

Emprego de Redes Neurais Artificiais RBF na previsão de nevoeiro no Aeroporto Internacional Afonso Pena

Maria Carolina Martins Pereira
PPGMNE
Universidade Federal do Paraná
Curitiba, Brasil
pereira.mcarolina@gmail.com

Ricardo Carvalho de Almeida
PPGEA
Universidade Federal do Paraná
Curitiba, Brasil
rcalmeida@ufpr.br

Resumo— Restrições de visibilidade severas causadas por nevoeiros são um grande empecilho à aviação, pois podem causar atrasos ou cancelamentos de voos, além de trazerem grandes riscos às operações aéreas. O Aeroporto Internacional Afonso Pena, localizado na cidade de São José dos Pinhais/PR, é o aeródromo brasileiro que mais sofre com estes fenômenos meteorológicos. Esta pesquisa objetiva verificar a eficácia do emprego de RNAs na previsão categórica do fenômeno nevoeiro para até três horas, bem como detectar quais as variáveis meteorológicas influenciam mais significativamente na eficácia da previsão. Comparados os desempenhos das previsões realizadas a partir das técnicas de climatologia, persistência e através do emprego das RNAs, chegou-se à conclusão de que o emprego de RNAs na previsão do fenômeno nevoeiro para o Aeroporto Afonso Pena é válido, pois esta técnica mostrou-se mais eficaz que as demais.

Palavras-chave—Redes Neurais Artificiais; nevoeiro; persistência; climatologia

I. INTRODUÇÃO

O Aeroporto Afonso Pena, localizado em São José dos Pinhais, Paraná, foi classificado pela Empresa Brasileira de Infraestrutura Aeroportuária (INFRAERO), com base no Anuário Estatístico Operacional de 2015 [4], como o sexto maior do país em movimentação de aeronaves, com cerca de 207 pousos e decolagens diárias. No entanto, em todo o país, segundo o mesmo Anuário, é o aeroporto que mais tem estas atividades interrompidas por restrição de visibilidade, causando grande prejuízo aos passageiros, que muitas vezes têm seus voos atrasados ou cancelados e às empresas aéreas, que acabam por ter despesas extras como com a alocação dos passageiros, reabastecimento das aeronaves, ocupação de pátio, entre outras; além de sobrecarregar o trabalho do Centro de Gerenciamento de Navegação Aérea (CGNA), cujo objetivo é gerenciar o fluxo de tráfego aéreo nacional. Conseguir prever este fenômeno contribui sobremaneira na organização da malha aérea do país, auxiliando empresas de transporte aéreo na otimização de tempo e custos, elevando a qualidade dos serviços prestados, além de aumentar a segurança nos pousos e decolagens.

As Redes Neurais Artificiais (RNA) são uma poderosa ferramenta computacional que permite o aprendizado de máquina, ou seja, o programa reconhece os padrões exibidos

em um banco de dados, e consegue prever o resultado quando novas informações lhe são apresentadas. O emprego desta tecnologia na previsão de nevoeiros pode produzir bons resultados, pois existe um modelo físico na gênese do fenômeno, mas este é demasiadamente complexo de ser analisado através de ferramentas tradicionais. Das diversas configurações de Redes Neurais Artificiais, as RNAs de Funções de Base Radial (*Radial Basis Function*- RBF) se destacam na previsão de fenômenos não-lineares e previsão de séries temporais [6]. Estes dois pontos são bastante significativos na previsão de nevoeiros, o que levou à escolha desta estrutura de RNA para este trabalho.

Esta pesquisa objetiva avaliar o desempenho do uso de Redes Neurais Artificiais na previsão de curtíssimo prazo do fenômeno meteorológico nevoeiro. Tomar-se-á por curtíssimo prazo o período compreendido por até três horas subsequentes, e tratar-se-á da previsão do fenômeno no formato categórico, considerando apenas a ocorrência ou a não ocorrência da condição. Para isso, se busca verificar quais variáveis locais, tais como dados de superfície, influenciam de maneira mais significativa no desempenho da Rede Neural na previsão deste fenômeno; se informações sinóticas, tais como as de ar superior, influenciam no desempenho da RNA, e quais destas informações são mais significativas. Avaliam-se também quantas horas de dados antecedentes são importantes para o bom desempenho da Rede Neural.

II. REVISÃO BIBLIOGRÁFICA

A. Nevoeiro

Nevoeiro é uma formação de grande concentração de gotículas de água (excepcionalmente, cristais de gelo), em massa de ar estável, sobre a superfície, provocando redução de visibilidade [5]. Para fins aeronáuticos, esta restrição deve ser inferior a 1000 metros [2].

B. Redes Neurais Artificiais

O termo “Redes Neurais Artificiais” advém do fato de que estas tecnologias buscam representar matematicamente as estruturas cerebrais biológicas. Da mesma maneira que os seres humanos buscam nos eventos passados os auxílios necessários para a resolução de novos problemas, as redes neurais

procuram padrões, em conjunto de dados, que auxiliem em classificações e previsões [1].

Basicamente, a Rede Neural Artificial detecta os padrões existentes no conjunto de dados utilizado para o treinamento da rede, “aprendendo” como eles se comportam, e quando apresentadas novas informações, ela deve ser capaz de tratá-los de acordo.

1) *RNAs de Função de Base Radial (RNA RBF)*: RNAs RBF são Redes Neurais Artificiais cuja função de ativação é qualquer função de distância. Estas funções geralmente são do tipo gaussianas. RNAs RBF estão dentre as RNA do tipo *feedforward*, e possuem três camadas: uma para receber os vetores com dados de entrada; uma camada oculta, onde cada neurônio possui uma RBF como função de ativação; e uma camada para os valores de saída, que pode conter um ou mais neurônios.

Além da função de ativação de base radial, outra característica importante das redes RBF é o processo de treinamento no ajuste dos centros e desvios padrões das gaussianas dos neurônios ocultos e dos pesos sinápticos entre os neurônios ocultos e os de saída.

Na primeira etapa do treinamento, são ajustados os pesos das conexões entre os neurônios de entrada e os da camada intermediária, alocando as funções de base radial de cada um dos neurônios ocultos.

Inicialmente, define-se o número de neurônios da camada oculta e seus respectivos representantes centrais. Para a classificação de padrões, a RNA RBF considera a distância radial de cada elemento apresentado como entrada ao centro da gaussiana, construindo hiperesferas delimitadoras em cada neurônio oculto, as quais definirão classes. Desta forma, um bom ajuste dos centros das gaussianas dos neurônios intermediários é fundamental para o bom desempenho da RNA.

Uma das maneiras de realizar este ajuste é através do método de k-médias. Os valores dos centros são iniciados aleatoriamente. Quando um vetor é apresentado, calculam-se as distâncias entre este vetor e cada um dos centros. O que apresentar menor distância é o vencedor, e o vetor passa a pertencer àquela classe. Após a classificação de todos os vetores, a posição do centro é ajustada tomando-se o ponto médio de cada uma das coordenadas entre a posição anterior do centro e o vetor apresentado. Ajustado o valor do centro, toma-se outro vetor pertencente à classe, e repete-se o processo até que todos os vetores sejam apresentados. Todos os vetores do grupo de treinamento são então novamente apresentados à RNA, e redistribuídos nas classes, agora com novos centros. Repete-se este processo de ajuste dos centros até que estes não se movimentem mais do que um valor pré-estabelecido. Este valor deve ser pequeno a ponto de significar a quase estagnação do processo de ajuste.

Quando os vetores centrais já estão bem definidos, a RNA calcula a variância dos valores que ficaram em cada classe, atribuindo-lhes uma curva gaussiana.

Na segunda etapa de treinamento são ajustados os pesos das conexões da camada de saída para obter a combinação linear final. Este processo é supervisionado e constitui-se dos

pares formados pelo valor de saída das funções gaussianas e as amostras do treinamento. O valor obtido das gaussianas é ponderado pelos pesos sinápticos, e submetido à função de ativação da camada de saída. As funções normalmente utilizadas nesta camada são as funções degrau, linear, sigmoideal ou tangente hiperbólica. O sinal resultante é então comparado ao sinal de referência, e os pesos sinápticos são ajustados. Após o ajuste, o conjunto de teste é apresentado à rede. Se o erro médio obtido for superior ao anterior, o treinamento é encerrado. Caso contrário, repete-se o processo de ajuste dos pesos, até alcançar-se o *overtraining*.

III. METODOLOGIA

A. Características do Aeroporto Afonso Pena

O Aeroporto Internacional Afonso Pena está localizado na cidade de São José dos Pinhais, região metropolitana na cidade de Curitiba/ PR. Encontra-se em uma região serrana próxima ao litoral. Este fator tem relação com a alta frequência de nevoeiros, pois ventos quentes e úmidos vindos desta região configuram condições que favorecerem a formação do fenômeno.

O aeródromo foi originalmente construído com o propósito de servir como base militar para o exército aliado na 2ª Guerra Mundial. Por este motivo, a área escolhida para sua construção tinha alta ocorrência de nevoeiro, pois isso facilitaria na camuflagem das aeronaves. No entanto, a partir do momento que passou a ser utilizado pela aviação civil, a alta frequência de ocorrência de nevoeiros passou a ser um empecilho às operações aeronáuticas [3].

B. Base de dados meteorológicos

Os dados meteorológicos utilizados neste estudo descrevem condições de superfície e altitude compreendidas no período de 01/01/1997 00h 00min, a 31/12/2014 23h 59min.

Os dados de superfície foram coletados na Estação Meteorológica de Superfície (EMS) de Curitiba, localizada no Aeroporto Internacional Afonso Pena, e obtidos do Sistema Automatizado de Registro e Gerenciamento das Observações Meteorológicas (WEBMET), através do Instituto de Controle do Espaço Aéreo (ICEA). Tomaram-se dados relativos à hora, data, visibilidade horizontal predominante (vis), precipitação acumulada (precip), temperatura do ar (temp), temperatura do ponto de orvalho (PO), cobertura total de nuvens (nuvem), pressão atmosférica ao nível da estação (pressão), direção do vento (dir), velocidade do vento (vel) e umidade relativa do ar (UR). Foram consideradas apenas as informações regulares. Os dados de informações especiais e locais foram ignorados. As informações de altitude foram coletadas do código TEMP através das radiossondagens realizadas na Estação Meteorológica de Altitude (EMA) de Curitiba, localizada no mesmo Aeroporto, e obtidas através do sítio da Universidade de Wyoming [8]. Foram coletadas os dados de velocidade do vento no nível de pressão de 250hPa (vel250), altitude geopotencial do nível de pressão de 850 hPa (GPH850), altitude geopotencial do nível de pressão de 500hPa (GPH500), temperatura do ar no nível de pressão de 700 hPa

(temp700), temperatura do ponto de orvalho no nível de pressão de 700 hPa (PO700), temperatura do ar no nível de pressão de 500hPa (temp500) e temperatura do ponto de orvalho no nível de pressão de 500hPa (PO500). Estes dados foram selecionados, por permitirem a indicação da presença de fenômenos de escala sinótica, que podem interferir nos ocorridos à superfície.

1) *Tratamento dos dados:* Os dados foram, inicialmente, tratados com a substituição de variáveis obviamente discrepantes “9999”, pois este valor não seria confundido com nenhuma medida real. Esta medida foi adotada, pois a simples exclusão da informação não era possível devido à perda da sequência temporal.

2) *Transformação dos dados:* A variável direção do vento é uma variável cíclica, e não determina uma quantidade em si, apenas uma orientação. Por esta razão foi necessária a polarização deste dado, de modo que a rede neural não hierarquiasse os valores. A direção do vento foi dividida em quatro grupos, de acordo com os quadrantes, designados por sequências de números 1 e -1, onde três dos valores são -1, e o valor 1, que ativa o neurônio cuja posição refere-se ao intervalo em que está a referida direção do vento. Assim, direções do vento forma polarizadas conforme apresentado na Tabela I, abaixo.

TABELA I. POLARIZAÇÃO DA VARIÁVEL DIREÇÃO DO VENTO

Direção do vento (em graus)	Sequência designadora
[000,090)	1 -1 -1 -1
[090,180)	-1 1 -1 -1
[180,270)	-1 -1 1 -1
[270,360)	-1 -1 -1 1

Como esta pesquisa trata da previsão categórica do fenômeno nevoeiro, as informações de visibilidade foram utilizadas não apenas em seu formato métrico, como também em seu formato categórico. Os casos em que se verificaram visibilidades inferiores a 1000m, e, por conseguinte, a presença do fenômeno, foram tomados como 1; e os casos em que a visibilidade registrada foi igual, ou superior, a 1000m, e, portanto, não configuraram ocorrência de nevoeiro, foram tomados como -1. Esta medida foi adotada no intuito de auxiliar o trabalho da RNA, pois simplifica a representação da variável em questão, atribuindo-lhe apenas dois valores possíveis: 1 e -1.

3) *Compatibilização da amostra dos dados de altitude:* Diferentemente das observações à superfície, as radiossondagens não ocorrem de hora em hora, e sim a cada 12 horas. Para que o número de informações de altitude fosse compatível com o número de informações de superfície, foi realizado o completamento da amostra dos dados sinóticos. Para tanto, cada informação de altitude foi repetida onze vezes, nas onze horas subsequentes ao horário efetivo da sondagem, mantendo, assim, a compatibilidade da série temporal entre as informações de superfície e altitude.

4) *Base de dados tratada:* Ao final do tratamento dos dados, a amostra de dados completa pode contar com 157683

informações de visibilidade horizontal predominante (vis), sendo 7300 destas, de visibilidades inferiores a 1000 metros, 155495 informações de precipitação acumulada (precip), 154733 informações de temperatura do ar (temp), 154733 informações de temperatura do ponto de orvalho (PO), 145145 informações de cobertura total de nuvens (nuvem), 153265 informações de pressão atmosférica ao nível da estação (pressão), 115820 informações de direção do vento (dir), 115820 informações de velocidade do vento (vel), 154337 informações de umidade relativa do ar (UR), 94440 informações de velocidade do vento no nível de pressão de 250hPa (vel250), 109776 informações de altura geopotencial do nível de pressão de 850hPa (GPH850), 108504 informações de altura geopotencial do nível de pressão de 500hPa (GPH500), 109836 informações de umidade relativa no nível de pressão de 700hPa (UR700), e 109704 informações de umidade relativa no nível de pressão de 500hPa (UR500), totalizando 1839292 dados.

Como o código TEMP não fornece, diretamente, dados de umidade relativa, trata-se, nesta pesquisa, como umidade relativa nos níveis de pressão de 700hPa e 500hPa, como a apresentação simultânea à RNA, dos dados de temperatura do ar e ponto de orvalho nos respectivos níveis de pressão.

C. Métricas para avaliação de desempenho de previsões de variáveis categóricas

Em relação à previsibilidade, um evento pode ser previsto ou não previsto. Em relação à ocorrência, ele pode ser observado ou não observado. Logo, dado um fenômeno qualquer, este certamente encontra-se em uma das quatro combinações possíveis de previsão e ocorrência indicados na Tabela II.

TABELA II. TABELA DE CONTINGÊNCIA

		Observado		
		Classe	Não-Classe	Total
Previsto	Classe	a	b	a+b
	Não-Classe	c	d	c+d
	Total	a+c	b+d	n=a+b+c+d

onde *a* representa o número de casos previstos e observados; *b* representa os casos em que o evento foi previsto, mas não observado; *c* representa o número casos não previstos, mas observados; *d* indica os casos não previstos e não observados; e *n* indica a soma total do número de casos.

1) *Atributos escalares:* O uso de atributos escalares na caracterização do desempenho de previsões facilita a compreensão dos resultados alcançados se comparado à utilização de frequências relativas ou absolutas [9]. Destacasse, abaixo, alguns dos atributos escalares mais utilizados na verificação de desempenho de previsões.

a) *Proporção Correta (PC):* Esta métrica avalia a proporção de acertos da previsão, tanto para os casos previstos e observados, quanto para os casos não previstos e não observados. É simplesmente, a fração entre o número de acertos da previsão, e os casos totais.

$$PC = \frac{a+d}{n} \quad (1)$$

Esta métrica pode não ser bem aplicada em casos de eventos raros, como o nevoeiro, pois a não-ocorrência do evento é muito mais provável que a ocorrência, o que torna estes casos bem mais difíceis de serem previstos.

b) *Índice de Sucesso Crítico (Threat Score- ISC)*: Esta métrica é bem mais ajustada a casos de eventos raros, pois desconsidera os casos de não-evento, mais abundantes e de maior previsibilidade.

$$ISC = \frac{a}{a+b+c} \quad (2)$$

c) *Viés (V)*: O Viés fornece a informação de quantas vezes o evento foi previsto, em relação a quantas vezes foi observado, sem levar em consideração a correspondência entre observações e previsões.

$$V = \frac{a+b}{a+c} \quad (3)$$

d) *Razão de Falso Alarme (RFA)*: A RFA representa a proporção de previsões que acabaram por não se concretizar.

$$RFA = \frac{b}{a+b} \quad (4)$$

Esta métrica assume valores entre 0 e 1. No entanto, como se refere a uma medida com significado desfavorável, pois indica erros, quanto mais próxima de zero, mais eficaz é a ferramenta de previsão.

e) *Probabilidade de Detecção (Hit Rate- PD)*: É, basicamente, a taxa de acerto da previsão relacionada apenas ao evento de interesse.

$$PD = \frac{a}{a+c} \quad (5)$$

Operacionalmente, é de suma importância que uma ferramenta de previsão seja eficaz em dois aspectos principais: capacidade de prever corretamente a ocorrência do fenômeno de interesse; e manter alta a confiabilidade das previsões, não emitindo falsos alarmes em demasia. Dada a motivação operacional desta pesquisa, as métricas de avaliação de desempenho da Rede Neural Artificial terão a seguinte hierarquia: como métrica mais importante, a que será primeiramente observada será a PD, pois este índice caracteriza bem o índice de acertos da ferramenta em relação aos casos nos quais o evento de fato ocorreu. Como segunda métrica, será utilizada a RFA, visto que é de interesse operacional não restringir as condições meteorológicas desnecessariamente. A RFA é um índice importantíssimo operacionalmente, pois é ela que garante a manutenção da

confiabilidade dos usuários em relação às previsões fornecidas. Como terceira métrica será utilizado o ISC, pois este índice destaca a previsibilidade de fenômenos raros, o que é o caso do fenômeno nevoeiro [9]. Como quarta, e última, medida será considerada a PC, pois para casos de fenômenos raros esta é a menos significativa.

D. Climatologia

Uma das maneiras de entender melhor um fenômeno meteorológico é estudar sua climatologia. Observando as condições reinantes quando na ocorrência do fenômeno nevoeiro, nas três horas anteriores ao evento, foi possível estabelecer um quadro situacional, o qual pode ser observado na Tabela III.

TABELA III. CONDIÇÕES FAVORÁVEIS À FORMAÇÃO DE NEVOEIRO

	Parâmetros da classificação	1	-1
Variável	mês	[5,8]	[1,5)U(8,12]
	hora	[21,8]	-8,21
	Temp (°C)	[10,15]	[-20,10)U[15,50)
	UR (%)	≥ 90	< 90
	Nuvem (oitavos)	[7,8]	[0,7)
	Pressão (hPa)	[915,920)	[900,915)U[920,952)
	Vel (nós)	[0,5)	≥ 5
	Precip (mm)	0	≠ 0

É possível realizar previsões através destas informações, inferindo que, na ocorrência concomitante destas condições, haverá a formação de nevoeiro. Este experimento é realizado posteriormente.

E. Persistência

Pode-se dizer que a meteorologia acontece de forma contínua, onde a consumação de um fenômeno guarda relações com as condições anteriores, e sua ocorrência interfere nas condições posteriores. Dado isso, pode-se dizer que as variáveis atmosféricas têm uma dependência estatística entre seus valores passado, presente e futuro [9]. A esta dependência entre condição e tempo, em meteorologia, dá-se o nome de Persistência. Ou seja, um fenômeno é persistente se seus valores anteriores e posteriores são estatisticamente dependentes. Uma forte relação de dependência, ou uma forte persistência, permite fazer inferências estatísticas tão melhores quanto maior for esta relação.

Fenômenos constatados persistentes possuem uma técnica de previsão bastante simples: a condição futura é inferida como a manutenção da condição atual. Para o caso desta pesquisa, onde a previsão do fenômeno nevoeiro está sendo abordada no formato categórico para até três horas, a previsão por persistência dá-se do seguinte modo: tomemos por x a variável hora, e por y a variável categórica visibilidade. Para visibilidades inferiores a 1000m, configura-se a formação de nevoeiro, e y recebe o valor 1; para visibilidades superiores ou iguais a 1000m, não há a condição de nevoeiro, e y recebe o valor -1.

Se em uma hora x qualquer existe a condição de nevoeiro, obtém-se o par $(x, 1)$. A previsão por persistência nos fornece a previsão para as horas $x+1$, $x+2$ e $x+3$ como sendo a

manutenção da condição observada na hora x , gerando então os pares $(x+1,1)$, $(x+2,1)$ e $(x+3,1)$. A inferência é realizada do mesmo modo para os casos em que não há a condição de nevoeiro, ou seja $y=-1$.

Pode-se dizer que o fenômeno nevoeiro apresenta comportamento bastante estável, no sentido de que uma vez instaurado, a condição de restrição de visibilidade se mantém por várias horas. Esta condição gera poucas variações na série temporal, o que pode proporcionar às previsões por persistência um bom desempenho para este fenômeno. Este experimento será realizado posteriormente.

F. Especificações da RNA utilizada

Foi utilizada uma RNA RBF, cuja função de base radial é do tipo Gaussiana, e a função de ativação na camada de saída é do tipo tangente hiperbólica. O número de dados do vetor de entrada, bem como o número de neurônios ocultos variou nas diferentes etapas do treinamento, como será descrito posteriormente. Na camada de saída, foi utilizado apenas um neurônio, visto que a previsão do fenômeno nevoeiro, nesta pesquisa, será realizada na forma categórica.

Foram usados 75% dos dados da amostra para o treinamento da rede, 15% para os testes, e 10% na validação. Os dados de treinamento foram normalizados através dos valores máximos e mínimos, cujo valor máximo assumiu o valor 1, e o valor mínimo assumiu o valor -1. Depois de normalizados os dados, foram utilizadas 10 iterações na validação cruzada. Ou seja, a rede separou a amostra completa em 10 partes, embaralhou a sequência dos vetores, e realizou um treinamento independente com cada uma das amostras menores, e, como desempenho final, apresentou a média dos desempenhos das validações.

O número de épocas de treinamento, na primeira e na segunda fase do treinamento da rede, variaram de acordo com o número de variáveis de entrada e o número neurônios ocultos, mas sempre foram estabelecidas de modo que a máxima movimentação da posição dos centros das classes, durante o processo de ajuste, fosse inferior, ou igual, a $1E-06$. O fator de escala dos pesos iniciais das sinapses foi considerado como 0,25, e as taxas de aprendizagem na correção das variâncias, na correção das posições dos centros e na correção dos pesos foi de 0,007.

O código do programa foi escrito em linguagem computacional FORTRAN 95, e fornecido pelo Professor Dr. Ricardo Carvalho de Almeida.

IV. EXPERIMENTOS

A. Previsão do fenômeno nevoeiro através de Climatologia

Para realizar a previsão de um fenômeno através de climatologia, adotou-se a abordagem de buscar, nos eventos passados, características comuns a todas as ocorrências. A partir disso, inferiu-se que, na configuração destas determinadas características, ocorreria o fenômeno em questão. A previsão para uma, duas e três horas foi então realizada, inferindo-se que, na ocorrência simultânea de tais condições, haveria a formação de nevoeiro nas três horas

subsequentes. O desempenho desta técnica de previsão pode ser observado na Tabela IV, a seguir.

TABELA IV. DESEMPENHO DA TÉCNICA DE PREVISÃO DE NEVOEIROS POR CLIMATOLOGIA

Horas de previsão	+1h	PC	0,95	ISC	0,04	V	6,46	RFA	0,96	PD	0,28
	+2h	PC	0,95	ISC	0,03	V	6,46	RFA	0,96	PD	0,24
	+3h	PC	0,95	ISC	0,03	V	6,46	RFA	0,97	PD	0,21

É possível perceber que a PC é bastante alta. No entanto, a RFA também o é, chegando próxima ao valor 1. Estes resultados indicam que, utilizando esta técnica de previsão, tende-se a superestimar a ocorrência de nevoeiros. Esta constatação fica clara através do índice V, pois seu valor mostra que o evento é previsto, aproximadamente, seis vezes mais frequentemente do que de fato ocorre.

B. Previsão do fenômeno nevoeiro através de Persistência

Foram utilizados, para a aplicação desta técnica de previsão, os dados de visibilidade contidos nas mensagens METAR, no período de 00h 00min do dia 01/01/2015 à 23h 00min do dia 31/12/2015, disponíveis na REDEMETS [7]. Da mesma forma que no tratamento dos dados para o treinamento, as informações especiais e locais foram descartadas, considerando-se apenas as informações regulares. No código meteorológico METAR, visibilidades superiores ou iguais a 10km são codificadas como "9999", e as visibilidades ausentes, codificadas como "////". Para não haver confusão com a notação adotada anteriormente, estas foram substituídas por 10km, e "9999", respectivamente. Para os casos de codificação "CAVOK", a visibilidade foi adotada também, como 10km. Todas as informações de visibilidade foram tomadas no formato bipolar, e a previsão para uma, duas e três horas consistiu, apenas, em repetir a condição presente.

A acurácia deste método de previsão pode ser observado na Tabela V, abaixo.

TABELA V. DESEMPENHO DA TÉCNICA DE PREVISÃO DE NEVOEIROS POR PERSISTÊNCIA

Horas de previsão	+1h	PC	0,99	ISC	0,61	V	1	RFA	0,24	PD	0,76
	+2h	PC	0,98	ISC	0,42	V	1	RFA	0,41	PD	0,59
	+3h	PC	0,97	ISC	0,32	V	1	RFA	0,52	PD	0,48

Nota-se que Proporção Correta é bastante elevada. Este resultado já era esperado, pois como já citado anteriormente, existe uma predominância das condições de não-nevoeiro em relação à ocorrência do evento. Observa-se também que o índice PD é bastante razoável, visto que esta é uma técnica de previsão sem destreza. A RFA, no entanto, é bastante elevada já para a primeira hora de previsão, e aumenta ainda mais na segunda e terceira horas. Um índice interessante nesta técnica de previsão é o Viés, pois, para todas as previsões, obteve-se valor 1. Este valor indica que o fenômeno foi previsto exatamente o mesmo número de vezes que ocorreu. Isso ocorre porque a previsão através do método de persistência consiste em, apenas, projetar a condição de interesse, sem

alterar seu valor, mantendo o número de previsões igual ao ocorrências.

C. Previsão do fenômeno nevoeiro através de RNA RBF

Para a realização das previsões para três horas subseqüente foram utilizadas três Redes Neurais distintas: uma para a previsão para uma hora (R1), uma para a previsão para duas horas (R2), e uma para a previsão para três horas (R3). Cada uma delas foi treinada e analisada de maneira independente das demais. O processo de treinamento das RNAs passou por várias etapas até se chegar nas configurações ótimas, conforme descrito a seguir.

1) *Treinamento das RNA:* O treinamento das RNA passou por várias etapas, as quais serão descritas a seguir. A melhor configuração obtida em cada uma das etapas passou para a etapa subseqüente, na busca pela configuração ótima das três redes.

a) *Etapa 1- Seleção dos dados de entrada:* Dentre todas as variáveis disponíveis na amostra de dados, era necessário escolher as mais relevantes para os experimentos com a RNA. O método adotado para a realização da escolha foi a correlação linear entre os valores de cada uma das variáveis, e os valores de visibilidade. Além disso, de todo o período de valores que compunha a amostra, era necessário determinar quão longas seriam as séries temporais utilizadas como valores de entrada para rede. Assim, foi calculada a correlação entre cada uma das variáveis e a variável visibilidade, de 0 a 12 horas anteriores. Obtiveram-se os resultados ilustrados na Fig. 1.

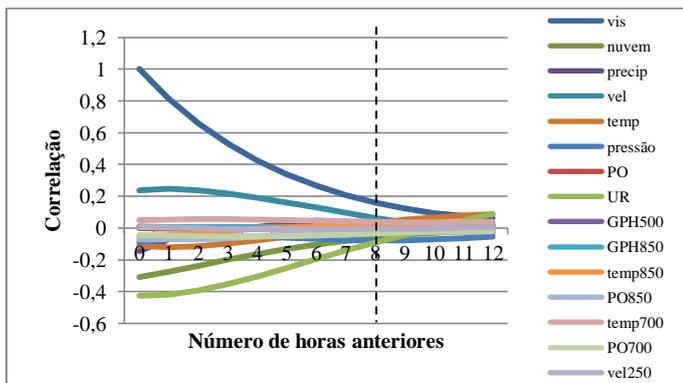


Fig. 1. Evolução temporal da correlação entre as variáveis componentes da amostra e a variável visibilidade

É possível notar que, das variáveis selecionadas, a que possui maior correlação com a visibilidade é a própria visibilidade, e que os valores de correlação mais significativos se concentram no período de 1 a 8 horas anteriores. Por isso, atribuiu-se o horário de 8 horas como o horário de corte para as séries temporais.

Os experimentos se iniciaram, então, com o treinamento empregando registros constituídos apenas de dados de visibilidade, nos formatos bipolar e métrico. Esses registros continham entre uma e oito informações de visibilidade, as quais são apresentadas à RNA no formato de um vetor n-

dimensional. A camada intermediária desta RNA inicial dispunha de cinco neurônios ocultos.

Os melhores resultados obtidos para R1, R2 e R3 foram séries temporais com quatro valores de visibilidade, de quatro horas anteriores, no formato bipolar.

b) *Etapa 2- Primeiro ajuste no número de neurônios ocultos:* Nesta etapa, cada uma das Redes R1, R2 e R3 foi novamente treinada, com os dados de entrada determinados na etapa anterior, mas com variações no número de neurônios ocultos. Os melhores desempenhos das Redes R1, R2 e R3, foram obtidos com a manutenção de cinco neurônios ocultos em todas as RNAs.

c) *Etapa 3- Inserção de novas variáveis:* Realizaram-se experimentos, combinando, uma a uma, todas as variáveis pertencentes à amostra à variável visibilidade. A inserção de novas variáveis foi realizada, arbitrando-se que as séries temporais de cada uma delas teria o mesmo tamanho da série temporal de visibilidades já definida. Para as três RNAs, a variável que apresentou melhor combinação com a variável visibilidade foi a variável precipitação. No entanto, apesar de desta combinação ter gerado resultados superiores às demais combinações, o desempenho das RNAs, em comparação à etapa de treinamento anterior foi um pouco inferior. Uma razão plausível para este resultado é que, com a inserção de mais informações de entrada, a rede necessite de mais neurônios para processá-los.

d) *Etapa 4- Segundo ajuste no número de neurônios ocultos:* Realizou-se um novo ajuste do número de neurônios ocultos, a fim de verificar-se a hipótese levantada ao final da etapa de treinamento anterior. Supondo que, com a inserção de mais informações, a rede necessite de mais neurônios para processamento, o número de neurônios inseridos na camada oculta foi variado, entre 5 e 10. No entanto, a inserção de mais neurônios ocultos não trouxe melhora no desempenho das RNAs.

Mesmo com a inserção de novas variáveis e ajuste no número de neurônios ocultos, as RNAs não superaram o desempenho obtido na etapa 2. A partir disso, supõe-se que, a inserção de mais variáveis, e novos ajustes na camada oculta também não trarão resultados significativos. Encerraram-se, então, os experimentos com inserção de novas variáveis. Assim, retornou-se à configuração obtida na etapa 2 como a configuração que passaria à próxima etapa.

e) *Etapa 5- Poda da amostra de dados de treinamento:* Por se tratar de um evento raro, o número de ocorrências do fenômeno nevoeiro é bastante inferior ao número de ocorrências de não-nevoeiro. O fato de haver um número muito maior de casos de não-nevoeiro (aproximadamente 20 vezes mais) poderia estar dificultando a identificação do fenômeno de interesse pela RNA. Para diminuir o número de casos de não-nevoeiro sem prejudicar a representatividade da amostra, foram tomados grupos de dados *n* vezes menores do que a amostra original, com *n* variando entre 2 e 20. As razões de poda indicam quantos elementos pertencentes a cada intervalo de visibilidade devem ser tomados. Por exemplo, para uma razão de poda igual a 2, a cada dois dados de cada intervalo, um irá fazer parte da nova amostra. A cada nova

amostra de dados, as RNAs foram treinadas novamente para se verificar seus desempenhos. Os melhores desempenhos obtidos para as RNAs foi a partir da poda de razão 20.

f) *Etapa 6- Terceiro ajuste no número de neurônios ocultos:* Com a poda da amostra de treinamento, o número de dados oferecidos às RNAs foi reduzido. Partindo-se da premissa de que o número de neurônios da camada oculta está diretamente ligado ao número de dados a serem processados, é possível que, após a poda da amostra, seja necessário um número menor de neurônios ocultos. Assim, as RNAs foram treinadas novamente, com as informações de entrada obtidas na etapa anterior, variando-se o número de neurônios ocultos entre 1 e 4. Os melhores resultados obtidos foram a partir de 2 neurônios dispostos na camada intermediária.

Assim, chegou-se às RNAs ótimas para a previsão do fenômeno nevoeiro para uma, duas e três horas: as séries temporais apresentadas como dados de entrada são constituídas apenas de dados de visibilidade de quatro horas anteriores, no formato bipolar; a camada intermediária deve conter dois neurônios ocultos; e, para o treinamento, devem-se podar os dados da amostra de maneira que os casos de nevoeiro e não-nevoeiro se mantenham compatíveis. O desempenho obtido a partir desta configuração ótima pode ser observado na Tabela VI, a seguir.

TABELA VI. MELHOR DESEMPENHO DA FASE DE TREINAMENTOS

	PC	ISC	V	RFA	PD
R1	0,85	0,71	0,77	0,05	0,74
R2	0,77	0,56	0,63	0,07	0,59
R3	0,71	0,45	0,52	0,09	0,47

2) *Operacionalização das RNAs ótimas:* O programa de operação possibilita que as configurações ótimas, obtidas na fase de treinamento da RNA, sejam, finalmente, utilizadas na previsão do fenômeno de interesse. Para os dados do conjunto a ser submetido à operação, foram tomadas as mesmas informações de visibilidade disponíveis nas mensagens METAR utilizadas na previsão através do método de persistência, as quais não foram empregadas no treinamento e ajuste das RNAs. Os desempenhos obtidos para as três RNAs pode ser observado na Tabela VII. Assim como o programa de treinamento, o programa de operação também foi cedido pelo Professor Dr. Ricardo Almeida, e a linguagem computacional utilizada foi FORTRAN 95.

TABELA VII. DESEMPENHO DA OPERACIONALIZAÇÃO DAS RNA

	PC	ISC	V	RFA	PD
R1	0,97	0,45	1,47	0,48	0,77
R2	0,97	0,33	1,47	0,59	0,61
R3	0,96	0,25	1,47	0,66	0,49

V. DISCUSSÃO E CONCLUSÕES

Analisando-se os desempenhos observados nos experimentos de previsão através das técnicas de climatologia, persistência e através do emprego de RNAs, podem-se destacar alguns resultados significativos das três técnicas.

Sobre a climatologia, pode-se concluir que esta técnica não produziu bons resultados para previsão do fenômeno nevoeiro. O método foi capaz de prever, aproximadamente, 20% dos eventos, e a razão de falso alarme foi de mais de 95%. Pode-se dizer que a climatologia é uma ferramenta bastante útil para a compreensão do comportamento das variáveis preditoras do fenômeno nevoeiro, tanto, que foi possível estabelecer um panorama destas variáveis quando na ocorrência do fenômeno. No entanto, assumir que, na ocorrência concomitante destas condições, haverá a formação do fenômeno, mostrou que a tendência é superestimar a ocorrência dos eventos.

Sobre a técnica de persistência, pode-se dizer que para o fenômeno nevoeiro esta técnica produz bons resultados, visto que o fenômeno em questão não sofre muitas alternâncias entre ocorrência e não-ocorrência. Desta forma, a previsão por persistência é uma forma simples de projeção do fenômeno nevoeiro, para até três horas, com índices de acerto razoáveis.

Sobre a previsão através de RNA é possível notar que previsões de restrições de visibilidade por nevoeiro podem ser feitas, apenas, a partir de séries temporais de 4 horas de outros dados de visibilidade. Informações que, inicialmente, se imaginavam ser importantes, como direção e velocidade do vento, pressão, cobertura de nuvens, e umidade relativa não se mostraram significativas quando combinadas às informações de visibilidade. Este comportamento, talvez se explique pelo fato de que o valor do alcance visual horizontal é diretamente influenciado pelo comportamento de outras variáveis meteorológicas. Assim, fornecer dados de visibilidade juntamente às informações que a geraram, pode ser, para a RNA, uma redundância de informações, e, a partir de informações repetidas a rede não é capaz de tomar novas decisões. Daí a estagnação no desempenho já na primeira combinação entre variáveis. Além disso, as forçantes sinóticas mostraram não afetar o desempenho da rede. Isso talvez se explique pelo mesmo motivo já citado: as forçantes sinóticas afetam as variáveis em superfície, que por sua vez afetam os valores de visibilidade. Por fim, todas as variáveis, de altitude ou superfície, acabam por se resumir na própria variável visibilidade.

Nota-se também, uma queda no desempenho apresentado entre o treinamento e a operação. Esse comportamento talvez se explique pelo tamanho do conjunto de dados utilizado, ou por uma possível atipicidade climática neste ano específico.

Comparando-se os resultados obtidos através das técnicas de climatologia, persistência e através do emprego de RNAs RBF na previsão de nevoeiros, é possível notar que a previsão através de climatologia resulta índices bastante inferiores aos índices obtidos através da técnica de persistência, e através do emprego da RNA. Estes últimos têm desempenhos bastante semelhantes, sendo que, seguindo a hierarquia entre as métricas de avaliação de desempenho estabelecida, pode-se dizer que os resultados obtidos a partir do emprego da RNA foram, ainda, superiores aos resultados da persistência.

Como sugestão para próximas pesquisas, propõe-se que sejam testadas outras combinações entre as variáveis meteorológicas, com séries temporais maiores, pois pode haver alguma combinação específica, que permita à RNA

compreender de forma mais refinada o comportamento do fenômeno em questão; além de se estudar o emprego de uma única RNA para realizar as previsões para os três prazos de validade simultaneamente.

REFERÊNCIAS

- [1] ALMEIDA, M. V. Aplicação de técnicas de Redes Neurais Artificiais na previsão de curtíssimo prazo da visibilidade e teto para o Aeroporto de Guarulhos- SP. Rio de Janeiro, 206 p. Tese (Doutorado em Engenharia Civil)- Faculdade de Engenharia Civil, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2009.
- [2] BRASIL, Comando da Aeronáutica. Departamento de Controle do Espaço Aéreo. Manual de Estações Meteorológicas de Superfície. MCA 105-2. Rio de Janeiro, 2015.
- [3] INFRAERO. Aeroporto Internacional Afonso Pena. Disponível em <http://www.infraero.gov.br/index.php/br/aeroportos/parana/aeroporto-afonso-pena.html>. Acesso em 07/08/2016 às 17:53.
- [4] INFRAERO. Anuário Estatístico Operacional 2015. Brasília, 2016.
- [5] OLIVEIRA, L. L.; VIANELLO, R. L.; FERREIRA, N. J. Meteorologia Fundamental. Erechim: EdiFAPES, 2001.
- [6] SILVA, I. N.; SPATTI, D. H.; FLAUZINO, R. A. Redes Neurais Artificiais para engenharia e ciências aplicadas. São Paulo: Artliber, 2010.
- [7] REDEMETS. METAR. Disponível em www.redemets.aer.mil.br.
- [8] Universidade de Wyoming. Sondagens. Disponível em <http://weather.uwyo.edu/upperair/sounding.html>
- [9] WILKS, D. S. Statistical Methods in the Atmospheric Sciences; 2nd ed; Cornell University: Department of Earth and Atmospheric Sciences. 2006.