

Sobre o Problema de Previsão de Poluentes Atmosféricos com Modelos Morfológicos-Lineares

Ricardo de A. Araújo^{1,2}, Adriano L. I. Oliveira², *Senior Member, IEEE*, Silvio Meira²

¹Departamento de Informática, Instituto Federal do Sertão Pernambucano, Ouricuri, PE, Brasil

²Centro de Informática, Universidade Federal de Pernambuco, Recife, PE, Brasil

Email: ricardo.araujo@ifsertao-pe.edu.br ou {raa,alio,srlm}@cin.ufpe.br

Resumo—Este trabalho apresenta um estudo sobre o problema de previsão de poluentes atmosféricos. Posteriormente, uma análise empírica é realizada investigando o desempenho preditivo de um modelo morfológico-linear com aprendizagem baseada em gradiente descendente, conhecido como neurônio crescente-decrescente-linear (*increasing-decreasing-linear neuron*, IDLN), para prever uma série temporal com observações da concentração diária média do poluente Ozônio (O₃), onde a estatística *U* de Theil (THEIL) e o erro percentual absoluto médio (MAPE) são utilizados para avaliação do desempenho preditivo.

I. INTRODUÇÃO

A poluição atmosférica (PA), causada por agentes que tornam o ar impróprio e nocivo a saúde humana, tem despertado um crescente interesse da sociedade, uma vez que esta impacta no bem estar e na qualidade de vida da população [1], [2]. A qualidade do ar depende de um conjunto de fatores que contribuem ou não para a dispersão dos poluentes, dentre os quais vale destacar as condições meteorológicas, a topografia e a magnitude das concentrações dos poluentes da região [3].

Dentre os poluentes encontrados na atmosfera, aqueles a apresentam um maior impacto na qualidade do ar são os aldeídos (RCHO), o dióxido de enxofre (SO₂), o dióxido de nitrogênio (NO₂), os hidrocarbonetos (HC), os materiais particulados (MP), o monóxido de carbono (CO), o ozônio (O₃), e os poluentes climáticos de vida curta (PCVC) [4]. Considerando o fato dos efeitos da PA não serem facilmente detectáveis, diversos estudos epidemiológicos apresentados na literatura indicam evidências da existência de relação entre a exposição a poluentes atmosféricos e doenças respiratórias e cardiovasculares [5].

Neste contexto, o O₃ é considerado um poluente secundário formado a partir de reações entre o NO₂ e compostos orgânicos voláteis na presença da radiação solar e, em altas concentrações, pode agravar os sintomas de diversas doenças [4]–[7]. Desta forma, a previsão da concentração de poluentes atmosféricos, em particular o O₃, representa um problema de fundamental importância para sociedade, uma vez que este possibilita a prevenção de eventos relacionados a altas concentrações do poluente, minimizando seus efeitos nocivos a saúde humana [4], [6].

Na literatura, é possível encontrar diversos trabalhos investigando modelos para previsão de fenômenos temporais relacionados a concentração de poluentes na atmosfera [1]–[4], [6], [8]. Entretanto, pesquisas recentemente realizadas por Mattos *et al.* [9] sugerem, através de um estudo empírico, que o fenômeno gerador deste tipo particular de série temporal pode ser representado através de um processo de passeio aleatório (*random walk*, RW) [10]. Este estudo investigou a utilização do procedimento de ajuste de fase (*phase fix procedure*, PFP), apresentado por Ferreira *et al.* [11], em um modelo de perceptron multicamadas (*multilayer perceptron*,

MLP) construído a partir de um método híbrido, composto de algoritmos evolucionários (*evolutionary algorithms*, EA) [12] e técnicas baseadas em gradiente descendente [13]–[16].

De acordo com Ferreira *et al.* [11], como o sucesso do PFP é condicionado a existência de distorções de fase temporais no fenômeno gerador da série temporal, caracterizado pelo comportamento de um processo RW, os resultados obtidos por Mattos *et al.* [9] indicam que o fenômeno gerador de séries temporais relacionadas a concentração de poluentes na atmosfera seguem um processo de RW. No entanto, nenhuma prova ou demonstração foram apresentadas para suportar tal conclusão, o que sugere a necessidade de um estudo adicional sobre o fenômeno gerador deste tipo particular de série.

Neste contexto, este trabalho apresenta um estudo sobre o problema de previsão de poluentes atmosféricos, onde são apresentadas evidências sugerindo que o fenômeno gerador deste tipo de série temporal não é gerado por um processo RW. Posteriormente, uma análise empírica é realizada investigando o desempenho preditivo de um modelo morfológico-linear com aprendizagem baseada em gradiente descendente, conhecido como neurônio crescente-decrescente-linear (*increasing-decreasing-linear neuron*, IDLN), para prever uma série temporal com observações da concentração diária média do poluente Ozônio (O₃), e as medidas estatística *U* de Theil (*U of Theil Statistics*, THEIL) e o erro percentual absoluto médio (*mean absolute percentage error*, MAPE) são utilizadas para avaliação dos resultados alcançados.

II. ANÁLISE DA SÉRIE TEMPORAL

Nesta seção será apresentado um estudo sobre a série temporal investigada neste trabalho

A. Ozônio O₃

A série temporal O₃ é composta por observações da concentração média diária, em partes por milhão (*ppm*), de ozônio na estação NAD27 da Agência de Proteção Ambiental do Estados Unidos (EPA), localizada em *Dania Beach* no estado americano da Flórida, no período de 01/01/1990 a 31/12/2000, que podem ser ilustradas na Figura 1.

De acordo com a Figura 1, não é possível identificar, apenas com a análise gráfica, a presença de componentes sazonais e de tendência nesta série. Devido ao fato do principal problema na caracterização do fenômeno gerador de uma série temporal ser, naturalmente, a escolha dos retardos temporais (dimensionalidade *n*), utiliza-se o gráfico *lagplot* [17] (apresentado na Figura 2) para determinar e analisar as relações entre os retardos temporais da série O₃.

A partir da análise da Figura 2, não foi possível identificar nenhuma estrutura que caracterizasse a presença de algum relacionamento de curto ou longo prazo presente nos retardos

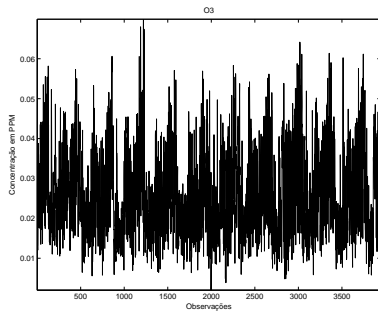


Figura 1. Gráfico da série O3.

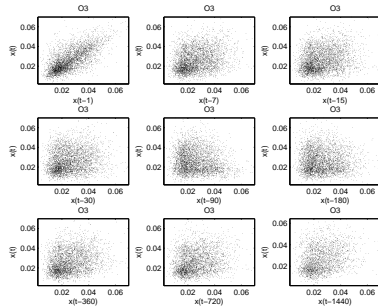


Figura 2. Lagplot da série O3.

temporais da série, o que poderia sugerir que o fenômeno gerador da série tende a um processo RW. No entanto, como o *lagplot* é fortemente dependente da interpretação humana dos gráficos e, em alguns casos, as relações contidas nestes gráficos não são humanamente simples de se identificar (a medida que a dimensionalidade n aumenta), outras técnicas devem ser consideradas. Desta forma, o teste de razão de variância (*variance ratio test*, VRT) [18] é utilizado na tentativa de verificar se o fenômeno gerador da série tende a um processo RW. O VRT estimado para a série O3 são $p=8.3027e-04$ e $h=1$, rejeitando a hipótese nula de um processo RW, a 5% de nível de significância.

Neste contexto, ambas as funções de autocorrelação (*autocorrelation function*, ACF) [19] e de autocorrelação parcial (*partial autocorrelation function*, PACF) [19], ilustradas nas Figuras 3 e 4, respectivamente, são utilizadas para analisar a existência de algum relacionamento linear entre os retardos temporais da série.

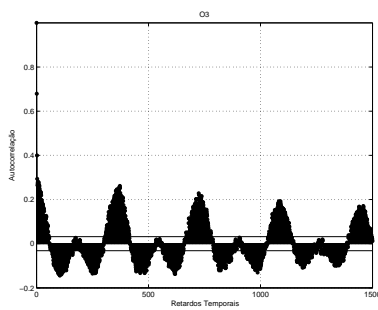


Figura 3. ACF da série O3.

Note que, de acordo com as Figuras 3 e 4, as ACF e PACF da série O3 apresentam uma combinação entre componentes de decaimento exponencial e senoidal amortecidos, o que pode vir a caracterizar algum relacionamento não-linear presente no

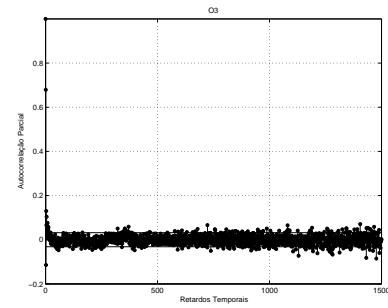


Figura 4. PACF da série O3.

fenômeno gerador da série. Desta forma, a informação mútua média (*mean mutual information*, MMI) [20], [21], ilustrada na Figura 5, é utilizada para verificar a existência de algum relacionamento não-linear entre os retardos temporais da série, uma vez que a MMI representa uma medida de dependência não-linear.

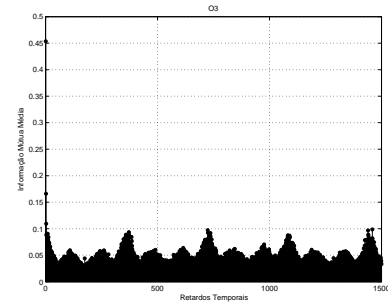


Figura 5. MMI da série O3.

De acordo com a Figura 5, é possível verificar a existência de dependência não-linear de curto prazo na série O3, uma vez que o primeiro mínimo da curva se encontra em retardos temporais de baixa ordem. Também, vale mencionar que a curva do gráfico sugere a existência de uma dinâmica periódica na série. Neste contexto, investigou-se o parâmetro de Hurst (*Hurst parameter*, HP) [22] para determinar, de maneira quantitativa, se o fenômeno gerador da série O3 tende a um processo auto-similar com dependência de curto ($0.5 < HP \leq 1$) ou longo ($0 \leq HP < 0.5$) prazos [23]. Vale mencionar que o HP também pode ser utilizado para verificar se o fenômeno gerador da série tende a um processo de RW ($HP=0.5$).

Neste contexto, o HP estimado para a série O3 foi de 0.6435, o que confirma a suposição que a série em questão é um processo auto-similar com dependência de curto prazo (série persistente, isto é, uma tendência crescente no tempo passado é mais provável de se manter em uma tendência crescente, e vice versa).

III. O MODELO MORFOLÓGICO-LINEAR

O modelo proposto, referido como neurônio crescente decrescente linear (*increasing decreasing linear neuron*, IDLN), consiste de uma combinação entre operadores não-lineares crescentes (dilatação e erosão), operadores não-lineares decrescentes (anti-dilatação e anti-erosão) e um operador linear (*perceptron linear*). A seguir será apresentada a definição formal do IDLN.

Seja $\mathbf{x} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ um padrão de entrada, representado pelos retardos temporais de uma série temporal,

sobre uma i -ésima janela deslizante, e seja y a saída, representada pela previsão, do IDLN, com regra de transformação local do tipo $\mathbf{x} \rightarrow y$ e formalmente definido por

$$y = \lambda\alpha + (1 - \lambda)\beta, \quad \lambda \in [0, 1], \quad (1)$$

onde

$$\beta = \sum_{i=1}^n x_i p_i, \quad (2)$$

e

$$\alpha = \theta\tau + (1 - \theta)\kappa, \quad \theta \in [0, 1], \quad (3)$$

em que

$$\tau = \varphi\delta + (1 - \varphi)\varepsilon, \quad \varphi \in [0, 1], \quad (4)$$

e

$$\kappa = \omega\bar{\delta} + (1 - \omega)\bar{\varepsilon}, \quad \omega \in [0, 1], \quad (5)$$

com

$$\delta = \delta_{\mathbf{a}}(\mathbf{x}) = \bigvee_{i=1}^n (x_i + a_i), \quad (6)$$

$$\varepsilon = \varepsilon_{\mathbf{b}}(\mathbf{x}) = \bigwedge_{i=1}^n (x_i + b_i), \quad (7)$$

$$\bar{\delta} = \bar{\delta}_{\mathbf{c}}(\mathbf{x}) = \bigwedge_{i=1}^n (x_i^* + c_i), \quad (8)$$

$$\bar{\varepsilon} = \bar{\varepsilon}_{\mathbf{d}}(\mathbf{x}) = \bigvee_{i=1}^n (x_i^* + d_i), \quad (9)$$

onde o termo n representa a dimensionalidade do padrão de entrada (\mathbf{x}), os termos $\lambda, \theta, \varphi, \omega \in \mathbb{R}$, e os termos $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}, \mathbf{p} \in \mathbb{R}^n$. O vetor \mathbf{p} representa os coeficientes (pesos) do operador linear. O termo β representa a saída do operador linear (módulo linear). O termo α representa o módulo não-linear, dado pela combinação linear (o termo de combinação é dado por θ) entre o módulo não-linear crescente (definido por τ) e o módulo não-linear decrescente (definido por κ). O termo τ representa a combinação linear (o termo de combinação é dado por φ) entre o operador morfológico de dilatação (definido por δ) e o operador morfológico de erosão (definido por ε). O termo κ representa a combinação linear (o termo de combinação é dado por ω) entre o operador morfológico de anti-dilatação (definido por $\bar{\delta}$) e o operador morfológico de anti-erosão (definido por $\bar{\varepsilon}$). Os vetores $\mathbf{a}, \mathbf{b}, \mathbf{c}$ e \mathbf{d} representam, respectivamente, os elementos estruturantes (pesos) dos operadores de dilatação ($\delta_{\mathbf{a}}(\mathbf{x})$) e erosão ($\varepsilon_{\mathbf{b}}(\mathbf{x})$) (empregados no módulo não-linear crescente) e de anti-dilatação ($\bar{\delta}_{\mathbf{c}}(\mathbf{x})$) e anti-erosão ($\bar{\varepsilon}_{\mathbf{d}}(\mathbf{x})$) (empregados no módulo não-linear decrescente). Note que a saída y do IDLN é dada por uma combinação linear entre o módulo linear e o módulo não-linear (o termo de combinação da saída do IDLN é dado por λ).

A. Processo de Aprendizagem

A partir da definição do modelo IDLN, é possível verificar que este requer o ajuste dos parâmetros $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}, \mathbf{p} \in \mathbb{R}^n$ e $\lambda, \theta, \varphi, \omega \in \mathbb{R}$. Portanto, o vetor de pesos \mathbf{w} (note que $\mathbf{w} \in \mathbb{R}^{5n+4}$) do modelo IDLN é definido por

$$\mathbf{w} = (\lambda, \theta, \varphi, \omega, \mathbf{p}, \mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}). \quad (10)$$

Na fase de treinamento, os pesos do modelo IDLN são ajustados de acordo com um critério de erro até a convergência, isto é, até o critério de parada ser alcançado. Portanto, é

necessário definir uma função objetivo, em termos dos pesos $J(\mathbf{w})$, a ser minimizada durante a fase de treinamento (representada pelo erro de previsão do modelo IDLN utilizando o vetor de pesos \mathbf{w}), sendo definida por

$$J(\mathbf{w}) = \frac{1}{M} \sum_{m=1}^M e^2(m), \quad (11)$$

onde M representa a quantidade de padrões de treinamento e $e(m)$ representa o erro instantâneo do modelo para o m -ésimo padrão de treinamento.

Neste trabalho, um método de gradiente descendente utilizando ideias do algoritmo de retropropagação do erro (*back propagation*, BP) [24] é proposto para ajustar o vetor de pesos do modelo. Portanto, o processo de aprendizagem do modelo IDLN consiste na atualização iterativa do vetor de pesos \mathbf{w} baseada no método do gradiente descendente. O ajuste do vetor de pesos \mathbf{w} para o m -ésimo padrão de treinamento é dado por

$$\mathbf{w}(i+1) = \mathbf{w}(i) - \mu \nabla J(\mathbf{w}), \quad (12)$$

em que $i \in \{1, 2, \dots\}$ e o termo μ representa a taxa de aprendizagem do modelo. O termo $\nabla J(\mathbf{w})$ é dado pelo gradiente de J em relação a \mathbf{w} e dado por

$$\nabla J(\mathbf{w}) = \frac{\partial J}{\partial y} \frac{\partial y}{\partial \mathbf{w}} = -2e(m) \frac{\partial y}{\partial \mathbf{w}}, \quad (13)$$

em que

$$\frac{\partial y}{\partial \mathbf{w}} = \left(\frac{\partial y}{\partial \lambda}, \frac{\partial y}{\partial \theta}, \frac{\partial y}{\partial \varphi}, \frac{\partial y}{\partial \omega}, \frac{\partial y}{\partial \mathbf{p}}, \frac{\partial y}{\partial \mathbf{a}}, \frac{\partial y}{\partial \mathbf{b}}, \frac{\partial y}{\partial \mathbf{c}}, \frac{\partial y}{\partial \mathbf{d}} \right) \quad (14)$$

O termo $\frac{\partial y}{\partial \lambda}$ é calculado por

$$\frac{\partial y}{\partial \lambda} = \alpha - \beta. \quad (15)$$

O termo $\frac{\partial y}{\partial \theta}$ é calculado por

$$\frac{\partial y}{\partial \theta} = \frac{\partial y}{\partial \alpha} \frac{\partial \alpha}{\partial \theta}, \quad (16)$$

em que

$$\frac{\partial y}{\partial \alpha} = \lambda, \quad (17)$$

e

$$\frac{\partial \alpha}{\partial \theta} = \tau - \kappa. \quad (18)$$

O termo $\frac{\partial y}{\partial \varphi}$ é calculado por

$$\frac{\partial y}{\partial \varphi} = \frac{\partial y}{\partial \alpha} \frac{\partial \alpha}{\partial \tau} \frac{\partial \tau}{\partial \varphi} = \lambda \frac{\partial \alpha}{\partial \tau} \frac{\partial \tau}{\partial \varphi}, \quad (19)$$

em que

$$\frac{\partial \alpha}{\partial \tau} = \theta, \quad (20)$$

e

$$\frac{\partial \tau}{\partial \varphi} = \delta - \varepsilon. \quad (21)$$

O termo $\frac{\partial y}{\partial \omega}$ é calculado por

$$\frac{\partial y}{\partial \omega} = \frac{\partial y}{\partial \alpha} \frac{\partial \alpha}{\partial \kappa} \frac{\partial \kappa}{\partial \omega} = \lambda \frac{\partial \alpha}{\partial \kappa} \frac{\partial \kappa}{\partial \omega}, \quad (22)$$

em que

$$\frac{\partial \alpha}{\partial \kappa} = 1 - \theta, \quad (23)$$

e

$$\frac{\partial \kappa}{\partial \omega} = \bar{\delta} - \bar{\varepsilon}. \quad (24)$$

O termo $\frac{\partial y}{\partial \mathbf{p}}$ é calculado por

$$\frac{\partial y}{\partial \mathbf{p}} = \frac{\partial y}{\partial \beta} \frac{\partial \beta}{\partial \mathbf{p}}, \quad (25)$$

onde

$$\frac{\partial y}{\partial \beta} = 1 - \lambda, \quad (26)$$

e

$$\frac{\partial \beta}{\partial \mathbf{p}} = \mathbf{x}, \quad (27)$$

em que \mathbf{x} representa o m -ésimo padrão de treinamento.

O termo $\frac{\partial y}{\partial \mathbf{a}}$ pode ser calculado por:

$$\frac{\partial y}{\partial \mathbf{a}} = \frac{\partial y}{\partial \alpha} \frac{\partial \alpha}{\partial \tau} \frac{\partial \tau}{\partial \delta} \frac{\partial \delta}{\partial \mathbf{a}} = \lambda \theta \frac{\partial \tau}{\partial \delta} \frac{\partial \delta}{\partial \mathbf{a}}, \quad (28)$$

em que

$$\frac{\partial \tau}{\partial \delta} = \varphi, \quad (29)$$

e

$$\frac{\partial \delta}{\partial \mathbf{a}} = \frac{Q_\sigma(\delta \cdot \mathbf{1} - (\mathbf{x} + \mathbf{a}))}{Q_\sigma(\delta \cdot \mathbf{1} - (\mathbf{x} + \mathbf{a})) \cdot \mathbf{1}^T}. \quad (30)$$

O termo $\frac{\partial y}{\partial \mathbf{b}}$ pode ser calculado por:

$$\frac{\partial y}{\partial \mathbf{b}} = \frac{\partial y}{\partial \alpha} \frac{\partial \alpha}{\partial \tau} \frac{\partial \tau}{\partial \varepsilon} \frac{\partial \varepsilon}{\partial \mathbf{b}} = \lambda \theta \frac{\partial \tau}{\partial \varepsilon} \frac{\partial \varepsilon}{\partial \mathbf{b}}, \quad (31)$$

em que

$$\frac{\partial \tau}{\partial \varepsilon} = 1 - \varphi, \quad (32)$$

e

$$\frac{\partial \varepsilon}{\partial \mathbf{b}} = \frac{Q_\sigma(\varepsilon \cdot \mathbf{1} - (\mathbf{x} + \mathbf{b}))}{Q_\sigma(\varepsilon \cdot \mathbf{1} - (\mathbf{x} + \mathbf{b})) \cdot \mathbf{1}^T}. \quad (33)$$

O termo $\frac{\partial y}{\partial \mathbf{c}}$ pode ser calculado por:

$$\frac{\partial y}{\partial \mathbf{c}} = \frac{\partial y}{\partial \alpha} \frac{\partial \alpha}{\partial \kappa} \frac{\partial \kappa}{\partial \bar{\delta}} \frac{\partial \bar{\delta}}{\partial \mathbf{c}} = \lambda(1 - \theta) \frac{\partial \kappa}{\partial \bar{\delta}} \frac{\partial \bar{\delta}}{\partial \mathbf{c}}, \quad (34)$$

em que

$$\frac{\partial \kappa}{\partial \bar{\delta}} = \omega, \quad (35)$$

e

$$\frac{\partial \bar{\delta}}{\partial \mathbf{c}} = \frac{Q_\sigma(\bar{\delta} \cdot \mathbf{1} - (\mathbf{x}^* + \mathbf{c}))}{Q_\sigma(\bar{\delta} \cdot \mathbf{1} - (\mathbf{x}^* + \mathbf{c})) \cdot \mathbf{1}^T}. \quad (36)$$

O termo $\frac{\partial y}{\partial \mathbf{d}}$ pode ser calculado por:

$$\frac{\partial y}{\partial \mathbf{d}} = \frac{\partial y}{\partial \alpha} \frac{\partial \alpha}{\partial \kappa} \frac{\partial \kappa}{\partial \bar{\varepsilon}} \frac{\partial \bar{\varepsilon}}{\partial \mathbf{d}} = \lambda(1 - \theta) \frac{\partial \kappa}{\partial \bar{\varepsilon}} \frac{\partial \bar{\varepsilon}}{\partial \mathbf{d}}, \quad (37)$$

em que

$$\frac{\partial \kappa}{\partial \bar{\varepsilon}} = 1 - \omega, \quad (38)$$

e

$$\frac{\partial \bar{\varepsilon}}{\partial \mathbf{d}} = \frac{Q_\sigma(\bar{\varepsilon} \cdot \mathbf{1} - (\mathbf{x}^* + \mathbf{d}))}{Q_\sigma(\bar{\varepsilon} \cdot \mathbf{1} - (\mathbf{x}^* + \mathbf{d})) \cdot \mathbf{1}^T}. \quad (39)$$

Vale mencionar a função de impulso suave $Q_\sigma(\mathbf{x}) = [q_\sigma(x_1), q_\sigma(x_2), \dots, q_\sigma(x_n)]$, é dada por

$$q_\sigma(x_i) = \exp\left[-\frac{1}{2} \left(\frac{x_i}{\sigma}\right)^2\right], \quad \forall i = 1, \dots, n. \quad (40)$$

IV. RESULTADOS EXPERIMENTAIS

A série temporal O3 foi normalizada no intervalo $[0, 1]$ e dividida em três conjuntos: treinamento (50% dos dados), validação (25% dos dados) e teste (25% dos dados). Vale mencionar que o procedimento de ajuste de fase foi utilizado diretamente no processo de aprendizagem do modelo IDLN para ajustar possíveis distorções de fase temporais no fenômeno gerador da série.

A medida estatística U de Theil (*U of Theil Statistics*, THEIL) foi utilizada para avaliar o desempenho preditivo, uma vez que esta permite avaliar o desempenho do modelo proposto em relação a um processo RW, e dada por

$$\text{THEIL} = \frac{\sum_{j=1}^{N-1} (x_j - \hat{x}_j)^2}{\sum_{j=1}^{N-1} (x_j - x_{j-1})^2}, \quad (41)$$

em que N representa a quantidade de padrões, e os termos x_j e \hat{x}_j representam a saída desejada e o valor previsto, respectivamente, para o j -ésimo padrão de entrada.

Também, o erro percentual absoluto médio (*mean absolute percentage error*, MAPE) foi utilizado para avaliação do desempenho preditivo, uma vez que este permite identificar precisamente os desvios percentuais do modelo, e dado por

$$\text{MAPE} = \frac{100}{N} \sum_{j=1}^N \left| \frac{x_j - \hat{x}_j}{x_j} \right|. \quad (42)$$

Para realização dos experimentos com o modelo proposto, é necessário definir uma arquitetura básica, dada pela notação IDLN (*lags;μ;σ*), em que o termo *lags* define os retardos temporais utilizados para representar o fenômeno temporal, o termo μ representa a taxa de aprendizagem, e o termo σ representa o fator de escala da função de impulso suave. A escolha dos retardos temporais foi baseada na análise da série, onde foram fixados os retardos temporais 2 até 53 para todos os experimentos, uma vez que o primeiro ponto de mínimo da MMI foi encontrado neste retardo temporal. Além disso, vale mencionar que o primeiro retardo temporal não é utilizado porque é necessário se criar uma estrutura de RW para a utilização do procedimento de ajuste de fase.

O valor de μ e σ são determinados empiricamente através de uma série de experimentos. Os valores iniciais do vetor de pesos do modelo IDLN são: $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}, \mathbf{p} \in [-1, 1]$ e $\lambda, \theta, \varphi, \omega \in [0, 1]$. No processo de treinamento do IDLN três condições de parada são utilizadas: i) quantidade máxima de épocas de treinamento (10^4), ii) $Pt \leq 10^{-6}$, e iii) $Gl \geq 5\%$. Todos os experimentos com o modelo IDLN foram desenvolvidos e implementados utilizando a ferramenta *Matlab*. Por fim, para cada arquitetura investigada, são realizados cinquenta experimentos, onde são calculadas as estatísticas básicas: raiz do erro quadrático médio (*root mean square*, RMS), mediana

(MEDIAN) e percentis 2.5% (PERC25) e 97.5% (PERC975). A seguir, serão apresentados os resultados experimentais com a série temporal investigada neste trabalho.

A. Série O3

O desempenho preditivo dos experimentos realizados com o modelo IDLN para a série O3 são apresentados nas Tabelas I e II. A Figura 6 ilustra o gráfico de regressão obtido pelo melhor experimento com o modelo IDLN para a série O3. O histograma, a ACF e a PACF dos resíduos obtidos pelo melhor experimento com o modelo IDLN para a série O3 são apresentados nas Figuras 7, 8 e 9, respectivamente. O teste de Kolmogorov-Smirnov ($p=1.6060e-215$ e $h=1$) indicou a rejeição da hipótese nula, a 5% de nível de significância, dos resíduos serem gerados a partir de uma distribuição normal padrão, bem como o teste Q de Ljung-Box ($p=7.0470e-02$ e $h=0$) indicou a inexistência de evidências para a rejeição da hipótese nula, a 5% de nível de significância, dos resíduos não serem autocorrelacionados.

Tabela I. RESULTADOS OBTIDOS DA MEDIDA THEIL PARA A SÉRIE O3 (CONJUNTO DE TESTE).

Parâmetros		THEIL			
μ	σ	RMS	MEDIAN	PERC25	PERC975
0.100	0,500	0,0035	0,0018	0,0017	0,0070
	0,050	0,0004	0,0004	0,0004	0,0005
	0,005	0,0002	0,0002	0,0002	0,0002
0.010	0,500	0,0048	0,0048	0,0043	0,0052
	0,050	0,0013	0,0013	0,0012	0,0014
	0,005	0,0009	0,0009	0,0009	0,0010
0.001	0,500	0,0072	0,0074	0,0058	0,0079
	0,050	0,0044	0,0044	0,0043	0,0046
	0,005	0,0041	0,0041	0,0041	0,0042

Tabela II. RESULTADOS OBTIDOS DA MEDIDA MAPE PARA A SÉRIE O3 (CONJUNTO DE TESTE).

Parâmetros		MAPE			
μ	σ	RMS	MEDIAN	PERC25	PERC975
0.100	0,500	1,0530	0,8042	0,7730	1,7494
	0,050	0,4227	0,4247	0,3567	0,5007
	0,005	0,1762	0,1772	0,1619	0,1863
0.010	0,500	1,8221	1,8178	1,7092	1,9152
	0,050	0,8164	0,8103	0,7695	0,8622
	0,005	0,5933	0,5819	0,5547	0,6357
0.001	0,500	2,4105	2,4607	2,0980	2,5509
	0,050	1,1794	1,1716	1,1208	1,2619
	0,005	1,0785	1,0801	1,0500	1,1237

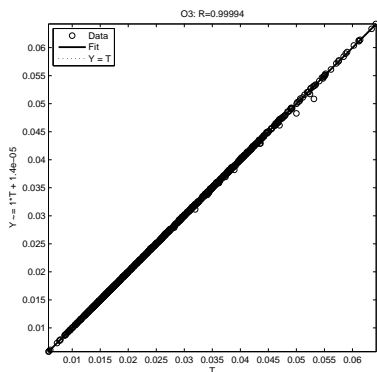


Figura 6. Gráfico de regressão da série O3 (conjunto de teste).

V. CONCLUSÕES

Este trabalho apresentou um estudo sobre o problema de previsão de poluentes atmosféricos. Neste contexto, foi utilizada a série temporal O3, composta por observações da

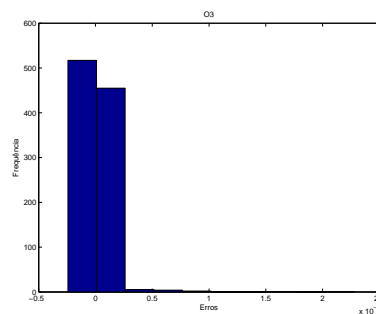


Figura 7. Histograma dos resíduos da série O3 (conjunto de teste).

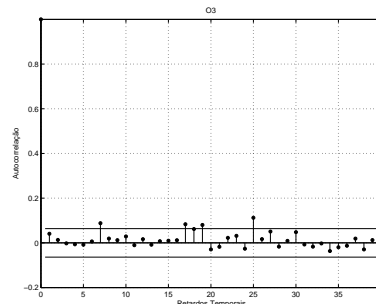


Figura 8. ACF dos resíduos da série O3 (conjunto de teste).

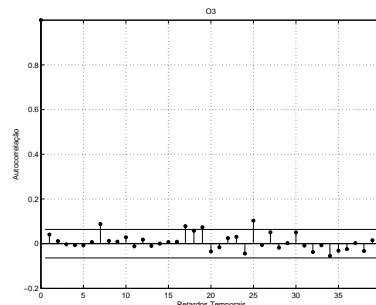


Figura 9. PACF dos resíduos da série O3 (conjunto de teste).

concentração média diária de ozônio na estação NAD27 da Agência de Proteção Ambiental do Estados Unidos (EPA), localizada em *Dania Beach* no estado americano da Flórida.

Inicialmente, este trabalho considerou realizar uma análise gráfica da série. Esta não foi capaz de indentificar quaisquer componentes de tendência ou sazonalidade. Como o principal problema na reconstrução do fenômeno gerador de uma série temporal é a caracterização do relacionamento entre os retardos temporais, este trabalho investigou o gráfico *lagplot* da série. Este não foi capaz de revelar a presença de nenhum relacionamento com dependência de curto ou longo prazo presente nos retardos temporais, o que poderia sugerir que o fenômeno gerador da série pudesse tender a um processo RW.

Considerando o fato que o *lagplot* é fortemente dependente da interpretação humana dos gráficos, este trabalho investigou o teste de razão de variância para verificar se o fenômeno gerador da série tende a um processo RW. Este rejeitou a hipótese nula de um processo RW, a 5% de nível de significância. Neste contexto, este trabalho analisou as funções de autocorrelação e autocorrelação parcial da série. Estas revelaram a existência de componentes de decaimento exponencial e senoidal amortecidos, sugerindo a possibilidade de haver algum relacionamento

não-linear presente no fenômeno gerador da série.

Posteriormente, foi analisada a informação mútua média da série. Esta revelou uma dependência não-linear de curto prazo, uma vez que o primeiro mínimo da curva pode ser encontrado em retardos temporais de baixa ordem. Além disso, a curva da informação mútua média apresentou indicações de uma dinâmica periódica na série. Por fim, o parâmetro de Hurst revelou, de maneira quantitativa, que o fenômeno gerador da série segue um processo auto-similar com dependência de curto prazo.

Baseado nas evidências encontradas na análise do fenômeno gerador da série O3, este trabalho investigou a utilização do modelo IDLN como solução para este tipo particular de problema. De acordo com os resultados obtidos com a medida THEIL, pode-se verificar que todos os experimentos realizados com o modelo proposto não tenderam a um processo RW, bem como os resultados alcançados com a medida MAPE mostraram que o modelo proposto obteve desvios percentuais inferiores a 2.5%.

Note que, analisando a medida THEIL, não é possível identificar precisamente quais os valores de μ e σ que otimizam o desempenho preditivo do modelo IDLN, uma vez que, independente da configuração do modelo, foram alcançados valores próximos de zero. Desta forma, os resultados obtidos com a medida MAPE serão utilizados para determinação dos melhores valores para μ e σ . A análise da estatística RMS demonstrou que a configuração $\mu=0.1$ e $\sigma=0.005$ apresenta os resultados mais estáveis e de melhor qualidade preditiva. Nesta configuração, a estatística PERC975 revelou que para maioria dos experimentos (97.5%), o valor da medida MAPE foi próximo de 0.186%. Também, a estatística PERC25 mostrou que para os melhores experimentos (2.5%), o valor da medida MAPE foi próximo de 0.161%. Por fim, a estatística MEDIAN revelou que pelo menos 50% dos experimentos alcançaram valores da medida MAPE próximos de 0.177%.

Neste contexto, o melhor experimento do modelo IDLN, com configuração $\mu=0.1$ e $\sigma=0.005$, foi escolhido para realização da análise de regressão e de resíduo. Inicialmente, foi investigado o gráfico de regressão para o conjunto de teste. Este não revelou dispersão significativa entre os valores reais e os preditos ($R=0.99994$), demonstrando um desempenho de previsão bastante acurado do modelo IDLN. Posteriormente, foi realizada a análise do histograma. Este sugeriu que os resíduos não seguem uma distribuição normal padrão. Para confirmar esta suposição, foi aplicado o teste de Kolmogorov-Smirnov, que rejeitou a hipótese nula dos resíduos serem gerados por uma distribuição normal padrão, a 5% de nível de significância. Posteriormente, foram investigadas as funções de autocorrelação e autocorrelação parcial. Estas sugeriram que os resíduos não são autocorrelacionados. Para confirmar esta suposição, foi aplicado o teste Q de Ljung-Box. Este não encontrou evidências para rejeitar a hipótese nula dos resíduos não serem autocorrelacionados. Desta forma, é possível verificar, pela análise de regressão e de resíduo, que o modelo IDLN é apropriado para estimar o fenômeno gerador da série O3, investigada neste trabalho.

Como trabalhos futuros, pretende-se realizar uma análise extensiva, utilizando outras séries temporais relacionadas a concentração de poluentes atmosféricos, para confirmar os resultados obtidos neste trabalho.

REFERÊNCIAS

- [1] D. Voukantsis, K. Karatzas, J. Kukkonen, T. Rasanen, A. Karppinen, and M. Kolehmainen, "Intercomparison of air quality data using principal component analysis, and forecasting of pm10 and pm2.5 concentrations using artificial neural networks, in thessaloniki and helsinki," *Science of the Total Environment*, vol. 409, pp. 1266–1276, 2011.
- [2] A. Vlachogianni, P. Kassomenos, A. Karppinen, S. Karakitsios, and J. Kukkonen, "Evaluation of a multiple regression model for the forecasting of the concentrations of nox and pm10 in athens and helsinki," *Science of the Total Environment*, vol. 409, pp. 1559–1571, 2011.
- [3] H. J. S. Fernando, M. C. Mammarella, G. Grandoni, P. Fedele, R. D. Marco, R. Dimitrova, and P. Hyde, "Forecasting pm10 in metropolitan areas: Efficacy of neural networks," *Environmental Pollution*, vol. 163, pp. 62–67, 2012.
- [4] S. Gocheva-Ilieva, A. Ivanov, D. Voynikova, and D. Boyadzhiev, "Time series analysis and forecasting for air pollution in small urban area: an sarima and factor analysis approach," *Stochastic Environmental Research and Risk Assessment*, vol. 28, no. 4, pp. 1045–1060, 2014.
- [5] N. Gouveia and T. Fletcher, "Time series analysis of air pollution and mortality: Effects by cause, age and socioeconomic status," *Journal of Epidemiology and Community Health*, vol. 54, no. 10, pp. 750–755, 2000.
- [6] M. Ismail, M. Z. Ibrahim, T. A. Ibrahim, and A. M. Abdullah, "Time series analysis of surface ozone monitoring records in kemaman, malaysia," *Sains Malaysiana*, vol. 40, no. 5.
- [7] J. N. Latosinska, M. Latosinska, and J. Bielik, "Towards analysis and predicting maps of ultraviolet index from experimental astronomical parameters (solar elevation, total ozone level, aerosol index, reflectivity). artificial neural networks global scale approach," *Aerospace Science and Technology*, vol. 43, no. 0, pp. 301 – 313, 2015.
- [8] K. Padma, R. S. Selvaraj, and B. M. Boaz, "Estimation of daily surface ozone using periodic and stochastic modeling in chennai region," *Indian Journal of Radio and Space Physics*, vol. 43, no. 4-5, pp. 293–302, 2014.
- [9] P. S. G. Mattos, F. Madeiro, T. A. E. Ferreira, and G. D. C. Cavalcanti, "Hybrid intelligent system for air quality forecasting using phase adjustment," *Engineering Applications of Artificial Intelligence*, vol. 32, pp. 185–191, 2014.
- [10] R. Sitte and J. Sitte, "Neural networks approach to the random walk dilemma of financial time series," *Applied Intelligence*, vol. 16, no. 3, pp. 163–171, May 2002.
- [11] T. A. E. Ferreira, G. C. Vasconcelos, and P. J. L. Adeodato, "A new intelligent system methodology for time series forecasting with artificial neural networks," in *Neural Processing Letters*, vol. 28, 2008, pp. 113–129.
- [12] F. H. F. Leung, H. K. Lam, S. H. Ling, and P. K. S. Tam, "Tuning of the structure and parameters of the neural network using an improved genetic algorithm," *IEEE Transactions on Neural Networks*, vol. 14, no. 1, pp. 79–88, January 2003.
- [13] M. Riedmiller and H. Braun, "A direct adaptive method for faster backpropagation learning: The rprop algorithm," in *Proceedings of the IEEE Int. Conf. on Neural Networks (ICNN)*, San Francisco, 1993, pp. 586–591.
- [14] M. Hagan and M. Menhaj, "Training feedforward networks with the marquardt algorithm," *IEEE Transactions on Neural Networks*, vol. 5, no. 6, pp. 989–993, November 1994.
- [15] M. F. Moller, "A scaled conjugate gradient algorithm for fast supervised learning," *Neural Networks*, vol. 6, pp. 525–533, 1993.
- [16] R. Battiti, "One step secant conjugate gradient," *Neural Computation*, vol. 4, pp. 141–166, 1992.
- [17] H. Kantz and T. Schreiber, *Nonlinear Time Series analysis*, 2nd ed. New York, NY, USA: Cambridge University Press, 2003.
- [18] A. W. Lo and A. C. MacKinlay, *A Non-Random Walk Down Wall St. Princeton*. Princeton University Press, 2001.
- [19] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, *Time Series Analysis: Forecasting and Control*, 3rd ed. New Jersey: Prentice Hall, 1994.
- [20] A. Fraser and H. Swinney, "Independent coordinates for strange attractors from mutual information," *Physical Review A*, vol. 33, no. 2, pp. 1134–1140, 1986.
- [21] A. Kraskov, H. Stgbauer, and P. Grassberger, "A new auto-associative memory based on lattice algebra," *Phys. Rev. E*, vol. 69, no. 6, 2004.
- [22] E. Hurst, "Long term storage capacity of reservoirs," *Transactions of the American Society of Civil Engineers*, vol. 116, pp. 770–799, 1951.
- [23] J. M. P. Menezes and G. A. Barreto, "Long-term time series prediction with the narx network: An empirical evaluation," *Neurocomputing*, vol. 71, no. 16-18, pp. 3335–3343, 2008.
- [24] S. Haykin, *Neural networks: A comprehensive foundation*. New Jersey: Prentice Hall, 1998.