 dias (48 pontos) e 1 dia (24 pontos). Análogo ao realizado para todas as outras estações, os critérios adotados para a análise das séries previstas foram o desvio da média real em relação à média do conjunto simulado e também os métodos do Erro Quadrático Médio (RMSE) e o Erro Absoluto Médio (MAE). Vale ressaltar, novamente, que os valores das Tabelas 4.18 e 4.19 devem ser analisados mediante consideração do que foi apresentado nas Figuras 4.41 e 4.42 para que as conclusões estejam, assim, baseadas na real fenomenologia descrita pela série avaliada.

Desta forma, iniciando pelo MP_{10} , verifica-se que os valores de desvio, RMSE e MAE são muito altos para quaisquer outras simulações maiores que 24 horas e que a melhor situação ocorreu para $k = 2$ (média prevista de $23,2 \mu\text{g}/\text{m}^3$ versus $22,46 \mu\text{g}/\text{m}^3$ apresentada pela real) sendo o desvio de 3%, $RMSE = 15,8$ e $MAE = 0,8 \mu\text{g}/\text{m}^3$. Considerando as características altamente dispersivas, variáveis, com fenomenologia sazonal subjacentemente caótica, este pode ser considerado um resultado muito bom.

Avaliando os resultados para o NO_x (Tabela 4.19), extrai-se que para $k = 3$, considerando uma previsão de 60 dias o desvio da média foi de 23%, $RMSE = 69,0$ e $MAE = 16,3 \mu\text{g}/\text{m}^3$. Para o mesmo valor de k , agora considerando uma média de 30 dias, o desvio foi de 42%, $RMSE = 74,0$ e $MAE = 27,5 \mu\text{g}/\text{m}^3$, para 15 dias o desvio foi 60%, $RMSE = 61,5$ e $MAE = 26,4 \mu\text{g}/\text{m}^3$, 1 semana (desvio = 76%, $RMSE = 70,4$ e $MAE = 31,7 \mu\text{g}/\text{m}^3$), 2 dias (desvio = 88%, $RMSE = 42,3$ e $MAE = 24,8 \mu\text{g}/\text{m}^3$) e, por fim, para 1 dia, desvio = 89%, $RMSE = 35,7$ e $MAE = 27,7 \mu\text{g}/\text{m}^3$. Via de regra, tem-se verificado que a diminuição do número de dias simulados implica em aumento do desvio, isso ocorre pelo fato de que quanto mais dias previstos maior será o número de valores para compor a média e, portanto, ocorre uma suavização do valor da média acarretando em menores erros. O NO_x para a estação de Cubatão – Centro foi uma das simulações que apresentou pior desempenho. Como já comentado anteriormente, a qualidade das previsões quando utilizada a metodologia k -NN está fortemente relacionada à série de treino. Como o k -NN é um método de aprendizado de máquina (parte da premissa de que a série reproduz seu próprio comportamento ao longo do tempo) a observação futura está fundamentalmente associada às anteriores (Parmezan e Batista, 2016), sendo assim, é evidente a sua relação

com a série de treino e, portanto, problemas no treino gerarão consequentemente erros nas previsões.

A falta de dados ao final das séries de treino é a principal causa da baixa eficiência do método em algumas situações, pois ao agrupar os dados, passa-se a juntar períodos diferentes, com características físicas e fenomenológicas diferentes, acarretando por fim em erros de aprendizado. Ainda assim, embora qualitativamente distante do valor real, o método se mostrou robusto o suficiente para a geração de séries temporais de variação da concentração horária de M_{10} e NO_x nas estações consideradas.

Com o objetivo de tornar genéricos os valores de vizinhos próximos (k) e dimensão de imersão I , de modo a fazer com que fique mais direta a aplicação do método k - NN , a seção a seguir traz a consideração do impacto da fixação de k e I generalizados na qualidade dos dados gerados, considerando as previsões de 24 horas.

Tabela 4.18 - Comparação entre as séries simuladas e reais para a concentração de MP_{10} – Cubatão - Centro (o autor, 2018).

	Média (60 dias) ($\mu\text{g}/\text{m}^3$)	Desvio	Média (30 dias) ($\mu\text{g}/\text{m}^3$)	Desvio	Média (15 dias) ($\mu\text{g}/\text{m}^3$)	Desvio	Média (1 semana) ($\mu\text{g}/\text{m}^3$)	Desvio	Média (2 dias) ($\mu\text{g}/\text{m}^3$)	Desvio	Média (1 dia) ($\mu\text{g}/\text{m}^3$)	Desvio
Real	19,66	-	19,98	-	19,39	-	20,32	-	19,23	-	22,46	-
k												
2	47,30	141%	39,20	96%	38,67	99%	33,57	65%	28,90	50%	23,21	3%
3	52,73	168%	50,47	153%	50,44	160%	42,97	112%	31,21	62%	28,54	27%
4	60,05*	205%	58,19	191%	52,96	173%	42,88	111%	31,80	65%	30,35	35%
5	69,20*	252%	68,86*	245%	68,17*	251%	66,59*	228%	59,20*	208%	49,23	119%
6	59,91*	205%	59,38*	197%	60,78*	213%	53,89	165%	40,82	112%	34,02	51%
7	57,20*	191%	56,87*	185%	55,32*	185%	50,69	150%	29,63	54%	27,83	24%
8	67,36*	243%	66,66*	234%	55,42*	186%	62,04*	205%	47,05*	145%	33,23	48%
9	66,59*	239%	63,95*	220%	55,52*	186%	46,55	129%	30,04	56%	26,41	18%
10	67,82*	245%	66,18*	231%	55,62*	187%	55,32	172%	30,55	58%	27,10	21%
	60 dias		30 dias		15 dias		1 semana		2 dias		1 dia	
k	RMSE ($\mu\text{g}/\text{m}^3$)	MAE ($\mu\text{g}/\text{m}^3$)	RMSE ($\mu\text{g}/\text{m}^3$)	MAE ($\mu\text{g}/\text{m}^3$)	RMSE ($\mu\text{g}/\text{m}^3$)	MAE ($\mu\text{g}/\text{m}^3$)	RMSE ($\mu\text{g}/\text{m}^3$)	MAE ($\mu\text{g}/\text{m}^3$)	RMSE ($\mu\text{g}/\text{m}^3$)	MAE ($\mu\text{g}/\text{m}^3$)	RMSE ($\mu\text{g}/\text{m}^3$)	MAE ($\mu\text{g}/\text{m}^3$)
2	37,8	27,6	28,1	19,2	27,5	193,3	22,4	13,3	20,5	9,7	15,8	0,8
3	39,5	33,1	37,1	30,5	37,0	31,0	28,8	22,7	20,2	11,9	18,6	6,0
4	46,1	40,4	45,5	38,3	42,8	33,6	29,5	22,6	20,5	12,6	19,5	7,9
5	51,6	49,5	51,0	48,9	51,2	48,8	49,9	46,3	47,3	40,0	39,5	26,7
6	43,6	40,2	43,6	39,5	46,2	41,4	39,5	33,6	30,2	21,6	22,7	11,5
7	40,8	37,5	40,7	36,9	41,0	35,9	37,1	30,4	19,5	10,4	20,3	5,3
8	50,0	47,7	49,1	46,7	48,9	45,9	46,5	41,7	38,3	27,8	25,8	10,8
9	49,7	46,9	47,4	44,0	44,2	39,3	32,9	26,2	20,8	10,8	21,0	3,9
10	50,7	48,2	49,2	46,3	47,7	43,5	41,5	35,0	22,0	11,2	21,2	4,7

*valores desconsiderados na análise devido à presença de convergência harmônica ou para um valor médio.

Tabela 4.19 - Comparação entre as séries simuladas e reais para a concentração de NOx – Cubatão - Centro (o autor, 2018).

	Média (60 dias) ($\mu\text{g}/\text{m}^3$)	Desvio	Média (30 dias) ($\mu\text{g}/\text{m}^3$)	Desvio	Média (15 dias) ($\mu\text{g}/\text{m}^3$)	Desvio	Média (1 semana) ($\mu\text{g}/\text{m}^3$)	Desvio	Média (2 dias) ($\mu\text{g}/\text{m}^3$)	Desvio	Média (1 dia) ($\mu\text{g}/\text{m}^3$)	Desvio
Real	55,16	-	49,52	-	43,78	-	41,58	-	27,97	-	24,27	-
k												
2	71,47	30%	77,01	56%	73,08	67%	74,06	78%	63,91	129%	46,02	90%
3	67,65	23%	70,08	42%	70,16	60%	73,27	76%	52,71	88%	46,03	89%
4	63,73*	16%	65,45	32%	53,26	22%	50,35	21%	54,24	94%	50,01	106%
5	66,50*	21%	73,46	48%	75,92	73%	72,92	75%	64,33	130%	63,52	162%
6	68,55*	24%	68,32*	38%	67,90*	55%	66,86*	61%	62,32	123%	56,24	131%
7	63,25*	15%	62,95*	27%	62,33*	42%	63,21	52%	71,00	154%	54,82	126%
8	72,00*	31%	69,31*	40%	62,40*	43%	63,70	53%	55,55	98%	50,61	109%
9	69,43*	26%	69,24*	40%	62,49*	43%	67,85*	63%	63,71	128%	57,52	137%
10	70,65*	28%	70,35*	42%	62,54*	43%	68,34	64%	61,50	120%	53,85	122%
	60 dias		30 dias		15 dias		1 semana		2 dias		1 dia	
<i>k</i>	RMSE ($\mu\text{g}/\text{m}^3$)	MAE ($\mu\text{g}/\text{m}^3$)	RMSE ($\mu\text{g}/\text{m}^3$)	MAE ($\mu\text{g}/\text{m}^3$)	RMSE ($\mu\text{g}/\text{m}^3$)	MAE ($\mu\text{g}/\text{m}^3$)	RMSE ($\mu\text{g}/\text{m}^3$)	MAE ($\mu\text{g}/\text{m}^3$)	RMSE ($\mu\text{g}/\text{m}^3$)	MAE ($\mu\text{g}/\text{m}^3$)	RMSE ($\mu\text{g}/\text{m}^3$)	MAE ($\mu\text{g}/\text{m}^3$)
2	69,0	16,3	74,0	27,5	71,5	29,3	66,4	32,5	57,4	36,0	36,1	21,8
3	62,7	12,5	61,4	20,6	61,5	26,4	70,4	31,7	42,3	24,8	35,7	21,7
4	56,9	8,6	55,6	15,9	39,7	9,5	37,7	8,8	41,8	26,2	34,1	25,7
5	58,2	11,3	62,8	24,0	65,4	32,1	63,3	31,3	49,7	36,3	49,5	39,2
6	46,5	13,4	43,8	18,8	40,5	24,1	38,8	25,3	38,1	34,3	36,9	31,9
7	46,8	8,1	45,0	13,4	44,4	18,6	45,5	21,6	56,4	43,0	37,2	30,6
8	50,7	16,8	50,1	19,8	49,2	20,8	44,9	22,1	42,0	27,5	37,9	26,4
9	47,0	14,3	44,4	19,7	41,0	25,0	39,9	26,3	39,6	35,7	38,3	33,2
10	47,1	15,5	44,6	20,9	41,5	26,0	39,7	26,8	37,2	33,6	33,8	29,5

*valores desconsiderados na análise devido à presença de convergência harmônica ou para um valor médio.

4.6.6 Generalização de k (vizinhos próximos) e I (dimensão de imersão)

O objetivo desta seção é o de apresentar um valor k e I generalizados para que não seja necessário um gasto computacional intenso como aquela dispendido nesta Tese. Evidentemente que estas generalizações teriam aplicação em um estudo puramente preliminar de modo a auxiliar a tomada de decisões no que tange as direções que serão adotadas. Inicialmente foi realizada a generalização do valor de k . O procedimento para tal foi simples, todas as previsões para as cinco estações e dois poluentes foram observadas e das tabelas foram extraídos os valores de k que forneceram os melhores resultados para a média de 24 horas para as quatro melhores situações de desvio, RMSE e MAE. Assim, o valor de k que mais se repetiu foi escolhido. A Tabela 4.20 apresenta, de forma ilustrativa, o valor de k que forneceu a melhor situação para a média de 24 horas. Lembrando que a escolha do melhor k generalizado se baseou na análise das quatro melhores situações e não apenas na primeira melhor como demonstrado na Tabela 4.20.

É fácil notar que o valor que mais se repetiu foi $k = 2$, porém ao imputar esse valor às demais estações os erros aumentam muito, tornando inviável essa generalização, seguindo com a análise, quando se considera todas as previsões apresentadas no estudo das seções anteriores, verifica-se que o $k = 3$ foi aquele que mais vezes se repetiu entre as melhores previsões. Sendo assim, nesta Tese, sugere-se o valor $k = 3$ como genérico e representativo das estações e poluentes aqui estudados. A Tabela 4.21 apresenta os resultados para o valor de k generalizado igual a 3 mantendo a dimensão de imersão igual àquela fornecida pela metodologia de Grassberger e Procaccia (1983). Verifica-se que para algumas estações os valores de desvio ficaram em faixas aceitáveis, como, por exemplo, Centro MP₁₀ (25%), Congonhas MP₁₀ (23%), Cubatão MP₁₀ (27%) e Ibirapuera MP₁₀ (21%). As demais situações apresentaram desvios que variaram de 33 a 89%, sendo a pior situação Cubatão NO_x. Torna-se claro, ao se comparar a Tabela 4.20 e 4.21 que a generalização de k prejudica sobremaneira a qualidade das previsões, tanto no que diz respeito ao desvio quanto ao RMSE e MAE. Como comentado anteriormente, esta é uma opção a ser aplicada em um estudo preliminar para auxiliar a tomada de decisão.

Tabela 4.20 - Melhor valor de k para cada uma das estações e poluentes (o autor, 2018).

Estação	Poluente	Treino	Previsto	I	Melhor k	Média real ($\mu\text{g}/\text{m}^3$)	Média prevista ($\mu\text{g}/\text{m}^3$)	Desvio (%)	RMSE ($\mu\text{g}/\text{m}^3$)	MAE ($\mu\text{g}/\text{m}^3$)
Centro	MP ₁₀	1998 - 2009	2010	13	10	34,3	32,9	4	39,4	1,5
	NO _x	1998 - 2005	2006	13	5	70,2	108,7	55	41,7	38,5
Congonhas	MP ₁₀	1998 - 2017	2018	13	6	12,8	11,4	11	7,8	1,8
	NO _x	1998 - 2017	2018	16	10	80,5	92,4	15	37,9	11,9
Cubatão	MP ₁₀	1998 - 2017	2018	9	2	22,5	23,2	3	15,8	0,8
	NO _x	1998 - 2017	2018	14	2	24,3	46,0	89	36,1	21,8
Guarulhos	MP ₁₀	1998 - 2008	2009	10	2	33,6	25,4	24	47,3	8,2
	NO _x	-	-	-	-	-	-	-	-	-
Ibirapuera	MP ₁₀	1998 - 2013	2014	13	4	18	15,9	12	16,5	2,1
	NO _x	1998 - 2017	2018	19	8	9,3	8,4	10	6,5	0,9

Tabela 4.21 - Previsões para $k = 3$ mantendo a dimensão de imersão de Grassberger e Procaccia (1983) (o autor, 2018).

Estação	Poluente	Treino	Previsto	Média Real ($\mu\text{g}/\text{m}^3$)	Média Prevista ($\mu\text{g}/\text{m}^3$)	Desvio (%)	RMSE ($\mu\text{g}/\text{m}^3$)	MAE ($\mu\text{g}/\text{m}^3$)
Centro	MP ₁₀	1998 - 2009	2010	34,3	25,6	25	37,6	8,7
	NO _x	1998 - 2005	2006	70,2	129,8	85	102,5	59,6
Congonhas	MP ₁₀	1998 - 2017	2018	12,8	15,7	23	10,7	2,4
	NO _x	1998 - 2017	2018	80,5	117,0	45	70,0	36,6
Cubatão	MP ₁₀	1998 - 2017	2018	22,5	28,5	27	18,6	6,0
	NO _x	1998 - 2017	2018	24,3	46,0	89	35,7	21,7
Guarulhos	MP ₁₀	1998 - 2008	2009	33,6	44,6	33	53,0	11,0
	NO _x	-	-	-	-	-	-	-
Ibirapuera	MP ₁₀	1998 - 2013	2014	18,0	21,8	21	18,8	3,7
	NO _x	1998 - 2017	2018	9,3	x	x	x	x

Seguindo com as propostas de generalizações, uma vez determinado o valor de $k = 3$, passa-se a escolha do valor de dimensão de imersão. Desta vez não foi aplicada a otimização utilizada por meio do algoritmo de Grassberger e Procaccia (1983). A dimensão de imersão generalizada foi determinada por meio da média (Equação 3.1) dos I de cada estação/poluinte. Sendo assim, sugere-se $I = 13$.

A Tabela 4.22 apresenta a análise das previsões com $I = 13$ e $k = 3$. Neste cenário, verifica-se que, com exceção da estação de Congonhas NOx, as demais apresentaram desvios e, em geral RMSE/MAE, maiores do que as séries com I otimizados por meio do algoritmo de Grassberger e Procaccia (1983). No caso de Congonhas NOx, cuja dimensão de Grassberger e Procaccia (1983) era 16, com o valor de $I = 13$ apresentou uma redução do desvio de 45% para 31%. Isso demonstra que nem sempre a dimensão obtida por meio do Grassberger e Procaccia (1983) é a melhor. Em linhas gerais, as previsões generalizadas apresentaram menor confiança que as particularizadas, pois, conforme observado, os valores de I e k exercem influência direta na qualidade dos dados gerados pelo código k -NN. Ainda assim, considerando-se as características do fenômeno e conhecendo a importância de previsões como estas, alguns dos resultados podem ser considerados muito bons. Conforme se verifica, tanto na Tabela 4.21 quanto 4.22, na linha referente à estação do Ibirapuera NOx, não existem dados disponíveis. Isso ocorreu pelo fato de ao se imputar o valor de $k = 3$ o algoritmo k -NN não gerou uma série temporal, apenas uma linha constante convergente para um valor médio desde o primeiro ponto.

Tabela 4.22 - Previsões para $k = 3$ e $I = 13$ (o autor, 2018).

Estação	Poluente	Treino	Previsto	Média Real ($\mu\text{g}/\text{m}^3$)	Média Prevista ($\mu\text{g}/\text{m}^3$)	Desvio (%)	RMSE ($\mu\text{g}/\text{m}^3$)	MAE ($\mu\text{g}/\text{m}^3$)
Centro	MP ₁₀	1998 - 2009	2010	34,3	25,6	25	37,6	8,7
	NO _x	1998 - 2005	2006	70,2	129,8	85	102,5	59,6
Congonhas	MP ₁₀	1998 - 2017	2018	12,8	15,7	23	10,7	2,4
	NO _x	1998 - 2017	2018	80,5	105,3	31	60,9	24,9
Cubatão	MP ₁₀	1998 - 2017	2018	22,5	24,2	8	19,7	1,8
	NO _x	1998 - 2017	2018	24,3	67,9	180	57,7	43,6
Guarulhos	MP ₁₀	1998 - 2008	2009	33,6	59,2	76	53,1	25,5
	NO _x	-	-	-	-	-	-	-
Ibirapuera	MP ₁₀	1998 - 2013	2014	18,0	21,8	21	18,8	3,7
	NO _x	1998 - 2017	2018	9,3	x	x	x	x

4.6.7 Proposta de melhoria das séries de treino

Conforme foi observado durante a discussão de todas as previsões geradas anteriormente, as lacunas relativas aos dados inválidos presentes nas séries temporais foram apontada como uma das principais causas da baixa eficiência de algumas previsões. Para a aplicação de ambas as metodologias, Caos e k -NN, é absolutamente necessário que as séries sejam contínuas, sendo assim, os espaços vazios foram lançados ao final da evolução temporal. Ao fazer isso os dados são todos agrupados e importantes características fenomenológicas, como a sazonalidade, são perdidas fazendo com que o a metodologia k -NN “aprenda errado” e gere, por fim, previsões equivocadas.

Neste sentido, esta seção tem por objetivo apresentar uma proposta de melhoria das séries de treino. O objetivo é preencher as lacunas presentes com o próprio código k -NN e, a partir desta nova série de treino completa, realizar previsões dos primeiros dois meses de 2018. Para tal os dados precedentes a cada lacuna foram utilizados como série treino e cada espaço vazio foi admitido como uma nova previsão. A metodologia k -NN foi eleita para o preenchimento dos dados faltantes, pois apresentou melhor eficiência que método ARIMA (Apêndice B). Uma explicação detalhada do procedimento adotado nesta etapa pode ser encontrada no Apêndice C.

Duas situações foram submetidas a este procedimento, Congonhas – MP₁₀ e Cubatão – NO_x. As Figuras 4.43 e 4.44 trazem as séries temporais de concentração horária de MP₁₀ (estação de Congonhas) e NO_x (estação de Cubatão) dos anos de 1998 a 2017.

Para o caso de Congonhas – MP₁₀, a série original apresentada na Figura 4.43 conta com 155534 pontos válidos, ao passo que deveriam ser 175320, sendo assim, verifica-se a presença de 19786 lacunas. A média calculada por meio da Equação 3.1 foi 45 ± 32 ($\mu\text{g}/\text{m}^3$) com $I = 13$. Já para Cubatão – NO_x, o mesmo número de pontos válidos deveria ser observado na série (175320), contudo a evolução apresenta 42903 dados inválidos. A média para a série original foi de 74 ± 67 $\mu\text{g}/\text{m}^3$ ($\mu\text{g}/\text{m}^3$) com $I = 14$.

Figura 4.43 – Série de treino (com lacunas) da concentração horária de MP_{10} de 1998 a 2017 na estação de Congonhas (o autor, 2018).

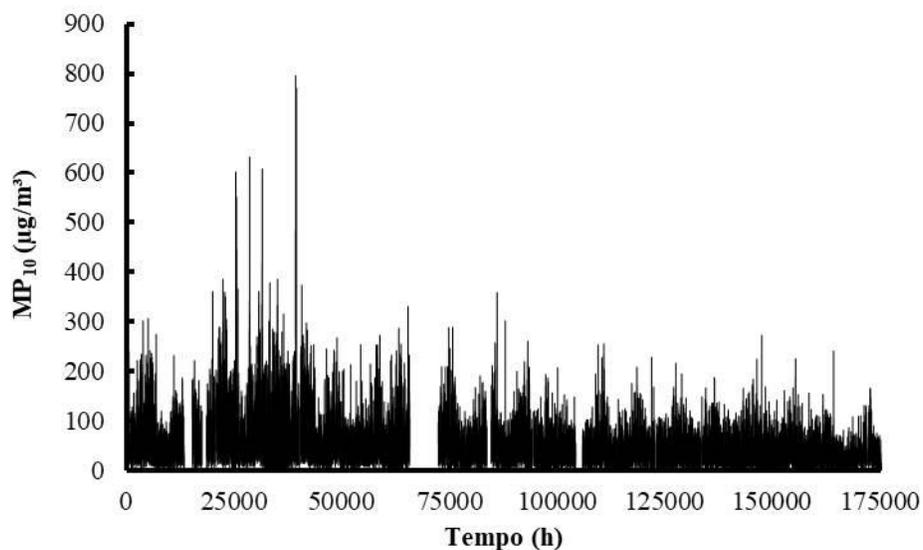
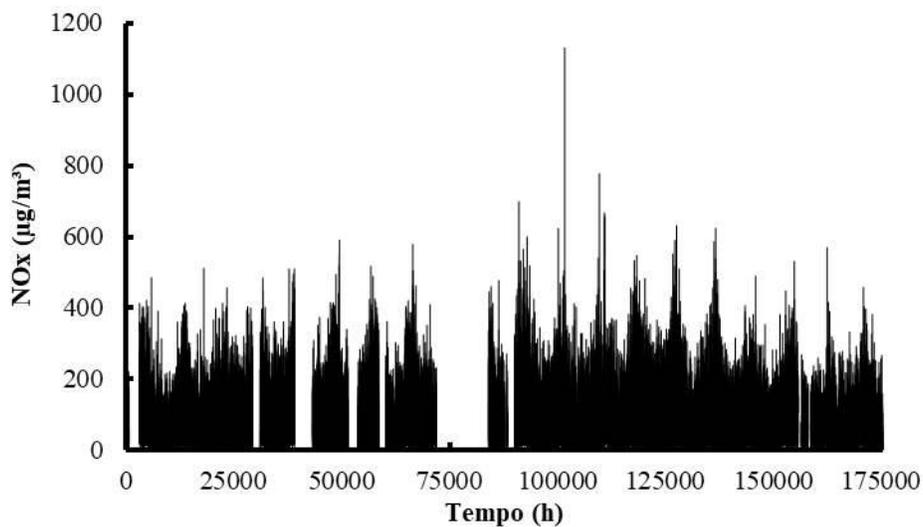


Figura 4.44 – Série de treino (com lacunas) da concentração horária de NO_x de 1998 a 2017 na estação de Cubatão (o autor, 2018).



As Figuras 4.45 e 4.46 apresentam, respectivamente, as séries de MP_{10} (Congonhas) e NO_x (Cubatão) com lacunas preenchidas por meio da metodologia k - NN .

Figura 4.45 – Série de treino da concentração horária de MP_{10} de 1998 a 2017 na estação de Congonhas, com lacunas preenchidas pelo método k - NN (o autor, 2018).

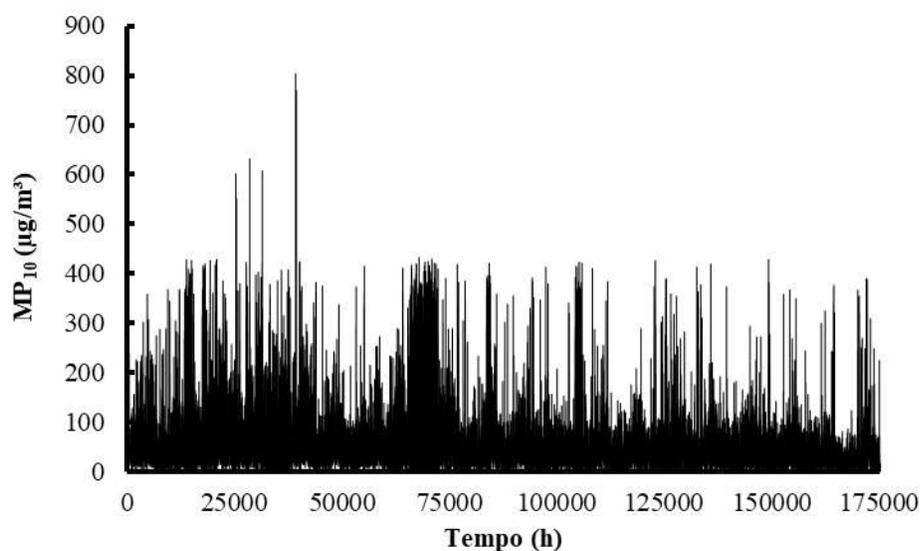
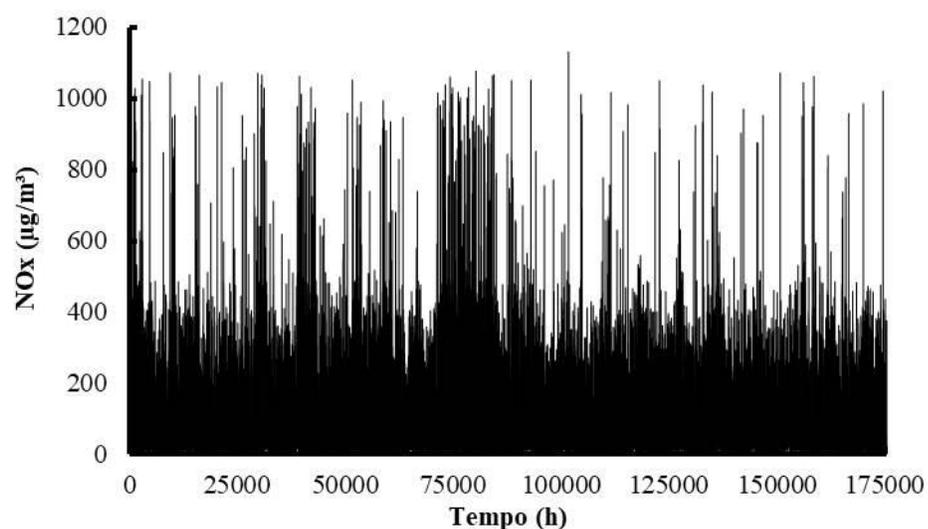


Figura 4.46 – Série de treino da concentração horária de NO_x de 1998 a 2017 na estação de Cubatão, com lacunas preenchidas pelo método k - NN (o autor, 2018).



As Figuras 4.47 e 4.48 são, respectivamente, os atratores reconstruídos para a série de treino concentração anual média de MP₁₀ (estação de Congonhas) e NO_x (estação de Cubatão) com as lacunas preenchidas pela metodologia *k-NN*. Na sequência, observa-se a Tabela 4.23 que detalha a análise do Caos.

A nova média obtida para a estação de Congonhas (MP₁₀) foi de 48 ± 40 ($\mu\text{g}/\text{m}^3$) com $I = 10$. No caso da média, o preenchimento promoveu um aumento de 6,67% e uma redução nos valores da dimensão de imersão, de dimensão de correlação, entropia de Kolmogorov. Isso ocorre pelo fato de que as interpolações, orientadas pelo algoritmo *k-NN*, tendem a apresentar uma evolução mais determinística, reduzindo, assim, a complexidade do sistema. Para o NO_x (Cubatão), a média encontrada foi 93 ± 87 ($\mu\text{g}/\text{m}^3$), exibindo, quando comparada à série original, um aumento de 25,68%. Já em relação à dimensão de imersão, a nova evolução temporal apresentou $I = 9$ e reduções nos valores de K_{ML} e D_2 novamente demonstrando que as interpolações por meio do algoritmo *k-NN*, mais determinísticas que o fenômeno real, promoveram uma queda na complexidade do sistema.

Uma constatação interessante é observada no estudo dos graus de liberdade (GL). Antes da interpolação, ambas as séries apresentavam 5 graus de liberdade e este número foi reduzido para 4 após o preenchimento das lacunas. Torna-se evidente que o quinto grau de liberdade, bastante comentado anteriormente como superestimado pelo arredondamento, é a própria presença de incertezas ocasionadas pelos vazios. O procedimento de agrupar os dados, de modo a tornar a série contínua, implica na junção de períodos sem conexão fenomenológica, acarretando na quebra de ciclos meteorológicos, fortes influenciadores da dinâmica da poluição atmosférica, e, por fim, aumentando o caráter caótico das séries.

O procedimento de preenchimento das lacunas é atividade bastante dispendiosa no que diz respeito ao tempo computacional. Para a série de Congonhas (MP₁₀) o tempo total foi de 108 horas, ao passo que para Cubatão (NO_x), o gasto foi de 221 horas.

Figura 4.47 – Atrator reconstruído a partir da série de treino da concentração horária de MP_{10} , com lacunas preenchidas por meio da metodologia k - NN , para os anos de 1998 a 2017 (Congonhas) (o autor, 2018).

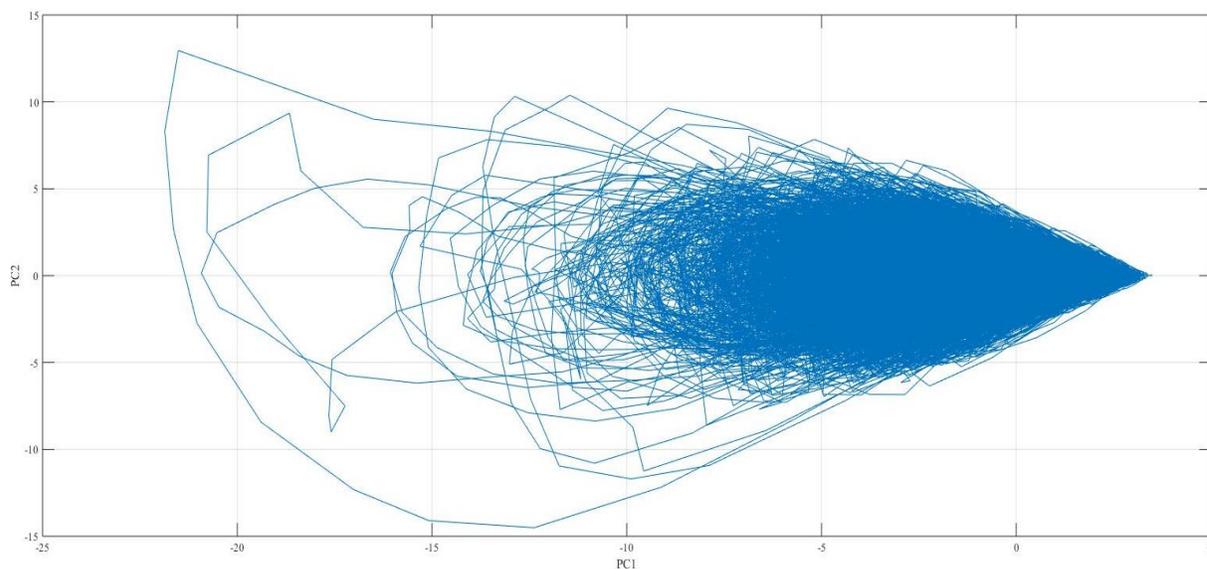


Figura 4.48 – Atrator reconstruído a partir da série de treino da concentração horária de NO_x , com lacunas preenchidas por meio da metodologia k - NN , para os anos de 1998 a 2017 (Cubatão - Centro) (o autor, 2018).

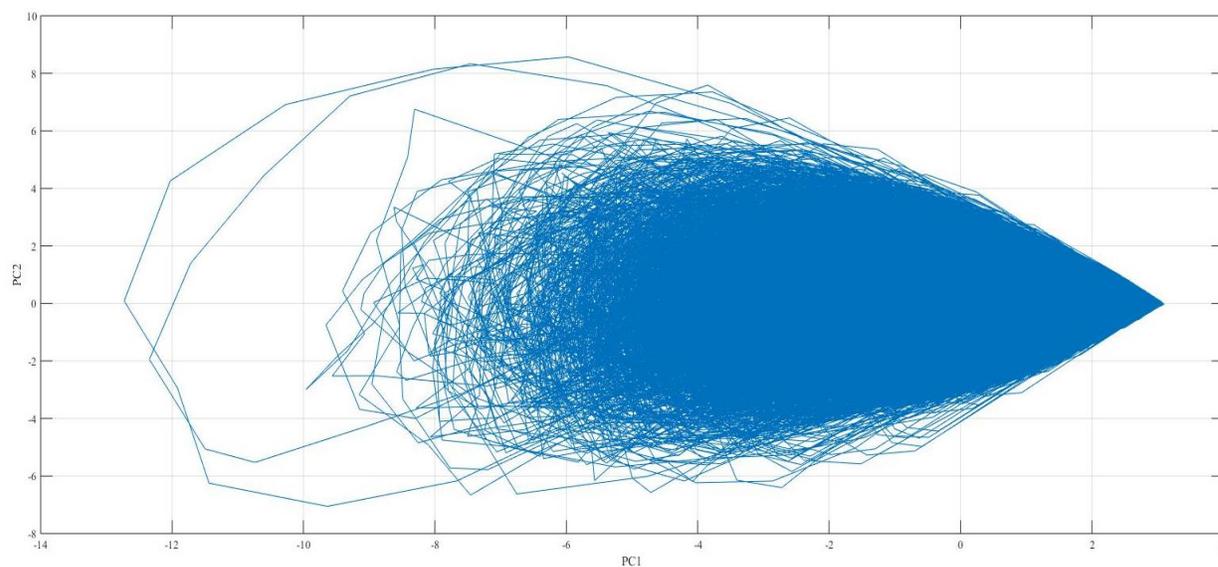


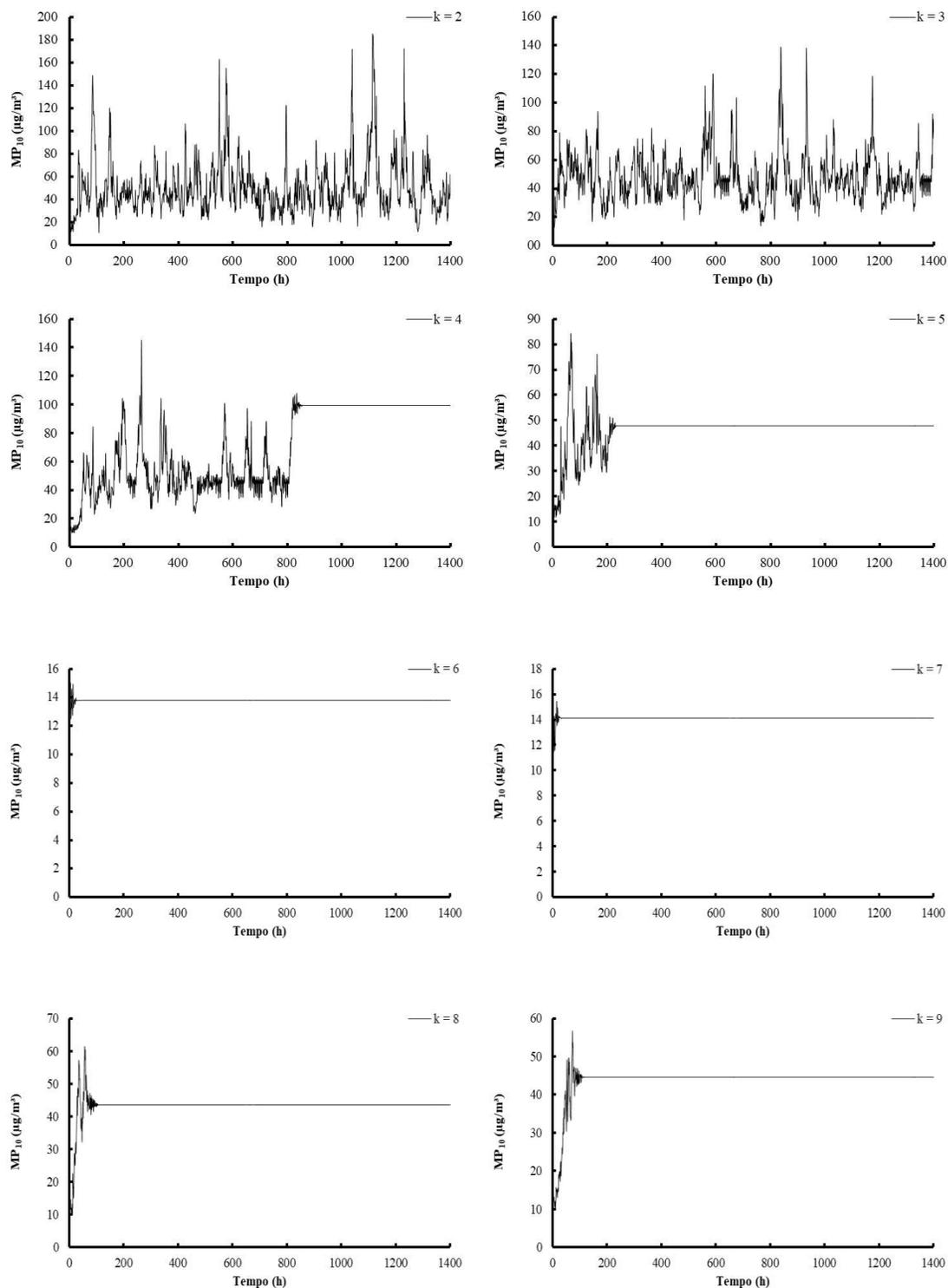
Tabela 4.23 – *Análise do Caos para as séries de treino originais e com preenchimento por meio da metodologia k -NN para as estações Congonhas (MP_{10}) e Cubatão (NO_x) (o autor, 2018).*

	Séries originais		Séries com preenchimento k -NN	
	Congonhas MP_{10}	Cubatão NO_x	Congonhas MP_{10}	Cubatão NO_x
I [-]	13	14	10	9
D_2 [-]	4,97	4,92	3,35	3,85
GL [-]	5	5	4	4
K_{ML} (bits/s)	0,1083	0,1201	0,0901	0,0956
λ [-]	0,15	0,16	0,19	0,22

Considerando as previsões, as Figuras 4.49 e 4.50 apresentam o estudo da variação de k (de 2 a 10) para, respectivamente, as novas séries de concentração horária de MP_{10} (Congonhas) e NO_x (Cubatão).

Para as previsões de MP_{10} (Congonhas), verifica-se que os valores de $k = 2$ e $k = 3$ promoveram séries completas (1440 pontos). Já para $k = 4$ observa-se convergência acima de 800 horas de previsão, para $k = 5$ a convergência é notada acima do ponto 200. Nos demais casos, nota-se apenas a geração de séries com no máximo 24 pontos. Observando a Figura 4.50 (NO_x – Cubatão), séries completas (1440 pontos) foram obtidas para os valores 2, 3, 4, 5 e 6 de vizinhos próximos. Problemas de convergência foram reportados acima de 1000, 500, 250 e 600 pontos para, respectivamente, $k = 7$, $k = 8$, $k = 9$ e $k = 10$. Comparando a Figura 4.42 com a Figura 4.50, é possível notar, qualitativamente, que o preenchimento da série de treino de NO_x (Cubatão) fez com que ocorressem menos problemas de convergência, sendo, inclusive, possível obter mais séries inteiras. Ao avaliar as Figuras 4.25 e 4.49, verifica-se, de forma mais discreta que o observado para o NO_x (Cubatão), que a nova série promoveu, também, a ocorrência de menos convergências para valores médios e variações harmônicas. Sendo assim, até este ponto, de forma qualitativa, é possível afirmar que o procedimento de preenchimento das lacunas das séries de treino acarretou em melhora significativa nas previsões.

Figura 4.49 – Previsão *out of sample* da concentração horária de MP_{10} nos primeiros dois meses de 2018 na estação de Congonhas utilizando o algoritmo k -NN e série de treino interpolada (o autor, 2018).



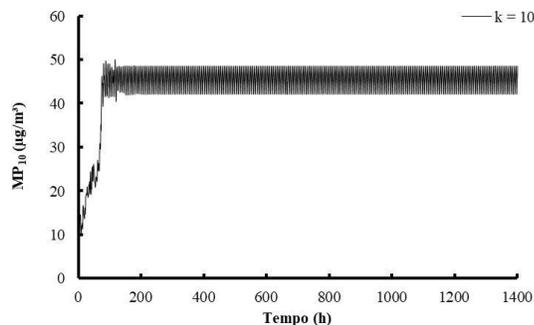
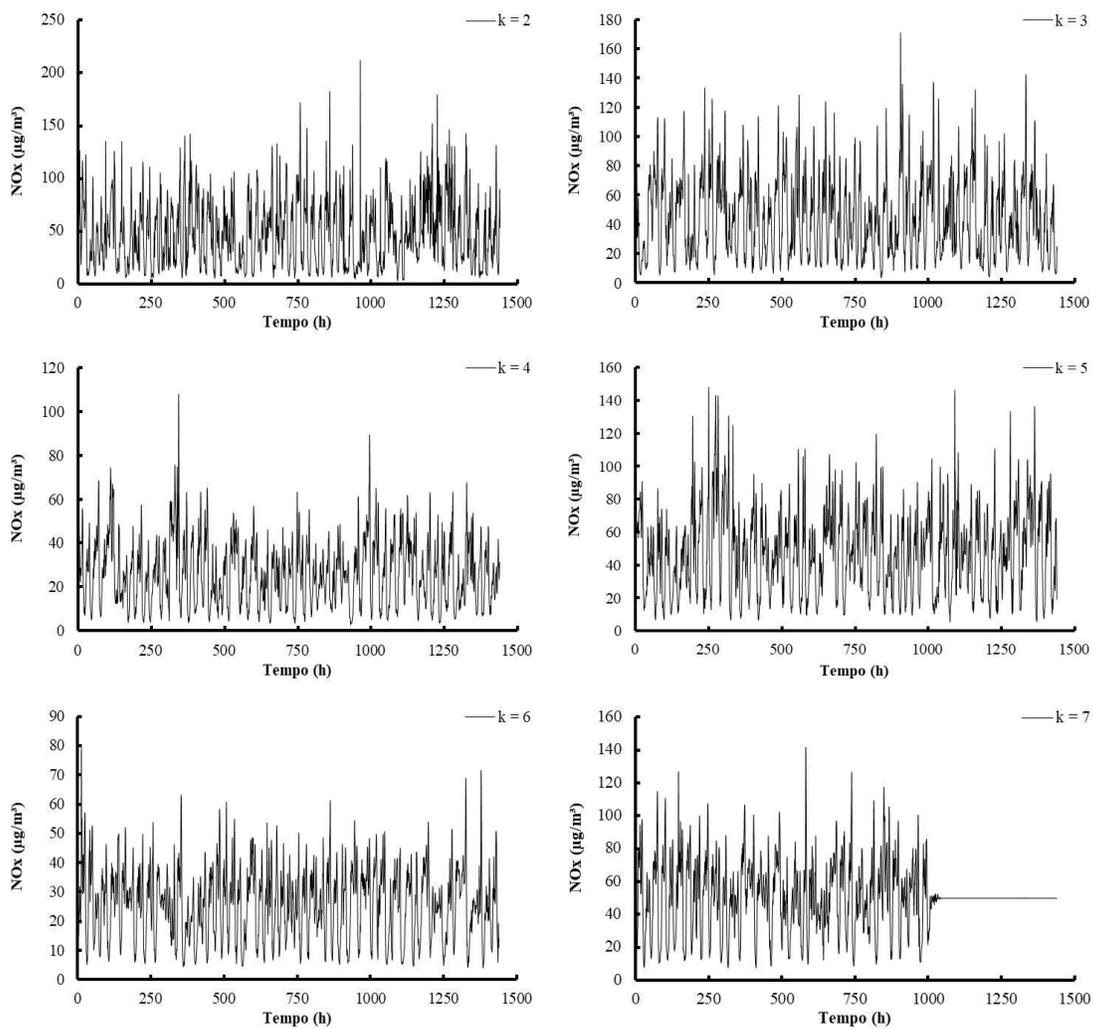
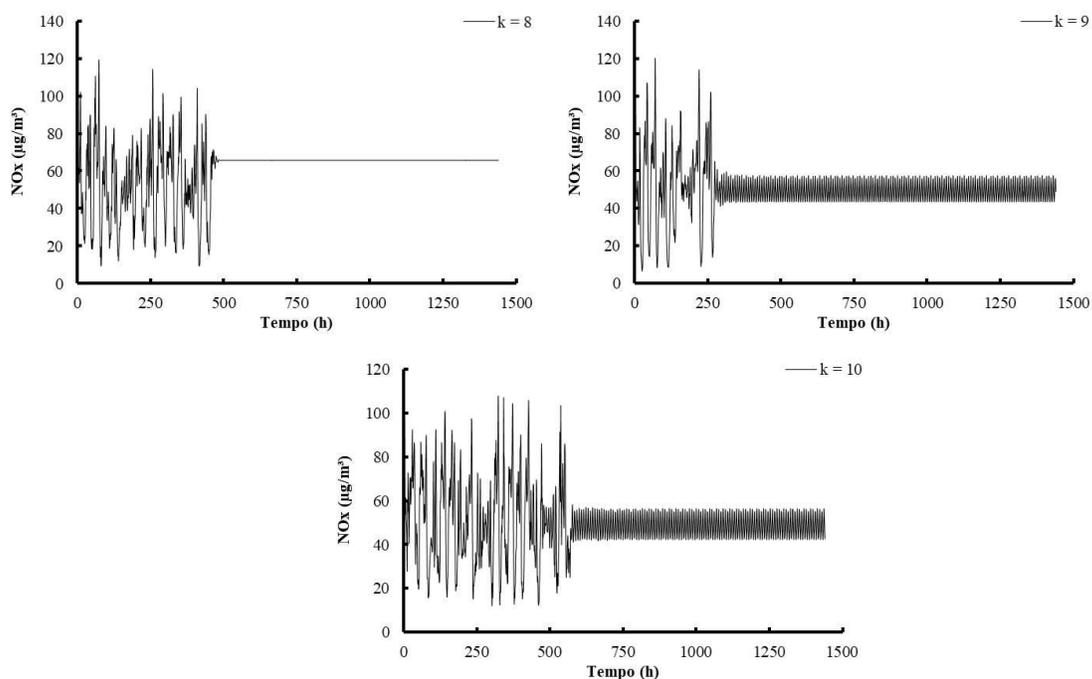


Figura 4.50 – Previsão *out of sample* da concentração horária de NOx nos primeiros dois meses de 2018 na estação de Cubatão utilizando o algoritmo *k-NN* e série de treino interpolada (o autor, 2018).





Objetivando corroborar esta constatação, as Tabelas 4.24 e 4.25 trazem, respectivamente, para o MP_{10} (Congonhas) e NO_x (Cubatão), a comparação dos dados reais com as previsões realizadas para 60 dias (1440 pontos), 30 dias (720 pontos), 15 dias (360 pontos), 7 dias (168 pontos), 2 dias (48 pontos) e 1 dia (24 pontos). Similar ao adotado anteriormente, a qualidade das previsões foi avaliada por meio do desvio da média real em relação à média do conjunto simulado e por meio o método da Raiz do Erro Quadrático Médio (RMSE) e o Erro Absoluto Médio (MAE). Ressalta-se, novamente, que as análises das Tabelas 4.23 e 4.24 precisam ser realizadas a luz das observações das Figuras 4.49 e 4.50 de modo a não se obter uma conclusão equivocada ao se considerar somente os valores das tabelas. É imperativo, nestes casos, verificar os limites para os quais o algoritmo convergiu de forma lógica para uma série temporal. A guisa de exemplo, não é possível asseverar com uma boa previsão o valor de desvio de 34% para o MP_{10} Congonhas (previsão de 60 dias), pois o algoritmo só foi capaz de gerar a uma série até o 24º ponto, logo, o resultado destacado não tem sentido físico e não será considerado nesta análise. Iniciando pela Tabela 4.24, verifica-se que para os casos em que se obteve séries completas (1440 pontos), $k = 2$ e $k = 3$, os valores dos desvios bem como RMSE e MAE, para todos os intervalos de previsão, se apresentaram muito altos. Quando se realiza a comparação entre a melhor previsão para o MP_{10} (Congonhas) antes do preenchimento da série,

verifica-se que esta ocorreu para $k = 6$ (média de 24 horas) com desvio de 11% $RMSE = 7,8$ e $MAE = 1,8 \mu g/m^3$, já para a nova série (comparando o valor de $k = 6 - 24$ horas), o desvio foi 5% $RMSE = 8,5$ e $MAE = 0,3 \mu g/m^3$. Extraindo a melhor situação de previsão utilizando a nova série de treino, verifica-se $k = 4$ (24 horas) com desvio de 2% $RMSE = 8,0$ e $MAE = 0,7 \mu g/m^3$. Passando à análise do NOx, praticamente para todas as situações de previsão o algoritmo k -NN, utilizando como treino a nova série preenchida, geraram dados com valores de desvio RMSE e MAE substancialmente menores que aqueles obtidos com a série sem preenchimento. Foram observados baixos desvios para casos que antes não podiam nem se quer ser considerados devido aos erros de convergência. Especificando, as melhor situações notadas para o NOx Cubatão utilizando a série de treino original foram para $k = 3$, em que se considerando uma previsão de 60 dias o desvio da média foi de 23%, $RMSE = 69,0$ e $MAE = 16,3 \mu g/m^3$. Ainda no mesmo valor de k , considerando uma média de 30 dias, o desvio foi de 42%, $RMSE = 74,0$ e $MAE = 27,5 \mu g/m^3$, para 15 dias o desvio foi 60%, $RMSE = 61,5$ e $MAE = 26,4 \mu g/m^3$, 1 semana (desvio = 76%, $RMSE = 70,4$ e $MAE = 31,7 \mu g/m^3$), 2 dias (desvio = 88%, $RMSE = 42,3$ e $MAE = 24,8 \mu g/m^3$) e, por fim, para 1 dia, desvio = 89%, $RMSE = 35,7$ e $MAE = 27,7 \mu g/m^3$. Já para as mesmas previsões e mesmo valor de k com série de treino preenchido, tem-se 60 (desvio da média de 13%, $RMSE = 52,4$ e $MAE = 10,6 \mu g/m^3$, 30 dias (desvio de 7%, $RMSE = 43,4$ e $MAE = 3,5 \mu g/m^3$), 15 dias (desvio foi 6%, $RMSE = 40,2$ e $MAE = 2,6 \mu g/m^3$), 1 semana (desvio = 13%, $RMSE = 28,3$ e $MAE = 5,5 \mu g/m^3$), 2 dias (desvio = 0%, $RMSE = 26,3$ e $MAE = 0,0 \mu g/m^3$) e para 1 dia (desvio = 9%, $RMSE = 26,3$ e $MAE = 2,2 \mu g/m^3$).

Tabela 4.24 - Comparação entre as séries simuladas e reais para a concentração de MP_{10} – Congonhas, utilizando a série de treino interpolada (o autor, 2018).

	Média (60 dias) ($\mu\text{g}/\text{m}^3$)	Desvio	Média (30 dias) ($\mu\text{g}/\text{m}^3$)	Desvio	Média (15 dias) ($\mu\text{g}/\text{m}^3$)	Desvio	Média (1 semana) ($\mu\text{g}/\text{m}^3$)	Desvio	Média (2 dias) ($\mu\text{g}/\text{m}^3$)	Desvio	Média (1 dia) ($\mu\text{g}/\text{m}^3$)	Desvio
Real	21,27	-	21,06	-	19,71	-	17,15	-	12,73	-	12,80	-
k												
2	49,90	135%	50,19	138%	47,42	141%	49,85	191%	32,95	159%	18,54	45%
3	47,47	123%	47,80	127%	46,21	134%	49,30	187%	38,59	203%	30,47	139%
4	70,80*	233%	49,02	133%	49,18	150%	36,48	113%	17,38	37%	12,52	2%
5	46,55*	119%	45,27*	115%	42,70*	117%	39,09	128%	21,47	69%	15,13	19%
6	13,79*	35%	13,78*	35%	13,77*	30%	13,74*	20%	13,60*	7%	13,42	5%
7	14,11*	34%	14,09*	33%	14,06*	29%	13,99*	18%	13,66*	7%	13,21	4%
8	43,16*	103%	42,79*	103%	14,07*	29%	40,35*	135%	30,27	138%	17,75	39%
9	43,74*	106%	42,88*	104%	14,08*	29%	37,28*	117%	20,64	62%	13,59	6%
10	44,40*	109%	43,12*	105%	14,10*	28%	34,71*	102%	17,71	39%	13,92	9%
	60 dias		30 dias		15 dias		1 semana		2 dias		1 dia	
k	RMSE ($\mu\text{g}/\text{m}^3$)	MAE ($\mu\text{g}/\text{m}^3$)	RMSE ($\mu\text{g}/\text{m}^3$)	MAE ($\mu\text{g}/\text{m}^3$)	RMSE ($\mu\text{g}/\text{m}^3$)	MAE ($\mu\text{g}/\text{m}^3$)	RMSE ($\mu\text{g}/\text{m}^3$)	MAE ($\mu\text{g}/\text{m}^3$)	RMSE ($\mu\text{g}/\text{m}^3$)	MAE ($\mu\text{g}/\text{m}^3$)	RMSE ($\mu\text{g}/\text{m}^3$)	MAE ($\mu\text{g}/\text{m}^3$)
2	38,7	28,6	37,7	29,1	36,9	27,8	44,3	32,8	29,7	19,7	11,7	5,1
3	32,6	26,2	32,6	26,8	31,6	26,5	36,6	32,1	31,6	25,5	23,8	15,3
4	57,4	49,5	34,1	27,9	37,1	29,6	25,4	19,1	11,3	3,7	8,0	0,7
5	27,7	25,3	27,3	24,2	27,4	23,0	28,9	21,9	14,4	8,4	9,7	1,9
6	12,3	7,5	12,1	7,3	11,2	5,9	9,0	3,4	6,9	0,7	8,5	0,3
7	12,1	7,2	11,9	7,0	11,0	5,6	8,9	3,2	7,1	0,7	8,8	0,0
8	24,1	21,9	24,1	21,7	25,0	22,4	26,3	23,2	23,9	17,2	12,6	4,0
9	24,7	22,5	24,2	21,8	24,3	21,5	23,6	20,1	13,9	7,3	9,0	0,2
10	25,4	23,1	24,6	22,1	24,0	20,9	21,5	17,5	9,8	4,6	9,1	0,6

*valores desconsiderados na análise devido à presença de convergência harmônica ou para um valor médio.

Tabela 4.25 - Comparação entre as séries simuladas e reais para a concentração de NO_x – Cubatão, utilizando a série de treino interpolada (o autor, 2018).

	Média (60 dias) (µg/m ³)		Desvio		Média (30dias) (µg/m ³)		Desvio		Média (15 dias) (µg/m ³)		Desvio		Média (1 semana) (µg/m ³)		Desvio		Média (2 dias) (µg/m ³)		Desvio		Média (1 dia) (µg/m ³)		Desvio		
Real	55,16	-	49,52	-	43,78	-	41,58	-	27,97	-	24,27	-													
k																									
2	48,19	13%	47,09	5%	45,30	3%	46,00	11%	50,16	79%	67,56	178%													
3	44,51	19%	45,99	7%	46,33	6%	47,08	13%	28,00	0%	26,50	9%													
4	26,32	52%	26,45	47%	28,10	36%	29,39	29%	24,25	13%	25,75	6%													
5	47,63	14%	47,25	5%	49,20	12%	40,02	4%	48,05	72%	65,52	170%													
6	25,66	53%	25,66	48%	26,93	38%	28,14	32%	31,26	12%	33,44	38%													
7	50,28*	9%	50,19	1%	48,62	11%	48,97	18%	47,00	68%	58,38	141%													
8	61,03*	11%	56,42*	14%	48,70	11%	49,45	19%	53,42	91%	48,72	101%													
9	49,73*	10%	50,08*	1%	48,76*	11%	49,08	18%	49,01	75%	42,87	77%													
10	49,25*	11%	49,84*	1%	48,81	12%	53,72	29%	56,18	101%	52,62	117%													
	60 dias		30 dias		15 dias		1 semana		2 dias		1 dia														
k	RMSE (µg/m³)	MAE (µg/m³)	RMSE (µg/m³)	MAE (µg/m³)	RMSE (µg/m³)	MAE (µg/m³)	RMSE (µg/m³)	MAE (µg/m³)	RMSE (µg/m³)	MAE (µg/m³)	RMSE (µg/m³)	MAE (µg/m³)	RMSE (µg/m³)	MAE (µg/m³)	RMSE (µg/m³)	MAE (µg/m³)	RMSE (µg/m³)	MAE (µg/m³)	RMSE (µg/m³)	MAE (µg/m³)	RMSE (µg/m³)	MAE (µg/m³)	RMSE (µg/m³)	MAE (µg/m³)	
2	53,5	7,0	51,6	2,4	45,5	1,5	45,2	4,4	43,4	22,2	54,9	43,3													
3	52,4	10,6	43,4	3,5	40,2	2,6	28,3	5,5	26,3	0,0	26,3	2,2													
4	55,6	28,8	47,8	23,1	40,2	15,7	34,5	12,2	23,1	3,7	25,8	1,5													
5	50,7	7,5	48,1	2,3	45,2	5,4	39,7	1,6	35,0	20,1	46,3	41,2													
6	54,8	29,5	47,4	23,9	38,5	16,8	34,3	13,4	21,3	3,3	22,5	9,2													
7	48,0	4,9	44,6	0,7	41,2	4,8	35,6	7,4	36,5	19,0	45,7	34,1													
8	46,2	5,9	43,7	6,9	39,5	8,7	35,3	7,9	38,8	25,4	40,7	24,4													
9	46,5	5,4	42,7	0,6	40,1	7,0	40,0	7,5	39,0	21,0	37,3	18,6													
10	45,8	5,9	41,8	0,3	37,1	6,4	36,0	12,1	35,4	28,2	36,0	28,4													

*valores desconsiderados na análise devido à presença de convergência harmônica ou para um valor médio.

É bastante evidente que para os dois casos analisados as interpolações das séries de treino promoveram melhores previsões, corroborando a assertiva de que a falta de dados impacta decisiva e negativamente na qualidade das previsões. Conforme mencionado anteriormente a qualidade das previsões quando utilizada a metodologia *k-NN* está diretamente relacionada à série de treino. Devido ao fato de o *k-NN* ser um método de aprendizado de máquina e, portanto, parte da premissa de que a série reproduz seu próprio comportamento ao longo do tempo, a observação futura está fortemente associada às anteriores (Parmezan e Batista, 2016), logo, é fundamental a sua relação com a série de treino e, sendo assim, qualquer problema no treino causará como consequência, erros nas previsões.

CAPÍTULO V

CONCLUSÕES

A poluição atmosférica é um assunto em pauta e bastante recorrente no meio científico, pois é evidente a relação entre o aumento da concentração dos poluentes atmosféricos e a incidência de diversos problemas de saúde da população, tornando, assim, a poluição um grave problema de saúde pública. Nas grandes áreas urbanas, a crescente frota de veicular e as emissões de poluentes tóxicos pelas chaminés industriais provocam altas concentrações de substâncias nocivas que são responsáveis pela baixa visibilidade e diversos problemas respiratórios nos seres vivos. Sendo assim, é evidente a necessidade de técnicas de análise da dispersão de poluentes atmosféricos bem como previsão. Nesta tendência, a presente Tese aplicou os invariantes caóticos entropia de Kolmogorov, dimensão de correlação e máximo expoente de Lyapunov a dados de variação temporal de concentração de MP_{10} e NO_x em cinco diferentes estações de monitoramento da CETESB com o objetivo de identificar características não-lineares. Realizou também a predição de séries temporais aplicando o método do vizinho mais próximo (k -NN).

Ao longo dos anos, o Caos determinístico está se solidificando como alternativa às metodologias tradicionais, como Fourier e DFA, quando o espectro de potências se apresenta de forma contínua sem a clara definição de uma frequência dominante. Nesta situação, a formulação de hipóteses simplificadoras prejudica o resultado final das análises tradicionais, sendo os fenômenos não lineares, sensíveis às condições e aperiódicos, portanto, passíveis da abordagem caótica. A avaliação da presença do Caos determinístico em séries temporais de concentração anual de NO_x e MP_{10} nas estações de monitoramento (CETESB) do Ibirapuera, Congonhas, Centro de São Paulo, Guarulhos e Cubatão, todas localizadas no Estado de São Paulo, demonstrou a natureza caótica dos sistemas, deixando claro que a poluição atmosférica apresenta, durante sua evolução temporal, atratores do tipo estranhos e, intuitivamente, uma dinâmica dissipativa, não linear, aperiódica a longo prazo,

determinística e extremamente sensível às condições iniciais. Por meio da aplicação do método da integral de correlação foi possível obter o número de variáveis que regem a dinâmica da poluição atmosférica para os poluentes avaliados, explicitando a presença de Caos nos dados estudados. As condições iniciais, informação vital na solução de sistemas dinâmicos, estão diretamente relacionadas aos graus de liberdade do fenômeno. Nos casos aqui estudados, tornou-se clara a presença de quatro graus de liberdade “fenomenológicos”, sendo o quinto identificado como sendo a própria indeterminação gerada pela presença de lacunas oriundas dos dados inválidos. Tais lacunas obrigam o agrupamento dos dados das séries, de modo a torná-la contínua, ocasionando a junção de períodos sem conexão, principalmente meteorológica, acarretando em quebra de ciclo e aumento de instabilidade, logo, mais Caos. A metodologia empregada nesta Tese descreve adequadamente, no que se refere aos poluentes considerados, a dinâmica da poluição do ar, indicando a sua aplicação no entendimento dos padrões, mecanismos e causas deste fenômeno.

No que se refere às previsões, observou-se que o método *k-NN* necessita de uma série de treino com a maior fidedignidade possível, pois seus fundamentos se alicerçam sob as características da referida evolução para que o método aprenda corretamente e, por fim, gere uma previsão adequada. Neste sentido, foram obtidas para algumas estações previsões muito boas, uma vez consideradas as características do fenômeno, até mesmo para 60 dias, mas a grande parte dos ensaios conduziu a bons desvios as previsões com alcance de 24 horas. Notadas as dificuldades do método em face da ausência de diversos dados, realizou-se o preenchimento das séries de treino (para as estações de Congonhas – MP_{10} e Cubatão – NO_x) por meio da própria metodologia *k-NN*. Os resultados corroboraram o postulado, a ausência dos dados prejudica sobremaneira as previsões. As novas séries, agora interpoladas e completas, acarretaram em previsões muito boas, bastante superiores àquelas que utilizaram como treino a série com dados inválidos agrupados, sendo, portanto, a sugestão desta Tese quando da ausência de informações. Assim, a metodologia *k-NN* pode ser aplicada satisfatoriamente no estudo preditivo das concentrações de MP_{10} e NO_x . Verificou-se também que a fidedignidade dos resultados é maior quando são avaliadas médias em períodos mais longos que a frequência de aquisição dos dados, como dias inteiros e/ou ainda semanas. Intuindo generalizar a aplicação da metodologia *k-NN* na

previsão da poluição atmosférica, esta Tese sugere, para estudos preliminares, que sejam aplicados $k = 3$ e $I = 13$.

Portanto, conclui-se que a junção do Caos determinístico, na identificação de padrões não lineares, com o algoritmo k - NN é uma excelente e viável alternativa às análises tradicionais no estudo preditivo da poluição atmosférica.

SUGESTÕES PARA TRABALHOS FUTUROS

A seguir, são apresentadas algumas sugestões para trabalhos futuros:

- Realizar o preenchimento, por meio da metodologia k - NN , de todas as séries temporais de cada ano/estação/poluinte e prosseguir com a análise do Caos e previsão;
- Testar outras metodologias de preenchimento de séries temporais;
- Ao realizar o preenchimento, avaliar o impacto dos preenchimentos por meio de um teste em branco.
- Aumentar o número de estações estudadas, abrangendo todas aquelas disponibilizadas pela CETESB;
- Estudar outros poluentes;
- Aperfeiçoar o estudo envolvendo os graus de liberdade e predição no que diz respeito a aplicação em políticas públicas, tendo em vista que o Caos se mostrou bastante adequado na identificação dos GL, assim como a predição;
- Aplicar outras metodologias de predição.

CAPÍTULO VI

REFERÊNCIAS BIBLIOGRÁFICAS

AMDUR, M. O; DOULL, J.; KLAASSEN, C. D. Casarett and Doull's toxicology: The basic science of poisons. 4th ed. Oxford: Pergamon Press; 1991.

AQI - Air Quality Index, Air Quality Forecasting: A Review of Federal Programs and Research Needs. Committee on Environment and Natural Resources - CENR, 2001.

BAI, L.; WANG, J.; MA, X.; LU, H. Air Pollution Forecasts: An Overview. *International Journal of Environmental Research and Public Health*, 15, 780; doi:10.3390/ijerph15040780, 2018.

BAI, Y.; LI, Y.; WANG, X.; XIE, J.; LI, C. Air pollutants concentrations forecasting using back propagation neural network based on wavelet decomposition with meteorological conditions. *Atmos. Pollut. Res.* 7, 557–566, 2016.

BORGES, A. S.; ANDRADE, F. M.; GUARDANI, R. Ground-level ozone prediction using a neural network model based on meteorological variables and applied to the metropolitan area of São Paulo. *International Journal of Environment and Pollution (IJEP)*, v. 49, No. 1/2, 2012.

BOX, G. E. P.; JENKINS, G. M.; REINSEL, G. C.; LJUNG, G. M. Time series analysis: forecasting and control. *Wiley Series in Probability and Statistics*. Wiley, New Jersey, United States of America, 5 edição, 2015.

BRAGA, A. L. F.; CONCEIÇÃO, G. M. S.; PEREIRA, L. A. A.; Kishi H.; PEREIRA, J. C. R.; ANDRADE, M. F.; et al. Air pollution and pediatric respiratory hospital admissions in Sao Paulo, Brazil. *J Environ Med.* 1999; 1:95-102.

BRAGA, A. L. F.; SALDIVA, P. H. N.; PEREIRA, L. A. A.; MENEZES, J. J. C.; CONCEIÇÃO, G. M. S.; LIN, C. A. Health effects of air pollution exposure on children and adolescents in São Paulo, Brazil. *Pediatr Pulmonol* 2001;31(2):106-13.

BRIONGOS, J.V; SOLER, J.G. Using free bed surface fluctuations in a 3-D fluidized bed for dynamic characterization. *AIChE Journal*, v. 50, pp. 3060-3067, 2004.

BROOMHEAD, D. S. and KING, G. P. Extracting qualitative dynamics from experimental data. *Physica D.*, v. 20, 217-236, 1986.

CAMELO, H. N.; LUCIO, P. S.; JUNIOR, J. B. V. L.; CARVALHO, P. C. Métodos de previsão de séries temporais e modelagem híbrida ambos aplicados em médias mensais de velocidades do vento para região nordeste do Brasil, *Revista Brasileira de Meteorologia*, v.32, n.4, 565 - 574, 2017.

Caracterização das Estações da Rede Automática de Monitoramento da Qualidade do Ar na RMSP - Estação Congonhas. Diretoria de Engenharia, Tecnologia e Qualidade Ambiental. Departamento de Tecnologia do Ar. Divisão de Tecnologia de Avaliação da Qualidade do Ar, CETESB, 2004.

CARVALHO, V. S. B.; FREITAS, E. D.; MARTINS, L. D.; MARTINS, J. A.; MAZZOLI, C. R.; ANDRADE, M. F. Air quality status and trends over the Metropolitan Area of São Paulo, Brazil as a result of emission control policies. *Environmental science & policy* 47 (2015) 68 – 79.

CARVALHO, V.; FREITAS, E.; MARTINS, L.; MARTINS, J.; MAZZOLI, C.; ANDRADE, M. Air quality status and trends over the Metropolitan Area of São Paulo, Brazil as a result of emission control policies . 47, pp. 68-79, 2015.

CASTILHO, G. J. Análise de caos em leito fluidizado circulante. Tese (Doutorado em Engenharia Química), Faculdade de Engenharia Química, Universidade Estadual de Campinas, Campinas-SP, 2011.

CASTILHO, G. J.; CREMASCO, M. A. Comparison of downer and riser flows in a circulating bed by means of optical fiber probe signals measurements. *Procedia Engineering*, v. 42, p. 326 – 333, 2012.

CATALANO, M.; GALATIOTO, F.; BELL, M.; NAMDEO, A.; BERGANTINO, A.S. Improving the prediction of air pollution peak episodes generated by urban transport networks. *Environ. Sci. Policy*, 60, 69–83, 2016.

CHEN, S. B.; XU, Y. L.; DING, C. H. Q.; LUO, B. A Nonnegative locally linear KNN model for image recognition. *Pattern Recognition*, v. 83, Pages 78-90, November, 2018.

CHEN, X. L.; WANG, P. H.; HAO, Y. S.; ZHAO, M. Evidential KNN-based condition monitoring and early warning method with applications in power plant. *Neurocomputing*, 2018.

CHERIF, F. Theoretical computation of Lyapunov exponents for almost periodic hamiltonian systems, *IAENG International Journal of Applied Mathematics*, 41:1, 2011.

CHOUDHURY, R.S. Stability conditions for the persistence, disruption and decay of two-dimensional dissipative three-mode patterns in moderately extended nonlinear systems and comparisons with simulations, in: L. Debnath, S.R. Choudhury (Eds.), *Nonlinear Instability Analysis, Advances in Fluid Mechanics*, Comp. MechanicsPubl., Southampton, pp. 43–91, 1997.

CORTES, P. L.; SOUZA, V. V. C.; CUEVAS, T. B. A Eficácia das Políticas Públicas na Redução das Emissões Veiculares na Cidade de São Paulo. XVII Encontro Internacional sobre Gestão Empresarial e Meio Ambiente (ENGEMA), 2015.

CREMASCO, M. A.; CASTILHO, G.J. ; ARAGON, J. M. ; Martin, L. . Experimental fluid dynamics study in a fluidized bed by deterministic chaos analysis. *Particulate Science and Technology*, v. 29, p. 179-196, 2011.

DAFILISA, P.M.; FRASCOLIA, F.; CADUSCHB, J.P.; LILEYA, T.J.D. Chaos and generalized multistability in a mesoscopic model of the electroencephalogram, *Physica D*, 238(13), pp. 1056–1060, 2009.

DALY, A.; ZANNETTI, P. 2007. An Introduction to Air Pollution – Definitions, Classifications, and History. Chapter 1 of AMBIENT AIR POLLUTION (P. Zannetti, D. Al-Ajmi, and S. Al-Rashied, Editors). Published by The Arab School for Science and Technology (ASST) (<http://www.arabschool.org.sy>) and The EnviroComp Institute (<http://www.envirocomp.org/>).

DAPPER, S. N.; SPOHR, C.; ZANINI, R. R. Poluição do ar como fator de risco para a saúde: uma revisão sistemática no estado de São Paulo. *Estudos Avançados* 30 (86), 2016.

DE SMITH, M. J. *Statistical analysis handbook*, the Winchelsea Press, Winchelsea, UK, 2015.

DERISIO, J. C. *Introdução ao controle de poluição ambiental*. São Paulo: CETESB; 1992.

DHANYA, C. T.; KUMAR, D. N. Multivariate nonlinear ensemble prediction of daily chaotic rainfall with climate inputs. *J. Hydrol.*, 2011, 403, 292-306, doi: 10.1016/j.jhydrol.2011.04.009, 2011.

DI, C.; WANG, T.; YANG, X.; LI, S. An improved Grassberger-Procaccia algorithm for analysis of climate system complexity. *Hydrology Earth System Sciences Discussions*, 2017.

DIVYA, T. L.; VIJAYALAKSHMI, M. N. Envisagation and analysis of air pollution caused by forest fire using machine learning algorithm. *International Research Journal of Engineering and Technology (IRJET)*, v. 2 Issue: 04, July, 2015.

DOCKERY, D. W.; POPE III, C. A. Acute respiratory effects of particulate air pollution. *Annu Rev Public Health*, 15:107-32, 1994.

DOCKERY, D.; POPE, C. A.; XU, X.; SPENGLER, J.; WARE, J.; FAY, M.; FERRIS, B.; SPEIZER, F. An association between air pollution and mortality in six U.S. cities. *New England Journal of Medicine*, 329, pp. 1753-1759, 1993.

DRAGOMIR, E. G. Air Quality Index Prediction using K-Nearest Neighbor technique. *Buletinul Universității Petrol – Gaze din Ploiești*, v. LXII, No. 1, p. 103 - 108, 2010.

DUCHIADE, M. P. Poluição do ar e doenças respiratórias: uma revisão. *Cad. Saúde Pública*, Set 1992, v. 8, no.3, p.311-330.

EL-GOHARY, A. Chaos and optimal control of cancer self-remission and tumor system steady states. *Chaos solitons and fractals*. v. 37 Issue: 5 Pages: 1305-1316, 2008.

ERBANO, G. H. Análise de séries de tempo financeiras. Uma aplicação da teoria do caos em finanças empíricas (Dissertação de Mestrado). Escola de Administração de Empresas de São Paulo, Fundação Getúlio Vargas, 72 páginas, 2004.

FENG, X.; LI, Q.; ZHU, Y.; HOU, J.; JIN, L.; WANG, J. Artificial neural networks forecasting of PM_{2.5} pollution using air mass trajectory based geographic model and wavelet transformation. *Atmos. Environ.* 107, 118–128, 2015.

FENG, Y.; ZHANG, W.; SUN, D.; ZHANG, L. Ozone concentration forecast method based on genetic algorithm optimized back propagation neural networks and support vector machine data classification. *Atmos. Environ.*, 45, 1979–1985, 2011.

FERRARA, N. F.; PRADO, C. P. C. *Caos uma introdução*. São Paulo: Edgard Blücher LTDA, 1994.

FRASER, A. M.; SWINNEY, H. L. Independent coordinates for strange attractors from mutual information. *Phys. Rev. A* 33, 1134-1140. 1986.

FU, M.; WANG, W.; Le, Z.; KHORRAM, M.S. Prediction of particular matter concentrations by developed feed-forward neural network with rolling mechanism and gray model. *Neural Comput. Appl.* 26, 1789–1797, 2015.

GLEICK, J. *Chaos: Making a new science*. Viking, New York, 1987.

GRASSBERGER, P.; PROCACCIA, I. Measuring the Strangeness of Strange Attractors. In *Physica D*, v. 9, 189-208, 1983.

GRELL, G.A.; SCHMITZ, P.R.; MCKEEN, S.A.; FROST, G.; SKAMAROCK, W.C.; EDER, B. Fully coupled “online” chemistry within the WRF model. *Atmos. Environ.* 39, 6957–6975, 2005.

GUO, H., 2011. Source apportionment of volatile organic compounds in Hong Kong homes. *Build Environ.* 46, 2280 a 2286.

GUPTA, V.; MITTAL, M. KNN and PCA classifier with Autoregressive modelling during different ECG signal interpretation. *Procedia Computer Science*, v. 125, Pages 18-24, 2018.

GUTIÉRREZ, J. L. O conceito de caos no mundo antigo. IV Colóquio de Filosofia Física e Matemática “Determinismo e Caos”, *Revista Primus Vitam* Nº 2, 2011.

HAN, D.; YANG, H.; WANG, X. Efficiency of the plate-number-based traffic rationing in general networks. *Transportation Research Part E: Logistics and Transportation Review* , 46, 1095-1110, 2010.

HARIKRISHNAN, K. P.; JACOB,R.; MISRA, R.; AMBIKA, G. Determining the minimum embedding dimension for state space reconstruction through recurrence networks. *Indian Academy of Sciences Conference Series* 1:1, DOI: 10.29195/iascs.01.01.0004, 2017.

HELLEBUST, S.; ALLANIC, A.; O’CONNOR, I. P.; WENGER, J. C.; SODEAU, J. R., 2010. The use of real-time monitoring data to evaluate major sources of airborne particulate matter. *Atmos. Environ.* 44, 1116 - 1125.

HOLT, C. C. Forecasting seasonals and trends by exponentially weighted moving averages. *International journal of forecasting*, 20(1):5–10, 2004.

HOLZFUSS, J.; MAYER-KRESS, G. Approach to Error-estimation in the Application of Dimension Algorithms. Los Alamos National Lab, NM (USA). 1985.

HUEBNEROVA, Z.; MICHALEK, J. Analysis of daily average PM10 predictions by generalized linear models in Brno, Czech Republic. *Atmos. Pollut. Res.* 5, 471–476, 2014.

HYNDMAN, R. J.; ATHANASOPOULOS, G. *Forecasting: principles and practice*. Monash University, Australia, Texts, 2nd, 380 p, 2014.

HYNDMAN, R. J.; KOEHLER, A. B.; SNYDER, R. D.; GROSE, S. A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting*, 18(3):439–454, 2002.

IMANDOUST, S. B.; BOLANDRAFTAR, M. Application of K-Nearest Neighbor (KNN) approach for predicting economic events: Theoretical Background. *International Journal of Engineering Research and Applications*, v. 3, Issue 5, pp.605-610, 2013.

Inventário Nacional de Emissões Atmosféricas da Aviação Civil, Agência Nacional de Aviação Civil, 2014.

ISA, N. E. M.; AMIR, A.; ILYAS, M. Z.; RAZALLI, M. S. The performance analysis of K-Nearest neighbors (K-NN) algorithm for motor imagery classification based on EEG signal. *ICEESI 2017 - MATEC Web of Conferences* 140, 01024, DOI: 10.1051/mateconf/201714001024, 2017.

ISLAM, M.; SIVAKUMAR, B. Characterization and prediction of runoff dynamics: a nonlinear dynamical view. *Advances in Water Resources*, 25(2):179–190, 2002.

KENNEL, M. B.; BROWN, R; ABARBANEL, H. D. I. Determining embedding dimension for phase space reconstruction using a geometrical construction. *Phys. Rev. A.*, v. 45, 3403-3411, 1992.

KRISHNAN, K. B.; MCA, S. S. S.; PHIL, M. Application of data mining in agriculture. *International Journal of Research In Computer Applications and Robotics*, v. 5 Issue 8, Pg.: 18-21, August, 2017.

KURT, A.; OKTAY, A.B. Forecasting air pollutant indicator levels with geographic models 3 days in advance using neural networks. *Expert Syst. Appl.* 37, 7986–7992, 2010.

KYOUNG, M. S.; KIM, H. S.; SIVAKUMAR, B.; SINGH, V. P.; AHN, K. S. Dynamic characteristics of monthly rainfall in the Korean Peninsula under climate change, *Stoch. Env. Res. Risk a.*, 25, 613-625, doi: 10.1007/s00477-010-0425-9, 2011.

LANA, X.; BURGUENO, A.; MARTINEZ, M. D.; SERRA, C. Complexity and predictability of the monthly Western Mediterranean Oscillation index, *Int. J. Climatol.*, 36, 2435-2450, doi: 10.1002/joc.4503, 2016.

LEE, C-K.; LIN, S-C. Chaos in air pollutant concentration (APC) time series. *Aerosol and Air Quality Research*, v. 8, No.4, p. 381-391, 2008.

LETELLIER, C.; AGUIRRE, L.A.; MAQUET, J.; GILMORE, R. Evidence for low dimensional chaos in sunspot cycles, *Astronom. Astrophys*, 449(1), pp. 379–387, 2006.

LI, H. Z.; MOULINE, Y.; CHOPLIN, L.; MIDOUX, N. Chaotic bubble coalescence in non-Newtonian fluids. *International Journal of Multiphase Flow*. v. 23, Issue 4, August 1997, Pages 713–723.

LI, T. Y; YORKE, J. A. Period three implies chaos. *The American Math. Monthly*, v. 82, N. 10, p. 985-992, 1975.

LIMA, E. A.; NOGUEIRA, J. S. *Coletânea Física Ambiental II. Baraúna (Edição Digital)*, 2013.

LONGLEY, P. A.; GOODCHILD, M. F.; MAGUIRE, D. J.; RHIND, D. W. *Sistemas e ciência da informação - Informação geográfica*. Bookman, 3ª Edição, 2013.

LORENZ, E. N. Deterministic non-periodic flow. *J. Atmos. Sci.*, v. 20, pp. 130-141, 1963.

LUO, X.; CAO, H. Evaluation of air quality using the CMAQ modeling system. *Procedia Environ. Sci.* 12, 159–165, 2012.

MACHADO, C. M. D.; ALLEN, A. G.; MCGONIGLE, A. J. S.; CARDOSO, A. A. *Química de óxidos de nitrogênio e enxofre na atmosfera da região da baixada santista*.

Sociedade Brasileira de Química (SBQ), 31a Reunião Anual da Sociedade Brasileira de Química, 2008.

MARQUES, R.; SANTOS, E. S. Redes de Monitoramento de Material Particulado Inalável, Legislação e os Riscos à Saúde. *Revista Brasileira de Geografia Médica e da Saúde*. ISSN: 1980-1726, *Hygeia* 8(14): 115 - 128, 2012.

MARTIN, A. E.; BRADLEY, W. H. Mortality, fog, and atmospheric pollution: an investigation during the winter of 1958–59. *Month Bull Min Hlth*. 19: 56. 1960.

MARTINS, L. C.; LATORRE, M. R. D. O; CARDOSO, M. R. A.; GONÇALVES, F. L. T.; SALDIVA, P. H. N.; BRAGA, A. L. F. Poluição atmosférica e atendimentos por pneumonia e gripe em São Paulo, Brasil. *Rev Saúde Pública*, 36(1):88-94, 2002.

MCCAULEY, J. L. Introduction to multifractal in dynamical system theory and fully developed turbulence. *Phys. Rev.* 189, 225, 1990.

MENEZES, A. L. C.; MENEZES, C. M. V. M. C; SILVA, J. C.; GRAUDENZ, G. S. A prevalência das políticas públicas como ferramenta na redução da poluição atmosférica, de origem veicular, no município de SP. XVII Encontro Internacional sobre Gestão Empresarial e Meio Ambiente (ENGEMA), 2015.

MIRMOMENI, M.; LUCAS, C. Model based method for determining the minimum embedding dimension from solar activity chaotic times series. *IJE Transactions A: Basics*. v. 21, No. 1, February, 2008.

MOLDAGULOVA, A.; SULAIMAN, R. B. Using KNN algorithm for classification of textual documents. 8th International Conference on Information Technology (ICIT), DOI: 10.1109/ICITECH.2017.8079924, 2017.

MONTEIRO, L. H. A. Sistemas dinâmicos. Editora Livraria da Física, 3ª Ed., São Paulo, 2011.

MONTGOMERY, D. C.; JENNINGS, C. L.; KULAHCI, M. Introduction to time series analysis and forecasting. Wiley Series in Probability and Statistics. Wiley, New Jersey, United States of America, 2ª Ed., 2015.

MONTGOMERY, D. C.; RUNGER, G. C. Estatística aplicada e probabilidade para engenheiros. 6ª Ed. Editora LTC, 2016.

MORAES, R.M.; DELITTI, W.B.C.; MORAES, J. Gas exchange, growth, and chemical parameters in a native Atlantic forest tree species in polluted areas of Cubatão, Brazil. *Ecotoxicol. Environ. Safety*, 54, 339, 2003.

MOREIRA, D. M.; TIRABASSI, T.; MORAES, M. R. Meteorologia e poluição atmosférica. *Ambiente & Sociedade*, Campinas v. XI, No. 1, p. 1-13, 2008.

MORETTIN, P. A.; TOLOI, C. M. C. Análise de séries temporais. Blucher, São Paulo, Brasil, 2 edição, 2006.

MOURA, V. B; SCATENA, R.; PRIETO, W. H.; CREMASCO, M. A. Preliminary study of k-Nearest Neighbor algorithm and ARIMA model in fluidized bed time series forecasting. XXXVIII Congresso Brasileiro de Sistemas Particulados. Universidade Estadual de Maringá, Maringá - Paraná, 2017.

MULLIN, T. The Nature of chaos, Clarendon Press, Oxford, 1993.

NIE, C. X.; SONG, F. T. Analyzing the stock market based on the structure of kNN network. *Chaos, Solitons & Fractals*, v. 113, p. 148-159, August, 2018.

Operação inverno 2005: relatório das atividades. Arquivo de texto: color., .pdf, 9,27MB.
Disponível em: < \\Srvfs01\etna\PCPV 2014\Referências >

PALUŠ, M.; DVORÁK, I. Singular-Value decomposition in attractor reconstruction: pitfalls and precautions. *Physica D.*, v. 55, p. 221-234, 1992.

PAN, L.; SUN, B.; WANG, W. City Air Quality Forecasting and Impact Factors Analysis Based on Grey Model. *Procedia Eng.* 12, p. 74–79, 2011.

PARMEZAN, A. R. S.; BATISTA, G. E. A. P. A. Descrição de modelos estatísticos e de aprendizado de máquina para predição de séries temporais. Instituto de Ciências Matemáticas e de Computação - USP, Relatório técnico 412, ISSN 0103-2569, 2016.

PCPV: Plano de Controle de Poluição Veicular 2017- 2019 / CETESB; Elaboração Antônio de Castro Bruni [et al.] - São Paulo: CETESB, 2017.

PEREIRA, L. A. A.; CONCENIÇÃO, G. M. D.; BRAGA, A. L. F. Cardiovascular effects of air pollution in adults in Cubatão, São Paulo, Brazil. *Epidemiology*, 15, S2, 2004.

POMMERANZENBAUM, I. R. Redes neurais artificiais na predição das principais séries do índice Ibovespa e suas Aplicações em sistemas automatizados de negociação – Rio de Janeiro: UFRJ/COPPE, 2014.

POPE, C. A. I.; BURNETT, R. T.; THUN, M. J.; CALLE, E. E.; KREWSKI, D.; ITO, K.; THURSTON, G. D. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *J. Am. Med. Assoc.* 287, 1132-1141. 2002.

PRIETO, W. H. Aplicação da teoria do caos em um leito fluidizado utilizando-se partículas A, B e D da classificação Geldart. Dissertação (Mestrado em Engenharia Química) - Universidade Estadual de Campinas, 2014.

PRIETO, W. H.; CREMASCO, M. A. Emprego da análise de caos na caracterização de regimes fluidizados para partículas do Grupo A da classificação Geldart. In: XXXVII Congresso Brasileiro de Sistemas Particulados - ENEMP 2015, 2015, São Carlos, Blucher Engineering Proceedings. São Paulo: Blucher, v. 1. p. 17-78, 2015.

RAGA, G. B.; LE MOYNE, L. On the nature of air pollution dynamics in Mexico City. 1. Nonlinear analysis. Atmos. Environ. 30, p. 3987 - 3993. 1996.

Relatório de Qualidade do Ar no Estado de São Paulo - 2013, Companhia de Tecnologia de Saneamento Ambiental – CETESB, 2014.

RISTANOSKI, G.; LIU, W.; BAILEY, J. A time-dependent enhanced support vector machine for time series regression. Em International Conference on Knowledge Discovery and Data Mining, páginas 946–954, Chicago, United States of America. ACM, 2013.

ROSENSTEIN, M. T.; COLLINS, J. J.; DE LUCA, C. J. A practical method for calculating largest Lyapunov exponents from small data sets. Physica D: Nonlinear Phenomena, 65(1-2):117–134, 1993.

RUELLE, D.; TAKENS, F. On the nature of turbulence, Commun. Math. Physics, v. 30, pp. 167-192, 1971.

SAEED, S.; IDRIS, A.; HUSSAIN, L.; AWAN, I. A. Analyzing the Dynamics of Particulate Matters (PM) using Nonlinear Dynamical Techniques and Predicting the Behavior Based on Robust Regression Models. IJCSNS International Journal of Computer Science and Network Security, v.17 No.11, November 2017.

SALDIVA, P. H. N.; POPE III, C. A.; SCHWARTZ, J.; DOCKERY, D. W.; LICHTENFELS, A. J.; SALGE, J. M.; et al. Air pollution and mortality in elderly people: a time-series study in Sao Paulo, Brazil. Arch Environ Health; 50:159 - 63. 1995.

SANTOS, F. C. Variações do método kNN e suas aplicações na classificação automática de textos. Dissertação de mestrado, Universidade Federal de Goiás, Instituto de Informática, Goiânia, 2009.

SAVI, M. A. Dinâmica não linear e caos. Editora e-papers, 2ª edição, 390 p., ISBN 978-85-7650-554-9, 2017.

SCATENA, R.; PRIETO, W. H.; CREMASCO, MARCO A. Avaliação da Ferramenta BRChaos para Análise de Caos na Identificação da Fluidodinâmica de Leito de Jorro. In: XXI Congresso Brasileiro de Engenharia Química, Fortaleza, v. 1. p. 1-1, 2016.

SCATENA, R.; MOURA, V. B.; PRIETO, W. H.; NUNHEZ, J. R.; CREMASCO, M. A. Theoretical study of chaotic invariants from pressure time series of fluidized bed CFD model for Gedalt-type B particle. In: XXXVIII Congresso Brasileiro de Sistemas Particulados - Enemp 2017, Maringá. Anais do XXXVIII Enemp. Campinas: Galoá, v. 1. p. 1-1, 2017.

SCATENA, R. Análise de caos em flutuações de pressão oriundas de modelagem CFD na condição de mínima fluidização em leito fluidizado. Dissertação de Mestrado, Universidade Estadual de Campinas, Faculdade de Engenharia Química, Campinas, SP : [s.n.], 2018.

SCHOUTEN, J. C.; TAKENS, F.; VAN DEN BLEEK, C. M. Maximum-Likelihood-Estimation of the Entropy of an Attractor. In *Physical Review E*, v. 49, pp. 126-129, 1994.

SHE, D.; YANG, X. A new adaptive local linear prediction method and its application in hydrological time series, *Math. Probl. Eng.*, doi: 10.1155/2010/205438, 2010.

SHUMWAY, R. H.; STOFFER, D. S. Time series analysis and its applications: with R examples. Springer Texts in Statistics. Springer Science & Business Media, New York, United States of America, 3ª edição, 2011.

SILIBELLO, C.; D'ALLURA, A.; FINARDI, S.; BOLIGNANO, A.; SOZZI, R. Application of bias adjustment techniques to improve air quality forecasts. *Atmos. Pollut. Res.* 6, 928–938, 2015.

SIVAKUMAR, B.; PERSSON, M.; BERNDTSSON, R., UVO, C. B. Is correlation dimension a reliable indicator of lowdimensional chaos in short hydrological time series? *Water Resour. Res.*, 38, 3-1, doi: 10.1029/2001WR000333, 2002.

SIVAKUMAR.B.; PHOON, K.K.; LIONG,S.Y.; LIAW,C.Y.A systematic approach to noise reduction in chaotic hydrological time series, *Journal of Hydrology*, v.219 (3/4), pp.103-135, 1999.

SIWEK, K.; OSOWSKI, S. Improving the Accuracy of Prediction of PM10 Pollution by the Wavelet Transformation and an Ensemble of Neural Predictors; Pergamon Press, Inc.: Oxford, UK, 2012.

SONG, Y.; QIN, S.; QU, J.; LIU, F. The forecasting research of early warning systems for atmospheric pollutants: A case in Yangtze River Delta region. *Atmos. Environ.* 118, 58–69, 2015.

SOUZA, J. Consequências da dimensão fractal da distribuição de dados no mapeamento gravimétrico da Bacia do Parnaíba. Dissertação de Mestrado, Ministério da Ciência e Tecnologia, Observatório Nacional, Coordenadoria de Geofísica, Rio de Janeiro, Fevereiro, 2002.

SPORTISSE, B. Pollution atmosphérique - Des processus à la modélisation. Springer-Verlag France, Paris, 2008.

STAM, C. J. Nonlinear dynamical analysis of EEG and MEG: Review of an emerging field. *Clinical Neurophysiology* 116, 2266–2301, 2005.

TRSTENJAK, B.; MIKAC, S.; DONKO, D. KNN with TF-IDF based framework for text categorization. *Procedia Engineering*, v. 69, Pages 1356-1364, 2014.

TSONIS, A. A. *Chaos: From Theory to Applications*. New York: Plenum Press, 274 p. 1992.

VAN DEN BLEEK, C. M.; COPPENS, M. O.; SCHOUTEN, J. C. Application of chaos analysis to multiphase reactors, *Chemical Engineering Science*, v. 57, pp. 4763-4778, 2002.

VAN DEN BLEEK, C. M.; SCHOUTEN, J. C. Can Deterministic chaos create order in fluidized-bed scale-up? *Chemical Engineering Science*, v. 48, No. 13, pp. 2367-2373, 1993.a.

VAN DEN BLEEK, C. M.; SCHOUTEN, J. C. Deterministic chaos: a new tool in fluidized bed design and operation. *Chemical Engineering Journal*, v. 53, pp. 75-87, 1993.b.

VANDER STAPPEN, MARIA, M. L. *Chaotic hydrodynamics of fluidized beds*, Delft: University of Technology, Delft – NL, 236 pp. PhD. Thesis, 1996.

VAPNIK, V. N. *The nature of statistical learning theory*. Information Science and Statistics. Springer Science & Business Media, New York, United States of America, 2 edição, 1999.

VIEIRA-FILHO, M. S.; LEHMANN, C.; FORNARO, A. Influence of local sources and topography on air quality and rainwater composition in Cubatão and São Paulo, Brazil. *Atmospheric Environment* 101 (2015) 200-208.

VILLEGAS, M.; AUGUSTIN, F.; GILG, A.; HMAIDI, A.; WEVER, U. Application of the polynomial chaos expansion to the simulation of chemical reactors with uncertainties. v. 82, Issue 5, p. 805–817, 2012.

VORMITTAG, E. M. P. A. A.; SALDIVA, P. H. N. Qualidade do ar no Estado de São Paulo - Sob a visão da saúde. Instituto de Saúde e Sustentabilidade, São Paulo, 2015.

WINTERS, P. R. Forecasting sales by exponentially weighted moving averages. *Management Science*, 6(3):324–342, 1960.

WOLF, A.; SWIFT, J. B.; SWINNEY, H. L.; VASTANO, J. Determining Lyapunov exponents from a time series. *Physica 16D*, 285, 1985.

YU, B.; HUANG, C.; LIU, Z.; WANG, H. WANG, L. A chaotic analysis on air pollution index change over past 10 years in Lanzhou, northwest China. *Stoch Environ Res Risk Assess* 25:643–653, 2011.

YU, H-L.; LIN, Y-C.; SIVAKUMAR, B.; KUO, Y-M. A study of the temporal dynamics of ambient particulate matter using stochastic and chaotic techniques. *Atmospheric Environment* 69, 37-45, 2013.

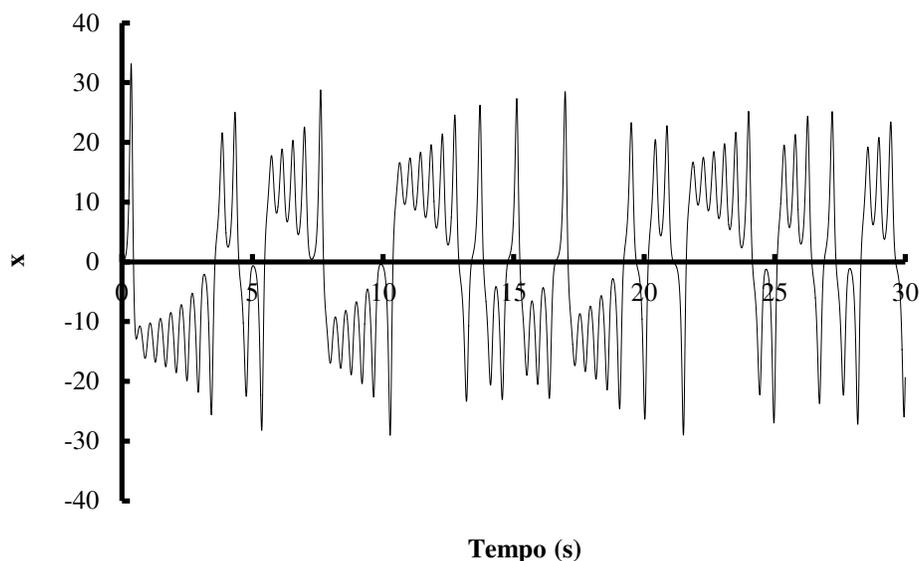
APÊNDICE A

AVALIAÇÃO DO CÓDIGO COMPUTACIONAL BRCHAOS

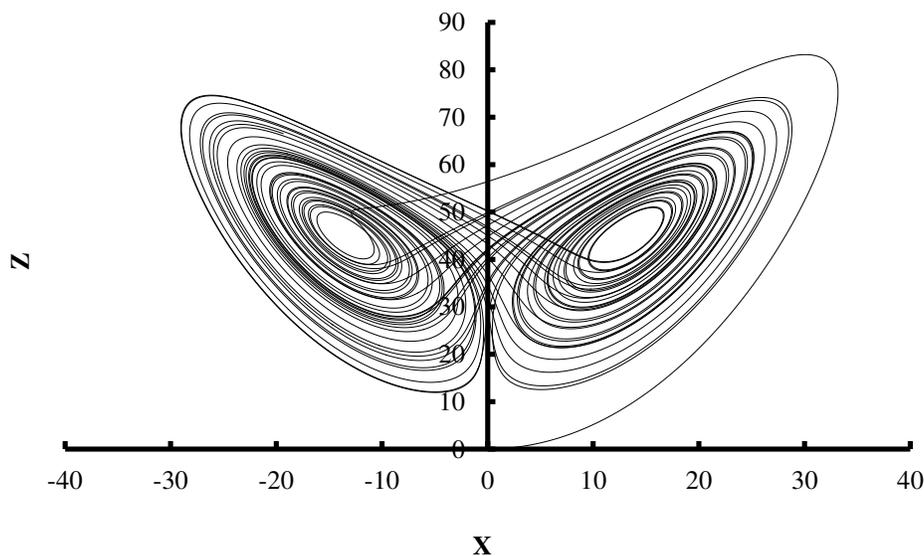
No que diz respeito à análise não linear de séries temporais, advindas de processos inerentes à Engenharia Química, o *software RRChaos* (Schouten *et al.*, 1992) é um dos mais aplicados para este fim. O pacote permite a reconstrução dos atratores, cálculo da dimensão de correlação e da entropia de Kolmogorov para séries temporais caóticas com ruídos. Muito embora seu uso seja bastante simples, o *RRChaos* é limitado quanto ao número de pontos utilizados na reconstrução do atrator e da série temporal, valores fixos de dimensão de imersão e incompatibilidade com os sistemas operacionais modernos havendo a necessidade de emulação de sua plataforma original tornando o processamento dos sinais extremamente lento.

Por estes motivos é que Moura (2013) desenvolveu um conjunto de códigos em linguagem MATLAB R2012a 64 bits que calcula a entropia de Kolmogorov (método de máxima verossimilhança (Schouten *et al.*, 1994), dimensão de correlação (algoritmo de Grassberger e Procaccia (1983)) e reconstrói o espaço de fases pelo método da decomposição em valores singulares (Broomhead e King, 1986). Nos anos seguintes à implementação destes códigos, o pacote sofreu modificações, como a inclusão do máximo expoente de Lyapunov pelo método de Rosenstein (1993), e recebeu o nome de BRChaos (*Basic Resolution of Chaos*).

Este apêndice tem o objetivo de aplicar o *BRChaos* na análise de sistemas consagrados, Lorenz (Lorenz, 1963) e Hénon, comparando os resultados com aqueles reportados pelos autores originais. Inicialmente o sistema de Lorenz (Equação 2.1) foi resolvido para $\sigma = 16$, $R = 45,92$, $b = 4,0$, $\Delta t = 0,01$ s por meio do método de Runge-Kutta de quarta ordem em *MATLAB R2016a*. A Figura A1 traz a série temporal da componente x pelo tempo (s).

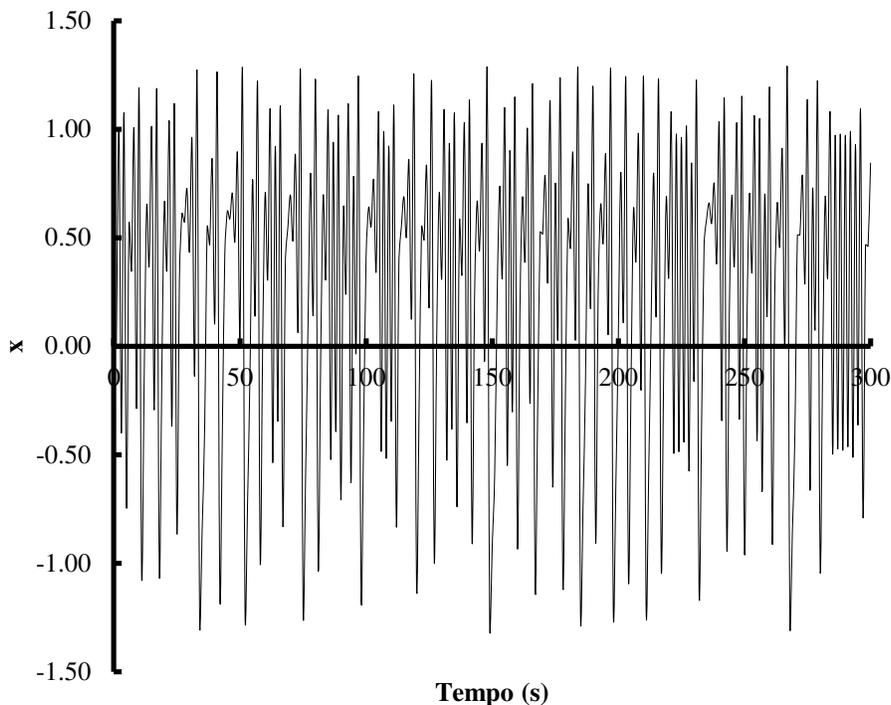
Figura A1 – Série temporal do sistema de Lorenz para a componente x versus o tempo (s) (o autor, 2018).

Ao plotar a componente z pela x , obtêm-se o conhecido atrator de Lorenz. A Figura A2 apresenta o atrator.

Figura A2 – Solução do atrator de Lorenz (o autor, 2018).

Já o sistema de Hénon foi resolvido, também pelo método de Runge-Kutta de quarta ordem, com $a = 1,4$ e $b = 0,30$, $\Delta t = 1$ s. A Figura A3 traz a série temporal da componente x pelo tempo (s). Já a Figura A4 apresenta a solução do mapa de Hénon.

Figura A3 – Série temporal de Hénon para a componente x versus o tempo (s) (o autor, 2018).



A Tabela A1 apresenta os dados de entropia de Kolmogorov e dimensão de correlação e máximo expoente de Lyapunov encontrados na literatura e os calculados por meio do *BRChaos* para o sistema de Lorenz e mapa de Hénon.

Os dados de dimensão de correlação e entropia de Kolmogorov (K_{ML}) foram comparados com os apresentados por Strogatz (1994) e Ferrara e Prado (1994) e, de modo a permitir a comparação, foram obtidos, para os dois sistemas, com dimensão de imersão (I) igual a 30. Já os valores de máximo expoente de Lyapunov (λ) teóricos foram extraídos do trabalho de Wolf *et al.*, 1985 e resolvidos para $I = 3$. A partir dos invariantes calculados, verifica-se que para o sistema de Lorenz a dimensão de correlação calculada pelo *BRChaos* apresentou um desvio de 0,5%, para K_{ML} foi de 0,1% e para o expoente de Lyapunov o desvio encontrado foi de 9,3%. Já para o sistema de Hénon, o erro para D_2 foi de 0,8%, para K_{ML} de 3,1% e, por fim, para λ 2,6%. Os baixos erros apresentados, quando comparados à valores teóricos de sistemas consagrados, pelos métodos utilizados no

BRChaos asseguram a eficiência do código, com a vantagem de ser de fácil execução com baixo custo computacional.

Figura A4 – Solução do mapa de Hénon (o autor, 2018).

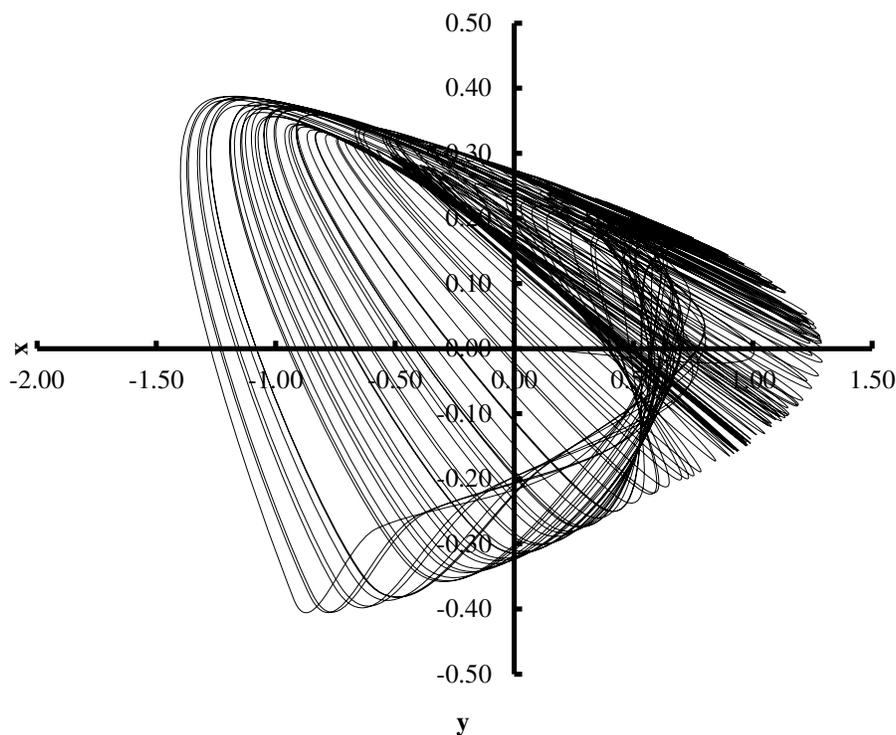


Tabela A1 – Comparação dos valores de entropia de Kolmogorov, dimensão de correlação e máximo expoente de Lyapunov obtidos por meio do BRChaos com os dados de literatura (o autor, 2018).

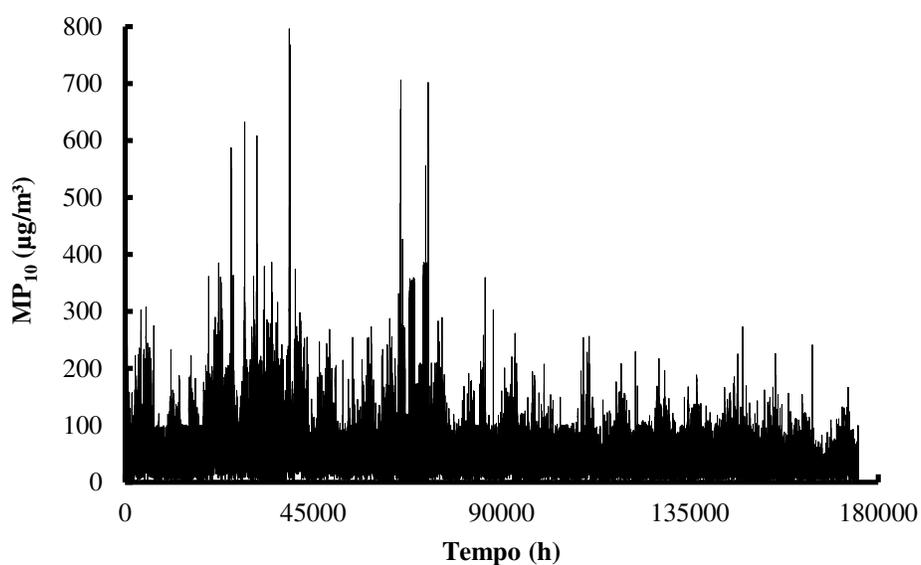
	Valores de literatura			Valores BRChaos		
	D_2 [-]	K_{ML} (bits/s)	λ	D_2	K_{ML} (bits/s)	λ [-]
Lorenz	2,06	2,480	1,50	2,07	2,482	1,64
Hénon	1,21	0,325	0,418	1,22	0,315	0,407

APÊNDICE B

INTERPOLAÇÃO ARIMA

Esta análise engloba a avaliação da influência do k na qualidade das previsões *out of sample* utilizando como treino a série temporal da concentração horária de MP_{10} obtida na estação de Congonhas com lacunas preenchidas por meio do método ARIMA. A Figura B1 apresenta a série histórica com os dados faltantes preenchidos por meio do método ARIMA. A Tabela B1 apresenta os resultados das previsões, sendo os pontos destacados os menores desvios encontrados.

Figura B1: Série histórica de concentração de MP_{10} de 1998 a 2017 obtida na estação de Congonhas com lacunas preenchidas por meio do método ARIMA (o autor, 2018).



A evolução temporal real (Figura 4.21) possui 155534 pontos válidos, contudo a evolução temporal apresentada na Figura B1 contém 175230 pontos, logo, foi necessário preencher os 19786 dados faltantes. O tempo utilizado nesse processamento foi de aproximadamente 95 horas. Uma vez completada a série, obteve-se a dimensão de imersão

por meio do método de GP. Para a série temporal apresentada na Figura B1, $I_{GP} = 11$. Novamente, constatou-se que, em relação à evolução original ($I_{GP} = 13$) houve uma queda no valor da dimensão de imersão. Isso ocorre pelo fato de que as interpolações, realizadas por meio da metodologia ARIMA, a semelhança do observado para a interpolação k-NN, tendem a apresentar uma evolução mais determinística, reduzindo, assim, a complexidade do sistema.

Os resultados apresentados na Tabela B1 são as previsões da concentração de MP_{10} na estação de Congonhas utilizando a série histórica cujas lacunas foram preenchidas por meio de interpolação utilizando metodologia ARIMA. Foram realizadas previsões para 60, 30, 15, 7 e 1 dia e as comparações foram feitas para os respectivos 60, 30, 15, 7 e 1 primeiros dias de 2018. A dimensão de imersão foi mantida constante e igual a 11.

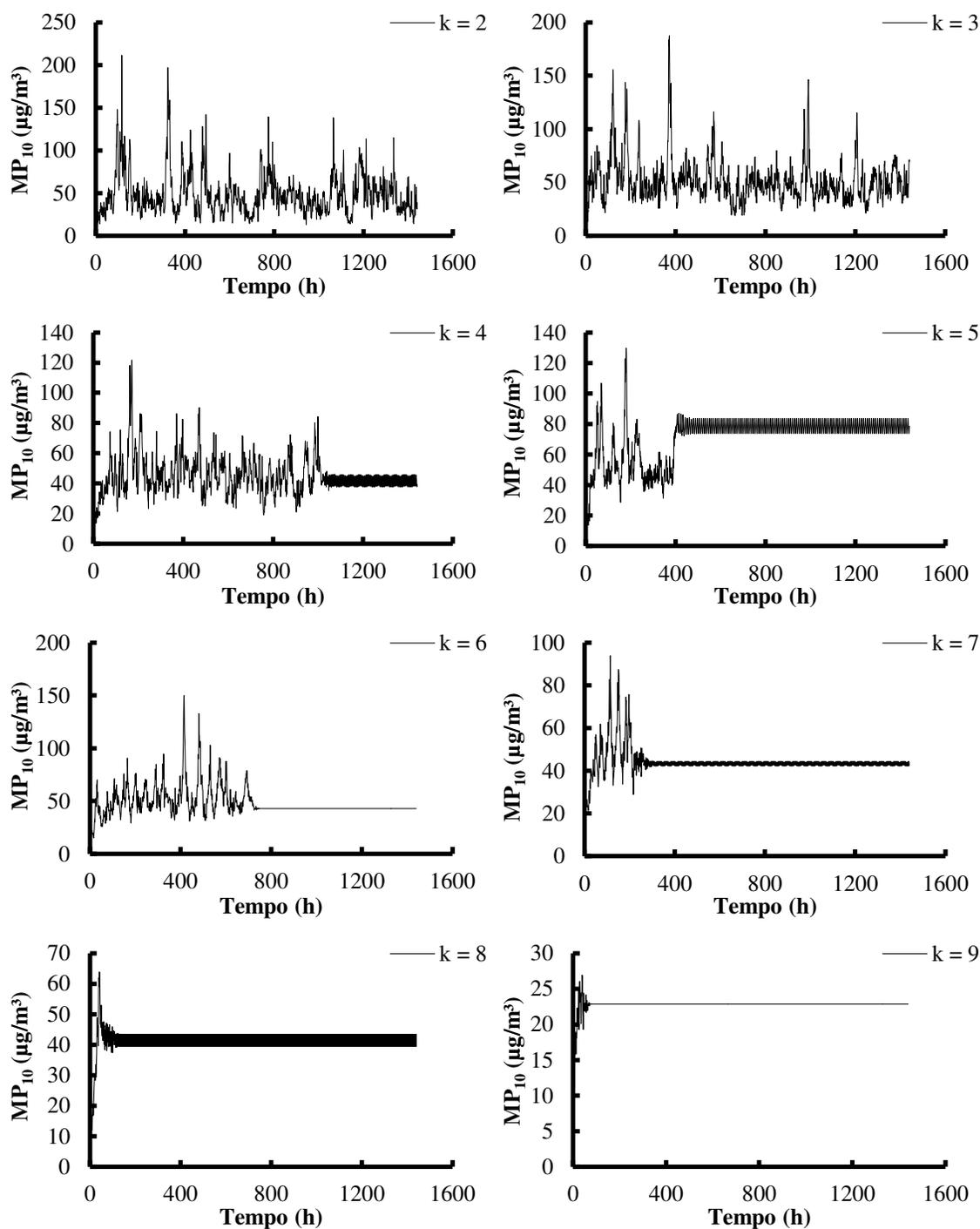
É importante esclarecer que os desvios de 7%, 8%, 131% e 29% encontrados respectivamente para $k = 9$ (média de 60 dias), $k = 9$ (média de 30 dias), $k = 4$ (média de 15 dias) e $k = 9$ (média de 7 dias) embora sejam os melhores valores para cada bloco de análise, não representam a dinâmica real, pois, conforme se observa na Figura B2, em todas essas situações, a evolução tende para um valor médio, fazendo com que a série perca as características originais. Assim, não são valores elegíveis para k .

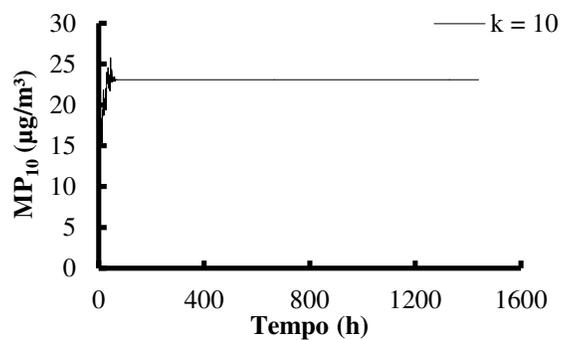
Por meio desta análise constatou-se que mesmo empregando o método ARIMA no preenchimento da série temporal, em nenhuma das situações a previsão realizada apresentou desempenho satisfatória, pois para as situações em que o desvio ficou baixo, ocorreu convergência para um valor médio e nas demais situações os desvios ficaram altos. Quando comparada à mesma análise com a série preenchida pela metodologia k-NN (Tabela 4.24), verifica-se que para praticamente todas as previsões válidas, os desvios calculados para as séries geradas por meio do treino preenchido com ARIMA ficaram bem superiores quando comparados ao k-NN. O tempo médio computacional para a obtenção de cada uma das 9 predições foi de aproximadamente 23 minutos, sendo o total de 3,45 horas.

Tabela B1: Resultados das previsões da concentração de MP_{10} na estação de Congonhas utilizando a série histórica com lacunas preenchidas por meio do método ARIMA para k variando de 2 a 10 nas situações de 60, 30, 15, 7, 2 e 1 dia de previsão (o autor, 2018).

	Média (60 dias) ($\mu\text{g}/\text{m}^3$)	Desvio	Média (30 dias) ($\mu\text{g}/\text{m}^3$)	Desvio	Média (15 dias) ($\mu\text{g}/\text{m}^3$)	Desvio	Média (1 semana) ($\mu\text{g}/\text{m}^3$)	Desvio	Média (2 dias) ($\mu\text{g}/\text{m}^3$)	Desvio	Média (1 dia) ($\mu\text{g}/\text{m}^3$)	Desvio
Real	21,27	-	21,06	-	19,71	-	17,15	-	12,73	-	12,76	-
k												
2	48,70	129%	49,32	134%	52,85	168%	55,21	222%	24,62	93%	21,44	68%
3	51,54	142%	54,40	158%	55,13	180%	56,98	232%	43,20	239%	34,29	169%
4	44,76	110%	46,55	121%	45,45	131%	41,93	144%	24,74	94%	17,97	41%
5	71,41	236%	64,50	206%	53,36	171%	51,19	198%	37,02	191%	24,04	88%
6	47,79	125%	52,69	150%	48,70	147%	43,58	154%	33,57	164%	22,05	73%
7	44,01	107%	44,69	112%	46,04	134%	46,10	169%	31,62	148%	22,59	77%
8	41,27	94%	41,04	95%	46,13	134%	39,59	131%	33,24	161%	19,56	53%
9	22,77	7%	22,69	8%	46,22	135%	22,14	29%	20,44	60%	17,70	39%
10	22,97	8%	22,86	9%	46,31	135%	22,19	29%	19,97	57%	17,17	35%

Figura B2: Séries temporais resultantes das previsões da concentração de MP_{10} na estação de Congonhas para k variando de 2 a 10 utilizando metodologia ARIMA no preenchimento dos dados faltantes (o autor, 2018).





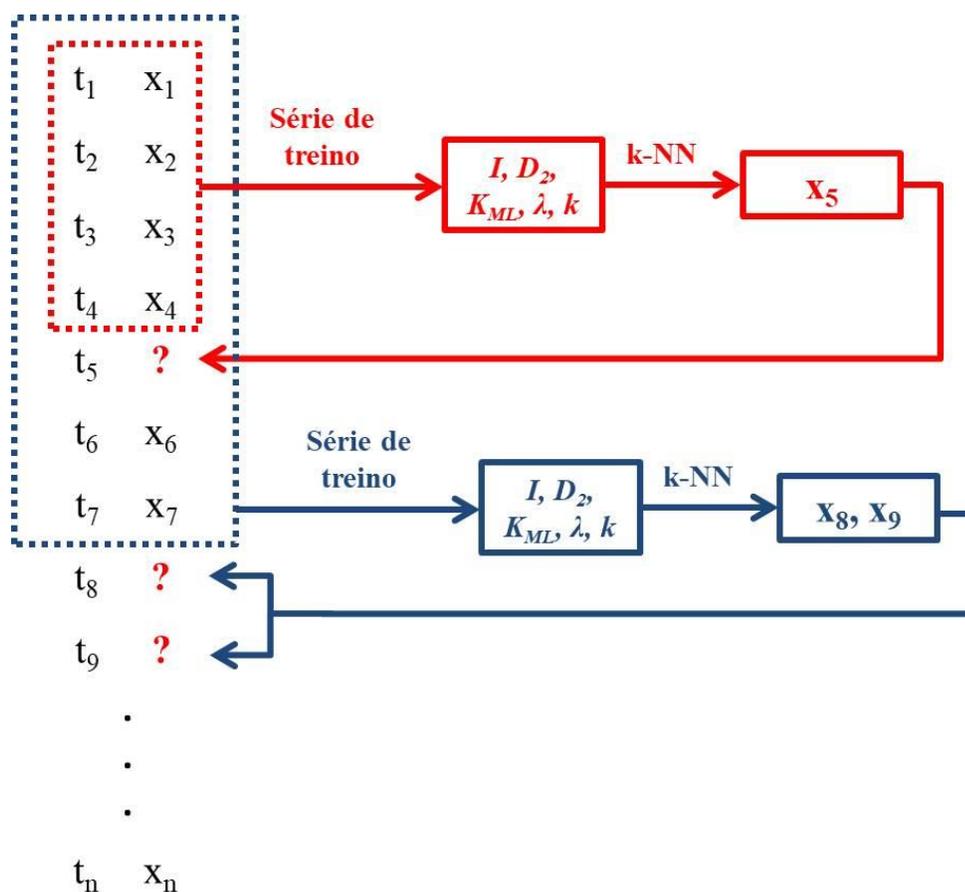
Devido ao elevado custo computacional (aproximadamente 98,45 horas) e dos resultados insatisfatórios obtidos por meio da interpolação utilizando a metodologia ARIMA, conclui-se que a técnica não apresenta vantagens quando comparada à mesma análise realizada por meio do *k*-NN.

APÊNDICE C

INTERPOLAÇÃO k -NN

Uma vez identificada a necessidade do preenchimento das séries de treino, o método de aprendizado de máquina dos vizinhos mais próximos, código k -NN (Perlin e Ceretta, 2007), foi aplicado para este fim. A Figura C1 apresenta o detalhamento da metodologia adotada.

Figura C1 – Preenchimento das séries temporais utilizando o método dos vizinhos mais próximos, k -NN (o autor, 2018).

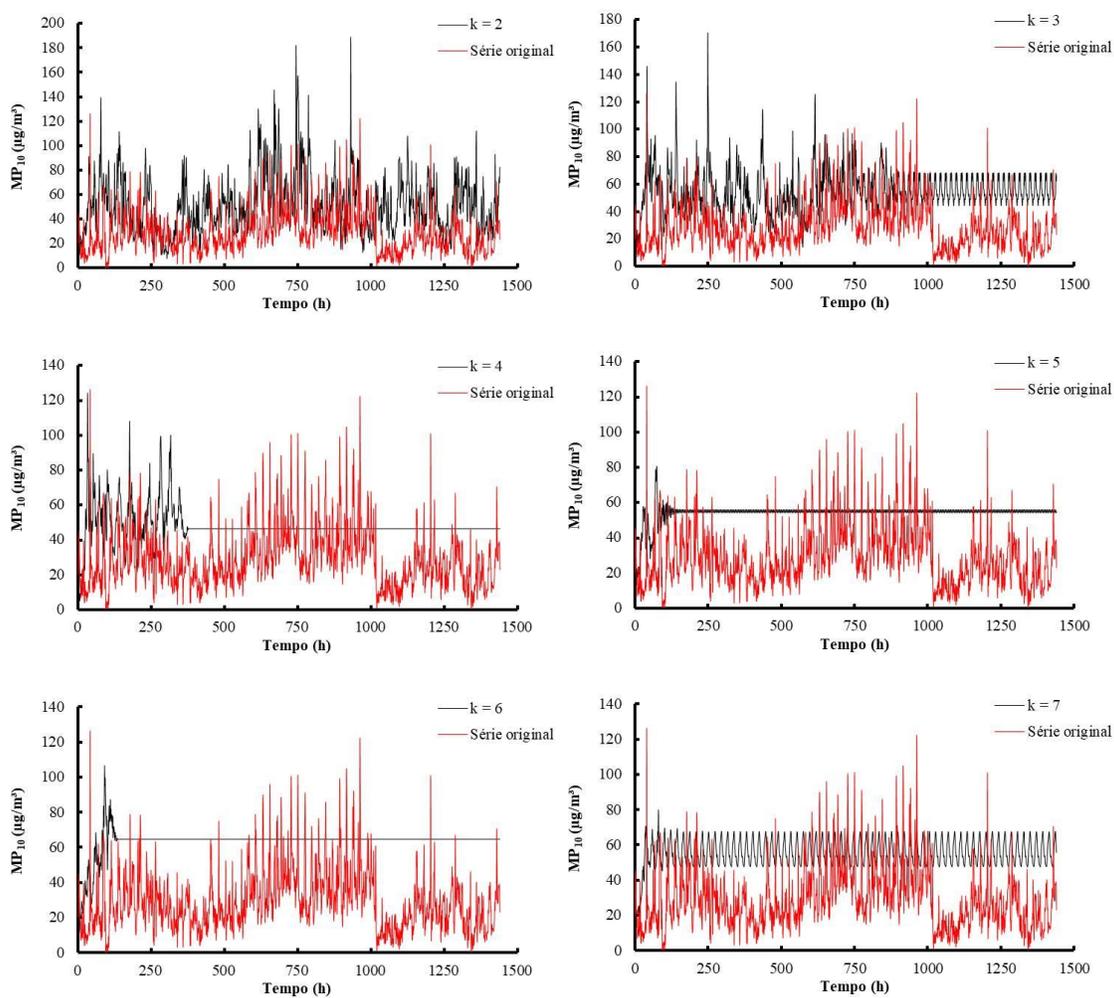


Na Figura C1 é possível verificar a existência de lacuna(s) identificada(s) pelo ponto de interrogação (?) que representa(m) o(s) dado(s) faltante(s). Os pontos imediatamente anteriores são utilizados na confecção da série de treino da qual são, inicialmente, extraídos os parâmetros do Caos (I , D_2 , K_{ML} e λ) que servirão para corroborar a aplicação da dimensão de imersão de Grassberger e Procaccia (1983) no método k -NN. Uma vez determinado o valor de I , realiza-se a otimização de k e, por fim, a geração/previsão *out of sample* do(s) dado(s) faltante(s). Uma vez gerado, o dado é associado à série temporal e integra a série de treino que será usada na previsão do próximo ponto faltante. No esquema da Figura C1 é possível observar esse passo na série de treino destaca em azul. Depois de prevista a lacuna relativa à posição temporal t_5 , o ponto x_5 se torna um dado da série de treino utilizada para se obter x_8 e x_9 . Este procedimento é repetido até o momento em que toda série temporal esteja completamente preenchida.

APÊNDICE D

COMPARAÇÃO DAS SÉRIES SIMULADAS E ORIGINAIS

Figura D1 – Comparação das previsões *out of sample* da concentração horária de MP_{10} com as séries originais nos primeiros dois meses de 2018 na estação do Ibirapuera utilizando o algoritmo k -NN para k variando de 2 a 10 (o autor, 2018).



Apêndice D – Comparação das séries simuladas e originais

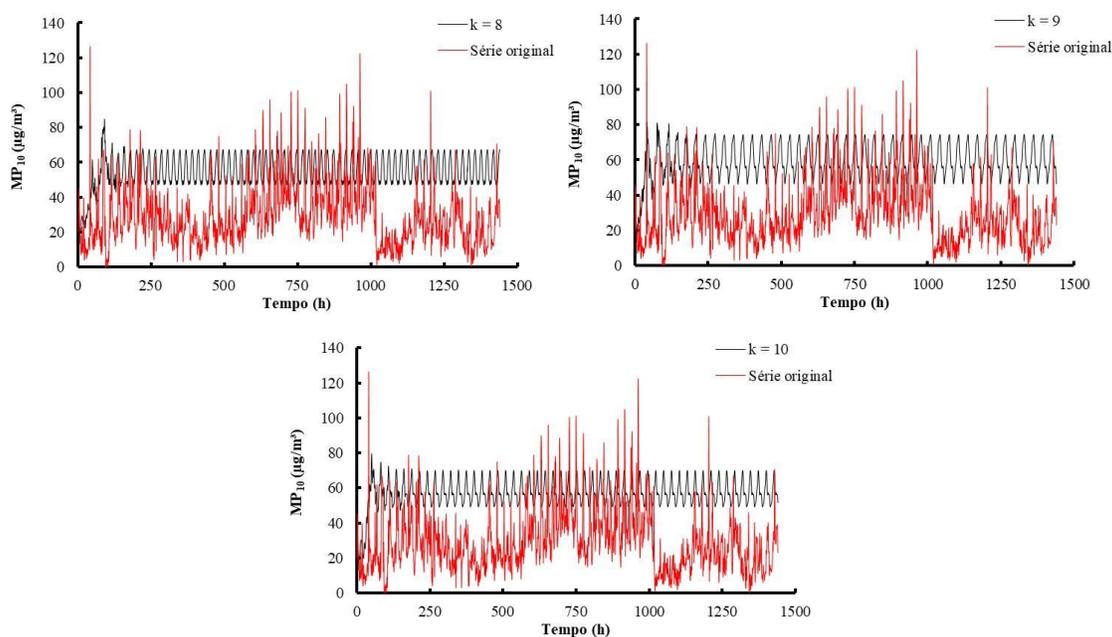
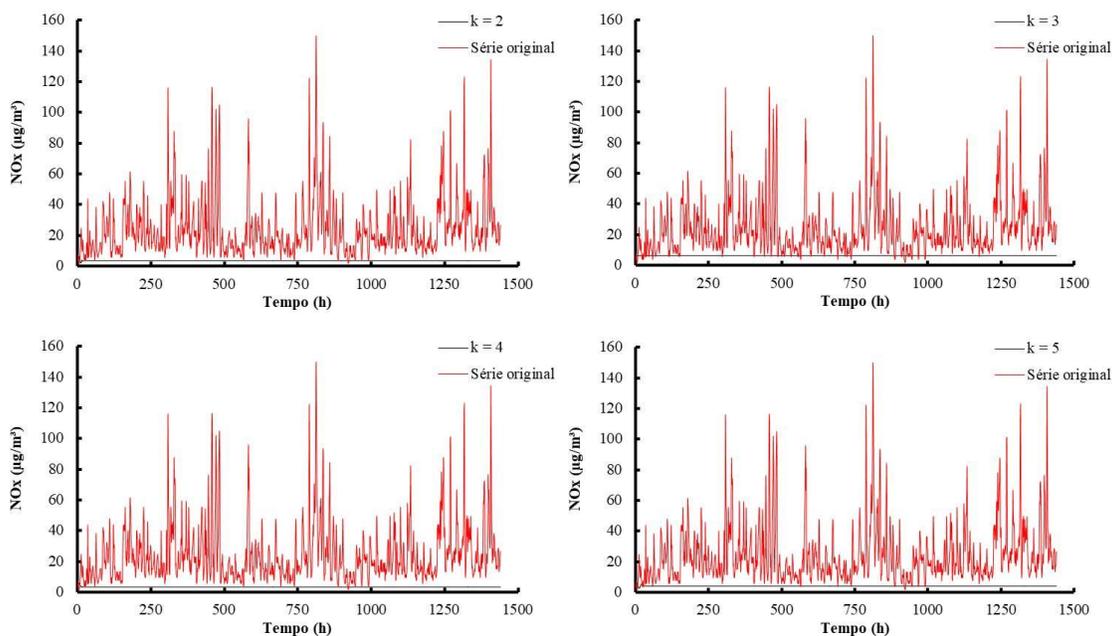


Figura D2 – Comparação das previsões *out of sample* da concentração horária de NO_x com as séries originais nos primeiros dois meses de 2018 na estação do Ibirapuera utilizando o algoritmo k -NN para k variando de 2 a 10 (o autor, 2018).



Apêndice D – Comparação das séries simuladas e originais

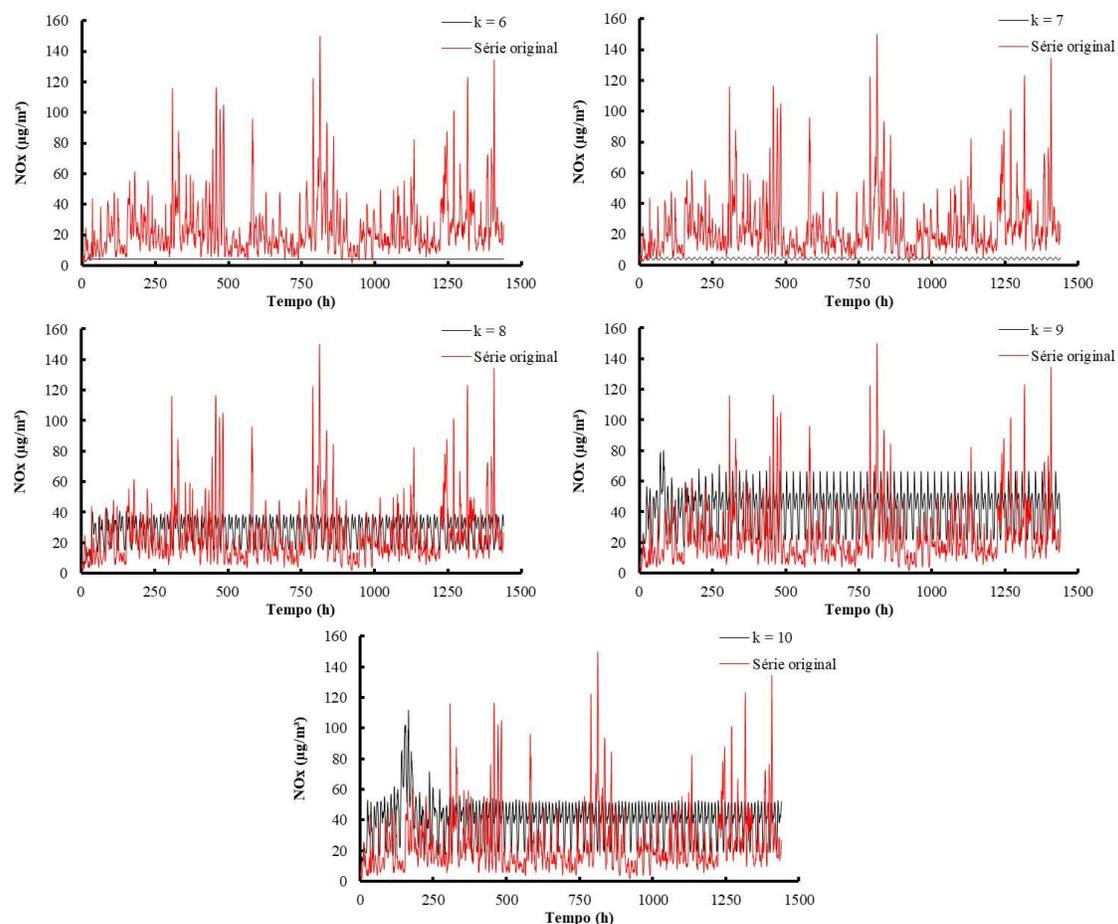
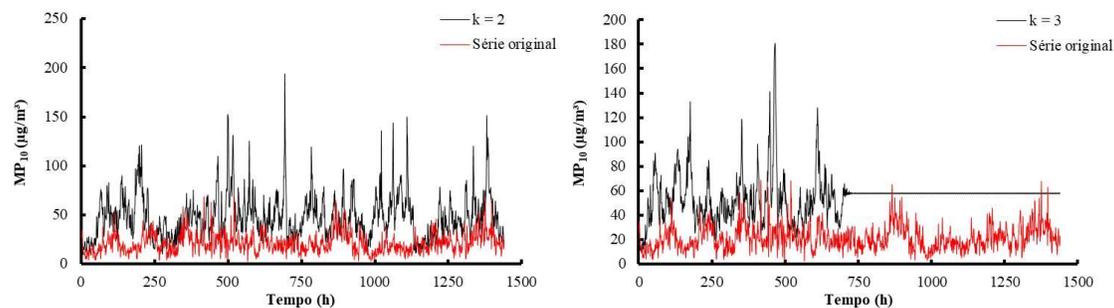


Figura D3 – Comparação das previsões *out of sample* da concentração horária de MP_{10} com as séries originais nos primeiros dois meses de 2018 na estação de Congonhas utilizando o algoritmo k -NN para k variando de 2 a 10 (o autor, 2018).



Apêndice D – Comparação das séries simuladas e originais

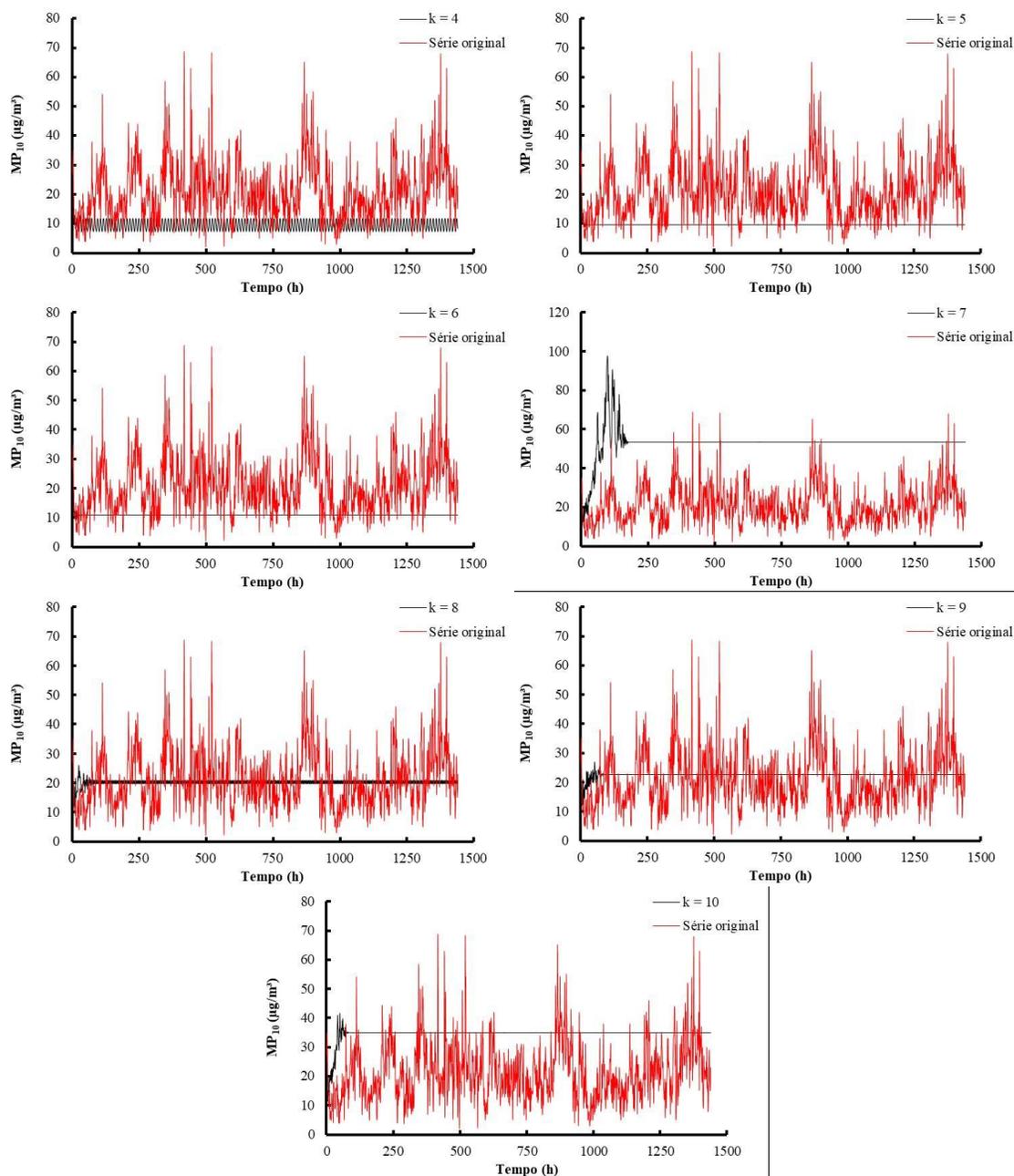
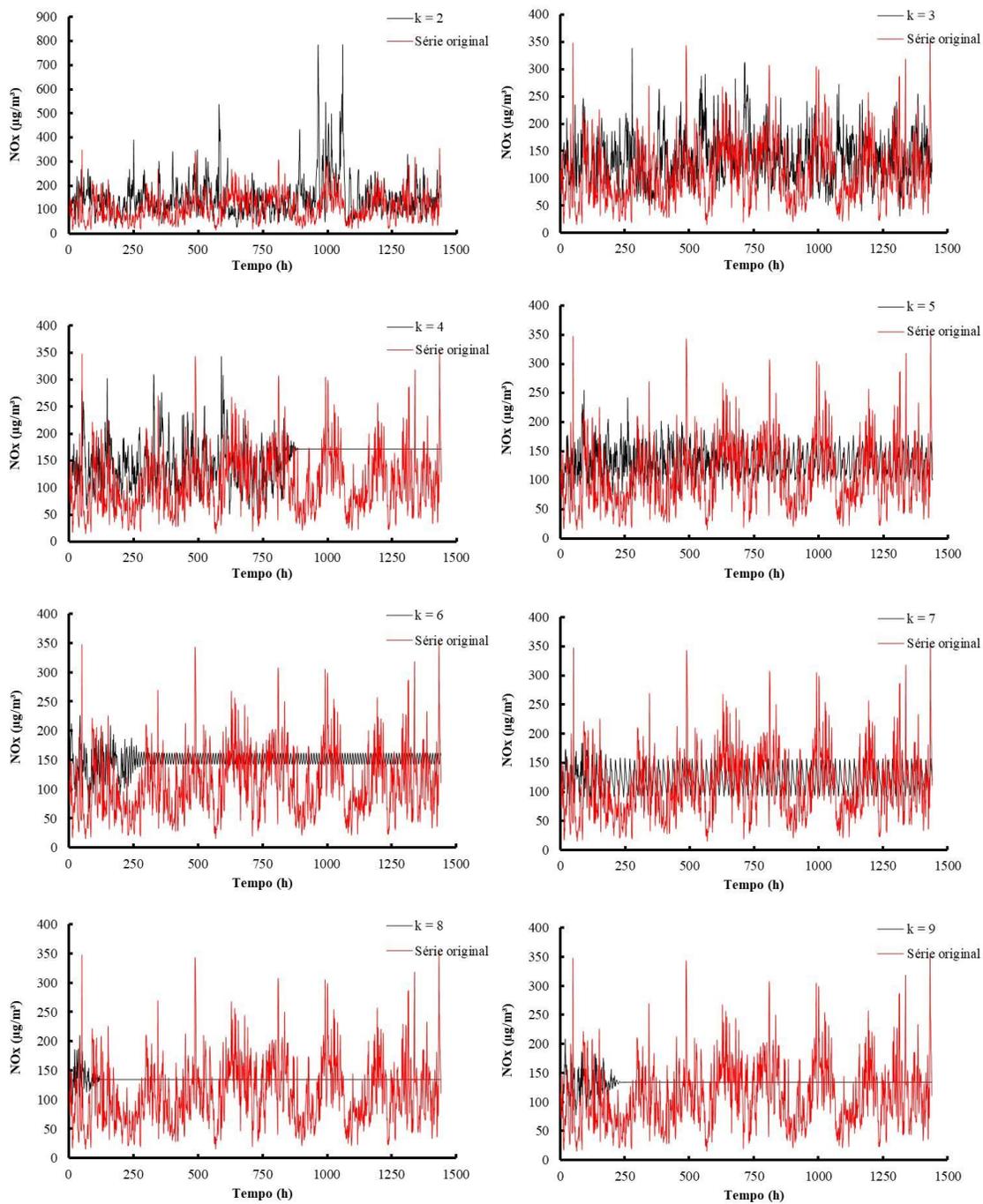


Figura D4 – Comparação das previsões *out of sample* da concentração horária de NOx com as séries originais nos primeiros dois meses de 2018 na estação de Congonhas utilizando o algoritmo *k*-NN para *k* variando de 2 a 10 (o autor, 2018).

Apêndice D – Comparação das séries simuladas e originais



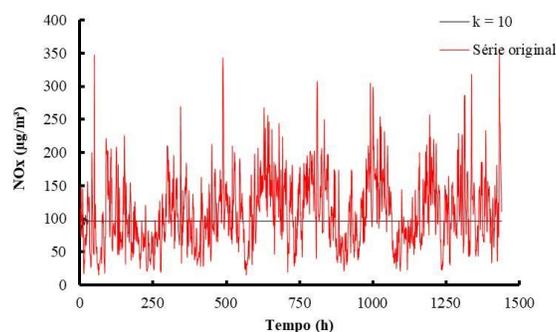
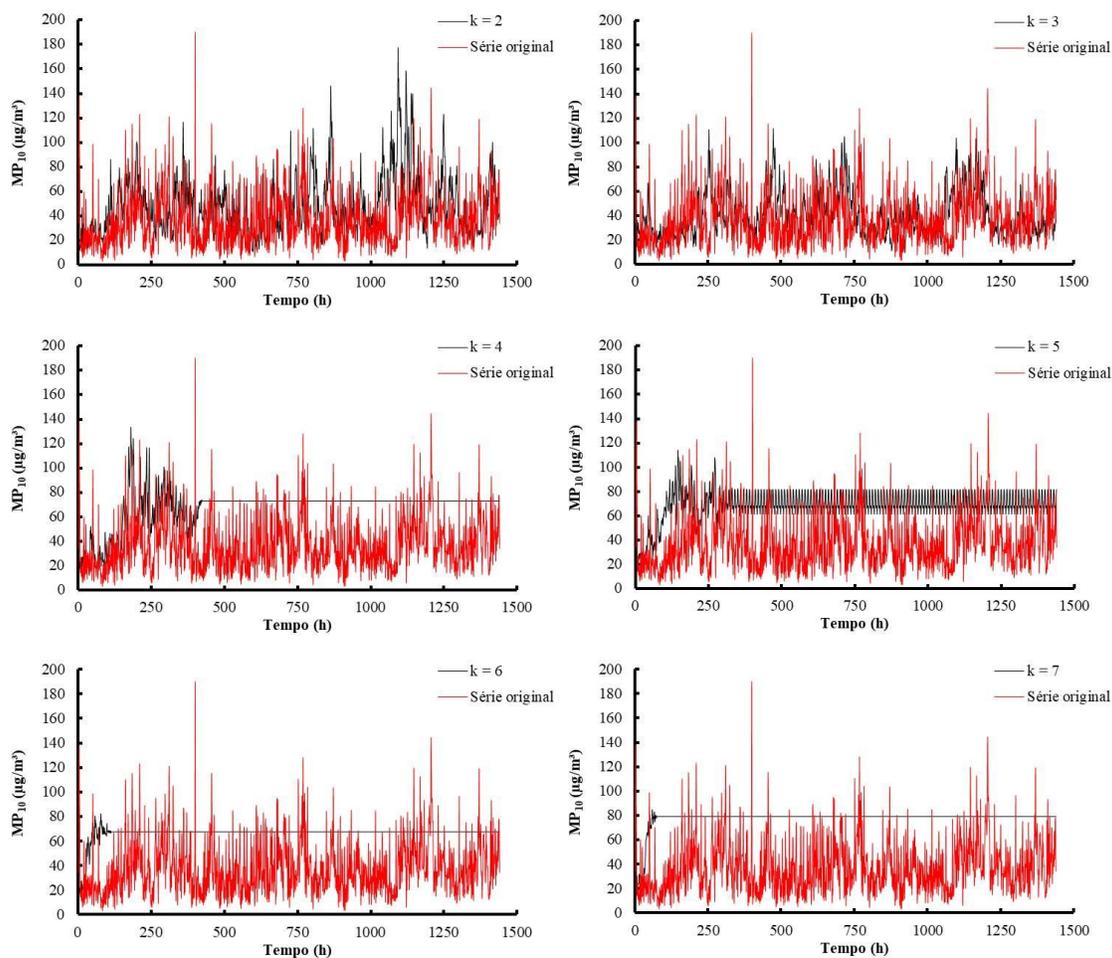


Figura D5 – Comparação das previsões *out of sample* da concentração horária de MP_{10} com as séries originais nos primeiros dois meses de 2010 na estação do Centro de São Paulo utilizando o algoritmo *k-NN* para *k* variando de 2 a 10 (o autor, 2018).



Apêndice D – Comparação das séries simuladas e originais

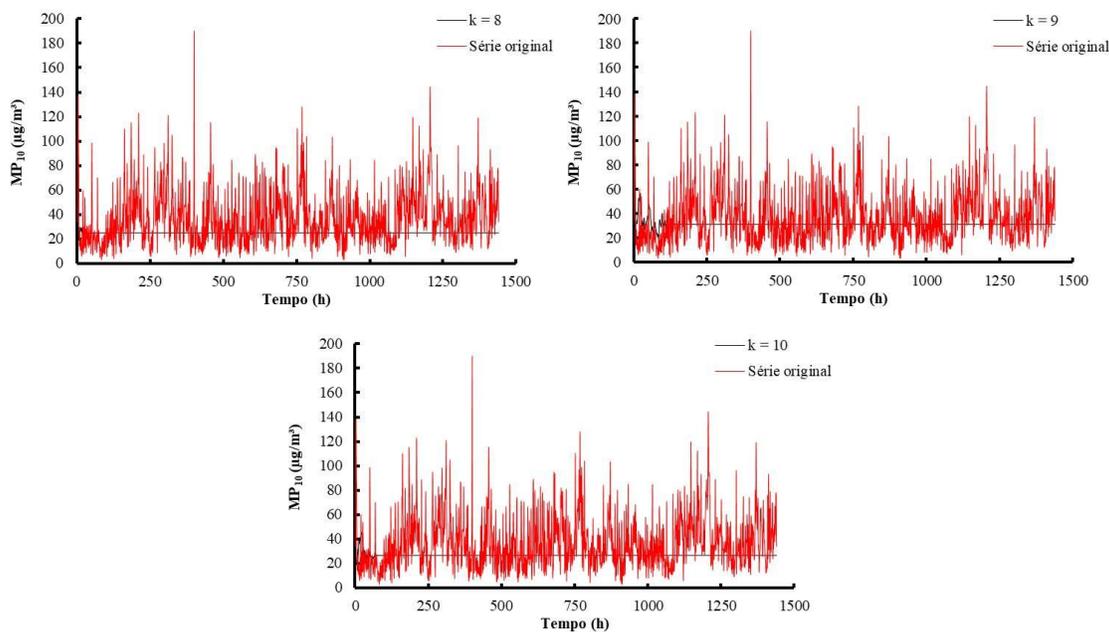
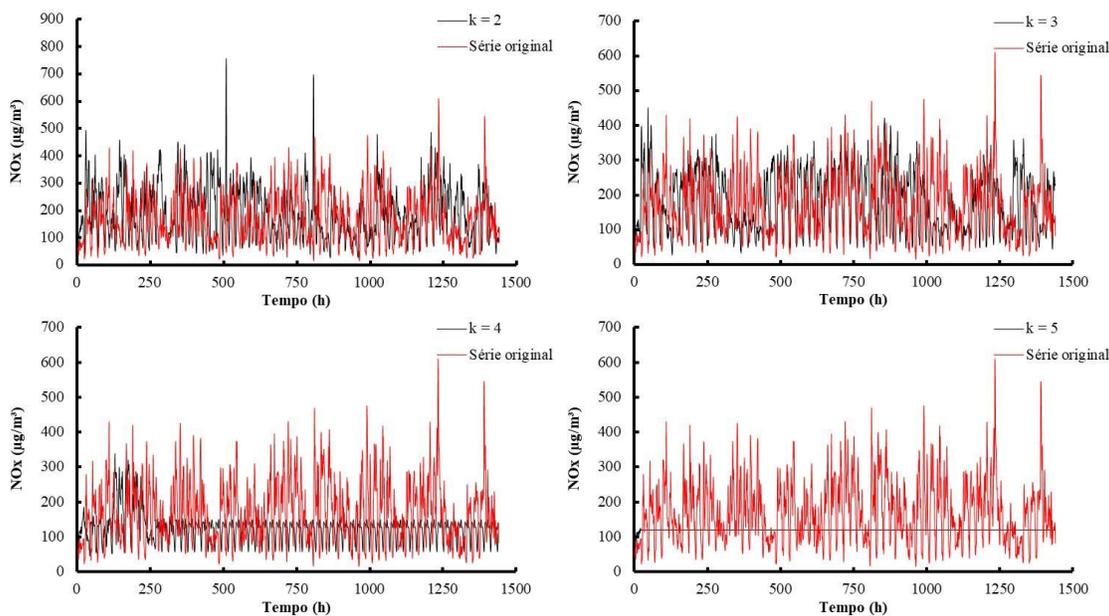


Figura D6 – Comparação das previsões *out of sample* da concentração horária de NO_x com as séries originais nos primeiros dois meses de 2000 na estação do Centro de São Paulo utilizando o algoritmo k -NN para k variando de 2 a 10 (o autor, 2018).



Apêndice D – Comparação das séries simuladas e originais

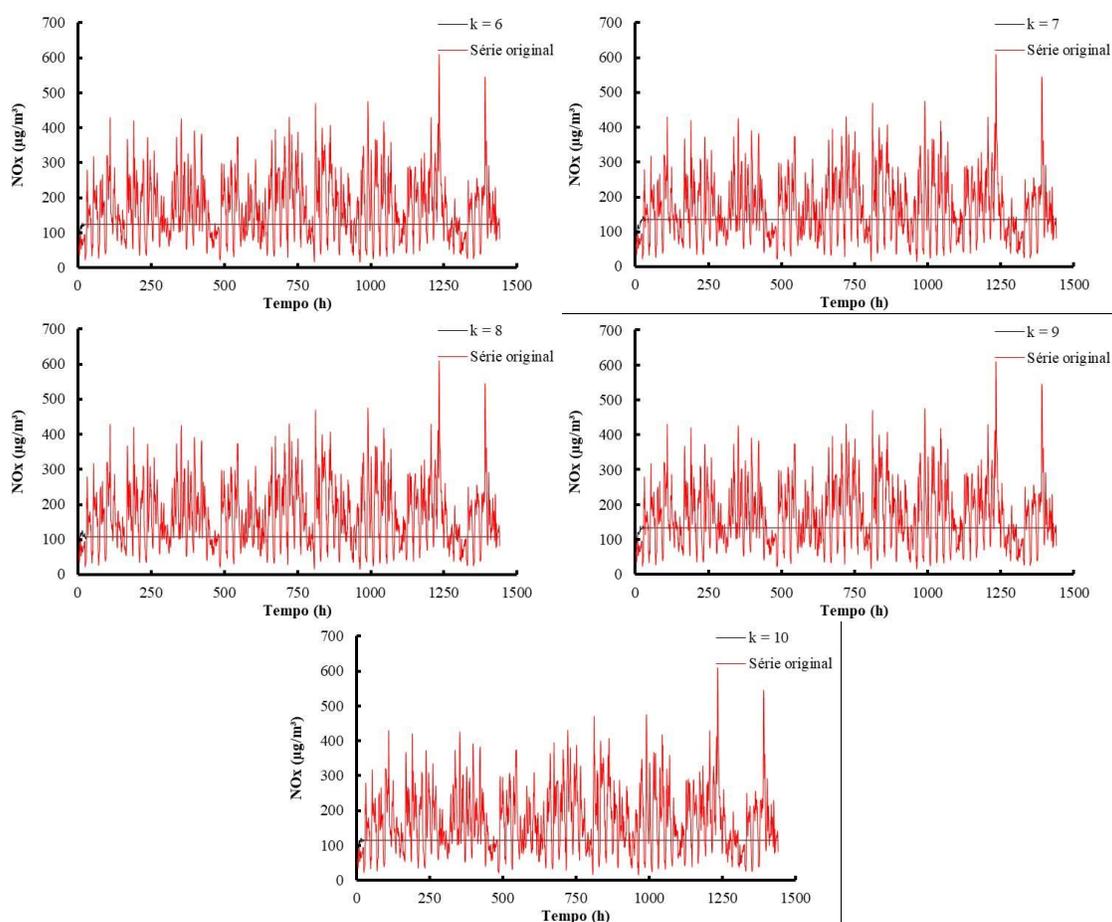
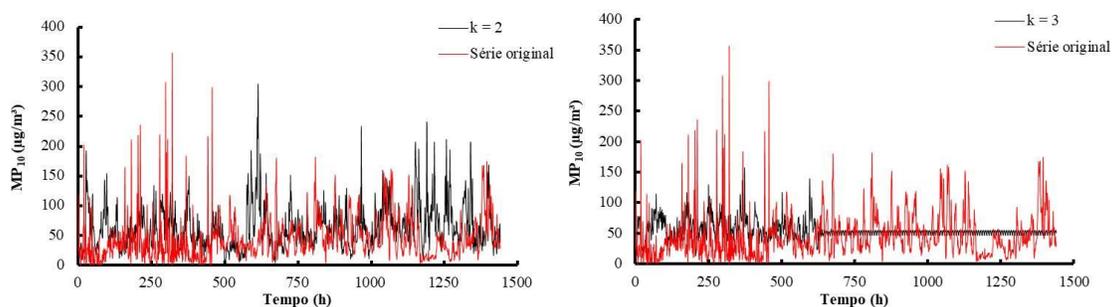


Figura D7 – Comparação das previsões *out of sample* da concentração horária de MP_{10} com as séries originais nos primeiros dois meses de 2009 na estação de Guarulhos utilizando o algoritmo k -NN para k variando de 2 a 10 (o autor, 2018).



Apêndice D – Comparação das séries simuladas e originais

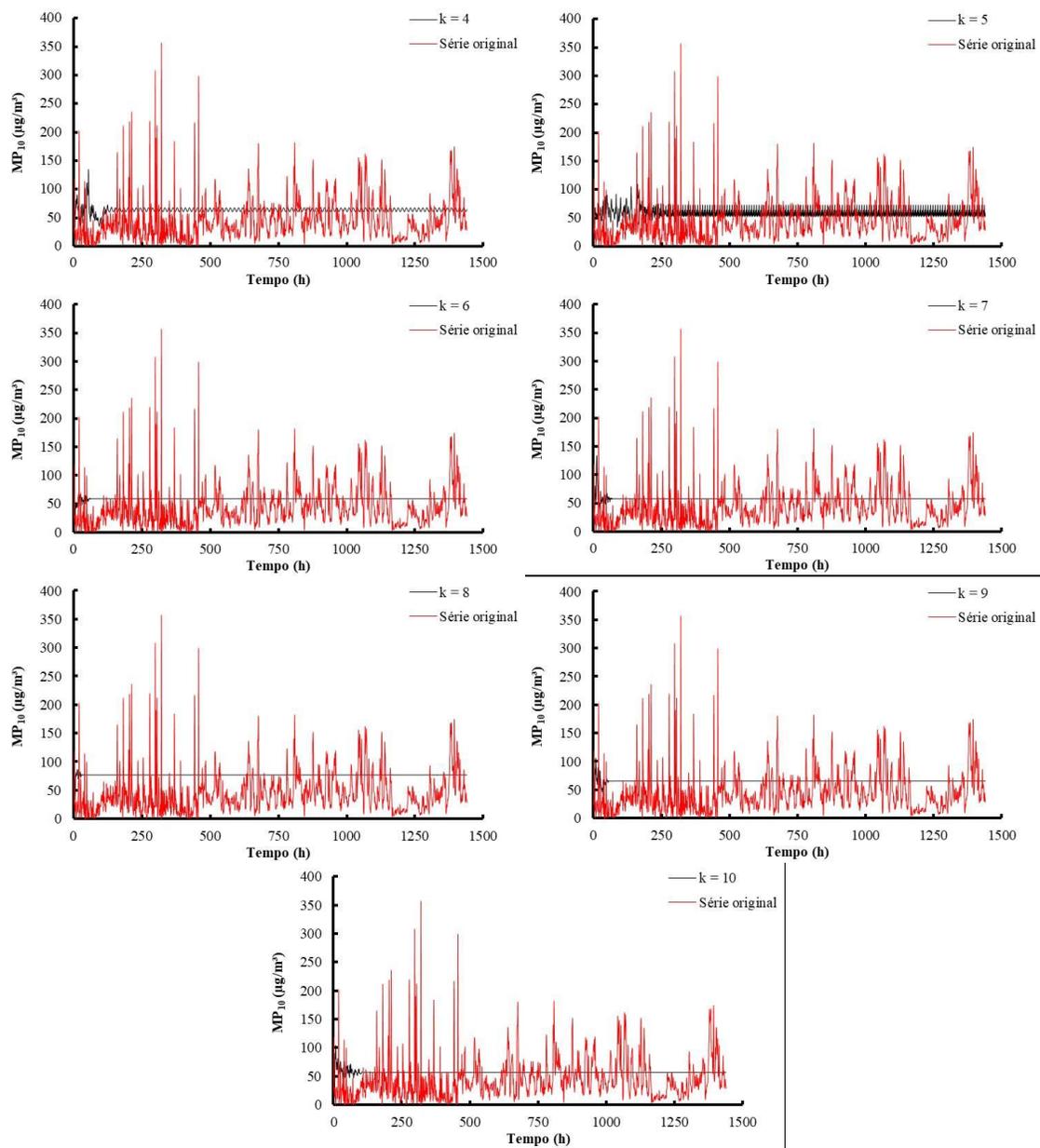
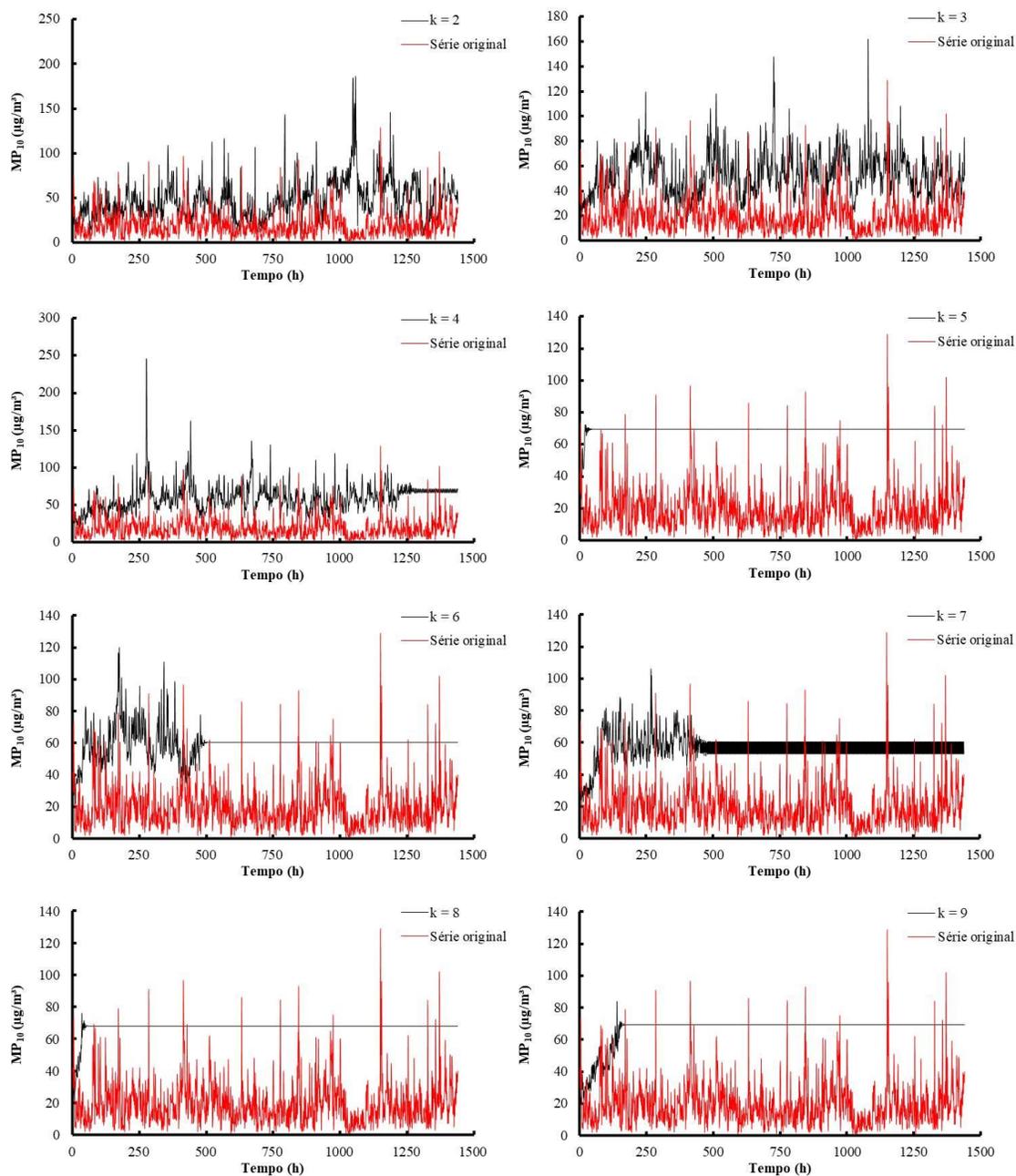


Figura D8 – Comparação das previsões *out of sample* da concentração horária de MP_{10} com as séries originais nos primeiros dois meses de 2018 na estação de Cubatão - Centro utilizando o algoritmo k -NN para k variando de 2 a 10 (o autor, 2018).



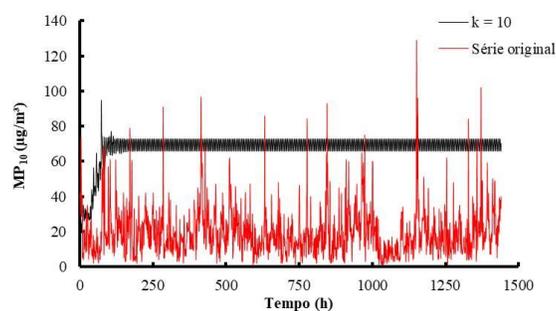
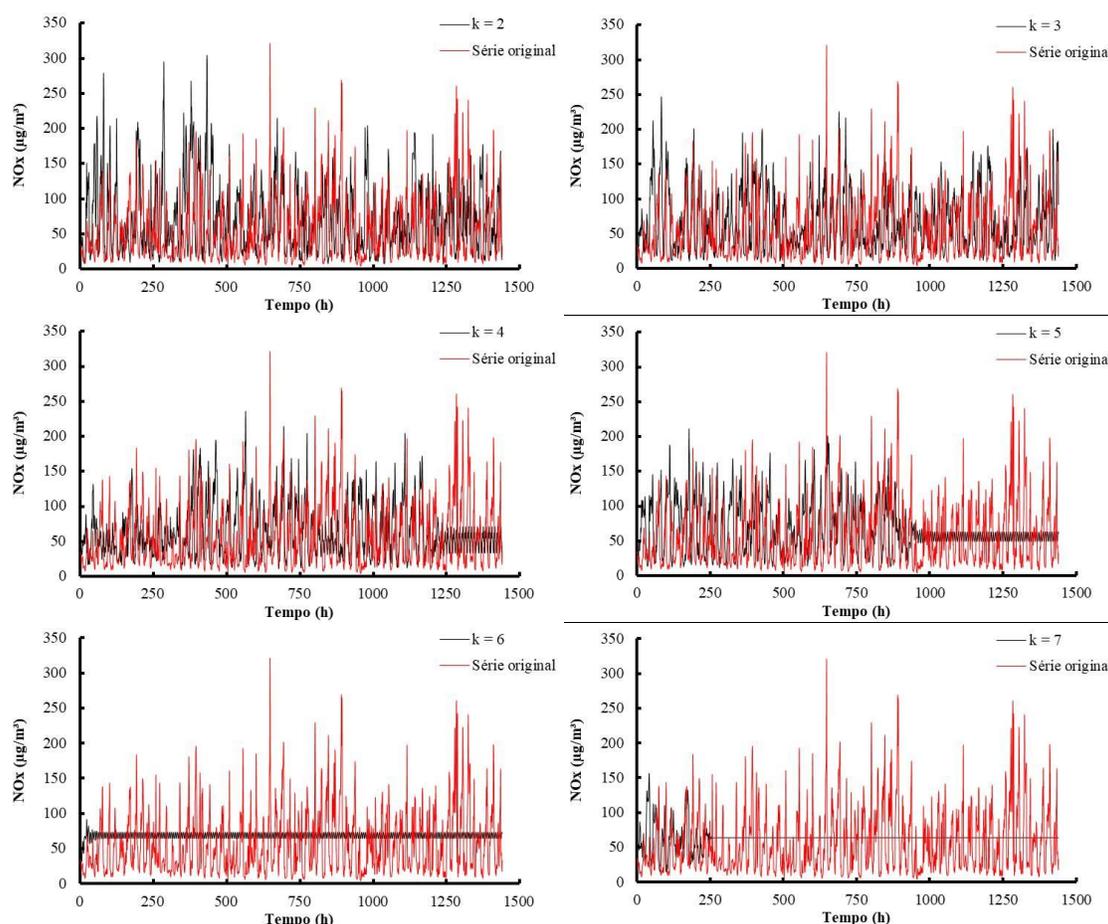


Figura D9 – Comparação das previsões *out of sample* da concentração horária de NO_x com as séries originais nos primeiros dois meses de 2018 na estação de Cubatão - Centro utilizando o algoritmo k -NN para k variando de 2 a 10 (o autor, 2018).



Apêndice D – Comparação das séries simuladas e originais

