

# CONJUNÇÃO DAS REDES NEURAIS ARTIFICIAIS COM A TRANSFORMADA WAVELET DISCRETA NO PROGNÓSTICO DA QUALIDADE DO AR.

B. C. V. da SILVA<sup>1</sup>, F. de O. CARVALHO<sup>1</sup>, L. F. Q. ROCHA e T. P. da SILVA<sup>1</sup>

<sup>1</sup>Universidade Federal de Alagoas, Centro de Tecnologia, Departamento de Engenharia Química

E-mail para contato: brunocarlos343@hotmail.com

**RESUMO** – *O ar, recurso indispensável para a sobrevivência da espécie humana, pode ter influência decisiva na qualidade de vida das populações. Dessa forma, o desenvolvimento de um sistema de monitoramento capaz de avaliar o quadro atual e, através de técnicas computacionais, fornecer prognósticos sobre a qualidade deste, torna-se uma peça chave dentro do sistema de gestão desse bem natural. Para a construção de um modelo com tais características, neste trabalho, foi avaliada a conjunção de técnicas baseadas em sistemas inteligentes, mais especificamente as redes neurais artificiais (RNA,) às ferramentas de processamento de sinais e tratamento de séries temporais, por meio da transformada wavelet discreta (DWT), no prognóstico da qualidade do ar. O objetivo é avaliar a utilização de um modelo composto, no qual a (DWT) decompõe a série temporal contendo os níveis de monóxido de carbono e os coeficientes resultantes da composição são utilizados como entradas na RNA. O modelo investigado se mostrou promissor na estimativa dos níveis do poluente quando se utilizou uma série temporal de 9358 dados, sendo a previsão realizada dentro do período de 168 horas. Um estudo mais aprofundado para escolha dos parâmetros da transformada wavelet discreta e da topologia da rede neural, utilizando técnicas específicas de otimização, pode estabelecer um modelo melhor em termos de acurácia.*

## 1. INTRODUÇÃO

A qualidade de vida de uma população é o reflexo das condições sob as quais esta está submetida, este quadro inclui conjunturas sociais, econômicas e também ambientais. Neste contexto, a qualidade do ar é de fundamental importância, por este ser um recurso indispensável à sobrevivência humana, podendo desempenhar um papel fundamental no que concerne à qualidade de vida das populações, em especial aquelas consideradas mais vulneráveis como crianças, idosos e indivíduos que possuam pré-disposição às doenças ligadas ao sistema cardiorrespiratório (OLIVEIRA e FERREIRA, 2004).

De acordo com o Conselho Nacional do Meio Ambiente – CONAMA, qualquer forma de matéria ou energia com intensidade e em quantidade, tempo ou características em desacordo com os níveis estabelecidos, com potencial de tornar o ar impróprio, nocivo ou ofensivo à saúde, pondo em risco o bem-estar público, a fauna e a flora, são considerados

poluentes atmosféricos. Os padrões de qualidade do ar estabelecidos no Brasil são de responsabilidade da resolução Conama nº 3/1990, que os divide em padrões primários e secundários. Segundo este mesmo documento, padrões primários de qualidade do ar são as concentrações de poluentes que, uma vez ultrapassadas, oferecem risco à saúde da população. Dentro deste contexto, então, é possível elencar alguns destes, tais como: monóxido de Carbono (CO), foco deste trabalho, dióxido de enxofre (SO<sub>2</sub>), dióxido de nitrogênio (NO<sub>2</sub>) e ozônio (O<sub>3</sub>).

Santi et al. (1999) afirmam que o monitoramento da qualidade do ar através da determinação quantitativa dos padrões primários e secundários pode ser utilizado para diversos fins, dentre eles, fornecer estatísticas com o intuito de ativar ações emergenciais durante períodos de estagnação atmosférica, acompanhar as tendências da qualidade do ar devido às alterações nas emissões de poluentes, avaliar a eficácia dos programas de controle da poluição do ar, assim como também permitir o desenvolvimento de estratégias e regulamentações específicas para o controle deste distúrbio.

Desta forma, uma estrutura de monitoramento eficaz, que consiga prover informações capazes de revelar um cenário pretérito e elucidar o quadro atual, aliada a uma ferramenta que traga consigo a capacidade de fornecer prognósticos, é essencial dentro de um sistema de gestão da qualidade do ar em diversas regiões. Para tal, os estudos de caso apresentados na literatura, através da análise de séries temporais e da criação de modelos preditivos baseados em sistemas inteligentes, a exemplo das RNA's, têm-se mostrado um instrumento poderoso na modelagem desse tipo de problema (ASHISH e RASHMI, 2011).

Uma das peças fundamentais para o desenvolvimento desses modelos são as redes neurais artificiais. Uma rede neural artificial (RNA) é um sistema de processamento de informação distribuída, maciçamente paralelo, que possui características de desempenho que se assemelham as do processamento de informação realizado pelas redes neurais biológicas do cérebro humano, sendo esta uma técnica computacional muito poderosa na modelagem de relações não-lineares complexas, particularmente em situações nas quais a forma explícita da relação entre as variáveis envolvidas é desconhecida (Haykin, 1999). Esses sistemas podem atuar tentando reconhecer regularidades e padrões em dados e são capazes de fazer generalizações baseadas nas experiências oriundas de seu treinamento frente a estes, habilidade que os tornam ferramentas ideais para previsão de parâmetros em sistemas abstrusos.

Como demonstrado por Ashish e Rashmi (2011), a utilização da transformada wavelet no pré-processamento dos dados de treinamento potencializou a habilidade de generalização das redes neurais, tornando-as um modelo mais robusto. Assim, o modelo “híbrido” forneceu resultados melhores de previsão em detrimento daqueles oferecidos pela generalização frente à apresentação convencional dos dados do caso em estudo.

Dessa forma, o presente trabalho propõe investigar modelo híbrido, que utiliza a transformada wavelet no pré-processamento da série temporal de poluição do ar, mais especificamente, dos dados horários de contração de monóxido de carbono no período de um ano, e produz um prognóstico a um passo a frente (uma hora à frente) para a qualidade do ar.

## 2. MATERIAIS E MÉTODOS

### 2.1. Obtenção do Banco de dados

O conjunto de dados utilizado neste trabalho está disponível em UCI – Machine Learning Repository e contém 9358 instâncias de dados aferidos por cinco sensores embutidos em um dispositivo de monitoramento da qualidade do ar. O dispositivo foi posto em uma área significativamente poluída em uma cidade da Itália. Os dados foram registrados de Março de 2004 a Fevereiro de 2005, sendo estes médias horárias, constituindo um conjunto de várias séries temporais ligadas à qualidade do ar daquele local.

As estratégias desenvolvidas para processamento dos dados utilizarão técnicas de tratamento de sinais, como a transformada Wavelet em conjunção com técnicas de sistemas inteligentes tais como Redes Neurais Artificiais, aplicadas em ambiente MATLAB.

### 2.2. Transformada Wavelet

A Transformada Wavelet é uma técnica matemática que transforma um sinal localizado no domínio do tempo para o domínio do tempo e da frequência. Uma wavelet é uma função ondulatória finita e cuja área sob o gráfico que a representa é igual à zero. Duas análises Wavelet amplamente utilizadas são a transformada wavelet contínua – CWT e a transformada wavelet discreta – DWT. A função de análise da transformada wavelet é definida no espaço bidimensional de escala e tempo e deriva de uma wavelet-mãe indicada por  $\psi(t)$ . Uma série de variantes desta pode ser obtida a partir da wavelet mãe por meio de dois processos, (1) translação, isto é, deslocamento da wavelet mãe no tempo e (2) dilatações, que atuam tanto no tempo quanto nas variáveis de escala. Se a variável  $b$  é usada para controlar o deslocamento no tempo e a variável  $a$  é usada para controlar a dilatação, então a série de wavelets pode ser descrita pela equação 1

$$\Psi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \Psi\left(\frac{t-b}{a}\right), a, b \in R, a \neq 0 \quad (1)$$

Sendo  $f(t)$  um sinal, então a transformada wavelet contínua é definida pela equação 2 como

$$W_f(a,b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{+\infty} f(t) \Psi\left(\frac{t-b}{a}\right) dt \quad (2)$$

A transformada wavelet discreta é a separação dos componentes de frequência do sinal de acordo com a escala. Estes componentes são divididos em detalhes  $D_i$  (componentes de altas frequências) e aproximação  $A_i$  (componentes de mais baixa frequência). Tais componentes serão utilizados como entradas para a RNA com o objetivo de se obter o prognóstico desejado.

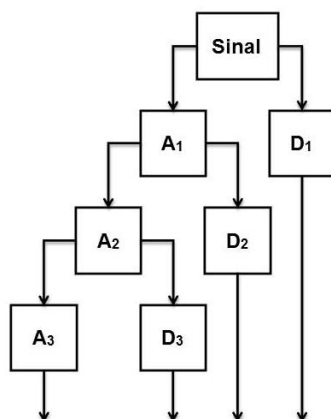


Figura 1 – Esquema da decomposição wavelet (DWT) com a separação dos componentes de mais alta frequência  $D_i$  e mais baixa frequência  $A_i$ .

## 2.3. Redes Neurais Artificiais

As redes neurais artificiais (RNAs) são dispositivos de regressão que contêm camadas de nós de computação, com notáveis características de processamento de informação. Eles são capazes de detectar não linearidades, tornando-os capazes de aprendizagem e adaptabilidade. Além disso, possuem alto paralelismo, robustez, generalização e tolerância ao ruído, o que os faz capazes de realizar tarefas de agrupamento, aproximação de função, previsão e associação. Os modelos neurais artificiais que possuem tais características são desejáveis porque: (1) a não linearidade permite melhor ajuste aos dados, (b) a insensibilidade ao ruído leva a predição exata na presença de dados incertos e erros de medição, (c) a aprendizagem e adaptabilidade permitem que o sistema atualize ou modifique sua estrutura interna em resposta à mudança do ambiente, e (d) a generalização permite a aplicação do modelo a dados desconhecidos. A representação gráfica de um neurônio artificial, elemento mais fundamental das RNAs, está exposta na figura 2.

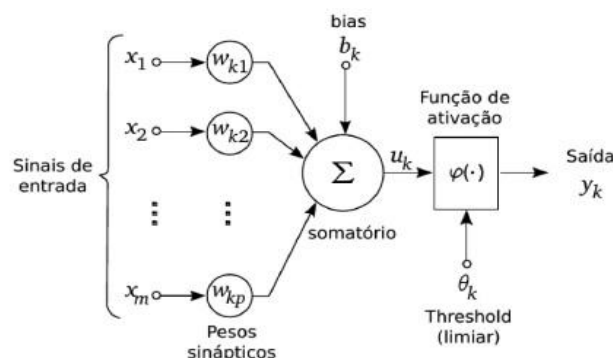


Figura 2 – Esquema ilustrativo de um neurônio artificial, proposto por McCulloch-Pitts.

Quando os neurônios artificiais estabelecem conexões entre si, propagando o sinal recebido, se configura o que se chama de rede neural. Os pesos são parâmetros ajustáveis da rede, que sofrem alterações à medida que o conjunto de aprendizado é apresentado ao sistema.

O ajuste destes se dá de tal forma que o valor da saída gerada seja o mais próximo possível do valor desejado. A este processo se dá o nome de treinamento.

Para realização do treinamento, que matematicamente, hoje, se trata de uma otimização, existem alguns algoritmos já estabelecidos que apresentam características peculiares nas suas utilizações. Dentre os algoritmos citados na literatura, pode-se destacar, pela frequência de utilização, dois algoritmos baseados em métodos de otimização diferentes: o de retropropagação do erro (*Backpropagation*), com a regra do delta generalizado, que é um método de gradiente descendente bastante utilizado e citado frequentemente na literatura; e o método de Levenberg-Marquardt, baseado no método quasi-Newton. Por se tratarem de algoritmos que envolvem o cálculo de derivadas, supõem-se a utilização de funções contínuas, principalmente as sigmóides.

## 2.4. Conjunção da Transformada Wavelet com a RNA

O modelo proposto utilizou as componentes de altas e baixas frequências, extraídas através da transformada wavelet discreta da série temporal em questão para treinamento supervisionado da rede neural artificial. A tarefa da RNA, neste caso, era prever em curto prazo, com um dado nível de exatidão, as concentrações de monóxido de carbono medidas em intervalos de uma hora, no período de uma semana. A metodologia utilizada nesse trabalho pode ser resumida no fluxograma abaixo ilustrada pela Figura 3.

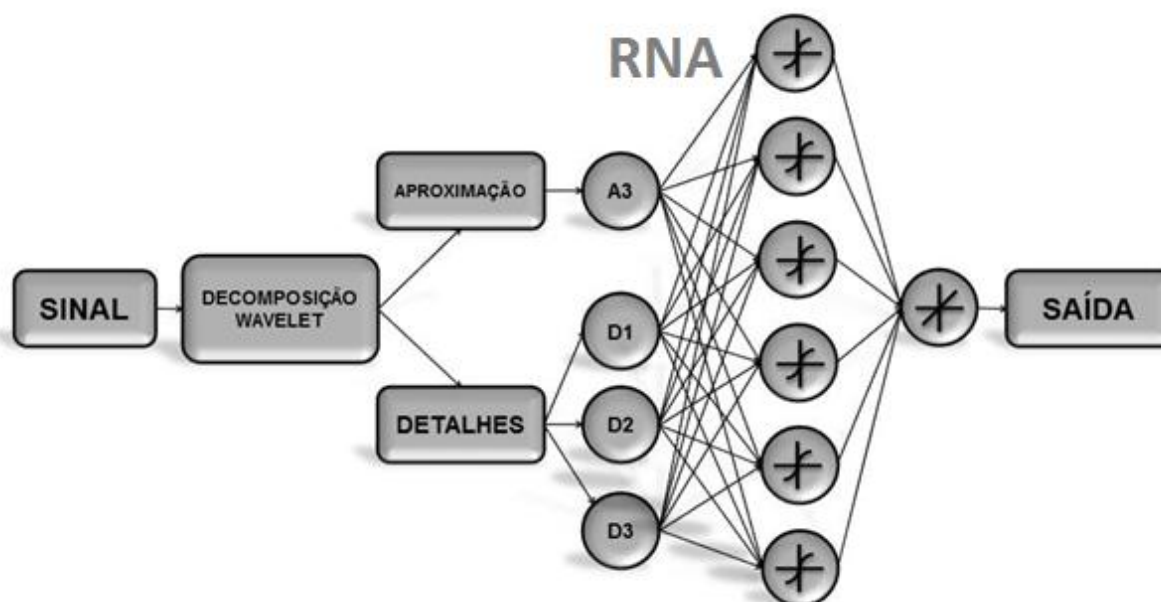


Figura 3 – Esquema da decomposição do sinal e posterior processamento deste através dos neurônios da rede neural.

## 3. RESULTADOS E DISCUSSÃO



Optou-se por escolher o gás monóxido de carbono como parâmetro para realização da previsão de poluição atmosférica, pois além deste ser um padrão primário, se constatou que este apresenta uma relação de proporcionalidade com os demais gases poluentes detectados pelo dispositivo de monitoramento, uma vez que havendo uma variação nos níveis de monóxido de carbono presente na atmosfera, haverá uma variação proporcional nos níveis dos demais gases poluentes citados neste estudo, mais especificamente dos níveis de dióxido de nitrogênio, ozônio e alguns hidrocarbonetos. Na figura 4 é possível visualizar esta relação.

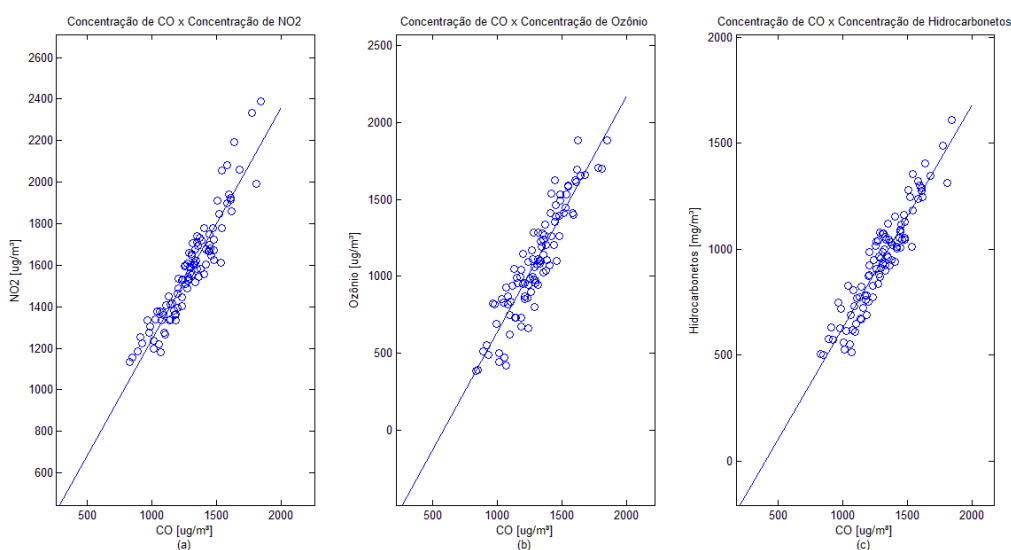


Figura 4 – Relações observadas entre as concentrações de CO e NO<sub>2</sub> (a), CO e O<sub>3</sub> (b) e CO e Hidrocarbonetos (c).

O sinal foi decomposto utilizando a transformada wavelet até o terceiro nível de detalhe. A wavelet utilizada foi da família symlet. Vale salientar que existe uma infinidade de famílias.

Uma vez realizado este processo, a série original pode ser vista como uma soma de outras séries, dessa forma, o sinal original  $S$  pode ser representado como  $S = A3 + D1 + D2 + D3$ .

A rede neural utilizada foi do tipo feed-forward, composta de uma camada de entrada com 4 neurônios, uma camada escondida composta de 6 neurônios tendo como função de ativação a tangente sigmoide e uma camada de saída com 1 neurônio linear. Além disso, o algoritmo de treinamento foi o de regularização bayesiana. Uma vez que  $A_i$  e  $D_i$  são vetores, estes foram utilizados como variáveis de entrada da rede, a qual deveria prever a concentração correspondente de monóxido de carbono. Na Figura 5 temos representada a sobreposição da série gerada pela previsão do modelo híbrido com a série real gerada pelo dispositivo de monitoramento da qualidade do ar.

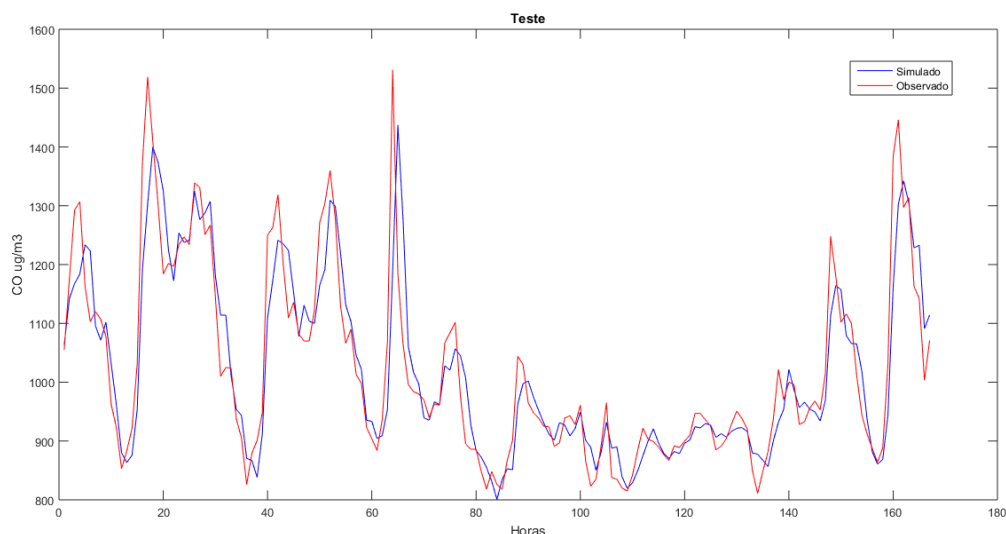


Figura 5 – Valores reais da concentração de monóxido de carbono e valores simulados pelo modelo.

A Tabela 1 traz os valores dos critérios de erro para o teste simulado pelo modelo escolhido (Wavelet – Neural-Network), esses valores de erro são o Erro médio quadrático (RMSE), o coeficiente de determinação ( $R^2$ ) e a média dos erros quadráticos (NSE). Esses parâmetros avaliam a precisão do modelo empregado na predição e nos indicam se o resultado pode ser tido como satisfatório.

Tabela 1 - Critérios de erros RMSE,  $R^2$  e NSE.

Modelo	Teste		
	RMSE m <sup>3</sup> /s	$R^2$	NSE
Wavelet – Neural-Network	69.3269	0.8040	0.8152

## 4. CONCLUSÃO

No ensaio realizado se propôs a realização da previsão dos níveis de alguns gases poluentes presentes na atmosfera durante o período de 168 horas, utilizando para isso a decomposição wavelet e as redes neurais artificiais. O modelo escolhido para a predição nos retorna um resultado dito satisfatório, visto que os critérios de erros obtidos se mostram relativamente próximos à idealidade, o que, de fato, atesta a predição dos índices de poluição aqui levantados. O gráfico que confronta os valores simulados e observados apresenta uma relativa similaridade nos sinais, esse comportamento pode demonstrar que o modelo desenvolvido pode ser utilizado para a predição dos índices de poluição de dias futuros, respeitando um devido intervalo de confiança imposto pelo erro inerente ao modelo. É válido ressaltar que um estudo mais aprofundado para escolha dos parâmetros da transformada wavelet discreta e da topologia da rede neural, utilizando técnicas específicas de otimização, pode estabelecer um modelo melhor em termos de acurácia.

## 5. REFERÊNCIAS

ASHISH, M., RASHMI, B. Prediction of daily air pollution using wavelet decomposition and adaptive-network-based fuzzy inference system. **International Journal of Environmental Sciences**, Volume 2, No 1, 2011.

HAYKIN, S., **Redes Neurais Princípios e prática**, editora bookman 2ª ed., Porto Alegre-RS, 1999.

OLIVEIRA, V. & FERREIRA, A. P. Poluição do ar e saúde ambiental na cidade do Rio de Janeiro: contribuição para a definição de estratégias de monitoramento. **Revista Eletrônica do Prodema**, 2007.

RESOLUÇÃO CONAMA nº 3, de 28 de junho de 1990. Publicada no DOU, de 22 de agosto de 1990, Seção 1, páginas 15937-15939.

SANTI, A. M. M.; ROSA, A. C.; SUZUKI, R. Y. Monitoramento da qualidade do ar: avaliação de metodologia baseada no licenciamento ambiental. **XXVII Congresso Interamericano de Engenharia Sanitária e Ambiental**, 1999, Porto Alegre.

UCI – Machine Learning Repository, Center for Machine Learning and Intelligent Systems. Disponível em: <<https://archive.ics.uci.edu/ml/datasets/Air+Quality>>. Acesso em: 23 de Abril de 2017.