

UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE ASTRONOMIA, GEOFÍSICA E CIÊNCIAS ATMOSFÉRICAS
DEPARTAMENTO DE CIÊNCIAS ATMOSFÉRICAS

Alessandro Santos Borges

**Previsibilidade do ozônio troposférico pela técnica de rede neural artificial:
estudo para a região metropolitana de São Paulo**

São Paulo
2008

Alessandro Santos Borges

**Previsibilidade do ozônio troposférico pela técnica de rede neural artificial:
estudo comparativo para a região metropolitana de São Paulo**

Tese de doutorado em Ciências Atmosféricas

Universidade de São Paulo
Instituto de astronomia, geofísica e ciências atmosféricas
Departamento de ciências atmosféricas

Orientação: Profa. Dra. Maria de Fátima Andrade
Co-orientação: Prof. Dr. Roberto Guardani

São Paulo
2008

Agradecimentos

Ao Instituto de Astronomia, Geofísica e Ciências Atmosféricas da Universidade de São Paulo, e seu Departamento de Ciências Atmosféricas.

À Faculdade de Engenharia Química – POLI – USP

À Coordenadoria de aperfeiçoamento de pessoal de nível superior (CAPES) pela concessão da bolsa de doutorado pelo programa PROCAD/CEPEMA.

À Companhia de Tecnologia de Saneamento Ambiental (CETESB), pela concessão de dados medidos das concentrações de poluentes e parâmetros meteorológicos das estações de qualidade do ar.

À Estação Climatológica do IAG, pelo fornecimento de dados meteorológicos.

Ao Laboratório de Meteorologia Aplicada a Sistemas de Tempo Regionais (MASTER), pelo fornecimento dos dados de previsão meteorológica.

Resumo

Este estudo trata do desenvolvimento de modelos de redes neurais artificiais (RNA) para previsão de concentrações de ozônio em superfície na Região Metropolitana de São Paulo (RMSP). Foram usados dados meteorológicos e de concentração de poluentes coletados entre 1999 e 2004 nas estações da rede de monitoramento de qualidade do ar mantida pela Cetesb (Companhia de Tecnologia e Saneamento Ambiental do Estado de São Paulo). Por meio de análises estatísticas multivariadas foram selecionadas as estações de monitoramento mais representativas para este estudo, para quais foram geradas RNAs respectivas a cada localidade. Tais modelos estatísticos utilizaram como entrada valores médios de parâmetros meteorológicos, relativos aos períodos da manhã (entre 08:00 e 12:00 horas) e da tarde (13:00h - 17:00h), obtendo-se como saída concentrações médias e máximas de ozônio para o período da tarde. Os modelos foram capazes de correlacionar as concentrações de ozônio com as variáveis de entrada dentro de níveis aceitáveis de variabilidade. A habilidade das RNAs como ferramentas de previsão foi testada por meio da aplicação de previsões meteorológicas resultantes do modelo de mesoescala RAMS (Regional Atmospheric Modeling System), fornecidas pelo Laboratório de Meteorologia Aplicada a Sistemas de Tempo Regionais (MASTER – DCA/IAG/USP). Os resultados indicam que os modelos representam uma ferramenta útil na estimativa de concentrações de ozônio, podendo então ser adotados para esse fim. A importância de cada variável para a previsibilidade do ozônio também foi avaliada para cada RNAs. As RNAs de Mauá e Santana se mostraram mais dependentes da variabilidade da temperatura, enquanto as demais foram mais influenciadas por variações no campo de vento. A umidade relativa, em geral, não foi uma variável que se destacou dentre as demais, porém foi importante para a qualidade do desempenho do modelo desenvolvido.

Abstract

This study reports the development of a neural network model to predict ozone concentrations in the Sao Paulo Metropolitan Area (SPMA). The database used to fit the model was composed of hourly average values collected from 1999 to 2004 by the air quality stations network from the Sao Paulo Environmental Protection Agency, CETESB. Through multivariate statistical analysis were selected the more representative monitoring stations for this study, from which were generated the respective RNAs to each place. Such statistical models considered average values of meteorological variables in the morning (08:00 and 12:00) and afternoon (13:00h - 17:00) to predict maximum and average ozone concentrations. The models were able to correlate ozone levels with the input variables within acceptable variability levels. The ability of RNAs as prediction tools were tested through the application of meteorological forecast from the RAMS (Regional Atmospheric Modeling System) mesoscale model, supplied by the 'Laboratório de Meteorologia Aplicada a Sistemas de Tempo Regionais' (MASTER - DCA / IAG / USP). The results indicate that the model can be adopted as a tool to predict ozone levels in the SPMA. The importance of each variable for the ozone previsibility was also evaluated for each RNA. The RNAs for Mauá and Santana showed more dependence on the temperature variability, while the others were more influenced by variations in the wind field. The relative humidity, in general, wasn't a variable that stood out among the others; however it was important for the quality of the developed model.

Sumário

INTRODUÇÃO	8
1.1 - APRESENTAÇÃO	9
1.2 - O POLUENTE OZÔNIO	10
1.3 - O OZÔNIO EM SÃO PAULO.....	13
1.4 - PREVISÃO DE POLUENTES.....	16
1.5 - OBJETIVO E ESTRUTURA DA TESE	21
TÉCNICA E METODOLOGIA DE APLICAÇÃO DAS REDES NEURAIAS ARTIFICIAIS	24
2.1 - DESCRIÇÃO DAS REDES NEURAIAS ARTIFICIAIS	25
2.2 - PRECEDENTES DA PESQUISA	30
DADOS UTILIZADOS	33
3.1 - ANÁLISE DESCRITIVA DOS DADOS	34
3.2 - CARACTERIZAÇÃO DAS VARIÁVEIS EM FUNÇÃO DAS CONCENTRAÇÕES DE OZÔNIO.....	43
3.2.1 - Óxidos de nitrogênio (NO e NO₂).....	43
3.2.2 - Temperatura do ar.....	47
3.2.3 - Umidade relativa	49
3.2.4 - Campo de vento horizontal.....	51
3.2.5 - Radiação Solar	54
3.2.6 - Altura da camada de mistura.....	56
3.2.7 - Pressão atmosférica.....	58
DESENVOLVIMENTO DA RNA PARA A ESTAÇÃO MONITORA DE QUALIDADE DO AR DO PARQUE DO IBIRAPUERA	61
4.1 – METODOLOGIA	62
4.2 - PRIMEIRO AJUSTE	63
4.3 - INCLUSÃO DA ALTURA DA CAMADA DE MISTURA.....	65
4.4 - REDE NEURAL COM SAÍDAS INDEPENDENTES.....	68
4.5 - SUBSTITUIÇÃO DOS DADOS DE VENTO DO SODAR PELOS DA ESTAÇÃO CLIMATOLÓGICA DO IAG	71
4.6 - INCLUSÃO DA PRESSÃO ATMOSFÉRICA	75
4.7 - ALTURA DA CAMADA DE MISTURA E PRESSÃO ATMOSFÉRICA COMO ENTRADAS	78
4.8 - TESTES DE MELHORIA DA REPRESENTATIVIDADE DOS DADOS	81

4.9 - EXPERIÊNCIAS VISANDO APRIMORAR A REPRESENTAÇÃO DAS CONCENTRAÇÕES ELEVADAS	83
4.10 - INFORMAÇÃO QUANTO AO DIA DA SEMANA	84
4.11- AJUSTES FINOS	90
4.12 - VALIDAÇÃO	92
4.13 - TESTE DE APLICAÇÃO DA REDE NEURAL	108
DESENVOLVIMENTO DE RNA PARA OUTRAS ESTAÇÕES DE QUALIDADE DO AR SELECIONADAS	113
5.1 - SELEÇÃO DAS NOVAS LOCALIDADES	114
5.2 - ESTAÇÃO DE MONITORAMENTO DE QUALIDADE DO AR EM MAUÁ	117
5.3 - ESTAÇÃO DE MONITORAMENTO DE QUALIDADE DO AR EM SÃO MIGUEL PAULISTA	130
5.4 - ESTAÇÃO DE MONITORAMENTO DE QUALIDADE DO AR EM PARQUE DOM PEDRO II	141
5.5 - ESTAÇÃO DE MONITORAMENTO DE QUALIDADE DO AR EM SÃO CAETANO DO SUL	145
5.6 - ESTAÇÃO DE MONITORAMENTO DE QUALIDADE DO AR EM SANTANA	157
5.7 - ANÁLISE GERAL	163
AJUSTES DE REDES NEURAIS ARTIFICIAIS INCLUINDO COMO ENTRADA AS CONCENTRAÇÕES DE POLUENTES ATMOSFÉRICOS	171
6.1 DESCRIÇÃO	172
6.2 ESTAÇÃO DE MONITORAMENTO DE QUALIDADE DO AR EM MAUÁ	172
6.3 ESTAÇÃO DE MONITORAMENTO DE QUALIDADE DO AR EM SÃO MIGUEL PAULISTA	177
6.4 ESTAÇÃO DE MONITORAMENTO DE QUALIDADE DO AR EM PARQUE DOM PEDRO II	181
6.5 ESTAÇÃO DE MONITORAMENTO DE QUALIDADE DO AR EM SÃO CAETANO DO SUL ..	184
6.6 ESTAÇÃO DE MONITORAMENTO DE QUALIDADE DO AR EM SANTANA	186
6.7 ESTAÇÃO DE MONITORAMENTO DE QUALIDADE DO AR EM IBIRAPUERA	189
CONCLUSÕES E PERSPECTIVAS PARA TRABALHOS FUTUROS	194
7.1 CONCLUSÕES	195
7.2 PERSPECTIVAS PARA TRABALHOS FUTUROS	199
REFERÊNCIAS	200

Introdução

1.1 - Apresentação

O ozônio troposférico é o principal constituinte da névoa fotoquímica característica de grandes centros urbanos com significativa frota veicular, tendo sido reconhecido como um dos principais poluentes que degradam a qualidade do ar. Por ser um poluente secundário, sua formação é decorrente de reações químicas entre seus precursores sob presença de radiação solar ultravioleta e sua concentração é regulada pelas condições meteorológicas (ZOLGHADRI et al., 2004). O ozônio tem sido motivo de preocupação em muitas regiões metropolitanas em todo o mundo, devido aos seus efeitos adversos ao sistema respiratório humano, além de prejuízos à vegetação (EPA - OAQPS, 1999). Em São Paulo é freqüente a ultrapassagem do padrão de qualidade do ar, especialmente em dias de alta insolação (CETESB, 2007), a qual está associada a admissões hospitalares por problemas respiratórios (FARHAT et al., 2005).

O desenvolvimento de modelos efetivos na previsão de concentrações de ozônio troposférico é bastante importante. A elaboração de estratégias de controle de emissões e de sistemas públicos de alerta dependem de previsões acuradas do poluente e o entendimento de sua dependência com as variáveis ambientais. Para isso, as redes neurais artificiais têm se mostrado ferramentas eficientes para classificação e regressão, representando uma metodologia alternativa que oferece maior flexibilidade do que as técnicas estatísticas convencionais usadas em previsão (HEO e KIM, 2004).

A proposta desta pesquisa foi explorar a possibilidade de se preverem concentrações médias e máximas de ozônio para o período da tarde (12:00h - 17:00h), com um intervalo de antecedência que depende da previsão meteorológica fornecida ao modelo. Esse estudo também visou um melhor entendimento das relações não lineares entre as variáveis meteorológicas no processo de formação do ozônio na baixa atmosfera.

Para tanto, foi aplicada a técnica de redes neurais artificiais, empregando fatores meteorológicos como variáveis precursoras das concentrações de ozônio. Os principais dados foram obtidos junto à Cetesb, que dispõe de uma rede de monitoramento de qualidade do ar na Região Metropolitana de São Paulo. Para os testes de aplicação recorreu-se aos prognósticos meteorológicos do laboratório Master, IAG-USP.

Este trabalho inseriu-se no âmbito do projeto de Pesquisa em Políticas Públicas: "Desenvolvimento de Tecnologia para Previsão de Ozônio na Baixa Atmosfera" (FAPESP, Processo: 98/14157-7), coordenado pelo Prof. Dr. Roberto Guardani do Departamento de Engenharia Química da Escola Politécnica da USP, que teve como objetivo desenvolver um modelo baseado em redes neurais artificiais para previsão de concentração de ozônio troposférico na RMSP, o qual foi implementado pela Companhia de Tecnologia de Saneamento Ambiental do Estado de São Paulo (CETESB).

1.2 - O poluente ozônio

Os problemas de poluição atmosférica causados pelo ozônio troposférico são de preocupação mundial e se manifestam de forma mais intensa sobre grandes centros metropolitanos e seus arredores, devido principalmente às emissões veiculares. Estudos epidemiológicos indicam que a exposição a elevadas concentrações de ozônio troposférico representa um perigo potencial à saúde humana, causando desde irritação nos olhos e vias respiratórias a sensações de aperto no peito, tosse e chiado na respiração (CETESB, 2007).

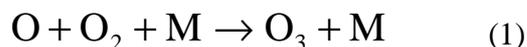
Um estudo sobre atendimento de crianças no pronto-socorro do Instituto da Criança da Faculdade de Medicina da USP mostrou um aumento das consultas por doenças respiratórias associado aos níveis de material particulado inalável (PM₁₀) e de ozônio (LIN, 1997). PEREIRA et al. (1998) mostraram também uma associação estatisticamente significativa entre NO₂ e mortalidade fetal tardia, a qual apresentou evidências de que este

poluente seja um marcador da poluição automotiva em geral, visto que um índice combinado de NO₂, CO, O₃ mostrou um efeito mais estável do que NO₂ isoladamente.

FARHAT et al. (2005) encontraram significantes associações positivas entre a poluição do ar e a morbidade respiratória em crianças. Os efeitos da poluição do ar foram detectados em termos de admissões hospitalares e atendimentos de emergência. Os casos de problemas respiratórios foram separados entre as doenças nas vias respiratórias e as que afetam o sistema pulmonar; em ambos a associação foi positiva.

As altas concentrações de ozônio também afetam a vegetação e o ecossistema adversamente (ZOLGHADRI et al., 2004), provocando danos às colheitas, à vegetação natural, plantações agrícolas e plantas ornamentais. Os efeitos observados em plantas referem-se à diminuição na taxa de crescimento, aumento da vulnerabilidade a insetos e problemas de pigmentação, devido à alterações no processo de fotossíntese (CARDOSO, 1997).

O ozônio é formado a partir de reações químicas que ocorrem no ar (WHO, 1976; ABDUL-WAHAB et al., 2004) pela reação (1) entre o oxigênio atômico e molecular:



Em que M é uma espécie que participa da reação, mas não tem sua estrutura alterada, podendo ser o N₂ ou O₂, cuja função é absorver o excesso de energia de reação e estabilizar o O₃. A principal fonte de oxigênio atômico na troposfera é a fotodissociação do dióxido de nitrogênio por radiação luminosa (hν) no intervalo de comprimento de onda entre 280 a 430nm (2):

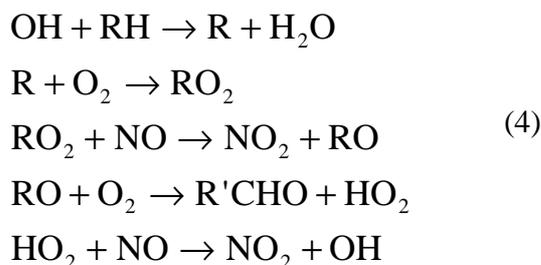


O ozônio é consumido principalmente pela reação com o NO, na qual o NO₂ e O₂ são formados novamente (3):



As três reações acima ocorrem rapidamente, estabelecendo uma concentração estável de ozônio. Entretanto, apenas estas não justificam o nível de ozônio encontrado em atmosferas urbanas poluídas. Na reação 3 é necessária uma molécula de ozônio para a geração de NO₂, consumindo NO. Uma reação que converta o NO em NO₂ sem necessidade de perda de O₃ pode levar ao acúmulo de ozônio.

Segundo FYNLAYSON-PITTS e PITTS (1997) e SEINFELD (1986) a conversão de NO em NO₂ sem consumo de ozônio é atribuída às cadeias de reações químicas iniciadas por reações entre o radical OH e compostos orgânicos voláteis (COVs), especialmente os hidrocarbonetos emitidos por automóveis e indústrias (4):



Em que *R* representa um grupo alquil. Na atmosfera o radical hidroxila pode ser formado por dissociação fotoquímica do ozônio, dos compostos carbonílicos ou do óxido nítrico, e pode ser removido através da reação com o NO₂ para formar ácido nítrico na sua forma gasosa. CARDOSO (1997), YNOUE (1999) e MARTINS (2002) apresentam os processos químicos de formação de ozônio com maiores detalhes.

Sendo assim, a formação do ozônio e de outras espécies secundárias não é instantânea. A ocorrência de picos de concentração de ozônio em regiões distantes das fontes emissoras é resultado da interação entre o transporte, a mistura turbulenta e as reações químicas que ocorrem simultaneamente (SEINFELD, 1986). A figura 1 ilustra a variação dos níveis de O₃, NO e NO₂, assim como da radiação solar ao longo do dia.

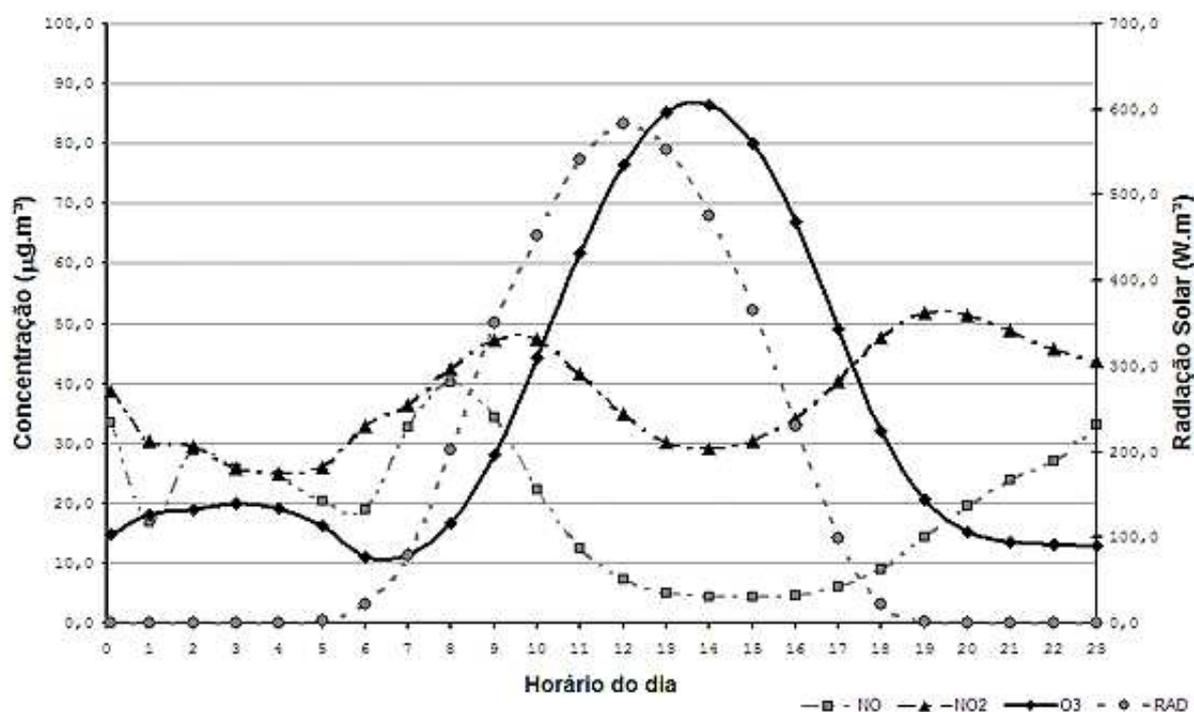


Figura 1: Variação média diurna de concentração de poluentes e radiação solar no Ibirapuera (1999-2004)

Medidas eficazes para redução do ozônio são bastante complexas, uma vez que este é um poluente formado na atmosfera a partir das emissões de diversas fontes, e sua taxa de formação, ou remoção, depende das condições meteorológicas. Sendo assim, o ozônio troposférico tem sido considerado um dos problemas mais sérios de poluição do ar (UNEP-WHO, 1994).

1.3 - O ozônio em São Paulo

A Região Metropolitana de São Paulo (RMSP) sofre com a deterioração da qualidade do ar, devida às emissões atmosféricas de cerca de 2000 indústrias de alto potencial poluidor e uma frota de aproximadamente 7,4 milhões de veículos, a qual representa 1/5 do total nacional. A região possui ao todo uma área de 8051 km² e uma população superior a 17 milhões de habitantes, que se encontra distribuída de maneira desordenada em uma área

urbanizada de 1747 km². Está a uma altitude média de 720 metros, distante 45 km do Oceano Atlântico, e sua topografia é dominada por colinas que variam de 650 a 1200 metros. A RMSP ocupa 0,1% do território brasileiro e é o terceiro maior conglomerado urbano do mundo, responsável por 1/6 do PIB nacional (CETESB, 2007; IBGE: <http://www.ibge.gov.br/>).

No caso do ozônio em superfície, o padrão de qualidade do ar ($160\mu\text{g}/\text{m}^3$ em 1h) e também o nível de atenção ($200\mu\text{g}/\text{m}^3$ em 1h) são freqüentemente ultrapassados, principalmente nos dias de alta insolação. O novo limite sugerido pela OMS ($120\mu\text{g}/\text{m}^3$ em 8h) também é ultrapassado em alguns eventos. A maior parte das estações medidoras da qualidade do ar na RMSP apresenta reduções nas concentrações de poluentes nos últimos anos, mas isto ainda não é um indicativo seguro de uma tendência (CETESB, 2007). De acordo com análises ainda preliminares houve uma tendência a aumento de ocorrência de números de violações de padrão horário de ozônio.

A figura 2 apresenta o número total de ultrapassagens do padrão e nível de atenção no período de 2002 a 2006, distribuídos por mês. O ozônio apresenta uma distribuição de episódios ao longo dos meses totalmente distinta dos poluentes primários. Como este poluente é formado na atmosfera por reações fotoquímicas que dependem da intensidade da radiação solar, dentre outros fatores, pode-se observar que a menor freqüência de episódios na RMSP ocorre nos meses de maio a julho, época de menores temperaturas e radiação solar. A partir de agosto, com o aumento da temperatura e da radiação, a freqüência de episódios de ozônio aumenta.

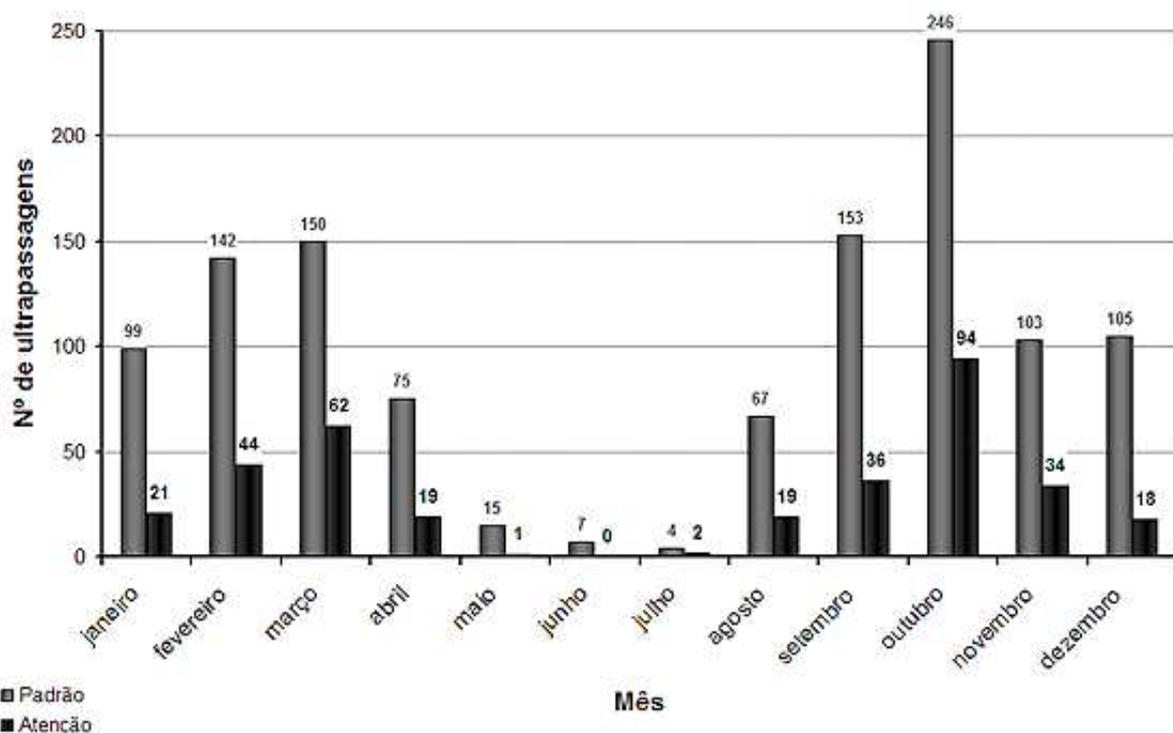


Figura 2: Número de ultrapassagens do padrão e níveis críticos de ozônio por mês (2002 a 2006) na RMSP. FONTE: CETESB (2007) - p36.

Nos meses mais quentes, como dezembro e janeiro, observa-se uma redução na frequência de episódios em comparação com os meses intermediários, o que pode ser explicado pelo aumento da nebulosidade no decorrer do dia, que reduz a radiação incidente nos baixos níveis da atmosfera (CETESB, 2007).

Vários trabalhos foram realizados para a descrição da formação e transporte do ozônio na Região Metropolitana de São Paulo (RMSP) com a utilização de modelos de qualidade do ar, como em ANDRADE et al., 2004; SANCHEZ-CCOYLLO et al., 2006 e MARTINS et al., 2006a. Os trabalhos realizados com os modelos de qualidade do ar tiveram como objetivo a compreensão e descrição da formação de ozônio na Região Metropolitana de São Paulo e identificação dos compostos orgânicos voláteis (COVs) que possuem maior potencial de formação do ozônio. Segundo MARTINS et al. 2006a, os COVs com maior efeito sobre a produção de ozônio na RMSP são os alcenos e os aromáticos. A acurácia do inventário de emissões foi verificada com simulações com o modelo fotoquímico euleriano CIT

desenvolvido no *Caltech* e na *Carnegie Mellon University* (McRAE et al., 1982a e 1982b; MILFORD et al., 1989). Os resultados indicam uma superestimativa de NO_x e uma subestimativa de compostos orgânicos voláteis (COVs), como descrito em VIVANCO et ANDRADE, 2006.

Na RMSP a principal fonte de emissão de COV é a emissão por veículos leves e pesados. Dada a grande incerteza no conhecimento de quais espécies de COVs são emitidas por essa fonte foram realizados experimentos de medidas de gases e partículas no interior dos túneis, com os resultados apresentados em SANCHEZ-CCOYLLO et al. (2008) e MARTINS et al. (2006b).

1.4 - Previsão de poluentes

Em decorrência dos males causados pelo ozônio troposférico à saúde, algumas agências de qualidade do ar vêm prevendo suas concentrações por vários anos (EPA - OAQPS, 1999). A previsão de poluição do ar vem sendo estudada há muito tempo, sendo que os primeiros modelos propostos incluíam simples correlações empíricas de dados (JORQUERA et al., 1998). A previsão da qualidade do ar, identificando a possibilidade de ocorrência de episódios críticos de poluição por ozônio, por exemplo, auxilia os órgãos de controle ambiental a avaliarem a situação e a determinarem medidas para controle de emissão de poluentes, preservando a saúde e qualidade de vida da população antes destas serem prejudicadas.

Nos Estados Unidos, por exemplo, o *South Coast Air Quality Management District* (<http://www.aqmd.gov>) prevê concentrações de ozônio e outros poluentes para 45 sub-regiões na região metropolitana de Los Angeles. Para cidades menores, algumas agências prevêm concentrações máximas para toda a cidade (como Charlotte, Carolina do Norte). As previsões

são geralmente formuladas durante a manhã ou início da tarde e então são comunicadas ao público.

Sistemas nacionais de informação, em operação em vários países da Europa, podem realizar previsões de poluentes com um dia de antecedência. Estes sistemas são baseados em modelos empíricos e/ou opiniões de especialistas, modelos estatísticos, casuais, ou a combinação desses (SCHLINK et al., 2003). A agência ambiental federal da Alemanha (*Umweltbundesamt*), por exemplo, faz previsões de ozônio e realiza medidas de qualidade do ar, disponibilizando os resultados na Internet (<http://www.umweltbundesamt.de/uba-info-daten/daten/aod.htm>).

MAFFEIS (1999) desenvolveu um modelo de previsão de CO para a região da Lombardia, na Itália. Foram usados dados de 35 estações de monitoramento de tráfego e um modelo (chamado *Forepoll*) constituído de três módulos básicos: determinístico, estocástico e bayesiano. O modelo funciona melhor para casos em que a poluição é mais intensa e difusa.

Algumas aproximações clássicas como, por exemplo, processos determinísticos ou estocásticos vêm sendo aplicadas para analisar sinais de ozônio, mas variabilidades temporais e espaciais inerentes do poluente tornam as equações básicas apenas uma aproximação cujo valor é condicionado a uma calibração apropriada de diversos parâmetros de ajuste, tais como informações de estabilidade atmosférica, coeficientes de difusão em direções laterais e vertical, mecanismos de reação e dados cinéticos, os quais não são muito bem entendidos (PASTOR-BÁRCENAS et al., 2005). Adicionalmente, devido à influência das condições meteorológicas sobre as concentrações de ozônio e a grande incerteza associada aos dados meteorológicos de entrada, é muito difícil obter uma boa concordância entre o modelo dinâmico e os dados observados (WANG et al., 2003).

Para a RMSP, FREITAS et al. (2005) examinaram o potencial do uso de um mecanismo químico simplificado para previsão de ozônio dentro de um modelo de

mesoescala, o *Regional Atmospheric Modeling System* (RAMS), construindo uma estrutura *in-line*. Em geral, os resultados mostraram boa concordância entre os valores modelados e os previstos, porém foram observados erros significativos de previsão e fraca correlação, que podem estar relacionados com a comparação feita entre os resultados do modelo em célula de grade e as medidas pontuais das estações de qualidade do ar.

Técnicas estatísticas, as quais se baseiam em padrões obtidos de observações de dados atmosféricos, se mostram rápidas e mais flexíveis. Durante a última década, tem havido considerável progresso no desenvolvimento de técnicas de modelagem estatística e tratamento de dados (SCHLINK et al., 2003).

KUKKONEN et al. (2003) usaram cinco modelos de redes neurais artificiais, um modelo estatístico linear e um sistema de modelagem determinística (DET) para avaliar a previsão de concentrações de NO₂ e PM₁₀ urbanos. Foram usadas concentrações horárias desses poluentes providas por duas estações de monitoramento no centro de Helsinque, Finlândia, entre 1996 e 1999. Em geral as redes neurais apresentaram melhor desempenho na previsão de NO₂ em comparação ao modelo determinístico, porém o resultado não foi tão bom para previsão de PM₁₀. Entretanto, as redes neurais se saíram melhores que os modelos estatísticos lineares para previsão de ambos os poluentes.

SCHLINK et al. (2003) realizaram um exercício de intercomparação utilizando 15 diferentes técnicas estatísticas para previsão de ozônio, aplicando-as em 10 séries de dados representando diferentes condições meteorológicas e de emissão na Europa. Esses modelos também foram comparados com um modelo determinístico de trajetória química. Os resultados demonstraram que, para a maioria das situações, os modelos de redes neurais artificiais são os mais recomendáveis para previsão de ozônio, por serem capazes de tratar associações não lineares e por serem facilmente adaptáveis às condições específicas de uma dada região.

JORQUERA et al. (1998) desenvolveram um estudo indicando a importância da previsão de concentrações máximas de poluentes atmosféricos, dado o sério impacto à saúde que elas causam. Nesse trabalho foram realizadas comparações entre diferentes sistemas de previsão de máxima diária de ozônio na cidade de Santiago, no Chile. Dentre as técnicas de modelagem utilizadas estavam as séries temporais lineares, redes neurais artificiais e a lógica *Fuzzy*, sendo que as duas últimas apresentaram melhores resultados.

CHALOULAKOU et al. (2003) desenvolveram um estudo comparativo com modelos de redes neurais artificiais e de regressão linear múltipla, para se prever em 24hs a concentração máxima horária de ozônio na bacia de Atenas, Grécia, em quatro estações de monitoramento representativas. Os resultados de desempenho indicaram que as redes neurais artificiais produziram melhores estimativas de ozônio, enquanto os modelos lineares foram menos eficientes quanto à acurácia de previsão de altas concentrações. A ultrapassagem do limite europeu, de $180 \mu\text{g}/\text{m}^3$, foi previsto com sucesso pelas redes neurais em 72% dos casos.

GARDNER e DORLING (2000) realizaram um estudo de caso utilizando dados do Reino Unido, demonstrando que modelos estatísticos, que trabalham com concentrações horárias de ozônio na superfície, necessitam de interações e relações não lineares entre as variáveis preditoras para descreverem o comportamento do ozônio com acurácia. Foram feitas comparações entre modelos de regressão linear, CART (*Classification and Regression Trees*) e redes neurais artificiais. Foi observado que modelos de redes neurais com múltiplas camadas conseguem capturar melhor essas relações fundamentais entre as variáveis. Entretanto, o CART mostrou-se mais interpretável fisicamente.

Modelos lineares não são adequados para lidar com a previsão de ozônio, pois produzem resultados relativamente pobres devido à complexidade induzida pela presença de uma grande variedade de processos subjacentes (PASTOR-BÁRCENAS et al., 2005). KOÇAK et al. (2000) utilizaram um método de predição local, através da teoria do caos, para

prever concentrações de ozônio na cidade de Istambul, na Turquia, em diferentes estações de medida. Segundo os autores, seus resultados mostram claramente que as concentrações de ozônio são governadas por um sistema caótico determinístico.

Os resultados de diversas pesquisas demonstram que o desempenho das redes neurais artificiais é geralmente superior em comparação com métodos estatísticos tradicionais, como a regressão múltipla, árvores de classificação e regressão (CART) e modelos auto-regressivos (GRIVAS e CHALOULAKOU, 2006). Tal situação encoraja o uso das redes neurais artificiais como ferramenta de estudo da previsibilidade do ozônio troposférico.

Modelos de redes neurais artificiais vêm sendo usados para previsão de concentração de vários poluentes em diferentes escalas temporais, obtendo-se resultados muito bons (GRIVAS e CHALOULAKOU, 2006). Entretanto, os modelos de redes neurais artificiais não são recomendados para analisar cenários futuros de poluição do ar, e também não são aplicáveis para previsões de distribuições espaciais de concentração em áreas urbanas (KUKKONEN et al., 2003).

GRIVAS e CHALOULAKOU (2006) avaliaram o potencial de diversos modelos de redes neurais artificiais para previsão de concentrações horárias de PM_{10} . O estudo envolveu quatro localidades com problemas de poluição por material particulado em Atenas, na Grécia. Os resultados foram satisfatórios, com coeficiente de determinação (R^2) variando entre 0,50 e 0,67. Estes resultados são superiores em comparação ao obtido pelos modelos de regressão linear múltipla desenvolvidos em paralelo (R^2 entre 0,29 e 0,35). O desempenho das redes neurais mostrou-se adequado para casos de altas concentrações.

PASTOR-BÁRCENAS et al. (2005) aplicaram a técnica de redes neurais artificiais para estudar a relação do poluente ozônio com seus precursores e os parâmetros meteorológicos em Valência, na Espanha. A análise indicou que a temperatura e a radiação solar foram os fatores mais importantes. Entretanto, as concentrações de NO e NO_2 foram as

menos relevantes, o que sugere que o ozônio registrado na estação de medida foi originado em outras localidades e transportado pelo vento.

ZOLGHADRI et al. (2004) desenvolveram um sistema público de alerta de poluição do ar em Bordeaux, na França, combinando técnicas de modelagem espacial de estado não-linear e redes neurais de múltipla camada. Variáveis meteorológicas foram inseridas como entrada para se estimar a máxima concentração de ozônio do dia seguinte e, adicionalmente, a duração do episódio de *smog*.

HEO e KIM (2004) desenvolveram um método de previsão de máximo diário de concentrações de ozônio para quatro pontos em Seul, na Coréia do Sul, utilizando as técnicas de lógica fuzzy e redes neurais artificiais. WANG et al. (2003) trabalharam com a própria técnica de rede neural artificial para poderem prever o máximo diário de concentração de ozônio em Hong Kong, na China.

RUIZ-SUÁREZ et al. (1995) desenvolveram uma ferramenta de previsão de ozônio utilizando duas estruturas de redes neurais artificiais. Uma chamada de Memória Associativa Bidirecional (*BAM*), e a outra de Memória Associativa Holográfica (*HAM*). Para esse estudo foram aplicados dados de cinco estações de monitoramento na Cidade do México. Os resultados mostraram que as duas redes foram capazes de prever razoavelmente as concentrações de ozônio, sendo que a rede *HAM* foi mais rápida que a *BAM*.

1.5 - Objetivo e estrutura da tese

Esta pesquisa teve como objetivo o estudo da previsibilidade de concentrações de ozônio troposférico na Região Metropolitana de São Paulo, avaliando os resultados obtidos pela aplicação de modelos de rede neural artificial. Foram empregados, principalmente, parâmetros meteorológicos como variáveis precursoras, cujos dados foram obtidos da rede de monitoramento de qualidade do ar mantida pela Cetesb.

A primeira etapa concentrou-se na aplicação das redes neurais artificiais para os dados da estação de qualidade do ar da Cetesb localizada no interior do Parque do Ibirapuera (1999 a 2003), por esta ser uma estação de referência ao estudo (item 2.2 - Precedentes da pesquisa). Os dados foram examinados, a fim de se estabelecerem as relações primárias entre as variáveis, como correlação, representatividade e similaridade (3.1 - Análise descritiva dos dados). Também foram feitas comparações das variáveis precursoras com as concentrações de ozônio, para entender o comportamento desta última em relação à variabilidade dos diferentes parâmetros individualmente (3.2 - Caracterização das variáveis em função das concentrações de ozônio).

O desenvolvimento da RNA para a estação de qualidade do ar no Ibirapuera envolveu uma série de testes de seleção de variáveis de entrada e ajustes de parâmetros do modelo estatístico. Inicialmente foram usadas médias de temperatura do ar, umidade relativa, radiação solar e campo de vento horizontal (expresso em suas componentes zonal e meridional).

Em seguida foi estudada a inclusão das médias de medida de altura da camada de mistura (item 4.3) e pressão atmosférica (item 4.6), mas a introdução isolada dessas variáveis não trouxe benefícios que justificassem sua implementação. Apenas com a adição de ambas (item 4.7) os resultados apresentaram uma melhora expressiva.

Foi estudada também a substituição do campo de vento em 50m por medidas feitas pela estação climatológica do IAG (item 4.5), mas os resultados não foram melhores com essa mudança. Já a adição de uma informação quanto ao dia da semana (item 4.10) contribuiu para a melhora dos resultados, sendo esta então adotada como uma das entradas da RNA.

Foram feitos outros testes de avaliação, como o uso de saídas independentes no modelo (item 4.4), melhoria da representatividade dos dados (item 4.8), e experiências visando enfatizar a importância das concentrações elevadas (item 4.9), mas nenhum desses testes implicou mudança da estrutura da RNA. Apenas com o ajuste fino (item 4.11), em que

parâmetros internos da rede foram variados, houve melhoria na qualidade das estimativas de concentrações de ozônio.

Sendo assim, a RNA para o Parque do Ibirapuera foi definida tendo como entradas medidas de: temperatura do ar, umidade relativa, radiação solar, campo de vento em 50m (obtido do sondador acústico SODAR e expresso em termos de componentes zonal e meridional), altura da camada de mistura e pressão atmosférica. Cada uma foi representada por valores médios, para os períodos da manhã (08:00-12:00h) e da tarde (13:00-17:00h). Além disso, foi inserida à entrada da rede uma informação binária quanto ao dia da semana (0- Seg. a Sex.;1- Sab. e Dom.). Como saída o modelo obtém estimativas de concentração média e máxima de ozônio para o período da tarde.

Esse modelo estatístico foi validado com dados de 2004 da estação Ibirapuera (item 4.12), com estudo dos desvios das estimativas (diferença entre previsto e observado). Também foi feito um teste de aplicação da RNA, fornecendo ao modelo previsões meteorológicas obtidas do laboratório Master, do IAG (item 4.13).

A etapa seguinte consistiu na aplicação da técnica de redes neurais para outras estações de monitoramento da Cetesb (item 5.1). Sendo assim, foram selecionadas as regiões de Mauá (item 5.2), São Miguel Paulista (5.3), Parque Dom Pedro II (5.4), São Caetano do Sul (5.5) e Santana (5.6). Ao final foi feita uma análise geral dos resultados (item 5.7), comparando o desempenho das diferentes redes neurais artificiais.

No capítulo 6 foram ajustadas novas redes neurais referentes a cada estação selecionada, incluindo medidas de concentração de poluentes à entrada dos modelos. Sendo possível, dessa forma, avaliar o impacto da química atmosférica sobre a qualidade das estimativas de ozônio em comparação às RNAs apresentadas nos capítulos 4 e 5.

As conclusões finais das interpretações dos resultados são apresentadas no capítulo 7.

Técnica e metodologia de aplicação das redes neurais artificiais

2.1 - Descrição das redes neurais artificiais

A teoria de redes neurais artificiais vem se consolidando, mundialmente, como uma nova e eficiente ferramenta para se lidar com o estudo de fenômenos complexos, em que extensas massas de dados devem ser modeladas e analisadas em um contexto multidisciplinar, envolvendo, simultaneamente, tanto os aspectos estatísticos e computacionais como os dinâmicos e de otimização. Uma referência geral sobre redes neurais pode ser vista em DAYHOFF (1990). Sua aplicação no tratamento de dados é especialmente relevante em casos que apresentem não-linearidades, devido à capacidade de tais modelos poderem reconhecer um padrão em um processo sem a necessidade de aplicação das leis físicas e químicas que governam o sistema. Tendo em vista a característica complexidade envolvida na modelagem atmosférica de concentrações de ozônio, com um número de fatores de diferentes naturezas atuando no sistema, o uso de um modelo de rede neural artificial representa uma ferramenta válida para o estudo de tendências de níveis de concentração desse poluente (GUARDANI et al., 1999).

As redes neurais têm sua origem na abstração de processos observados nos sistemas nervosos biológicos. A rede neural consiste de neurônios artificiais processadores, os quais são estruturas lógico-matemáticas que procuram simular a forma, o comportamento e as funções de um neurônio biológico. Assim, os dendritos dos neurônios foram substituídos por entradas, cujas ligações são ponderadas através de pesos, simulando os ganhos sinápticos, compondo assim canais de fluxo de informação entre neurônios, usualmente denominados interconexões. Os círculos na figura 3 representam os neurônios, nos quais as entradas da rede são processadas.

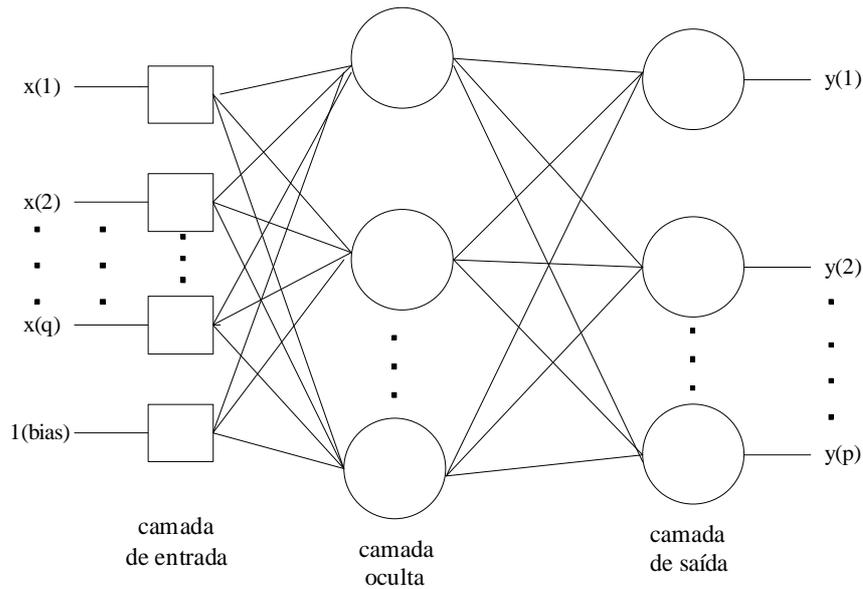


Figura 3: Ilustração de uma estrutura de rede neural artificial, de múltiplas camadas com sentido único (“*multilayer feedforward network*”)

Combinando diversos neurônios artificiais forma-se a rede neural artificial (figura 3). As entradas, simulando uma área de captação de estímulos, podem ser conectadas em muitos neurônios, resultando assim em uma série de saídas, em que cada neurônio representa uma saída. Essas conexões, em comparação com o sistema biológico, representam o contato dos dendritos com outros neurônios, formando assim as sinapses.

Os estímulos captados pelas entradas são processados por cada neurônio, ilustrado na figura 4, calculando a soma ponderada de todos os sinais interconectados vindos da camada anterior mais um termo independente de correção (bias):

$$S_j = \sum_{i=1}^N W_{i,j} x_j + W_{N+1,j} \quad (4)$$

Onde:

S_j = soma ponderada das entradas para um neurônio.

$W_{i,j}$ = Peso da variável i , no neurônio j .

x_j = variável normalizada de entrada i , no modelo de rede neural.

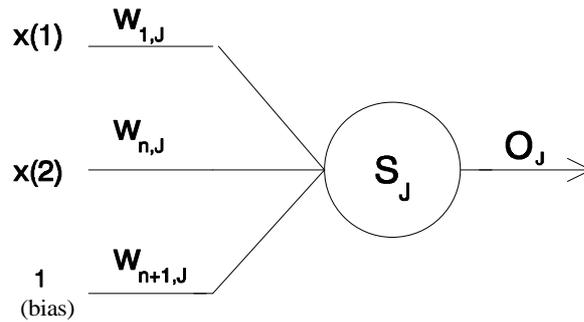


Figura 4: Ilustração de um neurônio processador j em uma rede neural artificial.

Assim, é gerada uma saída através de uma função de ativação, que corresponde ao limiar de disparo de um neurônio biológico. Essa função pode assumir diferentes formas, mas é comumente expressa como uma função sigmóide, também referida como função logística (equação 5), que possui a vantagem de ser simétrica e possuir todas as derivadas contínuas.

$$f(s_j) = \frac{1}{1 + e^{-s_j}} \quad (5)$$

A rede neural mais comumente empregada na modelagem e simulação de processos é a *feedforward network*. Nesta rede, os neurônios de uma camada são conectados a todos os neurônios da camada seguinte. Cada informação que sai de um neurônio de uma camada (i) é ponderada por um dado peso ($W_{i,j}$) e enviada a todos os neurônios da camada seguinte (j). A convergência da rede neural está ligada à melhor escolha possível do conjunto de pesos W .

A rede selecionada neste estudo possui três camadas: a camada de entrada de dados, a camada oculta (*hidden layer*) e a camada de saída. A camada de entrada não efetua nenhum processamento. A fim de eliminar problemas numéricos e diminuir as dificuldades de convergência, todos os dados do conjunto de aprendizagem são normalizados entre zero e um. Além dos dados de entrada, alimenta-se também um valor de referência (*bias*) à rede.

A fase de aprendizado ou treinamento é basicamente um problema de otimização e freqüentemente exige a apresentação da base de dados à rede por milhares de vezes. Desta forma, depara-se com as mesmas dificuldades associadas à otimização de processos, tais

como: convergência, existência de mínimos locais, esforço computacional, entre outras. A metodologia para a fase de treinamento de redes neurais foi basicamente consolidada por RUMELHART e McCLELLAND (1986), conhecida como método de retropropagação (*backpropagation*). Tal metodologia foi obtida do método de otimização do gradiente descendente e tem sido utilizada na grande maioria dos trabalhos publicados na área de redes neurais. Este é obtido a partir da minimização do resíduo quadrático (E) com respeito a $W_{i,j}$. Desta forma, o gradiente $\delta E/\delta W_{i,j}$ pode ser calculado e, conseqüentemente, a direção a ser seguida no processo de minimização do resíduo. Em cada passo da otimização (m) a atualização dos pesos pode ser dada pela equação (6):

$$W^{(m+1)} = W^{(m)} + \Delta W^{(m)} \quad (6)$$

Esta atualização pode ser feita, alternativamente, utilizando a técnica do “momento”, baseada na inclusão do termo de momento (α) proporcional à variação dos pesos, tal que $0 < \alpha < 1$:

$$W_{i,j}^{(m+1)} = (1 - \alpha)W_{i,j}^{(m)} + \alpha \cdot \Delta W_{i,j}^{(m)} \quad (7)$$

A diferença entre o método do gradiente descendente e o de retropropagação é que este último usa apenas um subconjunto de dados entrada/saída de cada vez, enquanto o método do gradiente tradicional emprega todos os dados do conjunto de aprendizado.

Para a camada de saída, a variação dos pesos referentes ao neurônio de saída k , incidentes sobre a saída do neurônio j , da camada oculta, a cada iteração m , é dada por (8):

$$\Delta W_{j,k}^{(m)} = \eta \cdot f'(S_k) (y_k^{(m)} - O_k^{(m)}) O_j^{(m)} \quad (8)$$

Para a camada oculta, a variação dos pesos referentes ao neurônio oculto j , incidentes sobre a entrada X_i , a cada iteração m , é dada por (9):

$$\Delta W_{i,j}^{(m)} = \eta \cdot f'(S_j) \left\{ \left[\sum_{k=1}^p f'(S_k) (y_k^{(m)} - O_k^{(m)}) W_{j,k}^{(m-1)} \right] X_i^{(m)} \right\} \quad (9)$$

Nas Equações (8) e (9), η representa o fator de aceleração, e f' é a derivada da função sigmoidal, dada por (10):

$$f'(z) = \frac{e^{-z}}{(1 + e^{-z})^2} \quad (10)$$

No algoritmo de retropropagação, em cada iteração, são calculados inicialmente os pesos para a camada de saída (O_k) e os pesos para a camada oculta (O_j), com os pesos estimados na iteração anterior, pelas equações (4) e (5), para os n neurônios da camada oculta e p neurônios da camada de saída; em seguida, as saídas da rede são utilizadas para o cálculo dos novos valores dos pesos, utilizando-se as equações (10), (9) e (8) para as variações e a equação (6) ou (7) para os novos valores. Isto é feito para cada apresentação dos pontos experimentais. Assim, inicialmente os erros são propagados para trás, na rede, e utilizados para estimar os novos valores dos pesos nas camadas anteriores.

Uma vez determinado o conjunto de pesos (W) que minimiza o resíduo (E), para simular o processo basta multiplicar um conjunto de dados de entrada qualquer pelos respectivos pesos, respeitando a estrutura da rede, obtendo-se o conjunto de resultados após a camada de saída da rede.

A obtenção de uma rede neural artificial confiável e robusta depende principalmente da escolha das variáveis envolvidas no processo, assim como da disponibilidade desses dados, além da definição de padrões internos como o número de neurônios na camada oculta, limites de normalização e parâmetros de aprendizado.

A principal característica na estrutura de redes neurais consiste em sua habilidade de adaptação e aprendizagem pelo ambiente. Portanto, os modelos de redes neurais podem lidar com dados imprecisos e situações não totalmente definidas. Uma rede treinada de maneira adequada tem a habilidade de generalizar quando é apresentada a valores das entradas que não estão presentes em dados usados em seu ajuste. A característica mais significativa das redes

neurais está em sua habilidade de aproximar qualquer função contínua não linear, o que as têm tornado úteis para modelar sistemas não lineares na combinação de controladores não lineares. Redes neurais podem ter várias entradas e várias saídas, sendo facilmente aplicáveis a sistemas com muitas variáveis.

2.2 - Precedentes da pesquisa

GUARDANI et al. (1999) testaram diferentes aproximações correlacionando os níveis de ozônio com precursores químicos e variáveis meteorológicas das estações de qualidade do ar da CETESB na Região Metropolitana de São Paulo, através de modelos de redes neurais artificiais.

Os algoritmos de preparação de dados, ajuste de redes neurais e simulação foram desenvolvidos pelo Laboratório de Simulação e Controle de Processos (LSCP) do Departamento de Engenharia Química, Escola Politécnica da Universidade de São Paulo. Diversos parâmetros da rede neural foram variados, tendo como critério de ajuste a medida do erro quadrático médio baseado no conjunto de dados selecionados para ajuste da rede (conjunto de aprendizado). A validação foi feita com um subconjunto de dados selecionado para tal fim (conjunto de teste).

No estudo de GUARDANI et al. (1999), os dados para ajuste e teste da rede neural foram obtidos de quatro estações de monitoramento da qualidade do ar da Cetesb na RMSP: Centro, Parque do Ibirapuera, Pinheiros e Mauá. Todos os dados se referem ao período de Junho a Dezembro de 1996, embora algumas partes do estudo tenham sido baseadas em períodos mais curtos.

As simulações indicaram uma resposta do ozônio em relação a diferentes poluentes atmosféricos e condições de tempo. Previsões utilizando o modelo demonstraram uma boa concordância com os valores medidos de concentração de ozônio. Sendo assim, a combinação

de redes neurais artificiais com prognósticos de parâmetros meteorológicos mostrou-se como uma ferramenta válida para previsão de ozônio, possibilitando seu controle em diferentes regiões da cidade.

GUARDANI et al (2003) empregaram um estudo estatístico para avaliar o comportamento das concentrações de ozônio em diferentes locais da RMSP. O estudo foi baseado em dados obtidos pela rede de estações de monitoramento da Cetesb, com ênfase nas concentrações de poluentes e variáveis meteorológicas, abrangendo locais com características diferentes de tráfego e atividades econômicas.

A aplicação combinada de análise de componentes principais com técnicas de agrupamento (*cluster*) sobre dados coletados entre 1997 e 2000 conduziu à identificação de relações implícitas entre variáveis associadas com processos dominantes relacionados à formação de ozônio em diferentes localidades. Similaridades entre diferentes regiões da cidade também foram detectadas e associadas com características locais.

Em termos de máximo diário de concentração de ozônio, a RMSP foi classificada em dois grupos através da técnica de agrupamento. O primeiro foi formado pelas estações Dom Pedro, Congonhas, Lapa e Osasco, as quais são caracterizadas por estarem próximas de vias de tráfego pesado; raramente apresentando índices elevados de ozônio. O segundo grupo consiste das estações Mooca, Parque do Ibirapuera, São Caetano do Sul, São Miguel Paulista e Mauá, as quais se localizam afastadas das vias de tráfego intenso; Ibirapuera e Mauá foram as que apresentaram os níveis mais elevados de concentração de ozônio. Mauá se dispõe de maneira mais isolada desse grupo de estações considerado, provavelmente por possuir um perfil peculiar, devido a sua proximidade a um importante complexo petroquímico.

Para identificar possíveis relações entre as variáveis medidas em estações representativas de cada grupo, foi empregada a análise de componentes principais (PCA) aos dados provenientes de: Dom Pedro, a qual foi representativa do primeiro grupo, em que as

estações se localizavam próximas de via de tráfego intenso; Parque do Ibirapuera, por aparecer agrupada com a maioria das estações da região metropolitana de São Paulo e por apresentar elevados níveis de ozônio; e Mauá, a qual permaneceu mais isolada das demais estações.

No caso de Ibirapuera, a primeira componente respondeu por aproximadamente 40% da variância dos dados, correspondendo às variáveis meteorológicas e o ozônio, indicando que o comportamento é regido principalmente pelas condições meteorológicas. A segunda componente ficou caracterizada pelos poluentes NO e NO₂, os quais são precursores do ozônio.

Tal comportamento difere da estação Dom Pedro, cujas variáveis dominantes no primeiro componente correspondem ao CO, hidrocarbonetos (exceto metano), NO e NO₂. As condições meteorológicas atuam mais nas componentes 2 e 3.

Embora em Mauá as variáveis meteorológicas também atuem na primeira componente principal, há influência da direção e velocidade do vento, além das concentrações de NO nas últimas componentes mais significativas.

Dados Utilizados

3.1 - Análise descritiva dos dados

Os dados utilizados nesta pesquisa, referentes aos parâmetros meteorológicos e concentrações de poluentes, foram obtidos da Companhia de Tecnologia de Saneamento Ambiental do Estado de São Paulo (CETESB), que possui uma rede de estações de monitoramento (figura 5 e tabela 3.1.1) cobrindo a Região Metropolitana de São Paulo (RMSP). Adicionalmente, para uma das etapas deste estudo (item 4.5), recorreu-se aos dados da estação climatológica do IAG, localizada no Parque do Estado, na Água Funda. Para testes de aplicação foram empregadas previsões meteorológicas resultantes do modelo de mesoescala RAMS, fornecidas pelo Laboratório de Meteorologia Aplicada a Sistemas de Tempo Regionais (MASTER – DCA/IAG/USP).

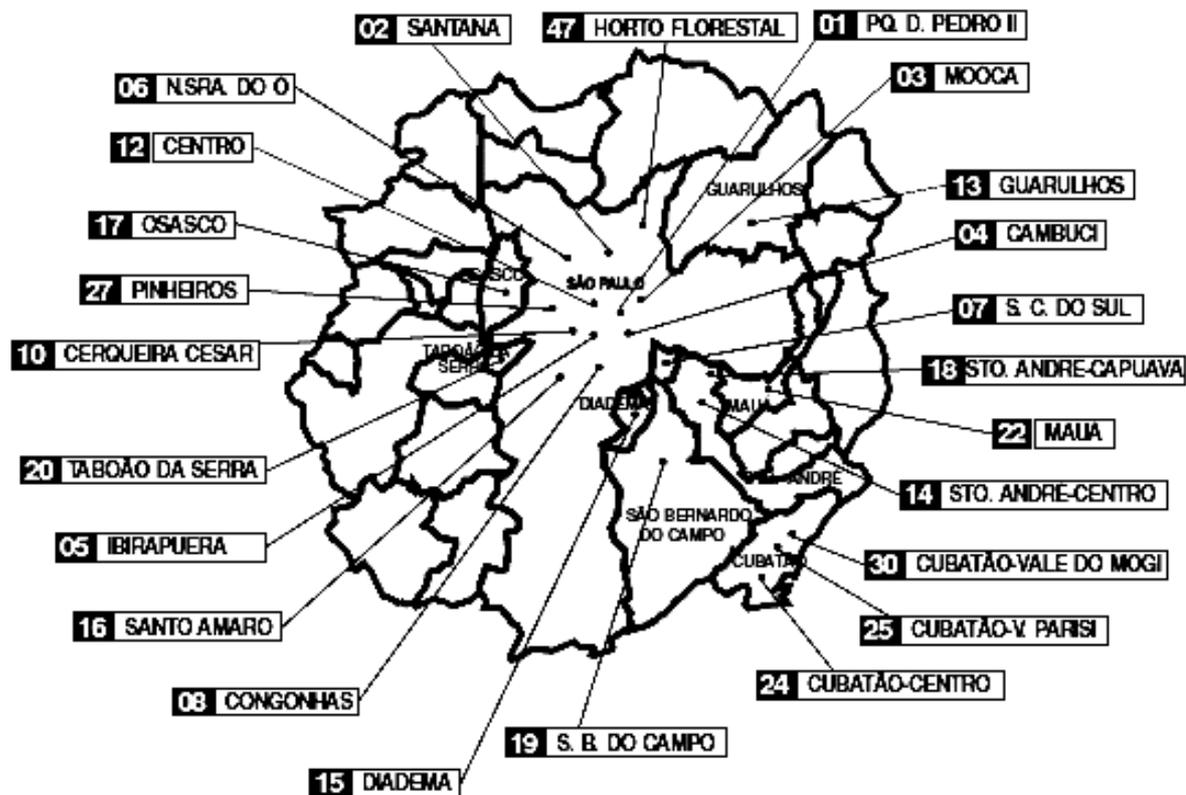


Figura 5: Localização das estações da rede automática da Cetesb na RMSP. FONTE: CETESB (2007) – p26.

Tabela 3.1.1: Demais parâmetros medidos pelas estações de monitoramento de ozônio consideradas nesse estudo.

Estação	Temp.	UR	Rad.	Press.	WD	WS	PI	NO	NO ₂	NO _x	NMHC	CH ₄	CO	SO ₂
Diadema							•							
Ibirapuera	•	•	•	•	•	•	•	•	•	•			•	•
Mauá							•	•	•	•				
Moóca					•	•	•							
P. D. Pedro II	•	•			•	•	•	•	•	•			•	•
Pinheiros	•	•			•	•	•	•	•	•			•	
Santana					•	•	•							
São Caetano do Sul	•	•			•	•	•	•	•	•	•	•	•	•
São Miguel Paulista	•	•			•	•	•							

De acordo com estudos prévios (item 2.2), verificou-se que a estação de monitoramento situada no Parque do Ibirapuera é a mais representativa para o estudo da previsibilidade do ozônio troposférico na RMSP, por apresentar maior incidência de elevados índices de ozônio, ser uma das mais completas em número de amostradores e servir, estatisticamente, de referência para toda a região (GUARDANI et al., 2003). Posteriormente foram adotadas outras estações que possuíam monitores em tempo real de ozônio (seção 5.1), porém, a primeira etapa da pesquisa concentrou-se no trabalho com os dados de Ibirapuera.

O período considerado neste estudo compreende os anos de 1999 a 2003, para o desenvolvimento da metodologia de previsão; e o ano de 2004 para validação da mesma. As principais variáveis consideradas foram: temperatura do ar, umidade relativa, radiação solar, campo de vento horizontal em 50m (obtidos do SODAR, instalado no Parque do Ibirapuera sob responsabilidade da CETESB, que disponibilizou os dados), altura da camada de mistura, pressão atmosférica e a concentração de ozônio. Medidas de concentrações de poluentes, como as de NO e NO₂, também foram apreciadas, mas não foram incorporadas ao modelo final, uma vez que ainda não há um sistema confiável para previsão de seus valores (capítulo 6). Os dados passaram por uma análise preliminar para eliminação de dados contendo erros e anômalos (*outliers*). A soma desses dados e as falhas de registro representaram cerca de 30% dos dados no período.

A escolha da medida de vento horizontal à altura de 50 m, obtida pelo sondador acústico (SODAR) da CETESB, deve-se ao fato das medições de vento, obtidas pela estação de monitoramento em si, estarem sujeitas a interferências de obstáculos. Por estar situada no interior do Parque do Ibirapuera, a estação se encontra próxima de árvores que obstruem a passagem do vento, tornando suas medidas inadequadas a este tipo de estudo. Para compensar este problema foram utilizados os dados do SODAR, que esteve em operação no Ibirapuera entre 1999 a 2004. O nível de 50m de altura demonstrou ser o mais apropriado, por representar um vento de superfície livre da maior parte dos obstáculos que poderiam gerar problemas de canalização. Também foi estudada a viabilidade de se empregarem os dados da estação climatológica do IAG. No entanto, um teste de ajuste de rede neural artificial indicou que essas medidas não são adequadas ao estudo (item 4.5).

Para caracterizar as variações dos parâmetros meteorológicos e das concentrações de poluentes, foram preparados gráficos com médias mensais, entre 1999 e 2004. A figura 6 mostra o comportamento dos poluentes ao longo do ano, em que se observa que os perfis de NO e NO₂ contrapõem o de O₃. Enquanto os óxidos de nitrogênio apresentam índices mais elevados nos meses de inverno, o ozônio encontra-se em seus níveis mais baixos, vindo a apresentar picos de altas concentrações nos meses de primavera e verão. Na figura 7 é ilustrada a relação do ozônio com a radiação solar, ficando clara a dependência do poluente com essa variável ao longo dos meses, conforme o esperado.

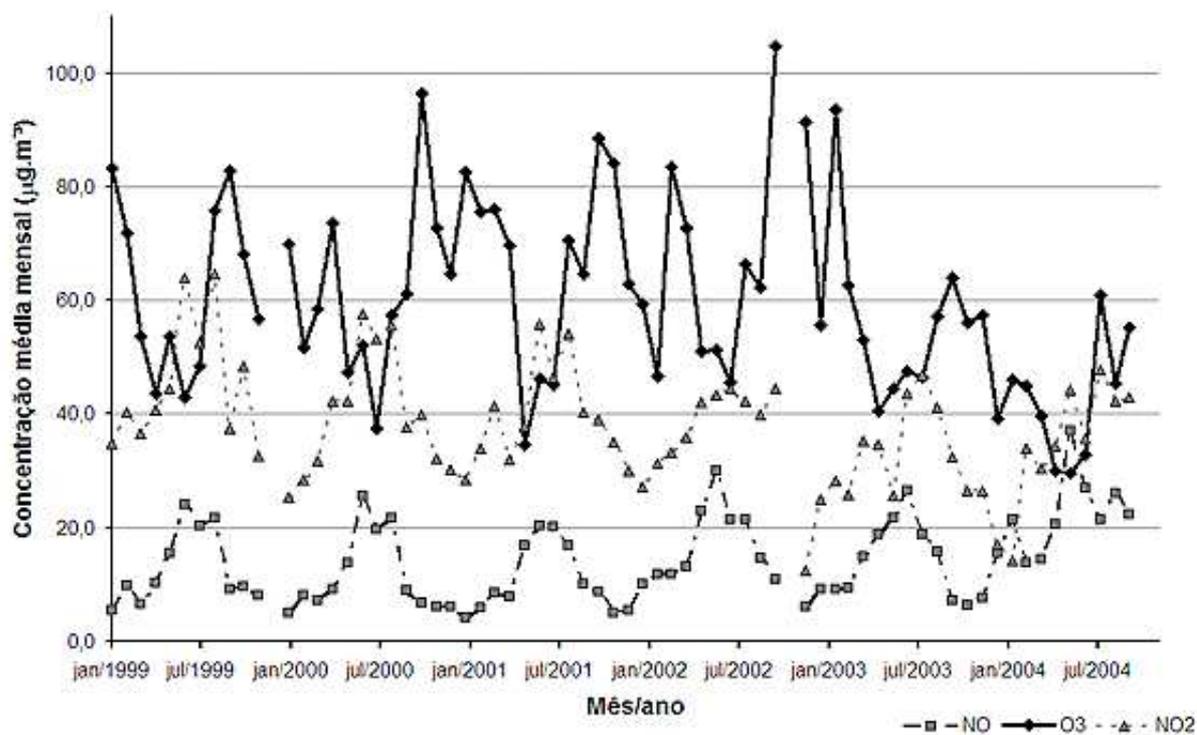


Figura 6: Variações das médias mensais de NO, NO₂ e O₃ entre 1999 e 2004.

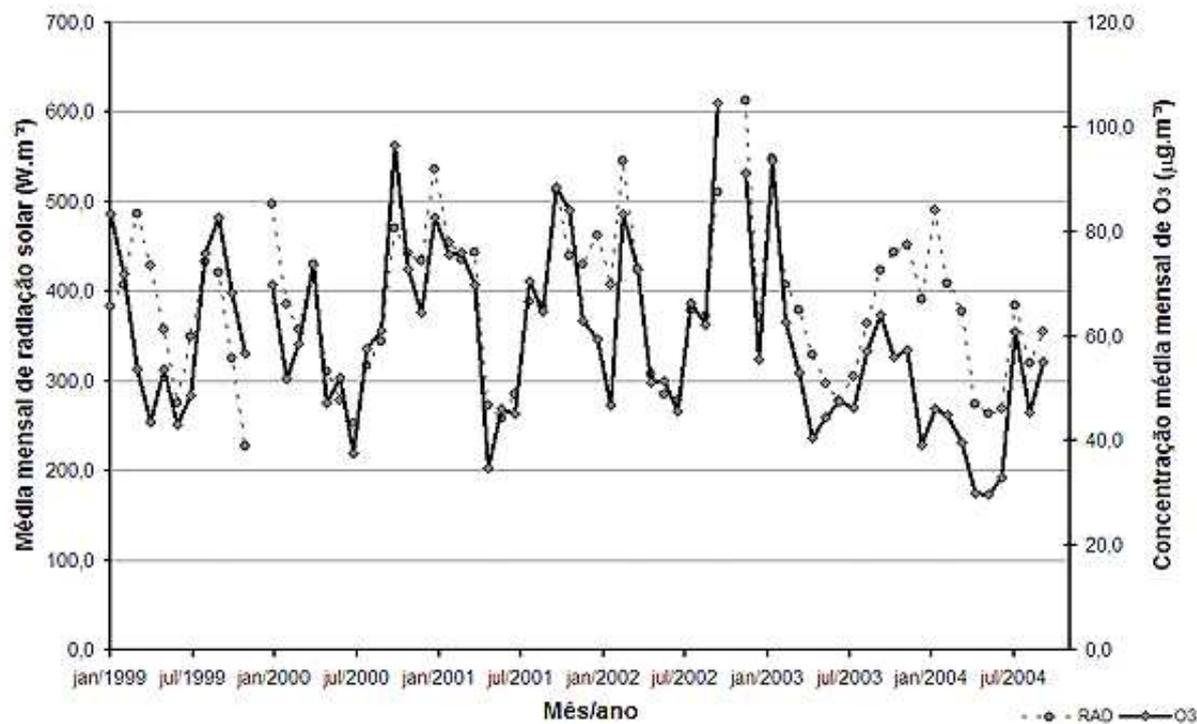


Figura 7: Variações das médias mensais de radiação solar e concentração de ozônio.

Outra consideração tomada nesse trabalho foi a adoção de valores médios representativos do período da manhã (8:00 - 12:00) e da tarde (13:00 - 17:00) para caracterizar o perfil diurno. Testes de ajustes de rede neural foram feitos para diferentes intervalos de tempo (seção 4.8), observando-se que a média ao longo das cinco horas (manhã e tarde) apresenta melhores resultados para estimativa de ozônio. Os intervalos que apresentaram falhas de medição foram descartados, ou seja, todas as médias se referem a períodos contínuos de dados. Algumas características dessas médias são apresentadas na tabela 3.1.2.

Tabela 3.1.2: Características dos dados, da estação de qualidade do ar do Ibirapuera, utilizados nos treinamentos de redes neurais artificiais.

	Manhã (08:00 - 12:00)				Tarde (13:00 - 17:00)			
	Média	Desvio	Mínimo	Máximo	Média	Desvio	Mínimo	Máximo
NO ($\mu\text{g.m}^{-3}$)	24,4	29,9	0,4	224,8	4,4	6,1	0,0	107,6
NO ₂ ($\mu\text{g.m}^{-3}$)	45,1	24,4	0,0	147,6	34,2	18,4	0,1	165,8
Temperatura (°C)	20,1	4,0	7,8	29,4	22,8	4,5	8,2	33,4
UR (%)	74,8	13,9	32,0	99,0	61,1	17,5	13,6	99,0
Rad. Solar (W.m^{-2})	402,8	187,2	0,0	827,0	366,6	176,5	0,0	926,1
u (m.s^{-1})	-1,2	2,1	-6,1	6,1	-0,3	2,5	-4,5	10,4
v (m.s^{-1})	-0,2	2,4	-7,9	7,4	1,3	3,2	-7,3	7,8
Cam.Mistura (m)	365,6	97,4	145,0	658,4	485,4	107,5	110,2	759,6
Pressão (hPa)	928,0	3,5	915,7	936,6	926,1	3,5	914,0	934,9
O ₃ ($\mu\text{g.m}^{-3}$)	43,3	23,6	0,5	180,0	78,9	39,5	1,4	274,7
O ₃ máx ($\mu\text{g.m}^{-3}$)	77,2	40,5	0,9	337,0	102,7	51,1	2,8	350,0

Para uma melhor descrição das relações entre as variáveis e os processos que atuam na fotoquímica do ozônio na RMSP, dentro do período de estudo considerado, os dados passaram por um estudo estatístico preliminar aos ajustes de redes neurais (tabelas 3.1.3 e 3.1.4).

O primeiro método utilizado foi a correlação linear de Pearson, dada pela razão entre a covariância existente entre duas variáveis e o produto de seus desvios-padrão. Os resultados obtidos, presentes na tabela 3.1.3, indicam que pela manhã a concentração média de ozônio está mais correlacionada com a radiação solar do que com as demais variáveis, porém à tarde a temperatura torna-se mais significativa. Em ambos os períodos o NO e a UR apresentam

correlação negativa com O₃, sendo que a umidade relativa é a mais importante nesse sentido.

A concentração máxima de ozônio apresenta o mesmo perfil de correlações, mas os índices de correlação são inferiores do que para a concentração média de ozônio.

Tabela 3.1.3: Correlações de Pearson entre as variáveis consideradas, medidas na estação de qualidade do ar do Parque do Ibirapuera. (a): correlações das médias das variáveis de manhã (08:00-12:00); (b): correlações das médias das variáveis à tarde (13:00-17:00); (c): correlações entre as médias da manhã e da tarde.

(a)

Manhã	NO	NO ₂	TEMP.	UR	RAD.	u	v	C.M.	Pressão	O ₃	O ₃ máx
NO	1,0										
NO ₂	0,7	1,0									
Temperatura	0,0	0,1	1,0								
UR	-0,1	-0,3	-0,4	1,0							
Rad. Solar	-0,1	0,1	0,6	-0,7	1,0						
u	0,2	0,1	0,3	0,0	0,0	1,0					
v	-0,3	-0,3	-0,5	0,3	-0,3	-0,4	1,0				
Cam. Mistura	-0,4	-0,4	0,3	-0,2	0,4	-0,1	0,1	1,0			
Pressão	0,0	0,1	-0,5	0,0	-0,2	-0,5	0,3	-0,2	1,0		
O ₃	-0,3	0,0	0,5	-0,5	0,7	-0,1	0,0	0,4	-0,1	1,0	
O ₃ máx	-0,1	0,2	0,6	-0,5	0,7	0,0	-0,2	0,3	-0,1	0,9	1,0

(b)

Tarde	NO	NO ₂	TEMP.	UR	RAD.	u	v	CM	Pressão	O ₃	O ₃ máx
NO	1,0										
NO ₂	0,4	1,0									
TEMP	-0,4	-0,2	1,0								
UR	0,4	0,0	-0,6	1,0							
RAD	-0,4	-0,4	0,7	-0,6	1,0						
u	-0,1	-0,1	0,5	-0,3	0,2	1,0					
v	0,0	-0,1	-0,5	0,4	-0,1	-0,7	1,0				
CM	-0,5	-0,4	0,4	-0,5	0,6	0,2	0,0	1,0			
Pressão	0,1	0,2	-0,5	0,1	-0,3	-0,5	0,3	-0,1	1,0		
O ₃	-0,5	0,0	0,7	-0,6	0,6	0,1	-0,2	0,4	-0,2	1,0	
O ₃ máx	-0,4	0,1	0,6	-0,5	0,5	0,0	-0,1	0,3	-0,2	0,9	1,0

(c)

		Tarde										
		NO	NO ₂	TEMP.	UR	RAD.	u	v	CM	Pressão	O ₃	O ₃ máx
Manhã	NO	0,1	0,3	0,1	-0,3	-0,1	0,3	-0,3	-0,1	0,0	0,0	0,0
	NO ₂	0,0	0,5	0,2	-0,4	-0,1	0,2	-0,3	-0,1	0,0	0,3	0,3
	TEMP.	-0,4	-0,2	0,8	-0,3	0,4	0,4	-0,3	0,3	-0,5	0,5	0,5
	UR	0,3	0,0	-0,3	0,7	-0,2	-0,1	0,2	-0,4	0,1	-0,4	-0,4
	RAD.	-0,4	-0,1	0,5	-0,5	0,5	0,1	-0,1	0,5	-0,2	0,6	0,6
	u	-0,1	-0,2	0,3	-0,1	0,1	0,7	-0,4	0,0	-0,5	0,0	-0,1
	v	0,1	-0,1	-0,6	0,5	-0,3	-0,6	0,8	-0,2	0,3	-0,3	-0,3
	CM	-0,2	-0,4	0,1	-0,1	0,4	-0,1	0,2	0,5	-0,2	0,2	0,1
	Pressão	0,1	0,2	-0,4	0,0	-0,2	-0,5	0,2	-0,1	1,0	-0,1	-0,1
	O ₃	-0,4	-0,1	0,3	-0,3	0,3	-0,1	0,2	0,4	-0,1	0,6	0,6
	O ₃ máx	-0,4	0,0	0,5	-0,4	0,3	0,0	0,1	0,4	-0,2	0,7	0,8

O segundo procedimento adotado foi a aplicação da análise de componentes principais (PCA), que tem como um dos propósitos a detecção de estruturas nas relações entre variáveis, através da interpretação dos componentes de acordo com os coeficientes atribuídos a cada uma das variáveis.

Cada componente principal é uma combinação linear dos pesos das variáveis observadas que não é correlacionada com as demais componentes. Elas são extraídas de forma que a primeira explique a maior parte da variância nos dados. Cada componente subsequente explica o máximo da variância total restante, não explicada pelas anteriores.

Pelos resultados da análise de componentes principais presentes na tabela 3.1.4, é possível identificar alguns processos que ocorrem em relação ao conjunto de observações. O primeiro fator responde por 33,7% de toda variância dos dados e está associado à fotoquímica do ozônio, em que radiação solar, temperatura e ozônio possuem o mesmo sinal e se configuram entre as variáveis de maior peso para o primeiro autovetor; o óxido de nitrogênio e a umidade relativa também apresentam coeficientes significativos, mas em sinal oposto. A segunda componente principal representa 17,1% da variância e está associada ao transporte pelo vento, tendo em vista que as componentes u e v possuem os maiores pesos, seguidas pelas concentrações médias de ozônio de manhã. Já a terceira componente (13,2% da variância) está relacionada à variação das concentrações de dióxido de nitrogênio ao longo do dia.

Tabela 3.1.4: Autovetores da matriz de correlação (seis primeiros PCAs), além de seus autovalores e proporções.

		Comp. 1	Comp. 2	Comp. 3	Comp. 4	Comp. 5	Comp. 6
Manhã (08:00-12:00)	NO	0,0	0,3	0,3	0,1	-0,1	-0,4
	NO ₂	-0,1	0,2	0,4	-0,2	-0,1	-0,3
	Temperatura	-0,3	0,1	-0,1	-0,2	0,0	0,0
	UR	0,2	0,1	-0,2	0,0	-0,5	0,1
	Rad. Solar	-0,3	-0,2	0,0	-0,1	0,3	0,1
	u	-0,1	0,3	-0,2	-0,1	0,1	-0,3
	v	0,2	-0,3	0,0	-0,1	-0,2	-0,3
	Cam.Mistura	-0,1	-0,3	-0,3	0,1	0,2	0,1
	Pressão	0,1	-0,2	0,4	0,3	0,1	0,1
	O ₃	-0,2	-0,3	0,1	-0,3	0,2	-0,1
	O ₃ máx	-0,3	-0,2	0,1	-0,3	0,0	-0,1
Tarde (13:00-17:00)	NO	0,2	0,1	0,1	-0,2	0,1	0,5
	NO ₂	0,1	0,1	0,4	-0,3	0,1	0,3
	Temperatura	-0,3	0,1	0,0	0,1	-0,2	0,1
	UR	0,3	-0,1	-0,2	-0,3	-0,1	-0,1
	Rad. Solar	-0,2	-0,1	-0,1	0,3	-0,3	0,2
	u	-0,2	0,4	-0,1	0,1	0,1	-0,1
	v	0,1	-0,4	0,0	-0,2	-0,2	-0,3
	Cam. Mistura	-0,2	-0,2	-0,1	0,4	0,1	-0,2
	Pressão	0,2	-0,2	0,3	0,3	0,1	0,0
	O ₃	-0,3	-0,1	0,2	0,0	-0,4	0,1
	O ₃ máx	-0,3	-0,1	0,2	-0,2	-0,3	0,1
Autovalor	7,4	3,8	2,9	1,7	1,2	1,0	
Variância total (%)	33,7	17,1	13,2	7,6	5,3	4,4	
percentual acumulativo	33,7	50,8	64,0	71,6	76,9	81,3	

Outra abordagem considerada para se compreender a relação entre as variáveis foi a análise de agrupamentos (*cluster*), pela técnica hierárquica de Ward, baseada na minimização do incremento da distância quadrática total, *ESS*, a cada agregação de um objeto a um dos grupos. Para cada grupo, a distância quadrática é calculada pela equação 11. A distância total é calculada pela soma das distâncias dos grupos. O resultado é representado pelo dendograma da figura 8. Neste caso foram adotadas médias diárias (08:00-17:00) e a concentração máxima de ozônio no mesmo intervalo; a distinção entre manhã e tarde não foi feita pois, para cada variável, tendeu-se a formar grupos muito próximos entre esses dois períodos.

$$ESS_k = \sum_{j=1}^{N_k} (X_j - \bar{X})^T (X_j - \bar{X}) \quad (11)$$

Em que N_k é o número de componentes do grupo k , X_j é um vetor de observações (dados multivariados) contido em k e \bar{X} é o centróide desse grupo. Assim, o total da soma dos quadrados dos desvios dos grupos é (12):

$$ESS = \sum_{j=1}^k ESS_j \quad (12)$$

O processo de agrupamento inicia-se com n grupos (igual ao número de observações). A cada passo do processo todos os pares de grupos (i, j) são considerados, e é selecionado para compor o novo grupo aquele que representar o menor incremento em ESS . Ou seja, por esse método, a matriz de similaridade é composta pelos valores de ESS correspondentes a cada par (i, j) . Em geral, esse método é considerado bastante eficiente, no entanto, tende a criar grupos (*clusters*) pequenos.

Como medida de distância entre dados, foi empregada a distância euclidiana, que é a distância geométrica em um espaço multidimensional (equação 13). Essa é, provavelmente, a medida de distância entre dados mais comumente utilizada.

$$D_{ij} = \left[\sum_{k=1}^p (X_{ik} - X_{jk})^2 \right]^{1/2} \quad (13)$$

Existem outras regras de ligação e medidas de similaridade entre dados, mas essa configuração se mostrou mais apropriada. Observou-se que, com a aplicação de outros métodos, obtêm-se conclusões similares.

Pela figura 8 há uma proximidade entre NO e temperatura, algo que pode ser observado nas correlações (tabela 3.1.3: b e c) e no primeiro PCA (tabela 3.1.4), para as médias da tarde. O ozônio se relaciona com a umidade relativa, provavelmente devido à importância desta última como processo de remoção do poluente da atmosfera. Por fim, a radiação solar, altura da camada de mistura e pressão atmosférica formam grupos distantes

das demais variáveis; posteriormente, pelos ajustes de redes neurais artificiais (item 4.7) foi verificado que elas influem de forma marginal na qualidade das estimativas de ozônio.

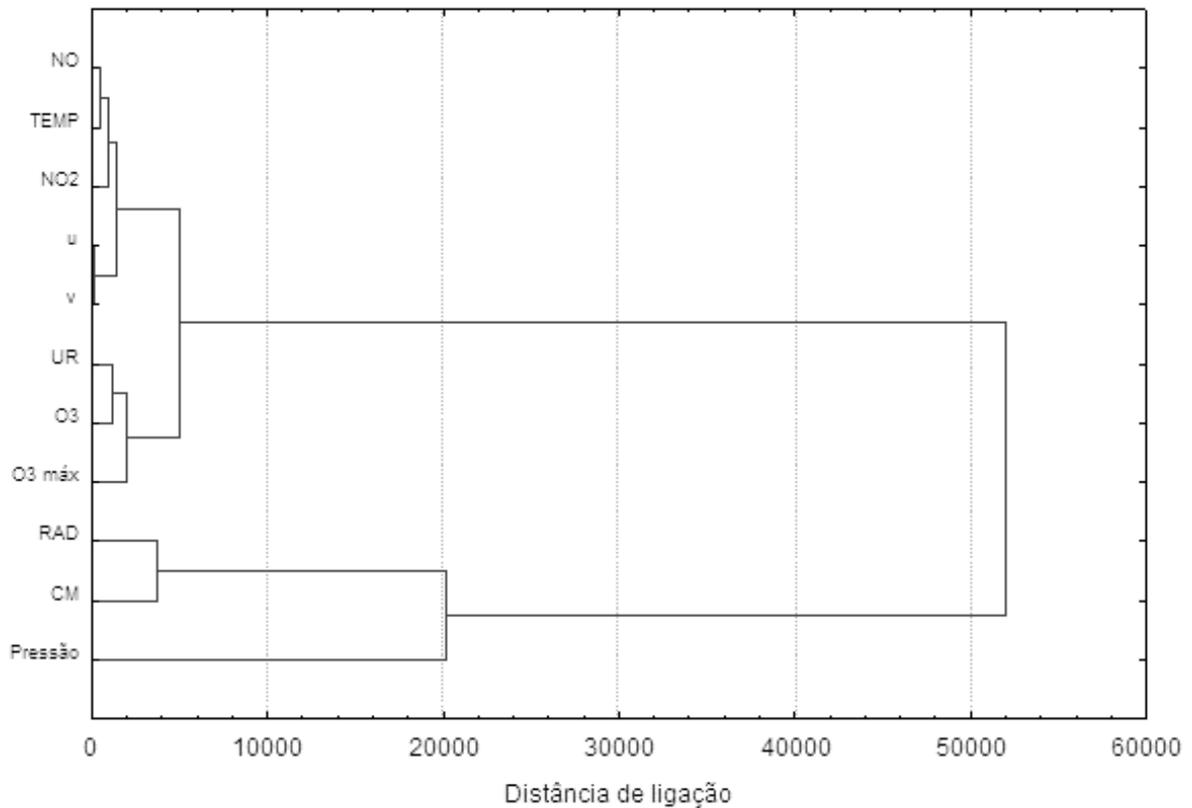


Figura 8: Dendrograma obtido da análise de agrupamento de dados, considerando o método de Ward e a medida de distância euclidiana.

3.2 - Caracterização das variáveis em função das concentrações de ozônio

3.2.1 - Óxidos de nitrogênio (NO e NO₂)

As distribuições das médias de concentração de NO e NO₂ são apresentadas nos histogramas das figuras 9 e 10 (manhã e tarde, respectivamente). Observa-se que durante a manhã as concentrações de NO apresentam seus índices mais elevados, mas estão sempre abaixo das de NO₂. No final da manhã e início da tarde, os processos fotoquímicos reduzem ainda mais suas concentrações, com a produção de ozônio.

Nas figuras 11 e 12 é feita uma comparação entre as concentrações médias dos óxidos de nitrogênio com as concentrações médias e máximas de ozônio, respectivamente, para o período da tarde. Não é possível definir um padrão claro de comportamento, o que se observa é uma grande dispersão dos pontos, com as concentrações mais elevadas de ozônio (acima de $160 \mu\text{g.m}^{-3}$) ocorrendo durante baixos níveis de NO, tanto para valores médios e máximos de ozônio.

Na figura 13 é apresentada uma relação entre NO, NO₂ e O₃. Nota-se que as concentrações mais elevadas de ozônio ocorrem, principalmente, para NO₂ próximo da faixa de 55 a 75 $\mu\text{g.m}^{-3}$, com NO abaixo de 5 $\mu\text{g.m}^{-3}$. Para este gráfico foi feita uma interpolação dos dados utilizando a técnica de Krigagem, a qual constitui uma das técnicas mais eficazes de interpolação. É um método geoestatístico que leva em consideração as características espaciais de autocorrelação de variáveis regionalizadas (LANDIM, 2000).

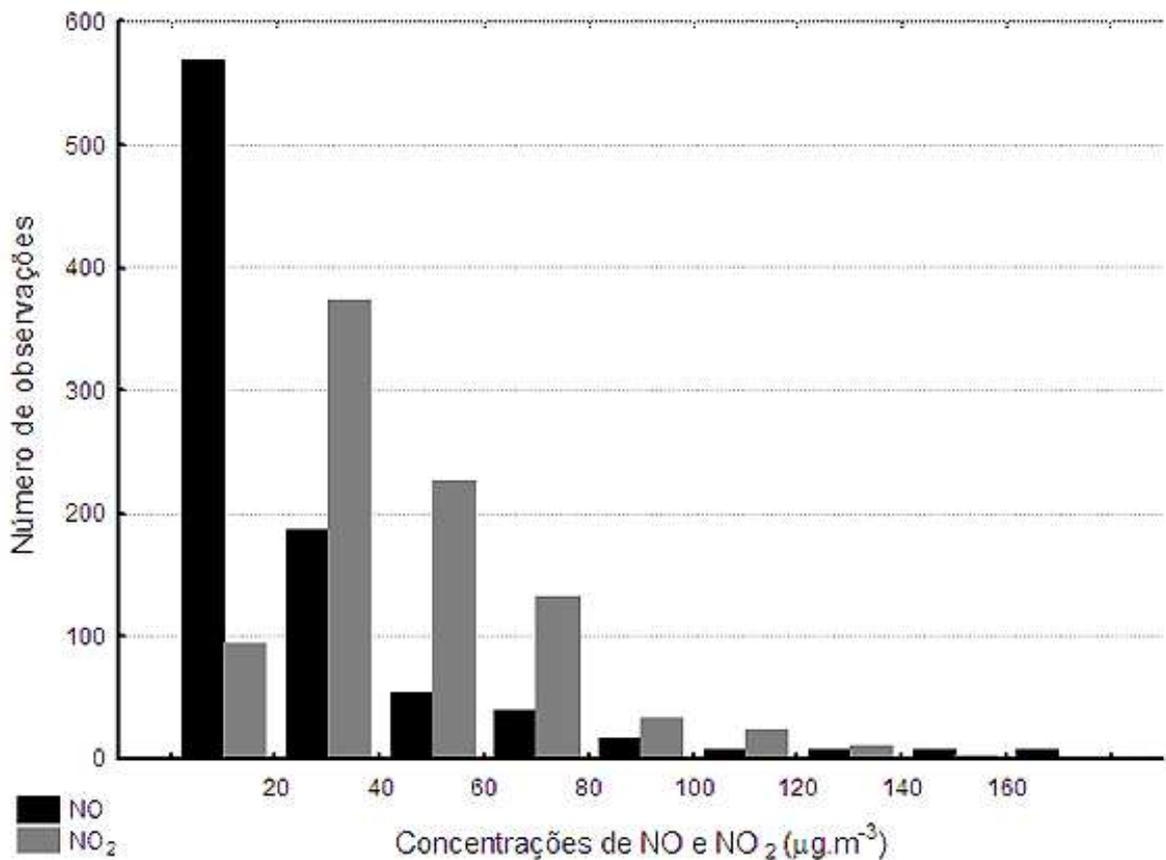


Figura 9: Histograma de distribuição das concentrações médias de NO e NO₂ no período da manhã.

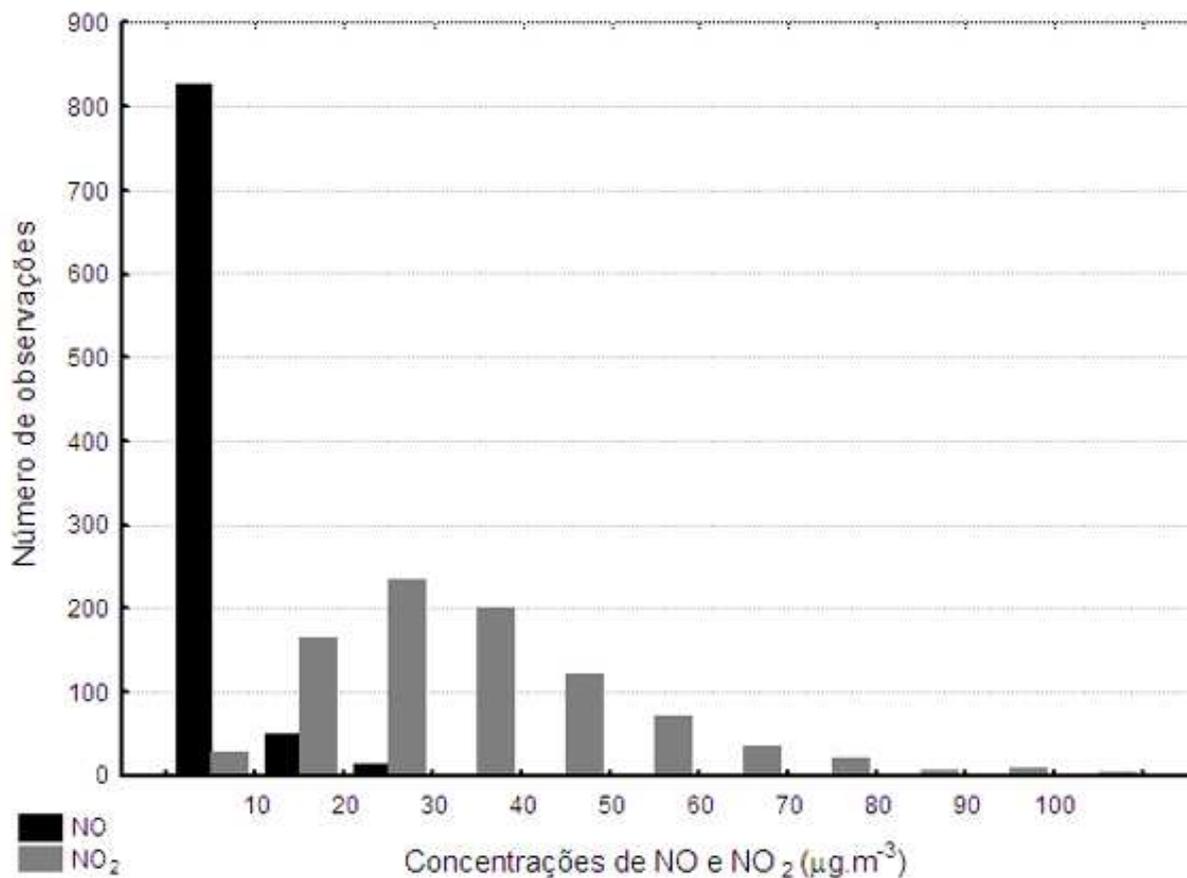


Figura 10: Histograma de distribuição das concentrações médias de NO e NO₂ no período da tarde.

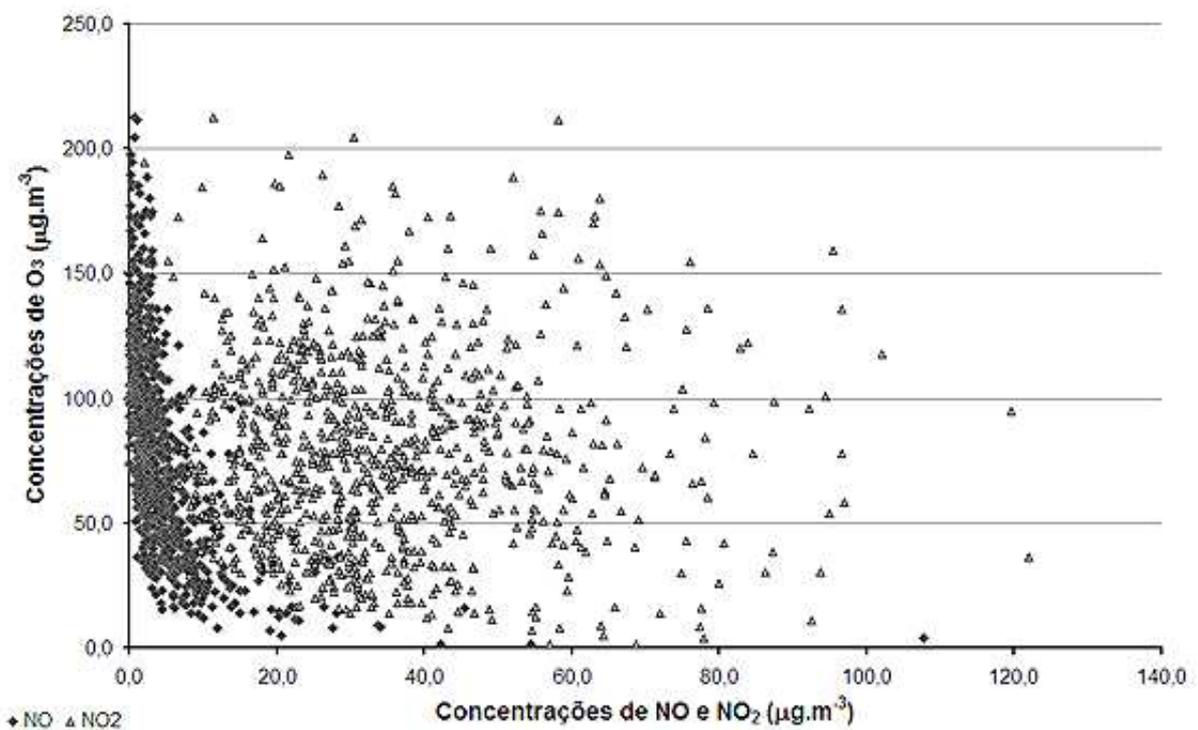


Figura 11: Variação das concentrações médias de NO e NO₂ em função da concentração média de ozônio para o período da tarde.

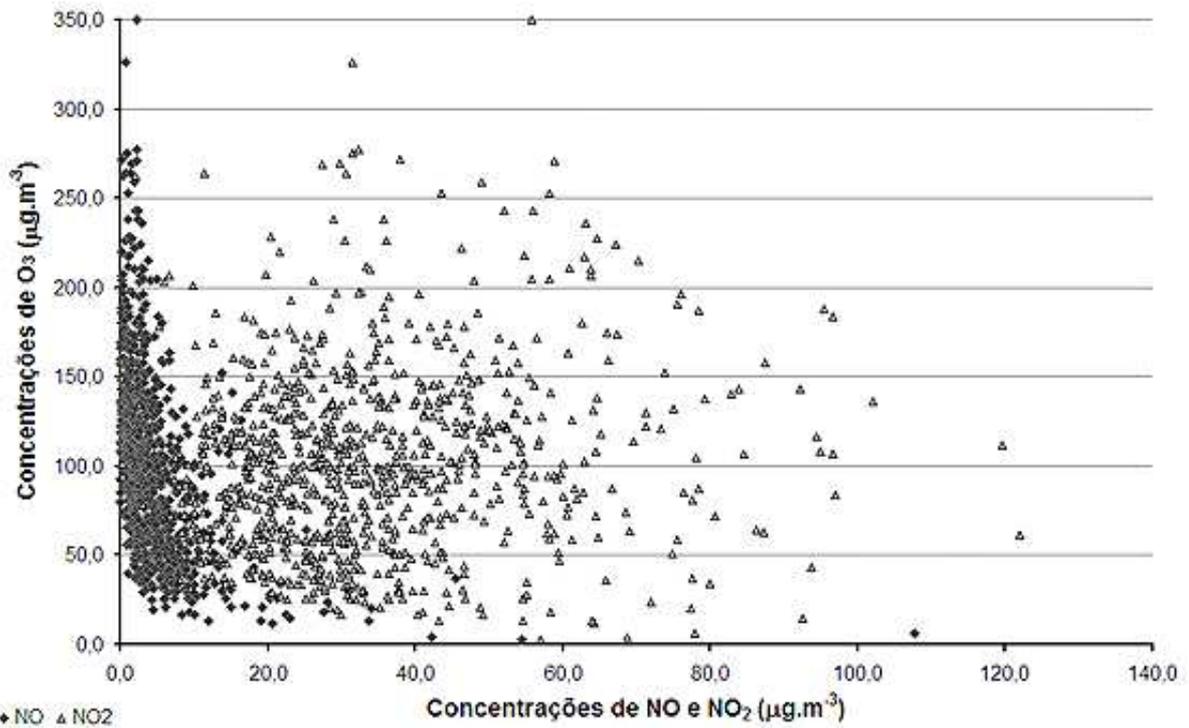


Figura 12: Variação das concentrações médias de NO e NO₂ em função da concentração máxima de ozônio para o período da tarde.

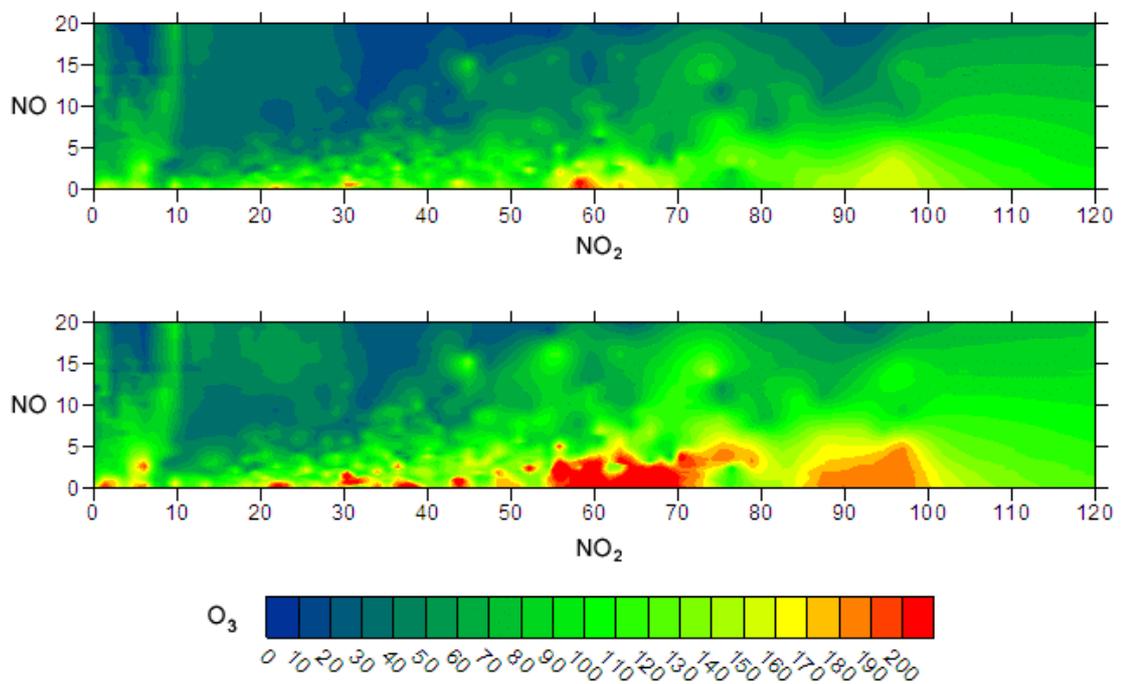


Figura 13: Concentrações médias de ozônio (em cima) e máximas (em baixo) com relação às concentrações de NO₂ e NO durante o período da tarde - concentrações em µg.m⁻³.

3.2.2 - Temperatura do ar

O histograma da figura 14 mostra as distribuições de temperatura média para os períodos da manhã e da tarde, com as respectivas curvas normais esperadas. Nas figuras 15 e 16 estão as comparações feitas para o período da tarde entre temperatura e as concentrações médias e máximas de ozônio, respectivamente. Observa-se que para cada concentração do poluente há uma faixa preferencial de temperatura, que vai se mantendo proporcional ao aumento das concentrações.

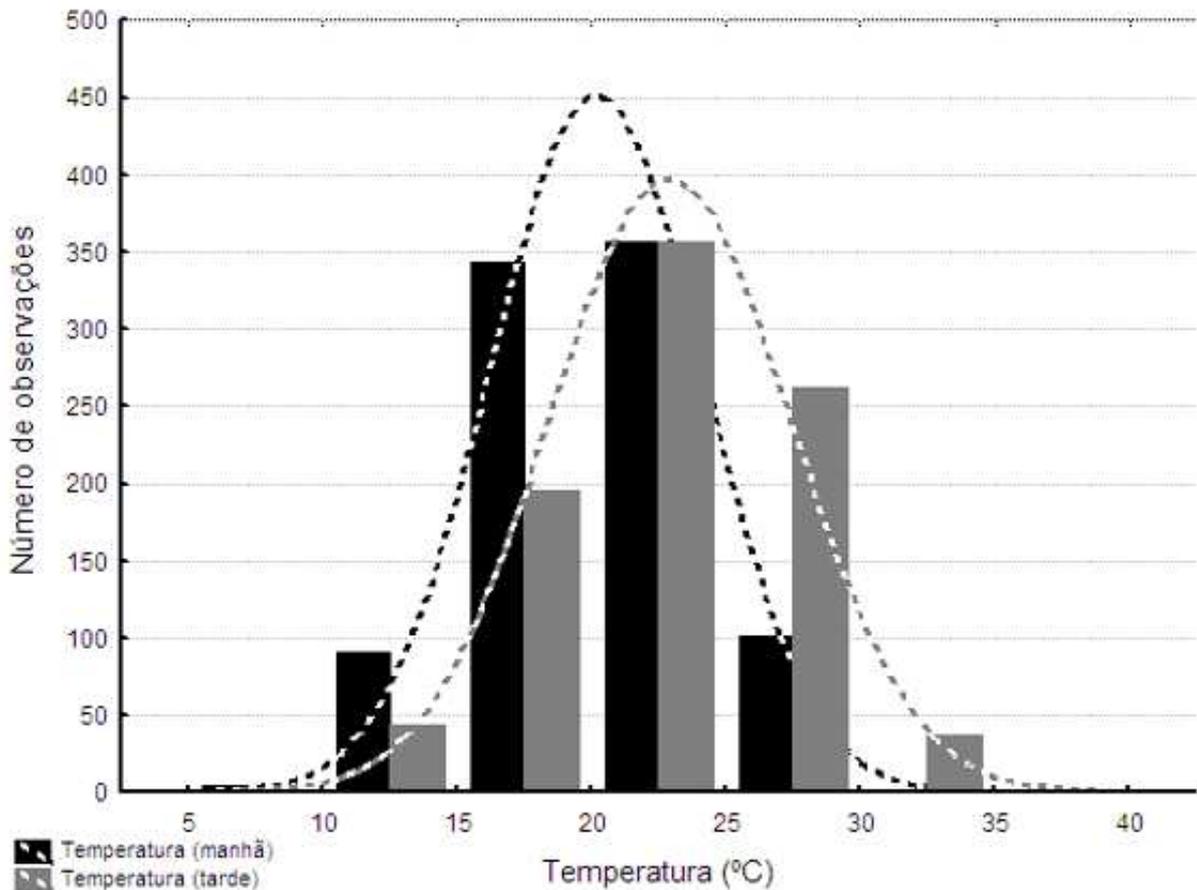


Figura 14: Histograma de distribuição das médias de temperatura do ar nos períodos da manhã (rosa) e da tarde (vermelho).

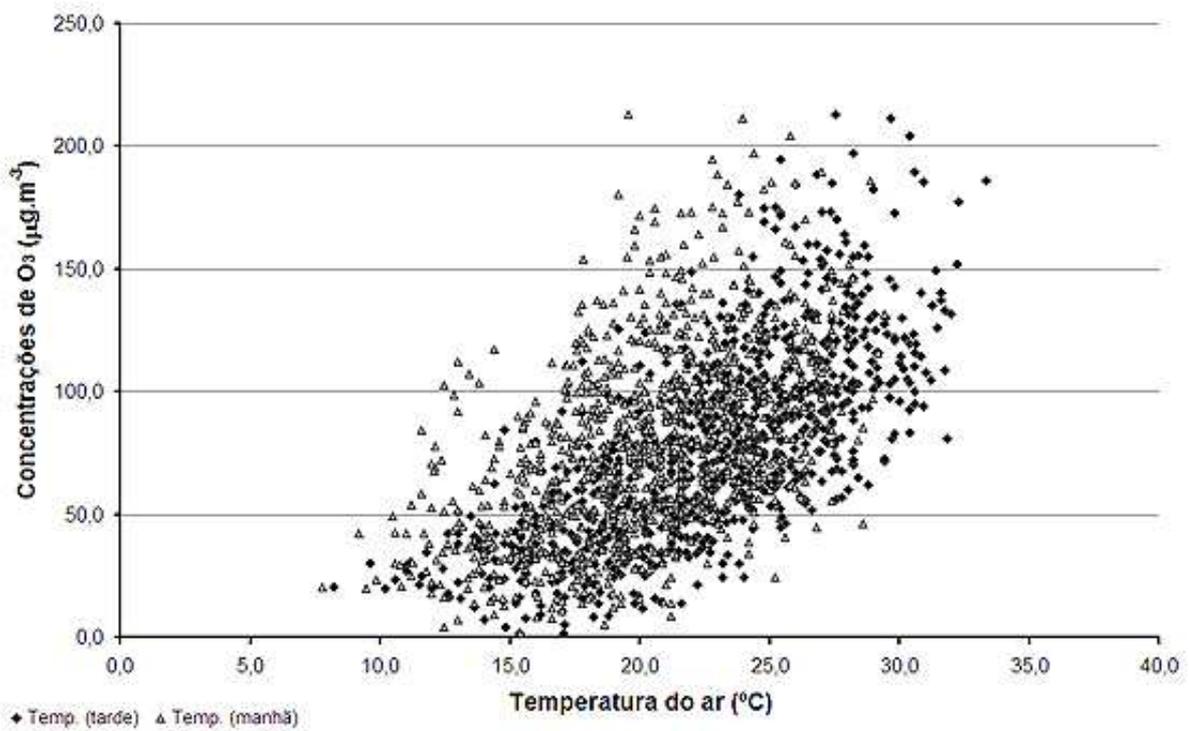


Figura 15: Variação das médias de temperatura em função da concentração média de ozônio para o período da tarde.

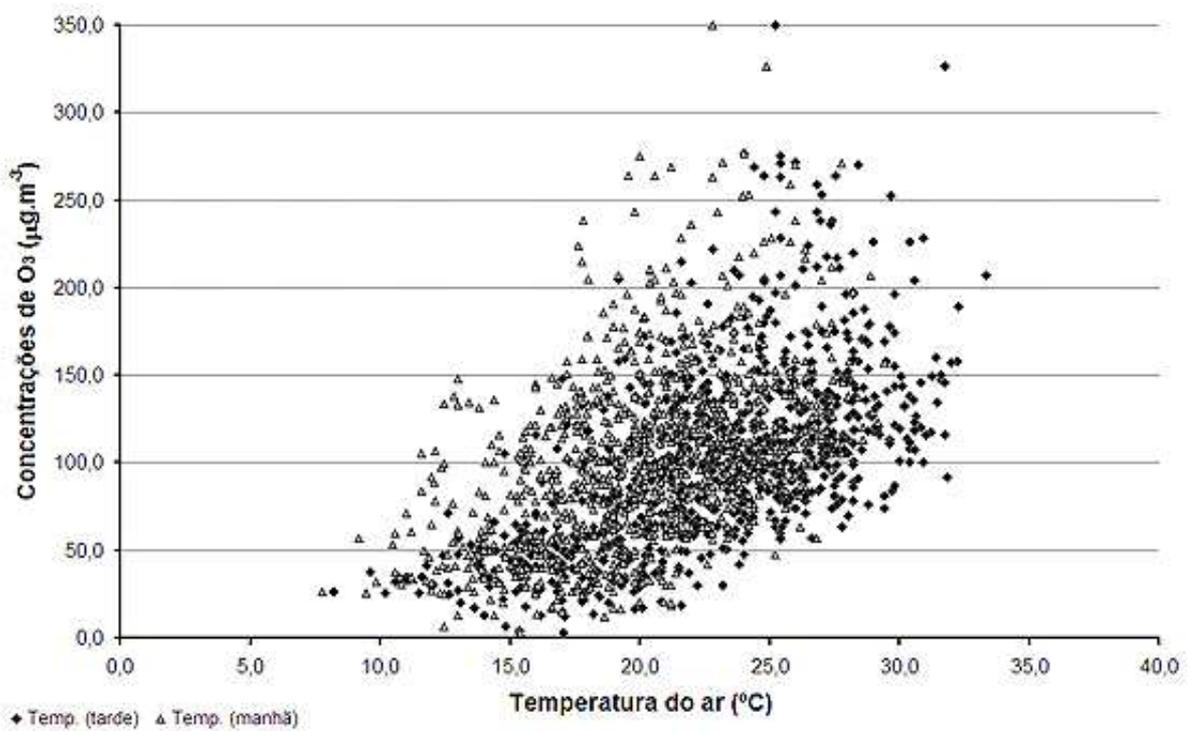


Figura 16: Variação das médias de temperatura em função da concentração máxima de ozônio para o período da tarde.

3.2.3 - Umidade relativa

Na figura 17 são apresentados os histogramas de frequência para as médias de umidade relativa de manhã e de tarde, com suas subseqüentes curvas normais esperadas. Nas figuras 18 e 19 a umidade relativa média da tarde é relacionada, respectivamente, com as concentrações médias e máximas de ozônio do mesmo período. Em geral as concentrações mais elevadas de ozônio estão associadas com uma baixa umidade relativa do ar, entretanto, diversos casos não se figuram nessa tendência.

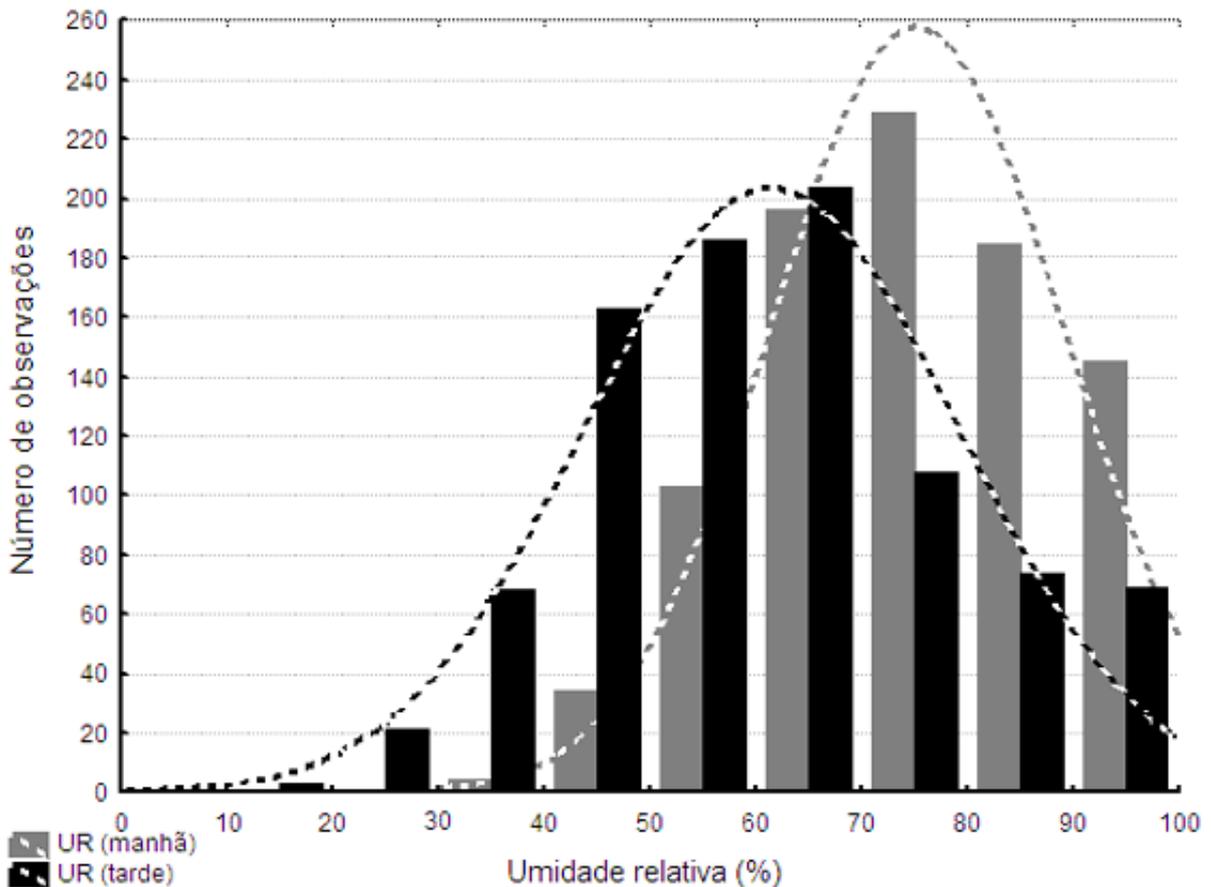


Figura 17: Histograma de distribuição das médias de umidade relativa nos períodos da manhã (azul claro) e da tarde (azul escuro).

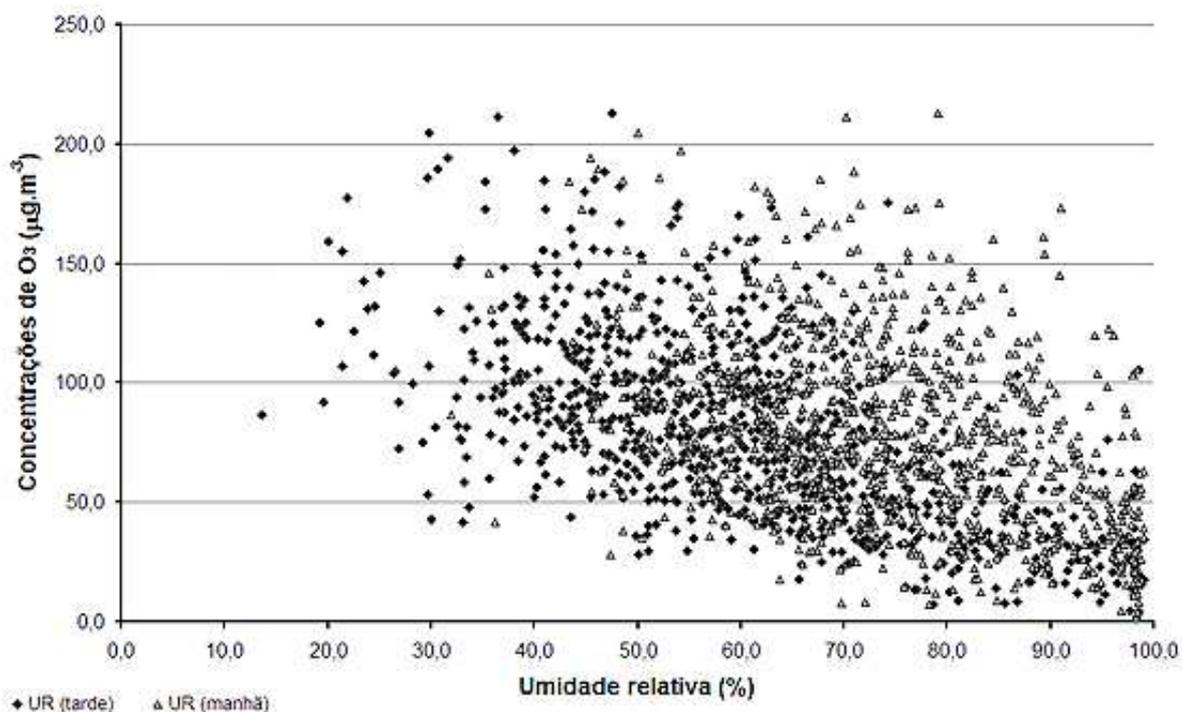


Figura 18: Variação das médias de umidade relativa em função da concentração média de ozônio para o período da tarde.

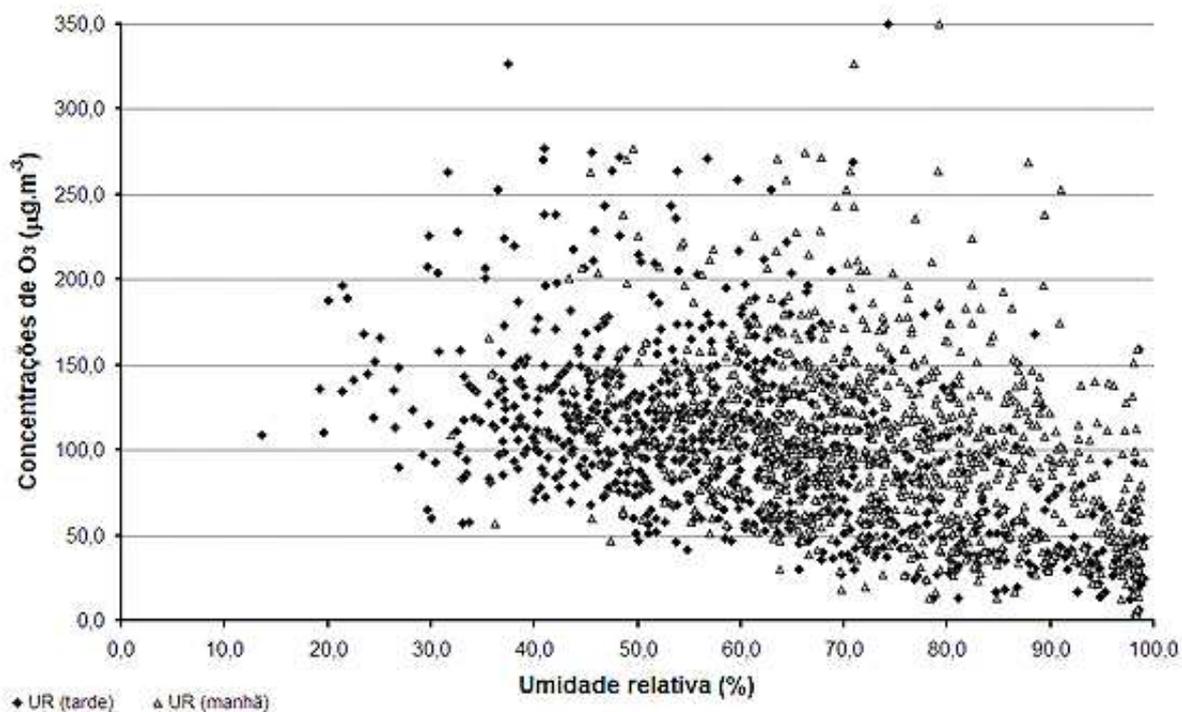


Figura 19: Variação das médias de umidade relativa em função da concentração máxima de ozônio para o período da tarde.

3.2.4 - Campo de vento horizontal

As distribuições das médias das componentes zonal (u) e meridional (v) do vento são apresentadas nas figuras 20 e 21, para manhã e tarde, respectivamente. Observa-se que há uma maior frequência de vento zonal na faixa de -2 a -3 m.s⁻¹, e vento meridional entre -1 e -2 m.s⁻¹, para o período da manhã; à tarde o vento zonal também contém maiores ocorrências dentro dessa faixa, e o vento meridional para valores positivos, entre 3 a 5 m.s⁻¹. Esse perfil é característico da ocorrência da entrada na RMSP da brisa marítima padrão (OLIVEIRA e SILVA DIAS, 1982), em que o vento passa de NE, no período da manhã, para SE à tarde.

As figuras 22 e 23 mostram a disposição das concentrações médias e máximas de ozônio, respectivamente, em relação às médias das componentes zonal e meridional do vento; os dados são representativos do período da tarde. O maior número de eventos representativos ocorre na presença de vento de sudeste, justamente por esse ser o padrão de circulação mais freqüente no período, mas há também muitas ocorrências de concentrações próximas de 100 µg.m⁻³ devido a ventos do setor NO. Alguns eventos bastante representativos (notavelmente para O₃ máximo) ocorrem para ventos próximos a SO, indicando o efeito do transporte. Para as demais direções não há ocorrências, exceto quando o vento está mais calmo (baixas velocidades nas componentes zonal e meridional).

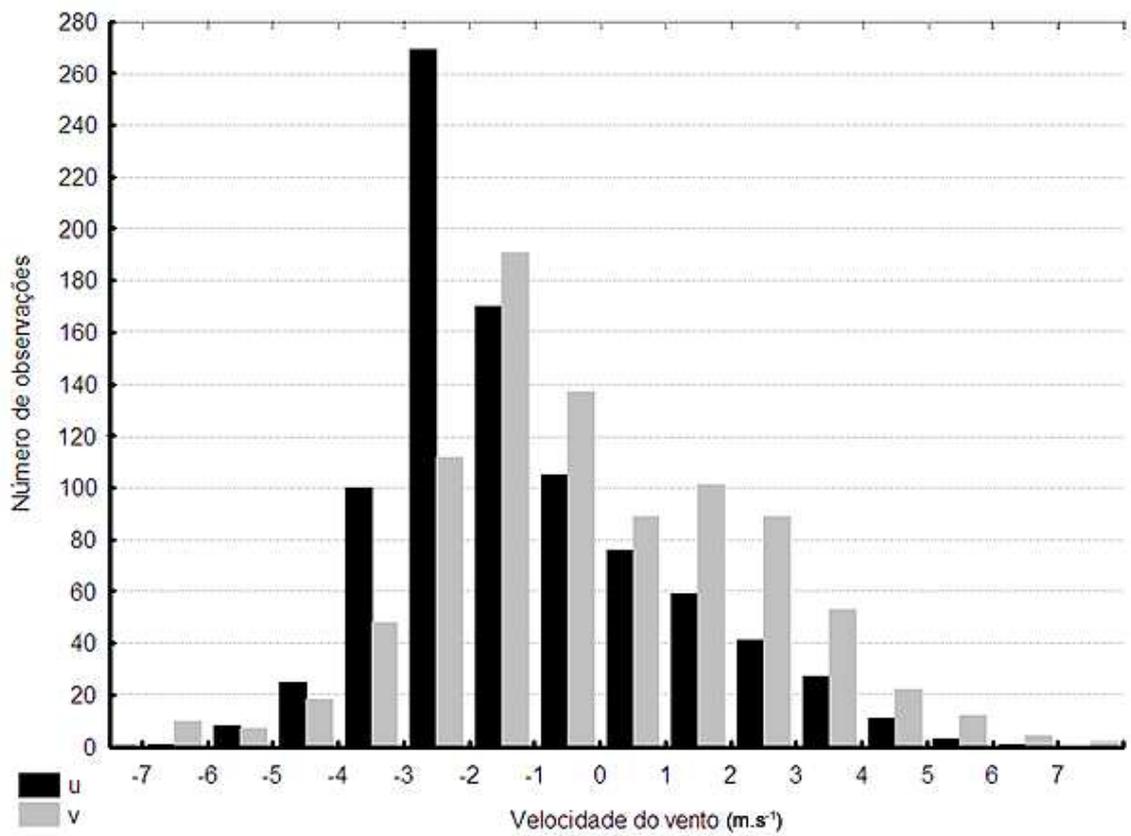


Figura 20: Histograma de distribuição das componentes da velocidade do vento (u,v) no período da manhã.

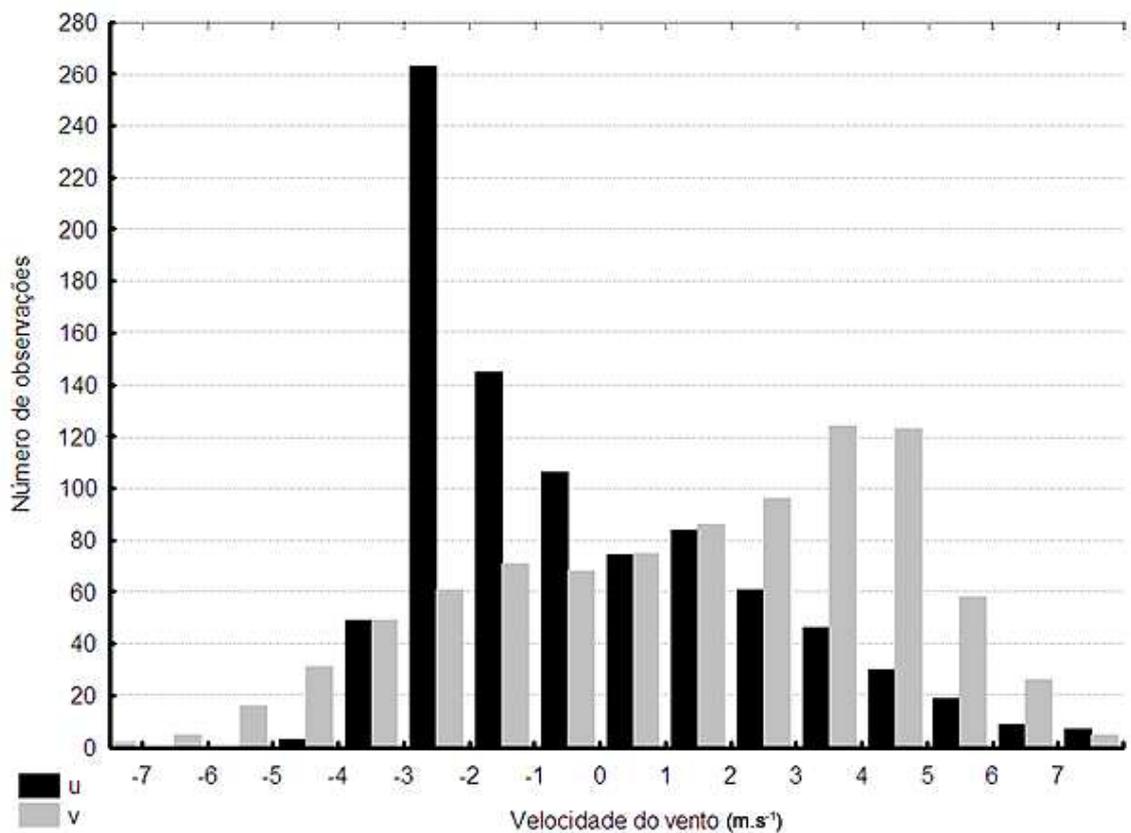


Figura 21: Histograma de distribuição das componentes da velocidade do vento (u,v) no período da tarde.

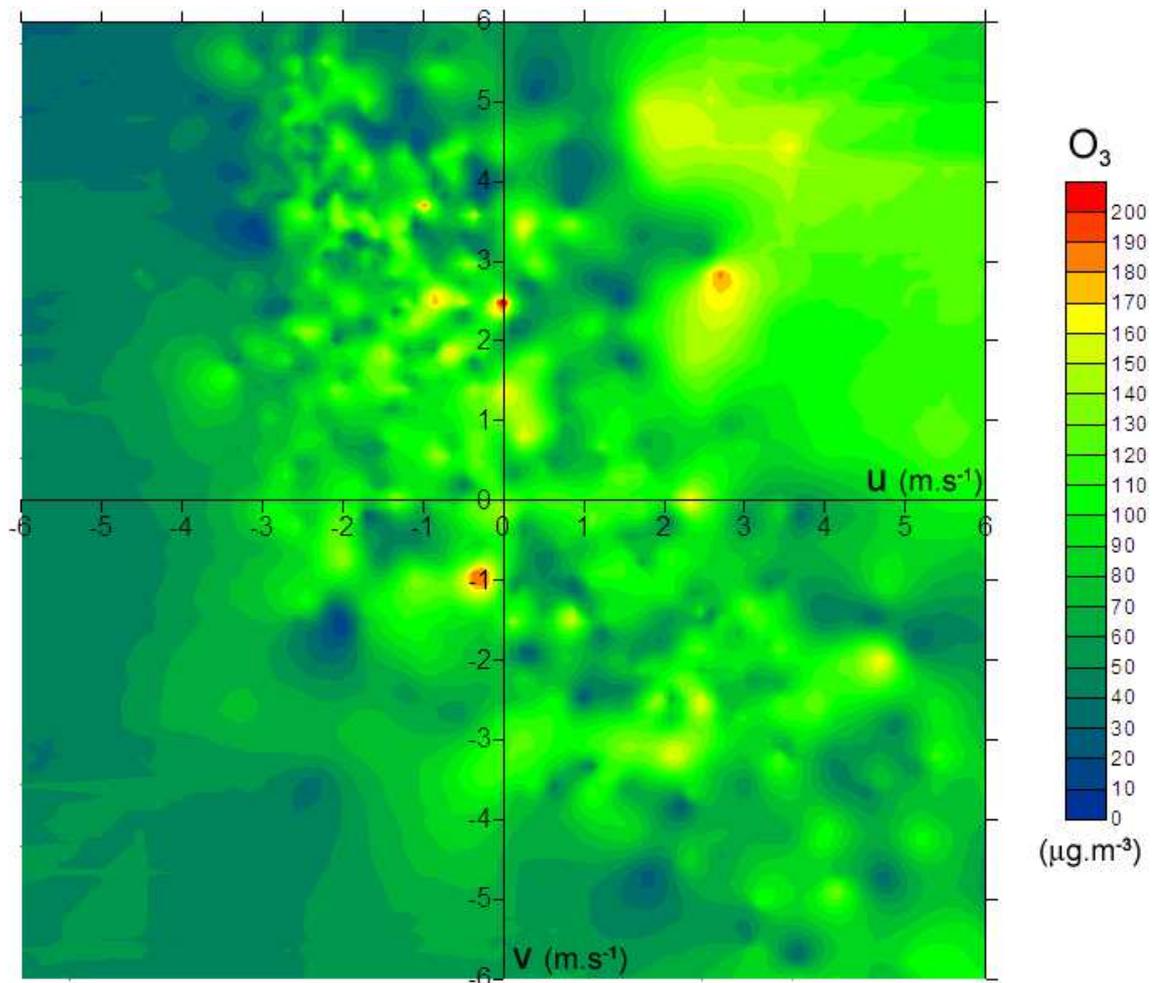


Figura 22: Concentrações médias de ozônio em relação às médias das componentes zonal (u) e meridional (v) do vento no período da tarde.

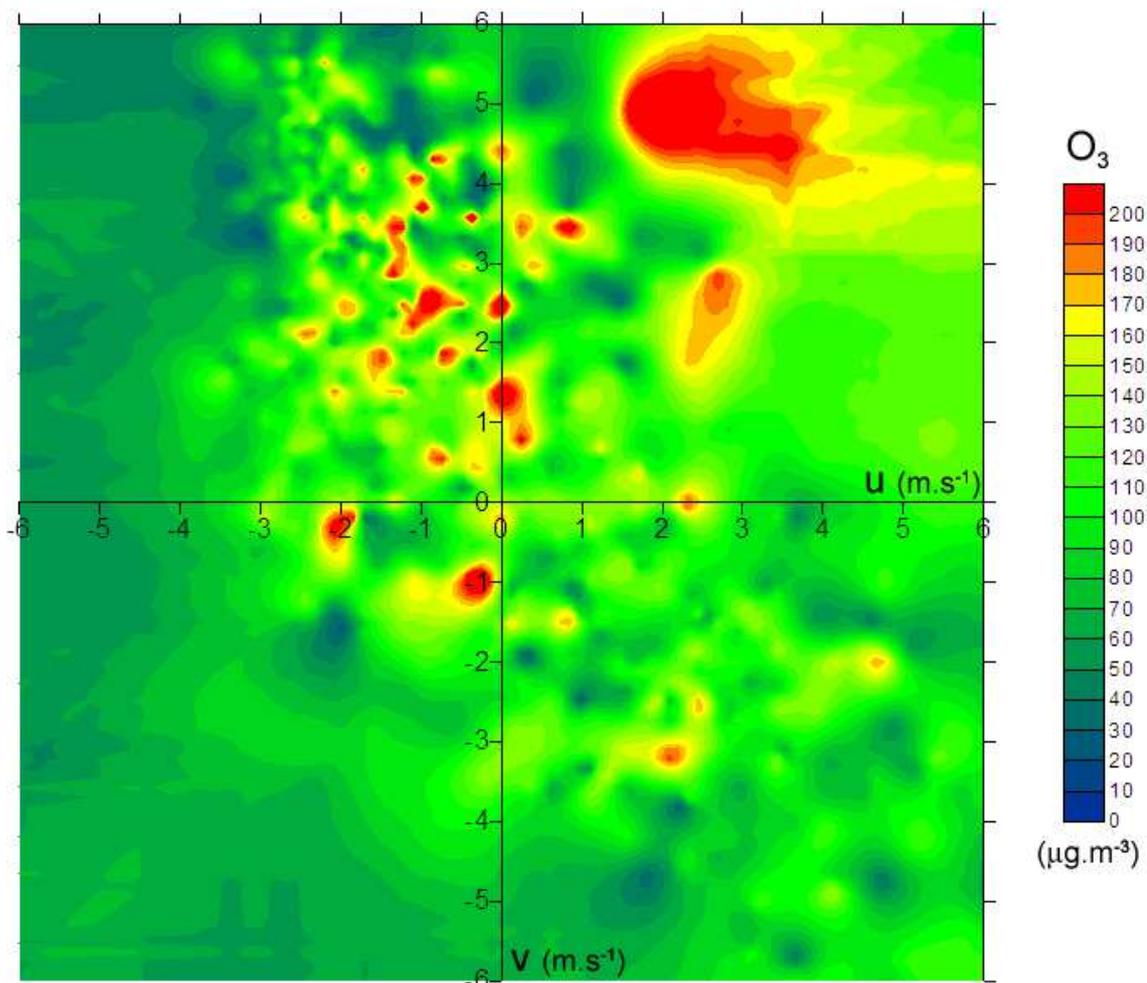


Figura 23: Concentrações máximas de ozônio em relação às médias das componentes zonal (u) e meridional (v) do vento no período da tarde.

3.2.5 - Radiação Solar

As distribuições dos dados médios de radiação solar, de manhã e tarde, são mostradas na figura 24. Nas comparações entre radiação solar e as concentrações médias e máximas de ozônio (figuras 25 e 26) observa-se que não há um padrão bem definido na disposição dos pontos. A característica principal é que a dispersão destes aumenta conforme as concentrações se elevam. A partir de $150 \mu\text{g.m}^{-3}$, para O_3 médio, e $200 \mu\text{g.m}^{-3}$, para O_3 máximo, essa dispersão se mantém na faixa de 200 W.m^{-2} a 800 W.m^{-2} .

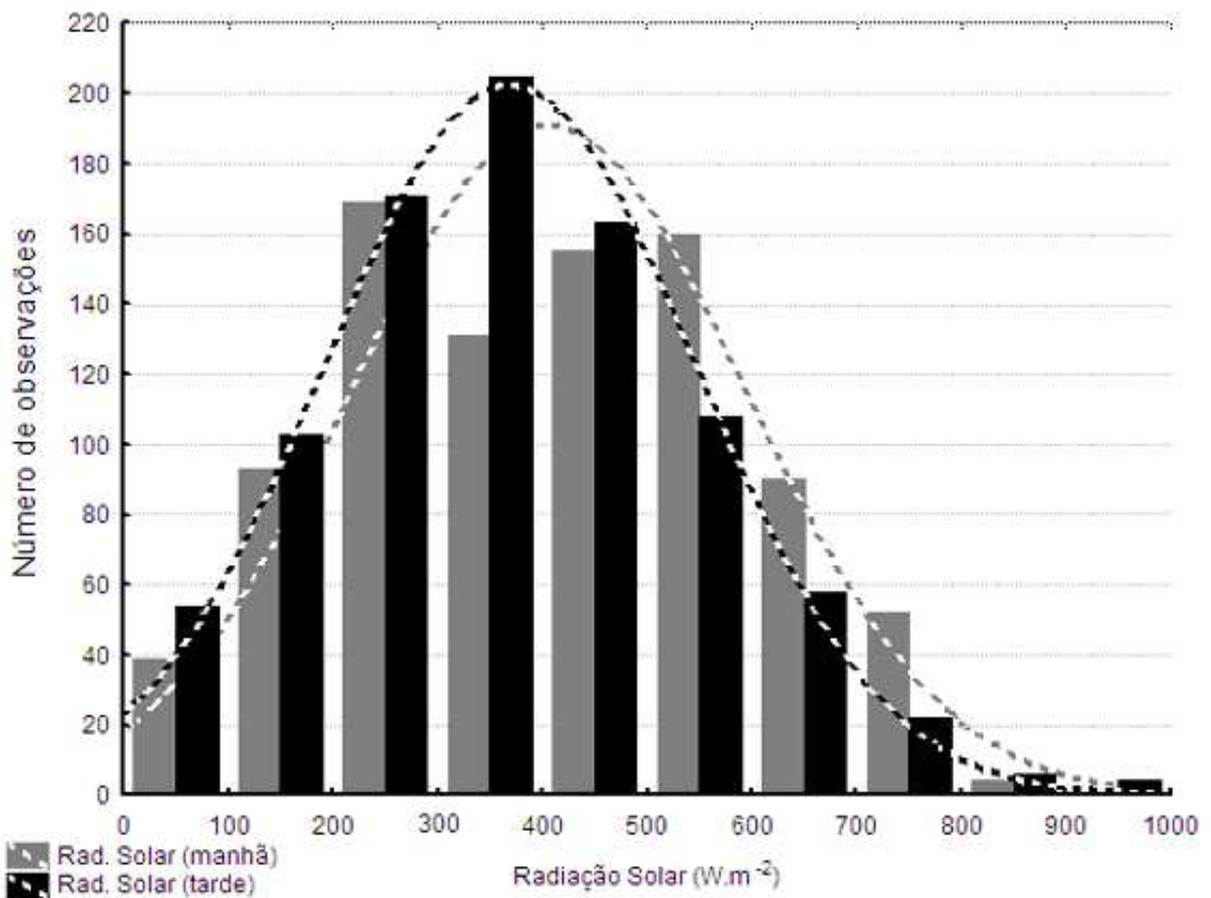


Figura 24: Histograma de distribuição das médias de radiação solar nos períodos da manhã (amarelo) e da tarde (vermelho).

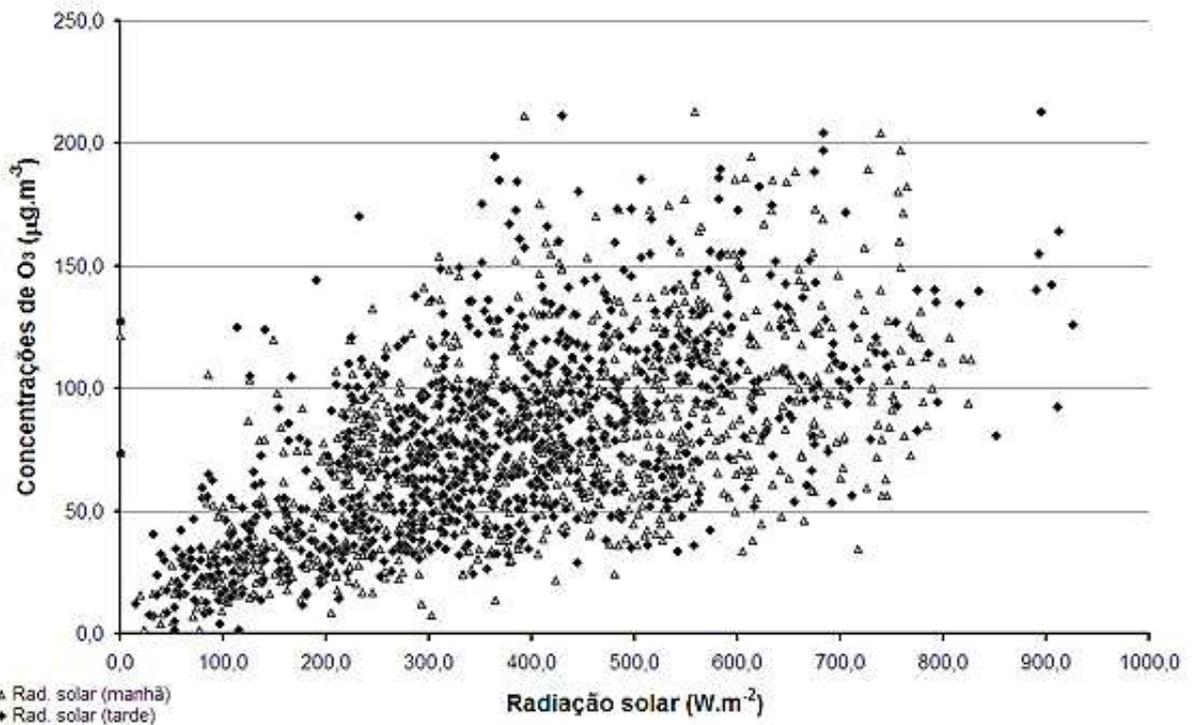


Figura 25: Variação das médias de radiação solar em função da concentração média de ozônio à tarde.

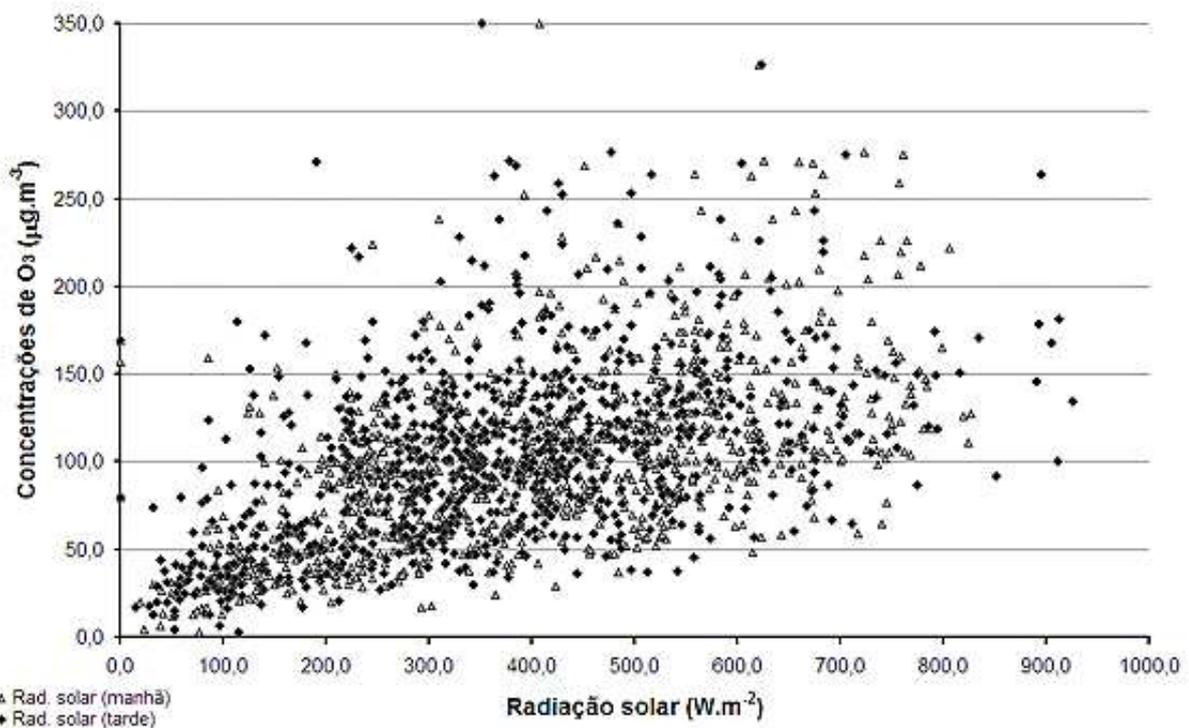


Figura 26: Variação das médias de radiação solar em função da concentração máxima de ozônio à tarde.

3.2.6 - Altura da camada de mistura

O histograma da figura 27 mostra as distribuições de altura da camada de mistura (médias de manhã e tarde) obtida com os dados do Sondador Acústico (SODAR). Nas figuras 28 e 29 as médias dessa variável são comparadas com as concentrações médias e máximas de ozônio à tarde, respectivamente. Por estas duas, não é possível identificar visualmente um padrão claro na distribuição dos pontos.

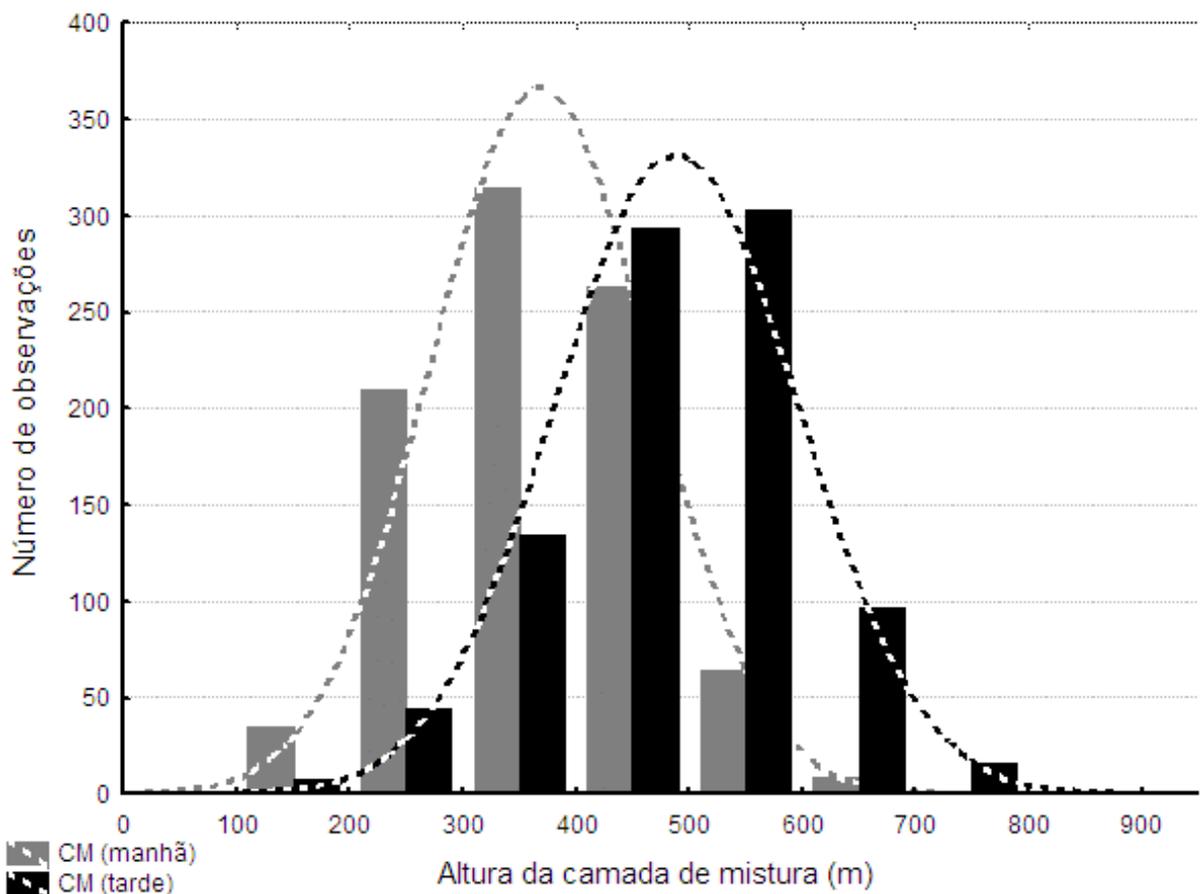


Figura 27: Histograma de distribuição das médias de altura da camada de mistura no período da manhã (verde claro) e da tarde (verde escuro).

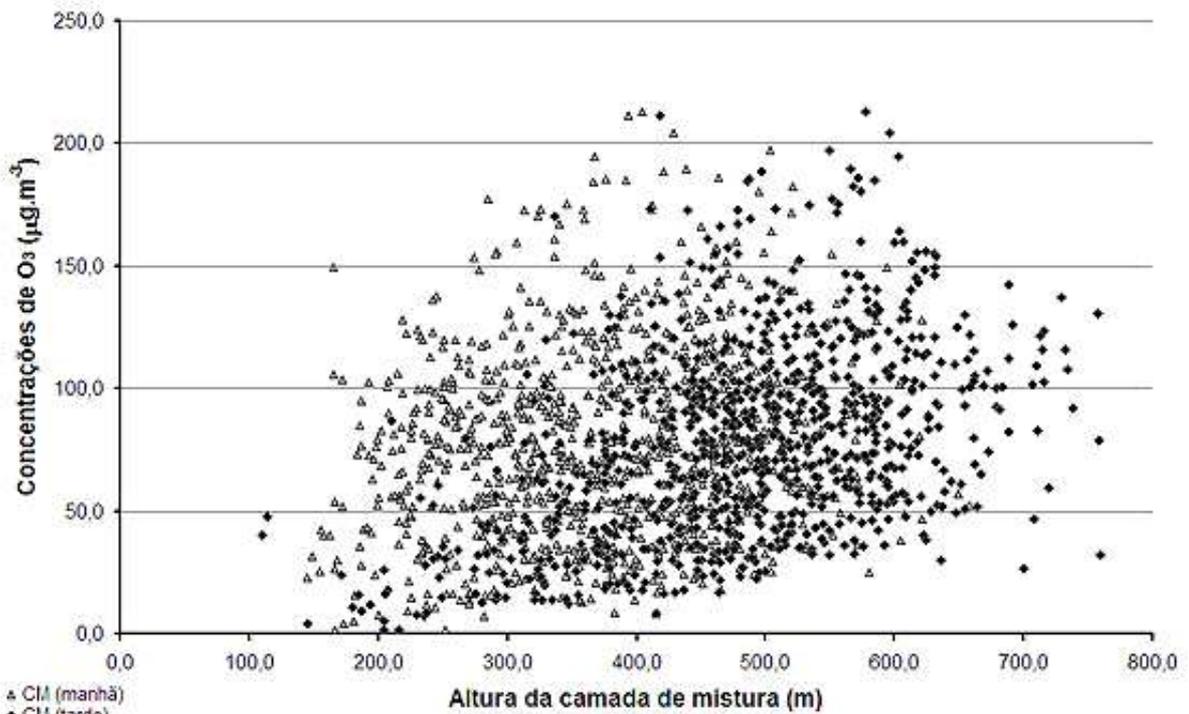


Figura 28: Variação das médias de altura da camada de mistura em função da concentração média de ozônio para o período da tarde.

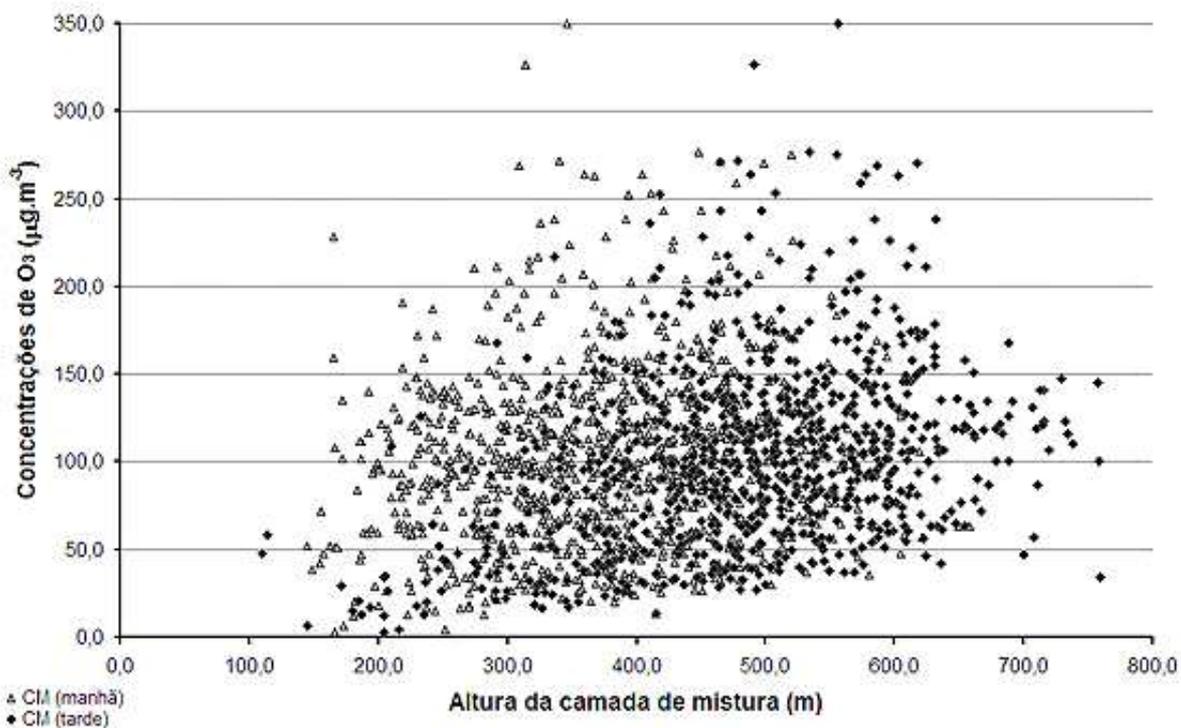


Figura 29: Variação das médias de altura da camada de mistura em função da concentração máxima de ozônio para o período da tarde.

3.2.7 - Pressão atmosférica

O histograma de frequência para pressão atmosférica é apresentado na figura 30. As correspondências entre essa variável e as concentrações médias e máximas de ozônio para o período da tarde estão presentes nas figuras 31 e 32. Por essas imagens não é possível identificar um padrão claro de comportamento do ozônio em função da pressão atmosférica..

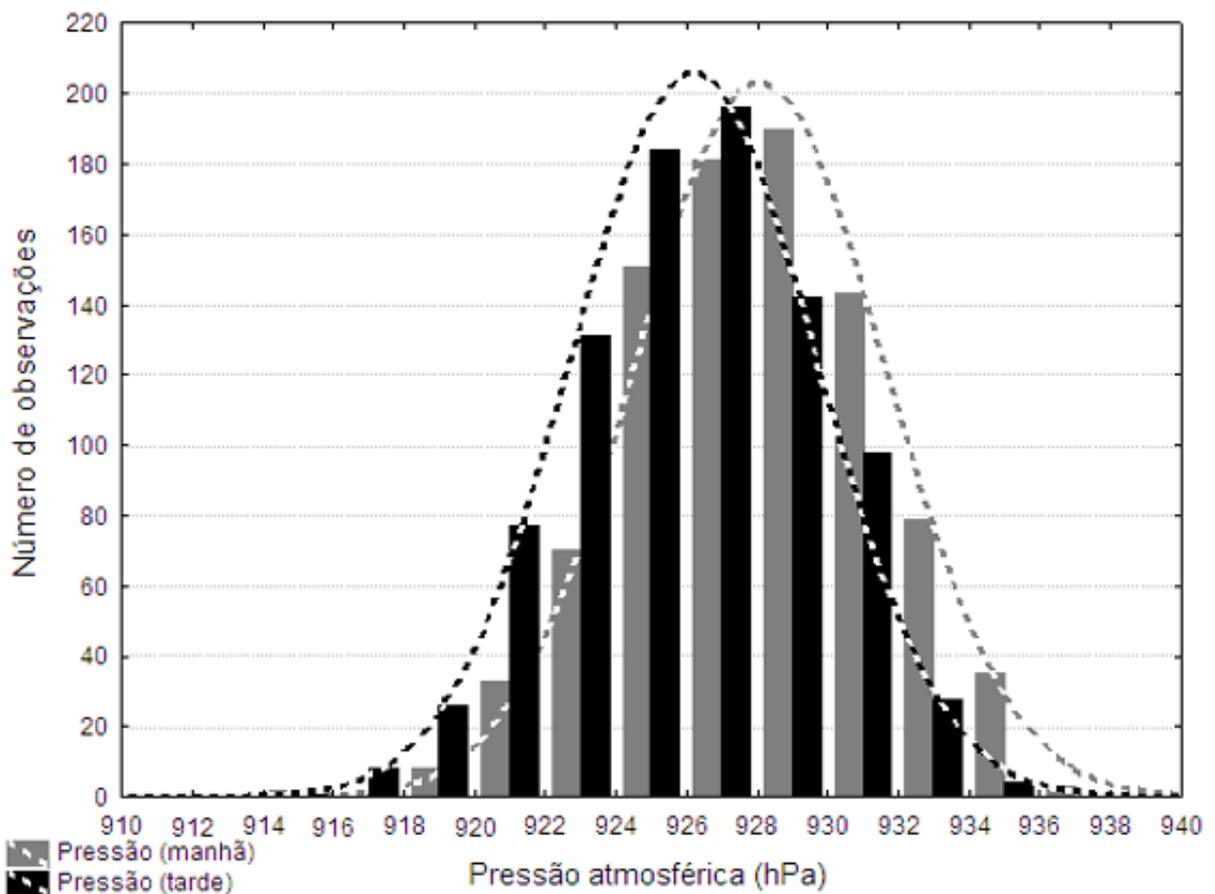


Figura 30: Histograma de distribuição das médias de pressão atmosférica no período da manhã (azul claro) e da tarde (azul escuro).

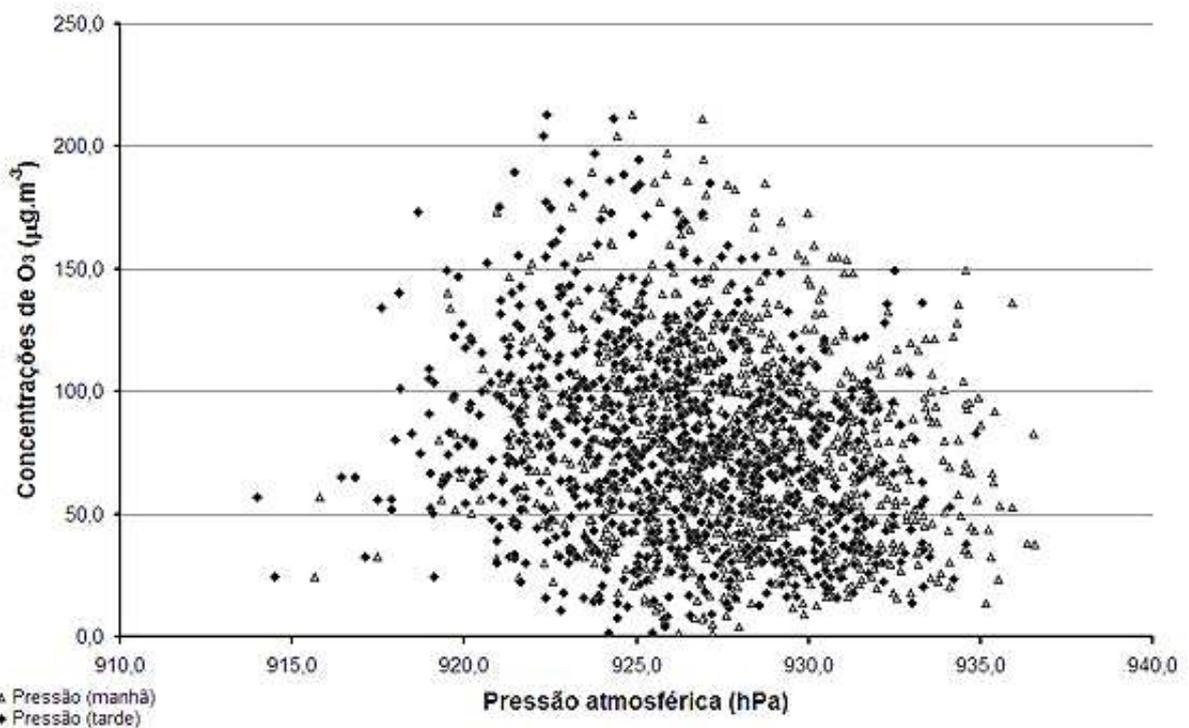


Figura 31: Variação das médias de pressão atmosférica em função da concentração média de ozônio para o período da tarde.

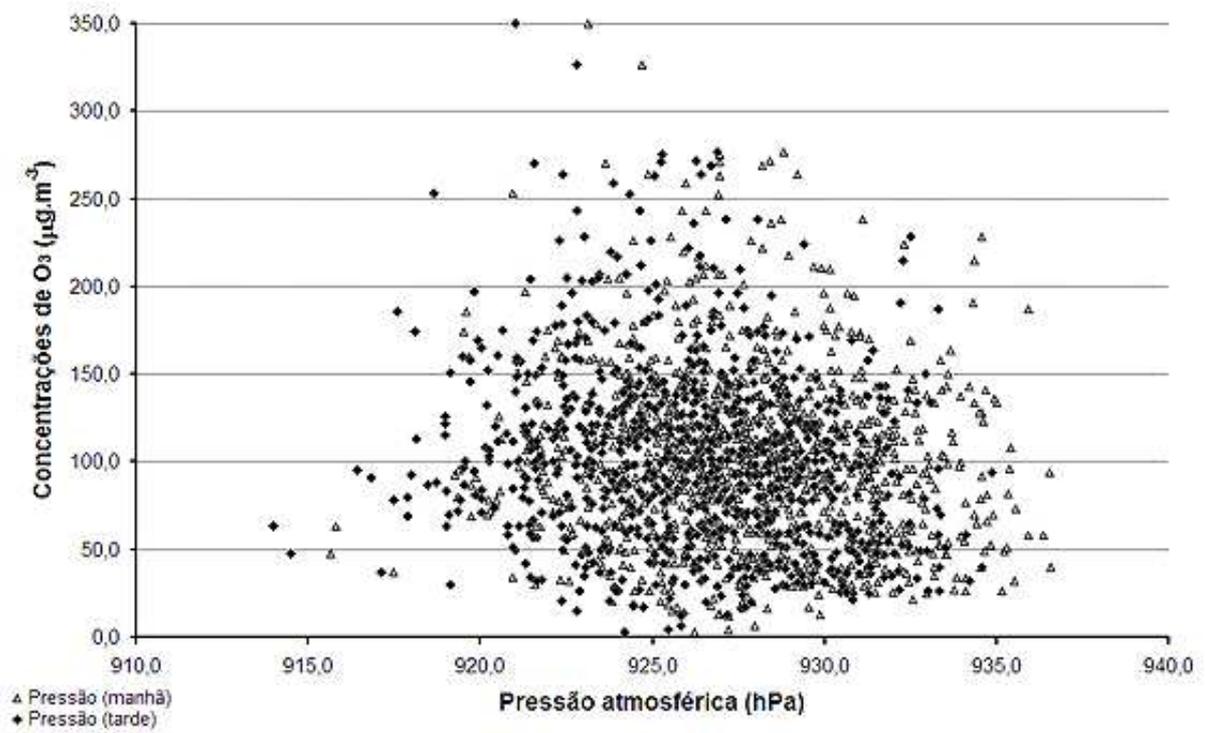


Figura 32: Variação das médias de pressão atmosférica em função da concentração máxima de ozônio para o período da tarde.

**Desenvolvimento da RNA para a estação monitora
de qualidade do ar do Parque do Ibirapuera**

4.1 – Metodologia

Foram feitos, inicialmente, testes comparativos para verificar a importância de determinadas variáveis para a melhoria do ajuste da rede neural. Os valores médios de temperatura do ar, umidade relativa, radiação solar e componentes do vento horizontal foram utilizados em todos os ajustes como entradas da rede, por serem variáveis reconhecidas como importantes no processo de formação, remoção e transporte do ozônio troposférico. Essas entradas são constituídas de médias para os períodos da manhã (8:00-12:00h) e da tarde (13:00-17:00h), conforme discutido na seção 3.1, e as saídas obtidas são estimativas dos valores de concentração média e máxima de ozônio para o período da tarde.

Como critério de avaliação da qualidade de cada ajuste, foram usados os parâmetros da reta de tendência da correlação entre valores calculados e medidos das variáveis de saída do modelo. Na comparação entre ambos os valores de cada variável de saída, o ideal é que os pontos, quando dispostos em um gráfico de dispersão dos valores calculados (y) em função dos valores medidos (x) de cada variável de saída estejam uniformemente dispersos em torno de uma reta $y = a.x + b$, com o coeficiente angular a e o coeficiente linear b tendendo a 1 e 0, respectivamente; e o coeficiente de determinação (R^2) o mais próximo possível de 1.

Para uma noção qualitativa da importância relativa das variáveis de entrada, foi determinado um índice S , obtido pela soma dos valores absolutos dos pesos entre variáveis de entrada e neurônios da camada oculta (equação 14). A comparação entre as somas dos valores absolutos dos pesos indica a importância relativa de cada entrada de acordo com a representatividade dos pesos.

$$S_i = \sum_{j=1}^{NH} |w_{ij}| \quad (14)$$

4.2 - Primeiro ajuste

Para este ajuste a rede foi configurada com 8 neurônios na camada oculta. O conjunto de dados foi dividido em 772 eventos para aprendizado e 151 para teste; esses dados foram apresentados 10 mil vezes à rede neural.

Na comparação entre a concentração média de ozônio calculado pelo modelo e observado (figura 33), os pontos se dispõem de forma pouco dispersa, próximos ao eixo de 45°. O coeficiente de determinação para o conjunto de aprendizado foi de 0,80. Na mesma comparação feita para a concentração máxima de ozônio (figura 34), os pontos se encontram de forma mais dispersa, inclusive com pontos bem afastados, especialmente para casos de concentração muito elevada que a rede neural subestimou, mas ainda assim o coeficiente R^2 esteve adequado em 0,75. Conforme a soma absoluta dos pesos apresentada na figura 35, a variável mais representativa nesse ajuste foi a umidade relativa no período da tarde, seguida pela temperatura no mesmo período e a componente zonal do vento pela manhã.

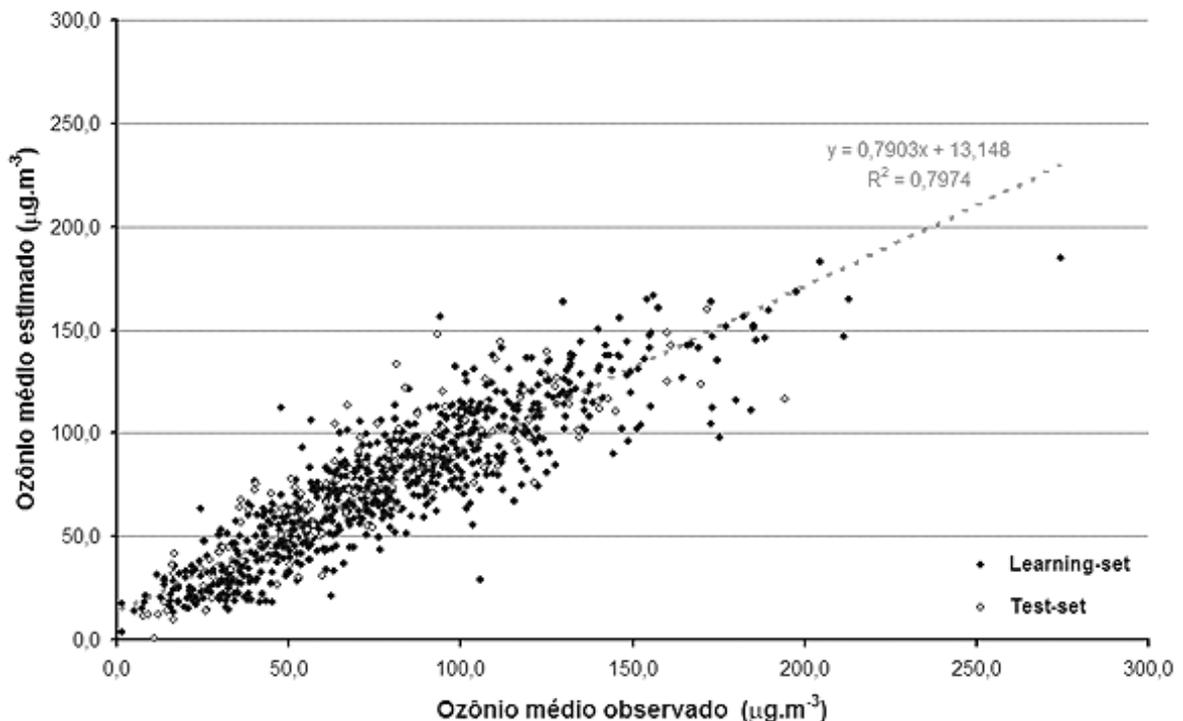


Figura 33: Relação entre valores calculados e medidos da concentração média de ozônio.

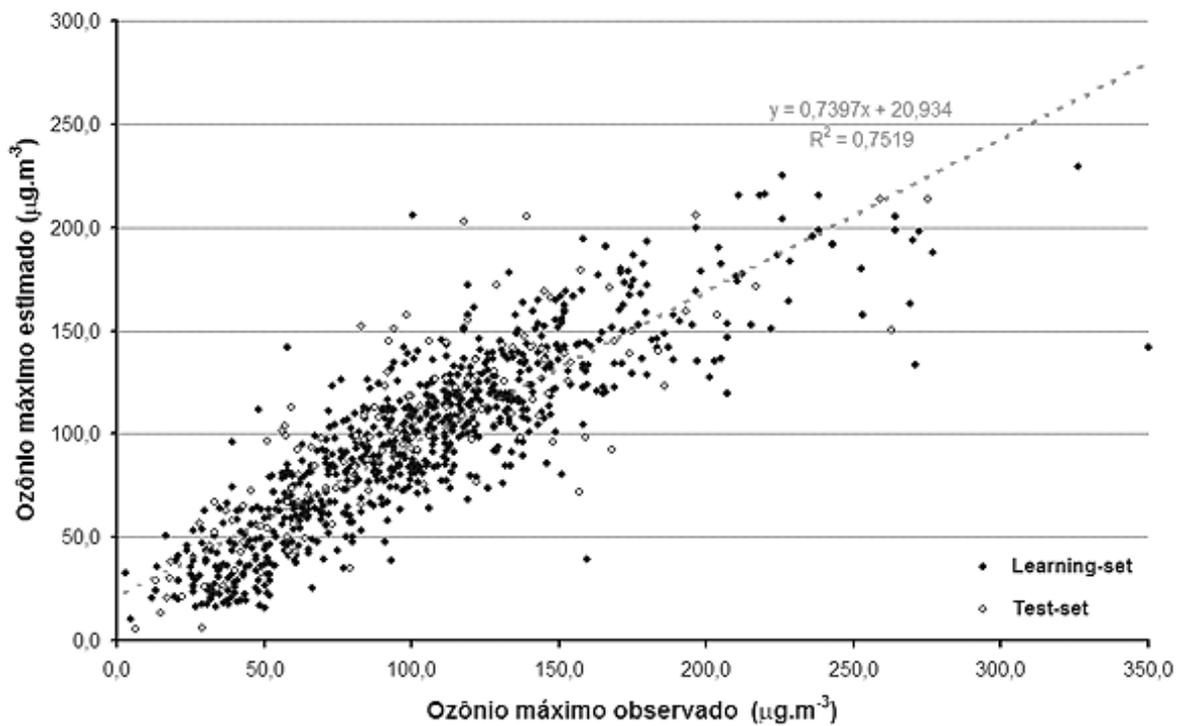


Figura 34: Relação entre valores calculados e medidos da concentração máxima de ozônio.

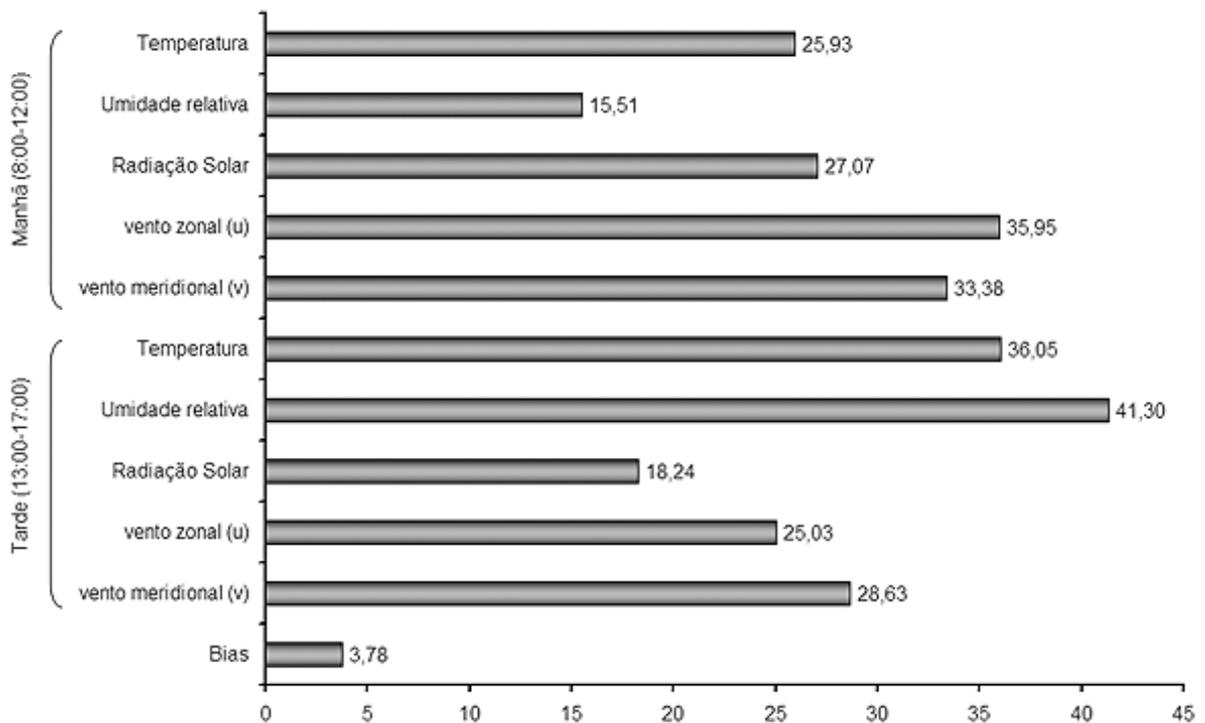


Figura 35: Soma dos valores absolutos dos pesos entre cada variável de entrada e os neurônios da camada oculta.

O primeiro ajuste se mostrou bem adequado, mas alguns pontos de concentrações elevadas não puderam ser previstos pela rede neural. Dada a importância desses eventos críticos sobre a qualidade do ar, foram feitos novos ajustes visando aprimorar a qualidade dessas estimativas.

4.3 - Inclusão da altura da camada de mistura

A altura da camada de mistura foi considerada como uma variável de entrada, junto com as demais variáveis do teste anterior, perfazendo um total de 12 entradas (médias de manhã e tarde para cada uma) e duas saídas (concentração média e máxima de ozônio à tarde). Essa rede foi configurada com 766 dados para aprendizado e 150 para teste, além de ter 8 neurônios na camada oculta e realizar 10 mil apresentações dos dados à RNA.

Comparando a concentração média de ozônio previsto e observado (figura 36), visualmente não se observam grandes mudanças em relação à figura 33 do teste anterior, mas o coeficiente de determinação subiu para 0,82. O mesmo pode-se dizer da comparação da concentração máxima de ozônio (figura 37), cujo R^2 foi de 0,77. A maior diferença ficou representada na figura 38, com a soma absoluta dos pesos das variáveis em cada neurônio. Neste caso, a componente zonal do vento de manhã foi a variável de maior importância para o ajuste; seguida pela umidade relativa à tarde, a qual havia aparecido em destaque no ajuste anterior. Tanto a temperatura, como a radiação solar, perderam um pouco de sua importância no ajuste; provavelmente porque a altura da camada de mistura reduziu suas representatividades, tendo em vista que essas três variáveis são bem correlacionadas (tabela 3.1.3 em 3.1).

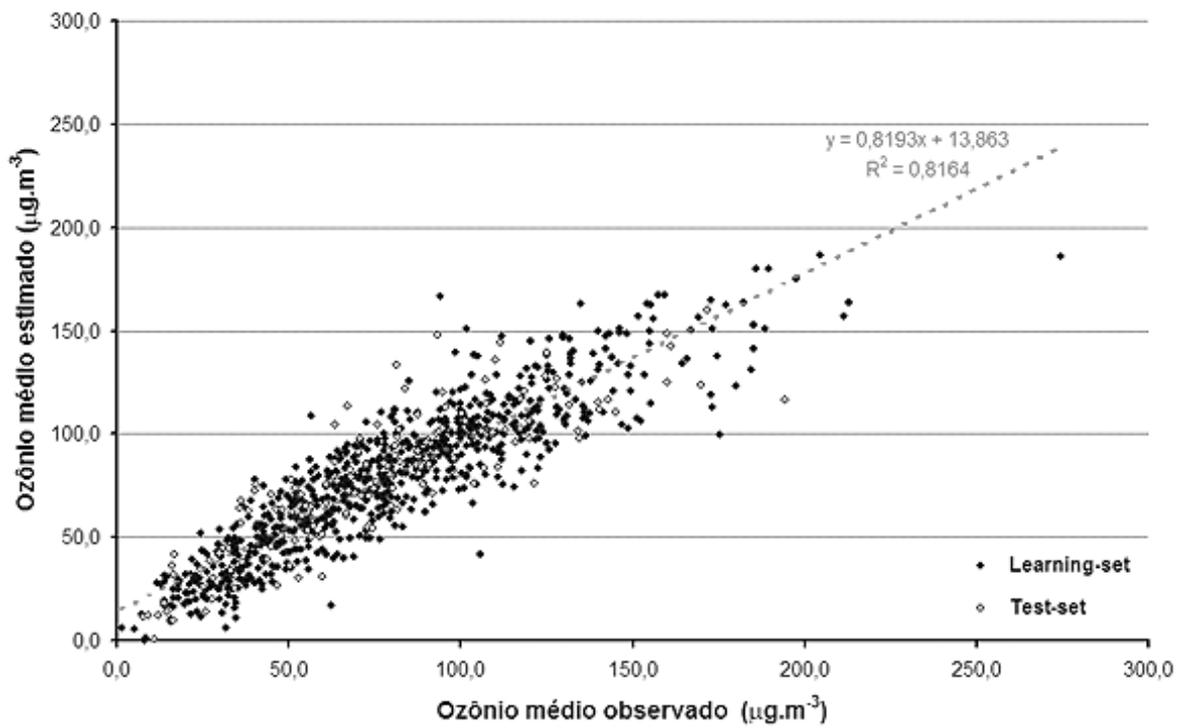


Figura 36: Relação entre valores calculados e medidos da concentração média de ozônio.

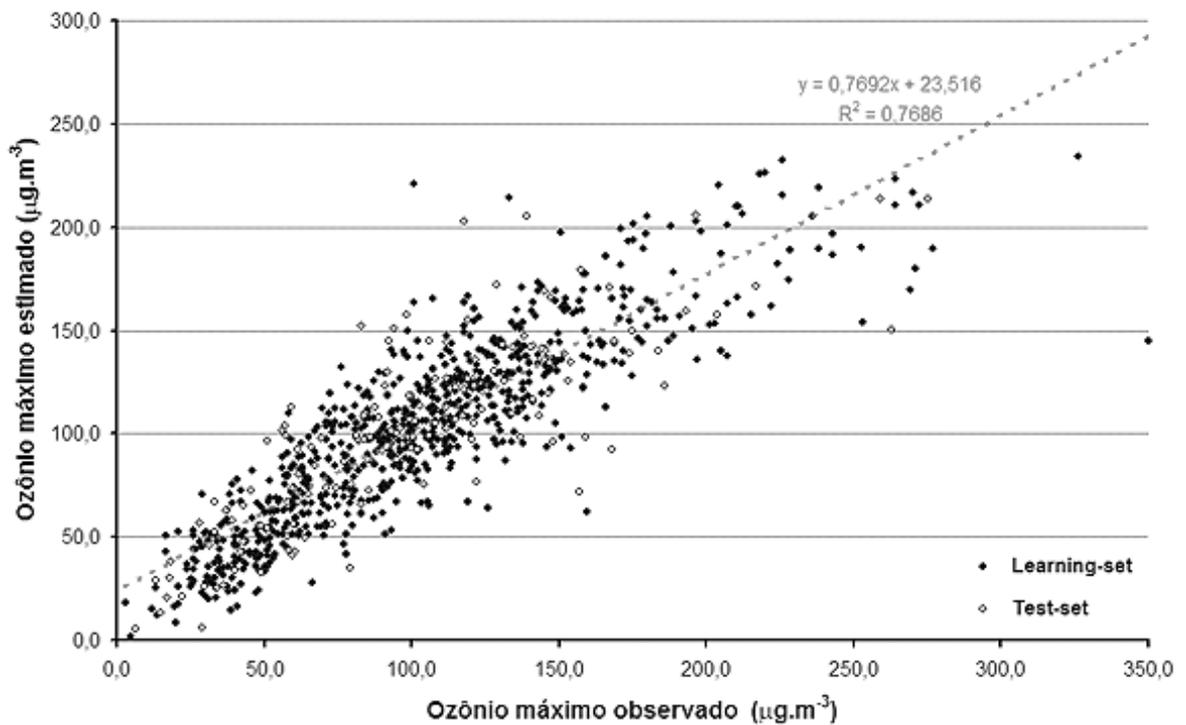


Figura 37: Relação entre valores calculados e medidos da concentração máxima de ozônio.

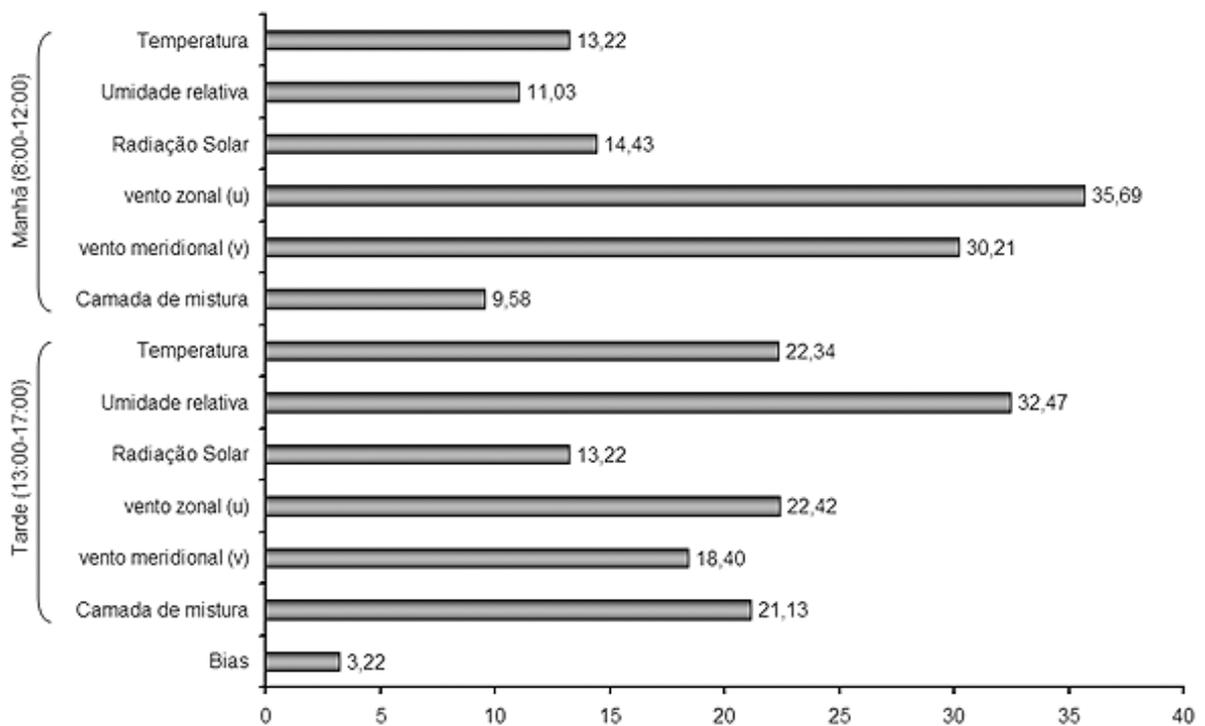


Figura 38: Soma dos valores absolutos dos pesos entre cada variável de entrada e os neurônios da camada oculta.

A radiação solar, que é importante na fotoquímica de formação do ozônio, já era bem caracterizada pela variabilidade da temperatura, dada sua correlação. Com a inclusão da medida de altura da camada de mistura foi reforçada a importância do efeito da radiação solar para a formação de ozônio, tendo em vista que essas variáveis são bem correlacionadas (tabela 3.1.3 – Seção 3.1). Isso levou à pequena melhoria observada nos resultados (figuras 37 e 38), porém, trouxe um grau de redundância na base de dados, que se refletiu em um maior custo computacional na realização do ajuste da rede.

Além disso, no sentido de aplicação prática, a inclusão da altura da camada de mistura torna o modelo dependente da informação de previsão de mais uma variável, a qual pode estar associada a erros que suprimam a pouca melhoria trazida pelo uso de mais esta entrada. Portanto, a inclusão dessa nova entrada parece não trazer benefícios expressivos na previsão das concentrações médias e máximas de ozônio.

4.4 - Rede neural com saídas independentes

Um dos testes verificados foi o da utilização de redes neurais distintas, uma para prever as concentrações médias de ozônio e outra para as máximas. As variáveis de entrada permanecem as mesmas, exceto pela altura da camada de mistura que não fez parte deste teste. Para ambas as redes foram utilizados 772 eventos para o conjunto de treinamento, 151 para o de teste, com 10 mil apresentações dos dados à rede que utiliza 8 neurônios em sua camada oculta.

Na comparação entre ozônio medido e calculado, apresentado nas figuras 39 e 40, respectivamente para concentrações médias e máximas de ozônio, praticamente não se observa diferença entre os resultados dos testes anteriores; os valores de R^2 foram iguais a 0,80 e 0,75, respectivamente. Pela soma absoluta dos pesos da figura 41, para a rede de concentrações médias de ozônio, o vento foi a variável mais importante no ajuste; a componente meridional no período da tarde foi a que apresentou o maior peso, seguida pelo vento zonal de manhã e as componentes restantes (u à tarde e v de manhã). Já para os máximos de ozônio (figura 42), a soma absoluta dos pesos indica o vento zonal de manhã como a variável mais importante no ajuste, seguida pela umidade relativa à tarde e o vento meridional de manhã.

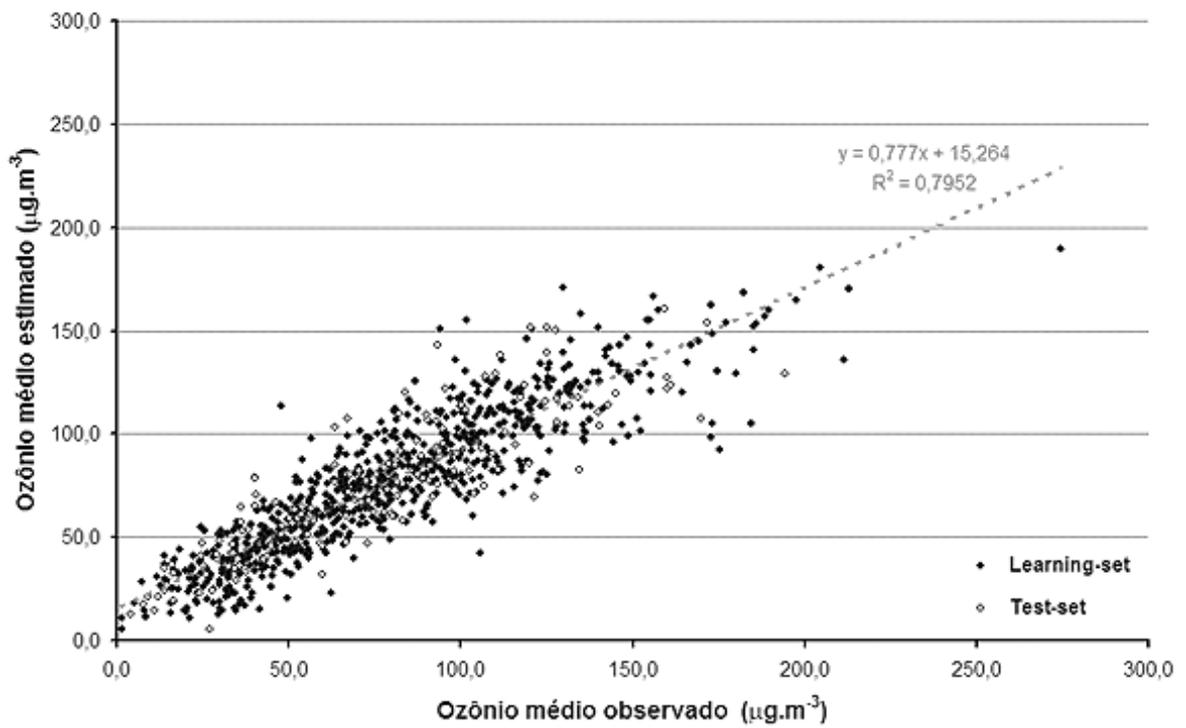


Figura 39: Relação entre valores calculados e medidos da concentração média de ozônio.

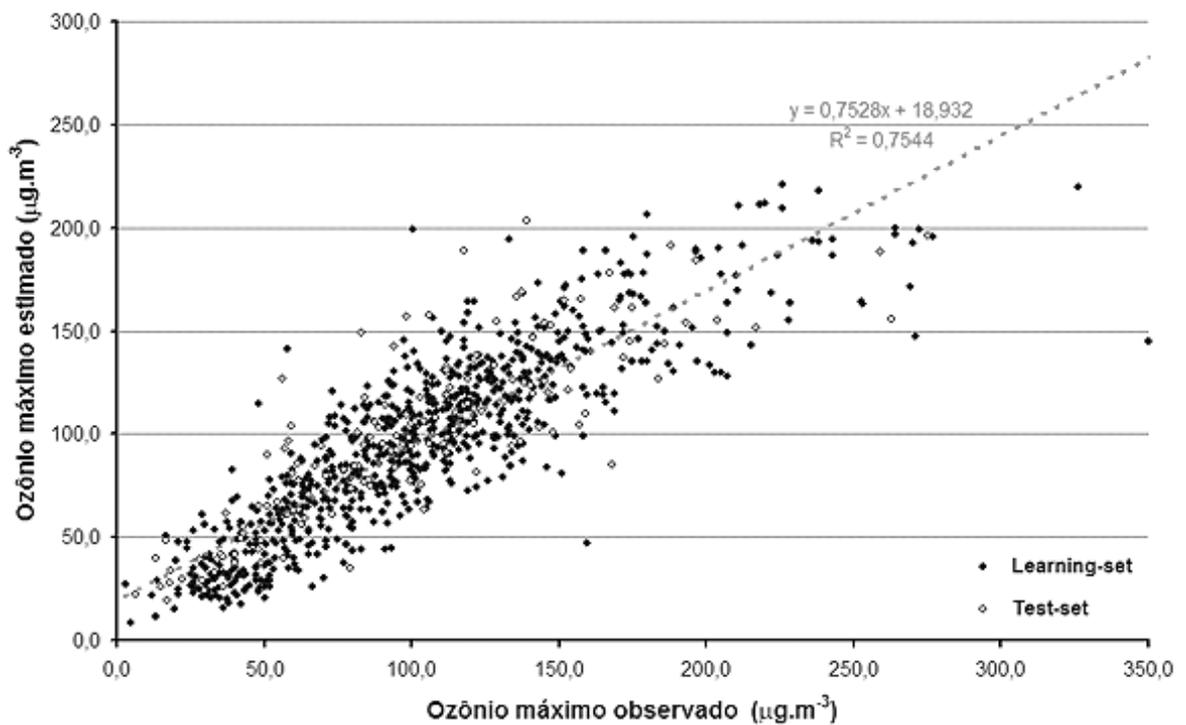


Figura 40: Relação entre valores calculados e medidos da concentração máxima de ozônio.

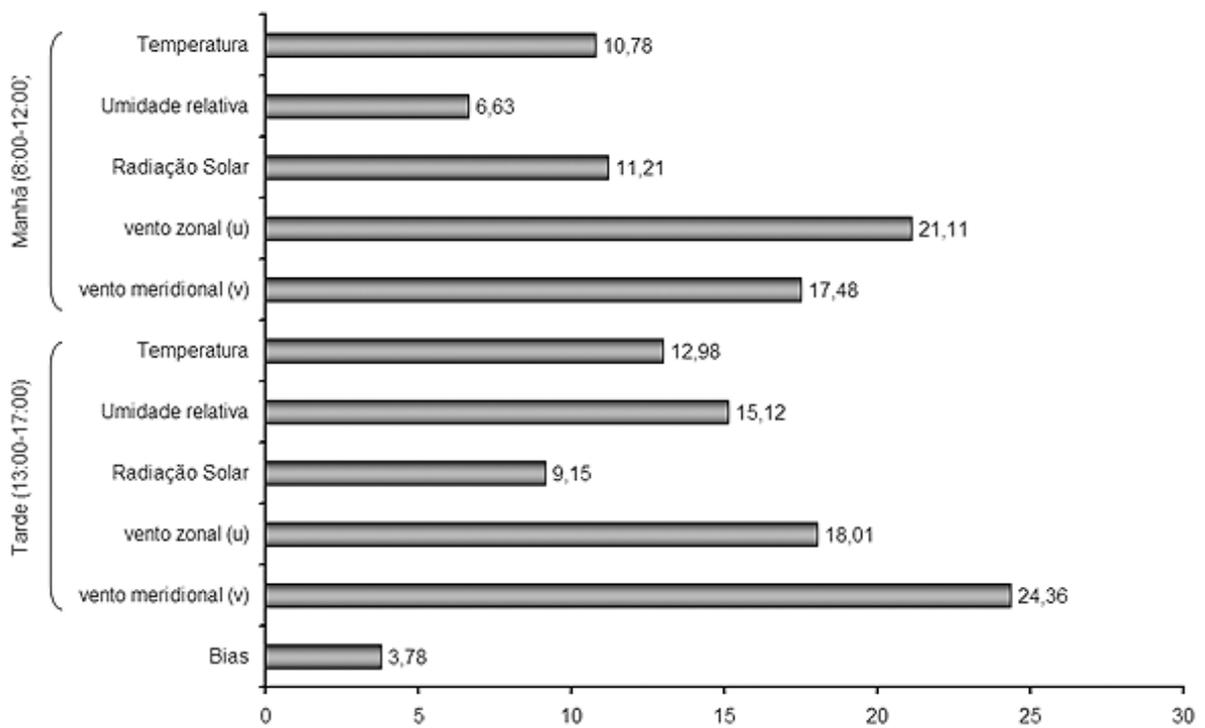


Figura 41: Soma dos valores absolutos dos pesos entre cada variável de entrada e os neurônios da camada oculta; para a rede de concentração média de O₃ como saída.

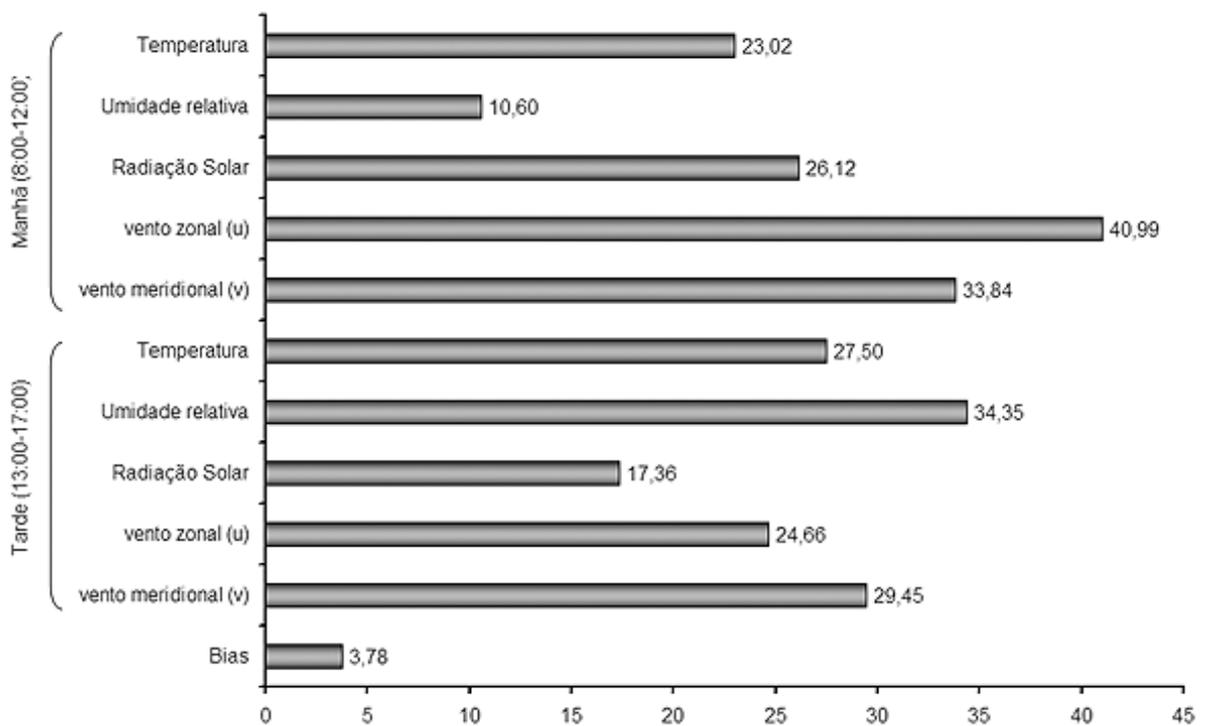


Figura 42: Soma dos valores absolutos dos pesos entre cada variável de entrada e os neurônios da camada oculta; para a rede de concentração máxima de O₃ como saída.

Esse teste não mostra diferenças significativas de qualidade em relação ao ajuste no qual são obtidas as duas saídas simultaneamente. Assim, do ponto de vista da previsão de ozônio, não há benefícios em se adotar uma rede neural para cada saída isoladamente. A diferença na qualidade é perceptível apenas pelo valor de R^2 , com alteração na terceira casa decimal (piorando para a concentração média de O_3 e melhorando para a concentração máxima de O_3).

Por outro lado, o uso de saídas distintas permite uma outra abordagem para a interpretação física dos resultados. Isso pode ser observado pelas figuras 41 e 42, com a soma absoluta dos pesos. Nota-se que as concentrações médias de ozônio à tarde são mais dependentes da circulação atmosférica do que dos demais parâmetros, embora a temperatura e umidade à tarde também apresentem pesos significativos. Já as concentrações máximas de ozônio mostram-se ser mais influenciadas pelo padrão de vento de manhã do que da tarde, quando temperatura e umidade desempenham papel mais relevante.

4.5 - Substituição dos dados de vento do SODAR pelos da estação climatológica do IAG

Nesse ajuste foi empregado o campo de vento da estação climatológica do IAG na Água Funda, em vez dos dados do SODAR no Ibirapuera. Apesar das medidas à altura de 50 metros se mostrarem adequadas ao ajuste de redes neurais, é preciso considerar que ao longo do período ocorreram diversas falhas de medição, o que levou ao descarte de eventos que possivelmente poderiam contribuir para a qualidade do treinamento da rede. A remoção dessas lacunas levou à redução de 16% do conjunto de dados horários válidos.

Como as medições feitas pela estação climatológica do IAG são menos sujeitas a falhas, foi estudada a possibilidade de se adotar essas medidas como substitutas dos registros do SODAR. De acordo com a tabela 4.5.1, há uma boa correlação (0,8) entre as componentes do vento de cada uma das estações. Essa relação se constata pelo dendograma da figura 43,

resultante da análise de agrupamento pelo método de Ward, utilizando distância euclidiana para medir a similaridade entre os dados para o agrupamento entre as variáveis.

Tabela 4.5.1: Correlações de Pearson para as componentes do vento obtido pelo SODAR (Ibirapuera) e a estação climatológica do IAG (Água Funda).

		Sodar		IAG	
		u	v	u	v
Sodar	u	1,00	-0,44	0,79	-0,11
	v	-0,44	1,00	-0,52	0,75
IAG	u	0,79	-0,52	1,00	-0,25
	v	-0,11	0,75	-0,25	1,00

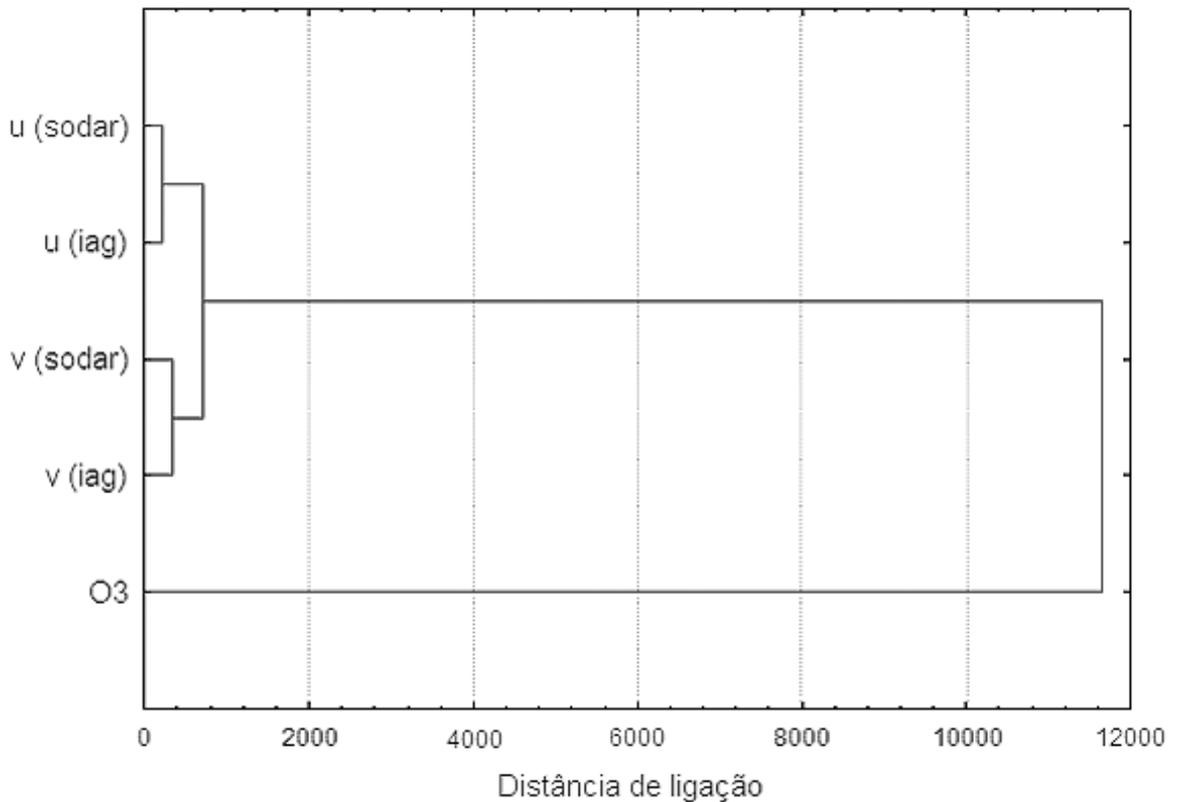


Figura 43: Dendrograma obtido da análise de agrupamento de dados de vento, considerando o método de Ward e a medida de distância euclidiana.

A rede neural foi configurada com 1401 eventos para o conjunto de aprendizado, 273 para o de teste, realizando-se 10 mil apresentações dos dados à rede e utilizando 8 neurônios na camada oculta. Para este caso em particular, foram usados dados de 1997 a 2003.

Comparando as concentrações médias de ozônio calculado e seu valor experimental (figura 44), os dados se encontram de forma um pouco mais dispersa do que o observado nos testes anteriores, especialmente a partir das medidas acima de $150 \mu\text{g.m}^{-3}$, resultando em R^2 de 0,78. Na comparação com concentrações máximas de ozônio (figura 45) a dispersão dos pontos é mais evidente; há muitos valores de concentração abaixo de $100 \mu\text{g.m}^{-3}$ que foram superestimadas, e outras acima de $200 \mu\text{g.m}^{-3}$ que foram subestimadas, ficando o R^2 em 0,71. De acordo com a soma absoluta dos pesos (figura 46), a componente zonal do vento de manhã foi a variável mais representativa no ajuste da rede neural, seguida pela média de temperatura à tarde e a componente meridional do vento de manhã.

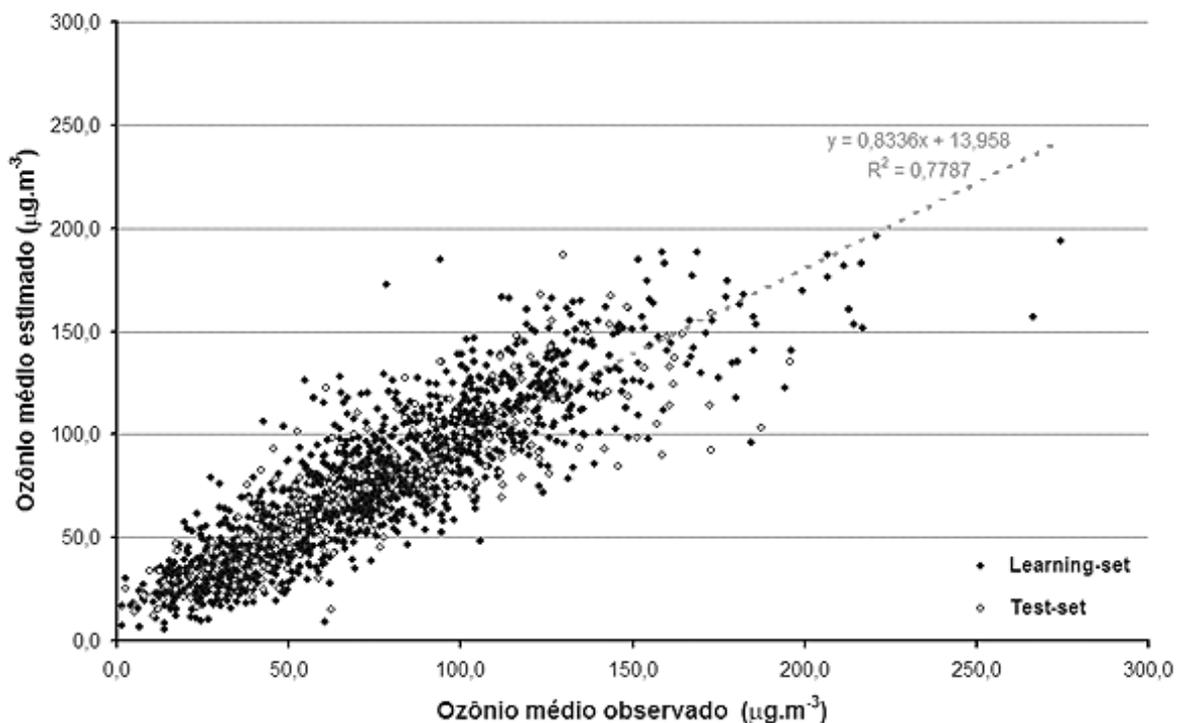


Figura 44: Relação entre valores calculados e medidos da concentração média de ozônio.

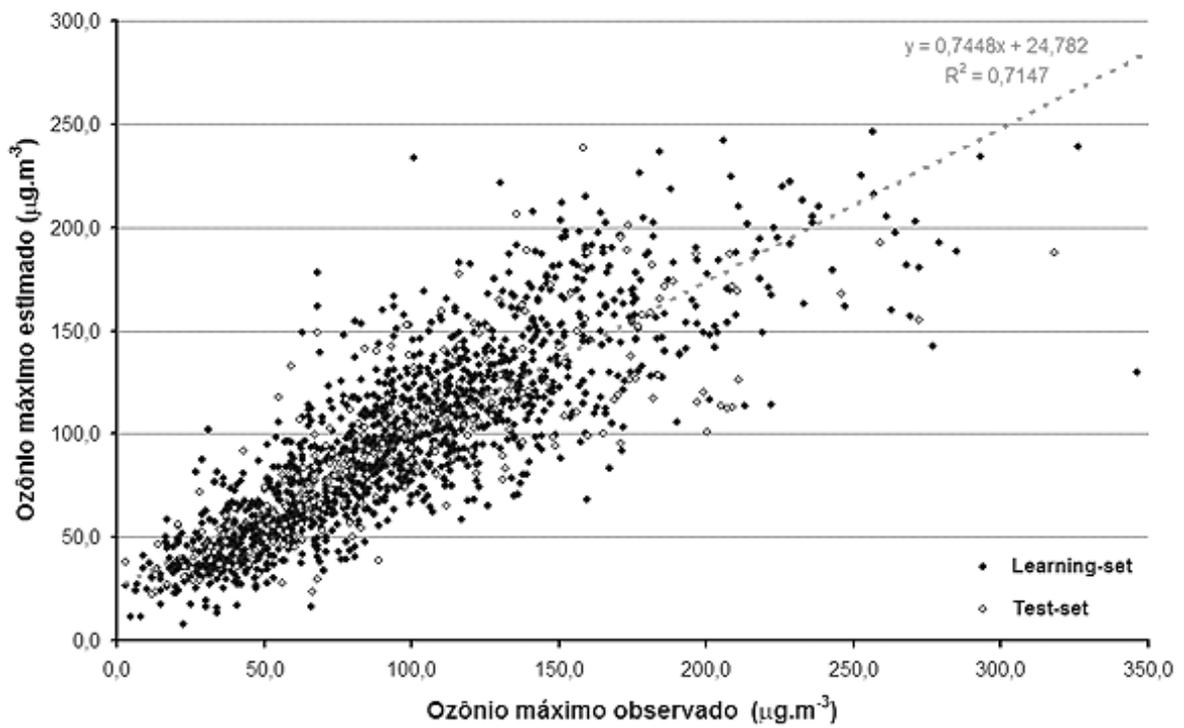


Figura 45: Relação entre valores calculados e medidos da concentração máxima de ozônio.

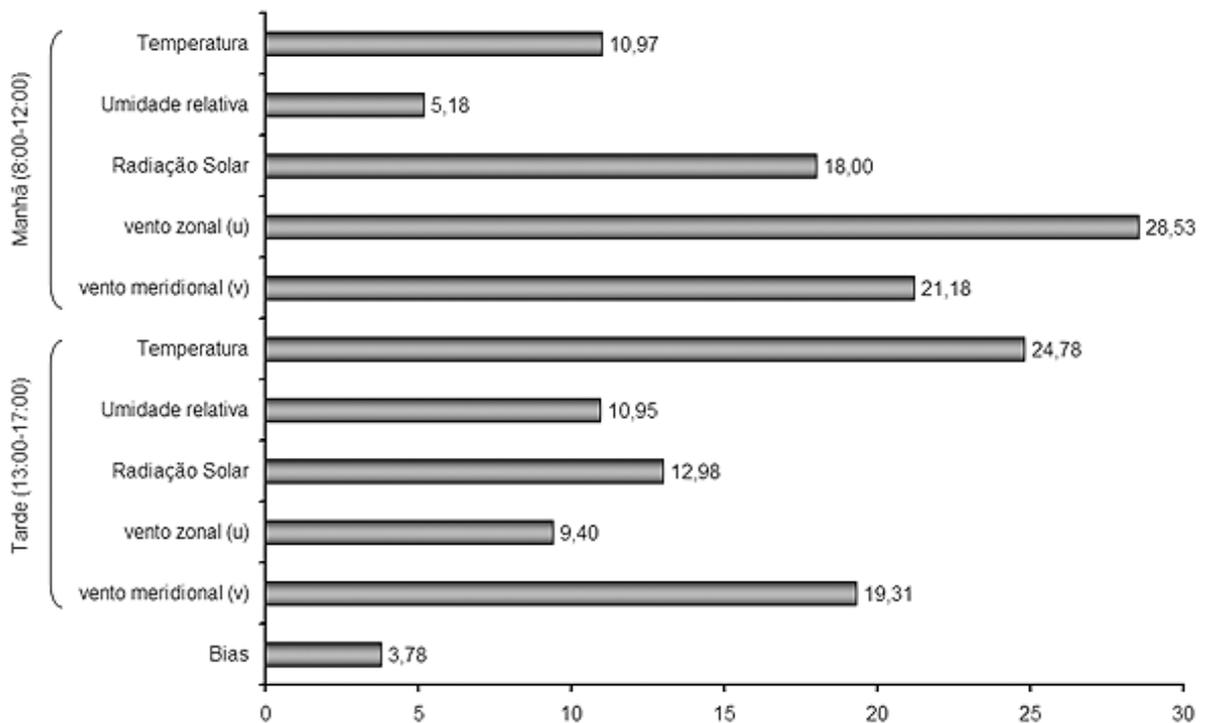


Figura 46: Soma dos valores absolutos dos pesos entre cada variável de entrada e os neurônios da camada oculta.

Assim, a substituição do vento do SODAR pelo vento do IAG não oferece vantagens ao ajuste da rede neural e, de fato, prejudica um pouco a qualidade das estimativas, especialmente para as concentrações máximas de ozônio, em que é observada uma grande dispersão dos pontos na comparação entre o calculado e o medido (figura 45). Qualitativamente (figura 46) a maior diferença em relação aos outros testes está na maior dependência do padrão do vento de manhã, o que chega a reduzir a importância de outras variáveis, como a umidade relativa, por exemplo.

4.6 - Inclusão da pressão atmosférica

Para este ajuste, a pressão atmosférica foi incluída como uma variável de entrada. As medidas foram obtidas da estação climatológica do IAG, por serem menos suscetíveis a falhas. Foram usados dados de 1999 a 2003. Essa rede foi configurada com 766 dados para aprendizado e 150 para teste, além de ter 8 neurônios na camada oculta, com 10 mil apresentações dos dados à RNA.

A comparação entre a concentração média de ozônio medida e calculada (figura 47) não é muito diferente do que fora constatado para o ajuste que utilizava altura da camada de mistura (item 4.3), resultando em coeficiente de determinação de 0,81. Também não há muitas diferenças em relação ao gráfico das concentrações máximas de ozônio (figura 48), cujo valor de R^2 foi de 0,75. Dentre as variáveis de maior peso neste ajuste, segundo a figura 49, estão o campo de vento pela manhã (componentes zonal e meridional, respectivamente) e a temperatura média no período da tarde.

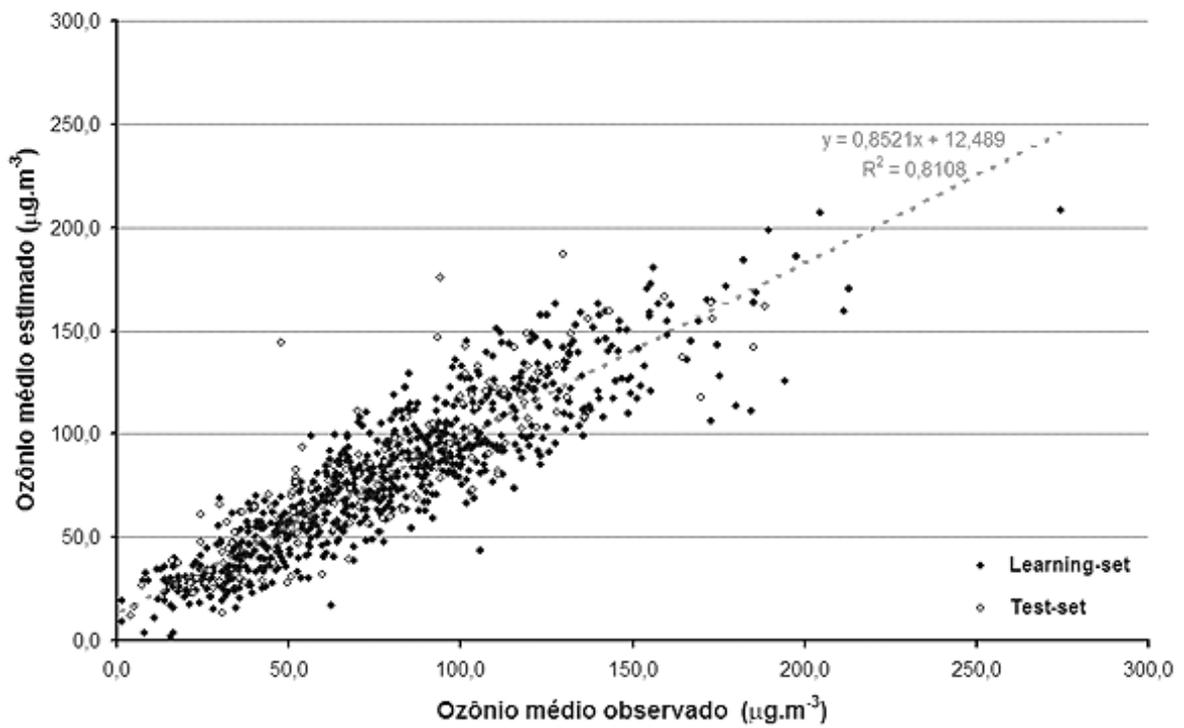


Figura 47: Relação entre valores calculados e medidos da concentração média de ozônio.

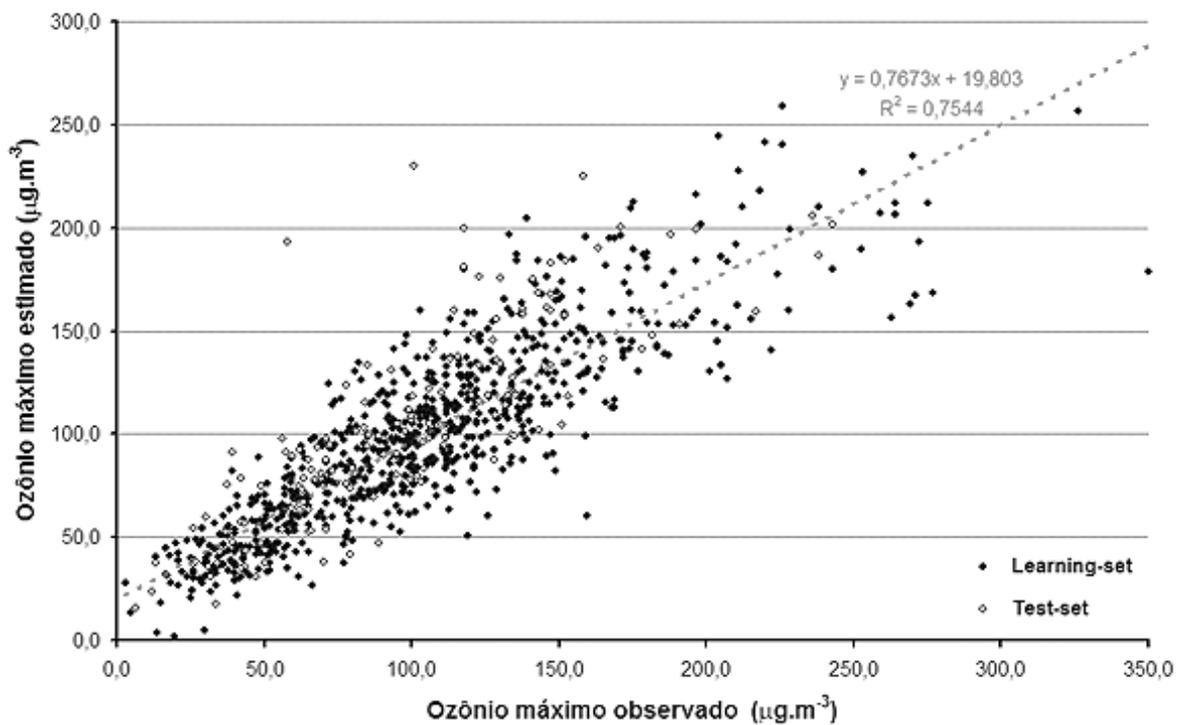


Figura 48: Relação entre valores calculados e medidos da concentração máxima de ozônio.

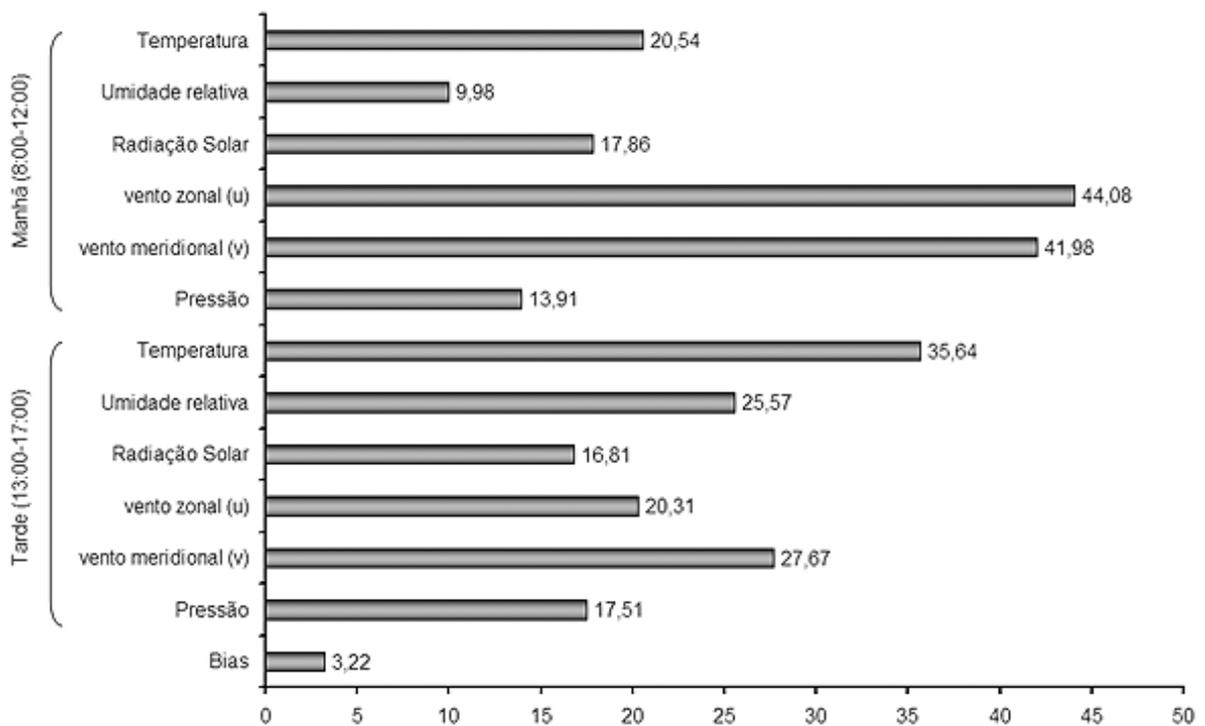


Figura 49: Soma dos valores absolutos dos pesos entre cada variável de entrada e os neurônios da camada oculta.

Pelos gráficos comparativos entre ozônio previsto e observado, não se observam melhorias significativas nos resultados, seja para as concentrações médias ou máximas de ozônio. A diferença só é perceptível pela variação do coeficiente de determinação (R^2). Em relação ao primeiro ajuste (item 4.2), houve um inexpressivo aumento desse índice após a inclusão da pressão atmosférica (1,7% para as concentrações médias de ozônio e 0,3% para as máximas); o que chega a ser inferior ao obtido com o ajuste incluindo a altura da camada de mistura. Sendo assim, no contexto da previsão de ozônio, não é muito vantajoso incluir mais essa variável; de fato, pode vir até mesmo a prejudicar os resultados, caso a previsão meteorológica não seja adequada, uma vez que estaria acrescentando mais uma fonte de incerteza à rede neural.

Pela interpretação física dos resultados (figura 49), a pressão atmosférica nesse ajuste desempenha um papel secundário na qualidade dos resultados. Porém, pela manhã ela tem importância equivalente à umidade relativa e radiação solar, e pela tarde tem pesos próximos

ao da radiação solar e a componente zonal do vento. O padrão de vento de manhã e a média de temperatura à tarde foram as entradas com maior importância para a qualidade do ajuste.

4.7 - Altura da camada de mistura e pressão atmosférica como entradas

Para este ajuste foram consideradas a altura da camada de mistura e pressão atmosférica como entradas da mesma rede neural, em vez de se adotá-las individualmente. O objetivo foi avaliar os possíveis benefícios que poderiam advir do uso dessas duas variáveis ao mesmo tempo, além de buscar um melhor entendimento da relação das diferentes variáveis para a previsibilidade do ozônio troposférico. Ao todo foram 14 entradas (médias de manhã e tarde) utilizando 767 eventos para o conjunto de treinamento, e 149 para o de teste, em 10 mil apresentações dos dados à rede que utiliza 8 neurônios na camada oculta.

Comparando as concentrações médias de ozônio calculado pelos valores observados (figura 50), houve uma pequena melhora no índice R^2 (igual a 0,83), em consideração aos ajustes anteriores. No gráfico dos máximos de ozônio (figura 51), observa-se que os pontos continuam muito dispersos para concentrações mais elevadas (acima de $200 \mu\text{g.m}^{-3}$); o resultado também foi um pouco melhor, com R^2 de 0,77. A soma absoluta dos pesos (figura 52) indica a média de temperatura à tarde como a variável mais representativa no ajuste, seguida pelo padrão de vento de manhã (médias de u,v) e a umidade relativa à tarde.

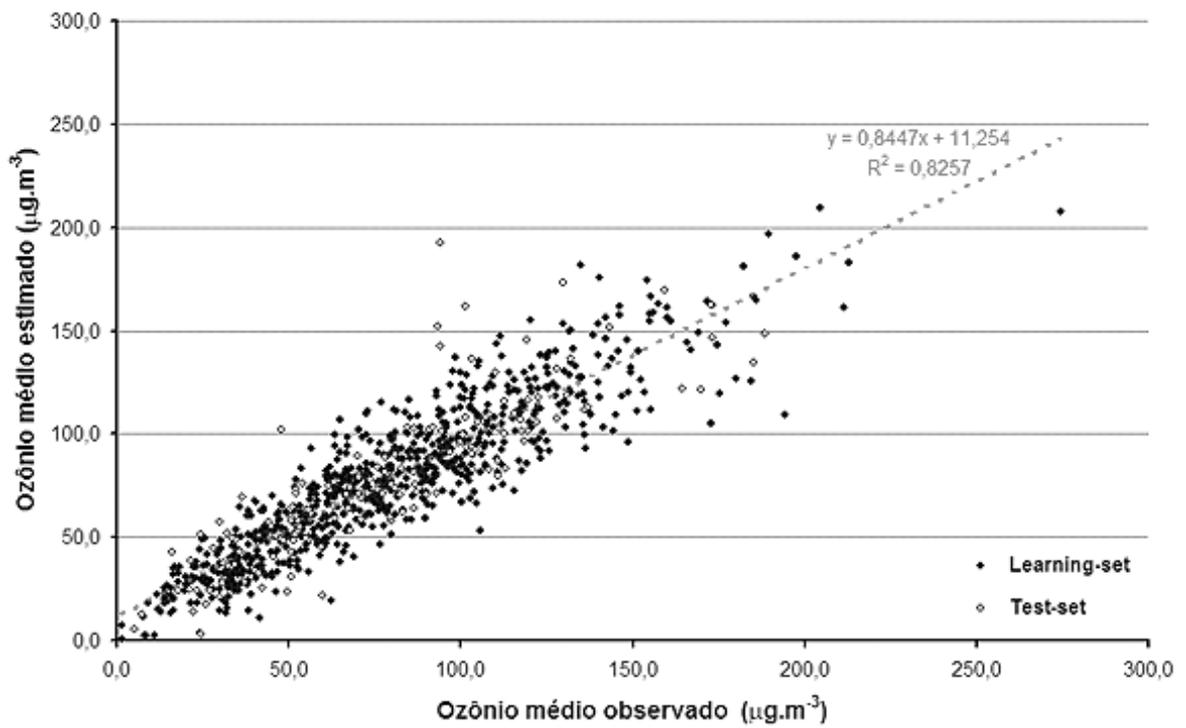


Figura 50: Relação entre valores calculados e medidos da concentração média de ozônio.

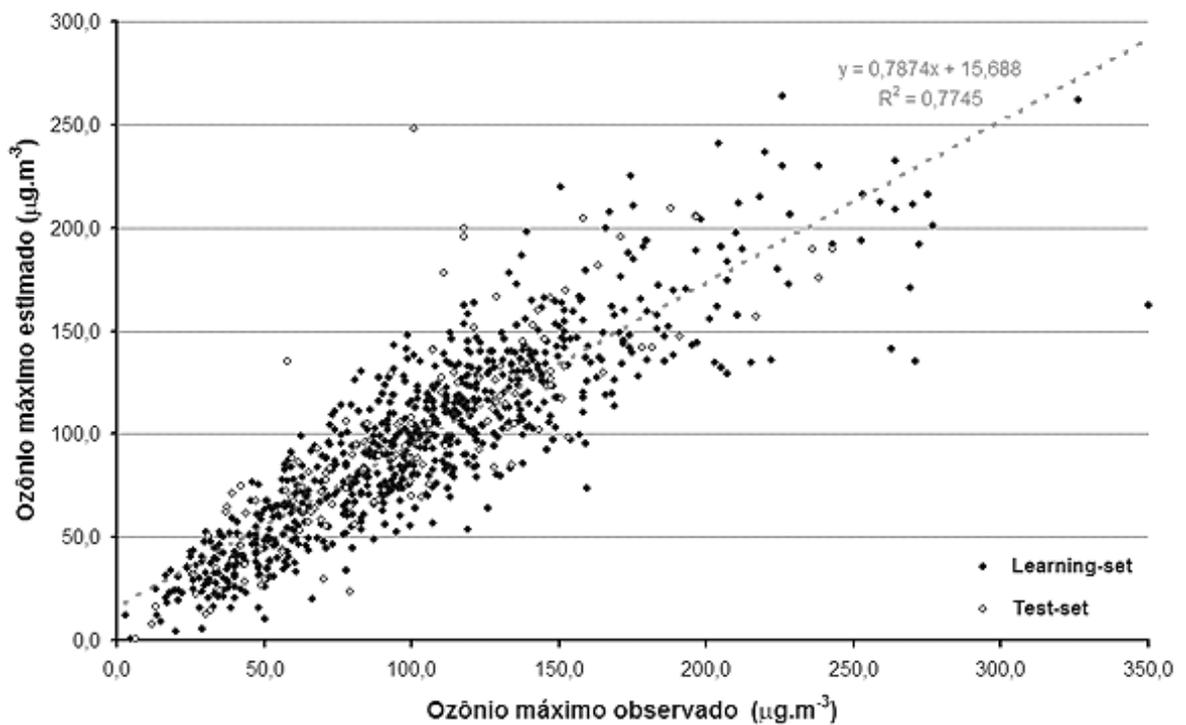


Figura 51: Relação entre valores calculados e medidos da concentração máxima de ozônio.

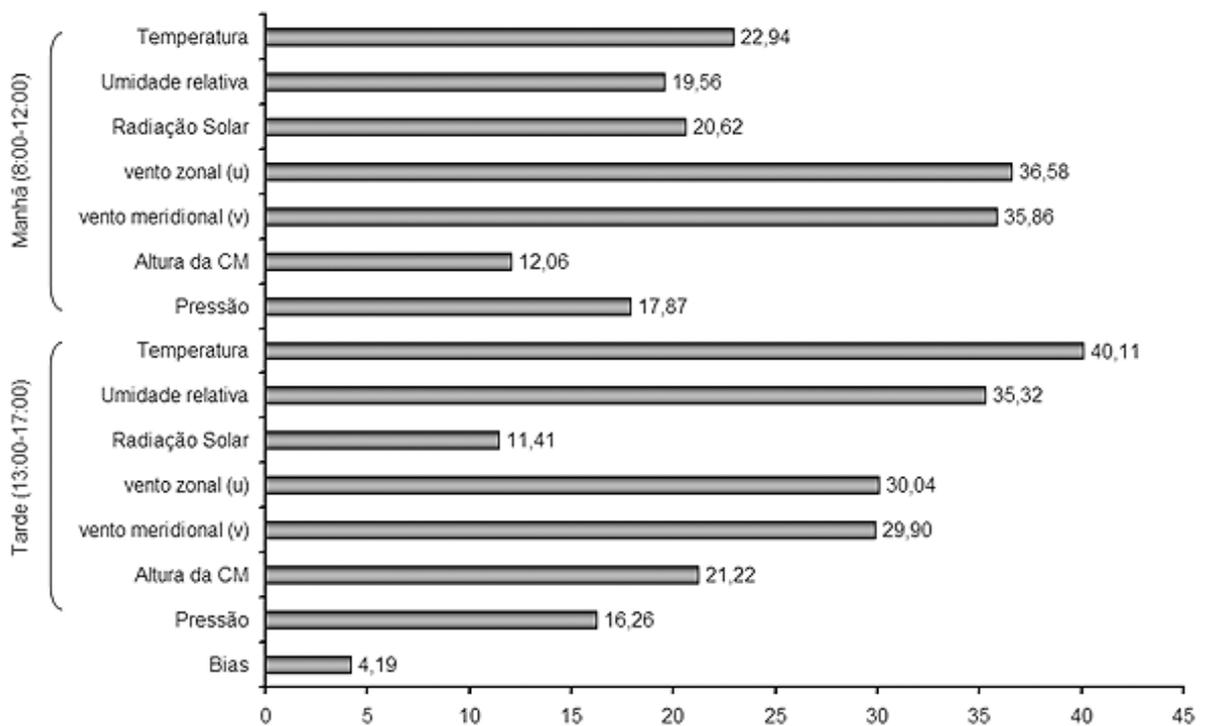


Figura 52: Soma dos valores absolutos dos pesos entre cada variável de entrada e os neurônios da camada oculta.

Em comparação ao primeiro ajuste feito, as melhorias notadas nos coeficientes de determinação foram de 3,5% e 3,0%, para as concentrações médias e máximas de ozônio, respectivamente; superiores ao obtido pelo uso das variáveis isoladamente, porém, ainda pouco expressivas. Como instrumento de previsão, é preciso ter cautela no uso dessa rede, uma vez que as incertezas nas entradas podem prejudicar a qualidade das estimativas de ozônio. Por outro lado, no sentido de interpretação física, este ajuste permitiu uma visão mais ampla da relação entre as variáveis para a previsibilidade do ozônio (figura 52).

Na figura 53 a soma absoluta dos pesos da figura 52 foi expressa em termos percentuais e comparada com a proporção média obtida dos ajustes anteriores.

É interessante notar que a radiação solar, altura da camada de mistura e pressão atmosférica foram variáveis que apresentaram, proporcionalmente, os menores pesos. Isso porque, conforme visto na análise de agrupamento do item 3.1 (figura 8), essas variáveis se agrupam de maneira distante das demais. Mesmo a radiação solar, importante na fotoquímica

do ozônio, não é uma das mais relevantes entradas da rede neural, já que outras entradas, como a temperatura e velocidade do vento, indiretamente representaram sua variabilidade.

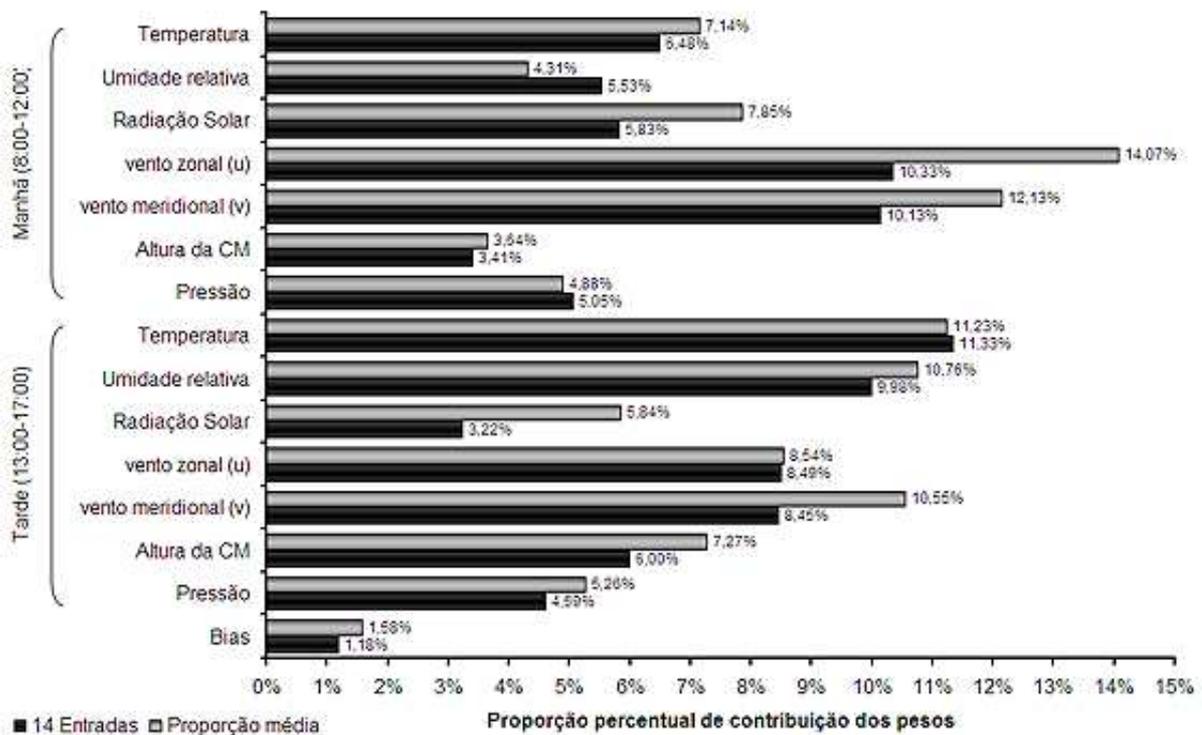


Figura 53: Proporção percentual de contribuição dos pesos de cada variável para o ajuste de rede neural com 14 entradas, e para a média dos ajustes anteriores.

4.8 - Testes de melhoria da representatividade dos dados

Foram feitos alguns testes visando aumentar o número de eventos representativos apresentados no treinamento da rede neural. Até então os casos inseridos na entrada do modelo eram compostos de médias de manhã e tarde, para intervalos contínuos, ou seja, sem falhas de medição (ausência de registro) ao longo de cada um dos períodos de cinco horas. Entretanto, por esse critério a base de dados ficou constituída de relativamente poucos eventos, uma vez que as falhas foram frequentes.

Para o primeiro teste (A), as médias foram calculadas mesmo na ausência de alguns registros horários ao longo do período. Dessa forma a base de dados aumentou em 43%.

Outra abordagem (B) foi a ampliação do intervalo, passando das 8:00-12:00h para 07:00-12:00h de manhã, e das 13:00-17:00h para 13:00-18:00h à tarde. Neste caso foi mantido o critério de se calcularem médias em intervalos consecutivos. Com isso, a base de dados foi reduzida em 7,8%. O objetivo foi verificar se as médias desse novo intervalo são mais apropriadas para caracterizar os perfis das variáveis da manhã e da tarde.

Outro teste realizado (C) baseou-se nos dois anteriores, ampliando o intervalo de 5 para 6 horas e considerando os períodos que apresentaram algumas ausências de medidas; dessa vez reduzindo a tolerância para, no máximo, duas por período. Isso levou a um aumento de 50,8% do número de eventos na base de dados em relação ao teste anterior, ou 39,1% em relação ao original.

Por fim, foi feito um teste (D) semelhante ao primeiro, porém, adotando o critério de tolerância do teste anterior, ou seja, calculando as médias quando houvesse, no máximo, duas ausências em cada intervalo de 5 horas.

A configuração das redes neurais é a mesma, exceto pelo número de dados separados para o conjunto de aprendizado e de teste (vide tabela 4.8.1). Todas utilizaram 14 entradas, tal como no item 4.7, com 10 mil apresentações dos dados à rede e tendo 8 neurônios na camada oculta.

Os resultados obtidos estão resumidos na tabela 4.8.1. Foi observado que, em todas as situações, não houve melhoria na qualidade dos resultados; todas as redes ajustadas tiveram desempenho inferior ao ajuste do item 4.7. O teste B foi o que apresentou melhor resultado para concentrações médias de ozônio, mas foi o segundo pior para as máximas do poluente. O teste A teve o melhor ajuste para concentrações máximas de ozônio, mas ainda assim inferior ao obtido no item 4.7.

Tabela 4.8.1: Parâmetros de configuração e indicadores de qualidade de cada ajuste.

Teste	Número de eventos		R^2		Reta ajustada	
	Aprendizado	Teste	O3 médio	O3 máximo	O3 médio	O3 máximo
A	1097	213	0,8187	0,7670	$0,8118x+12,366$	$0,7527x+21,567$
B	712	133	0,8213	0,7521	$0,8193x+10,837$	$0,7404x+22,410$
C	1068	206	0,8013	0,7422	$0,7907x+12,857$	$0,7152x+26,367$
D	1041	201	0,8073	0,7614	$0,7999x+16,944$	$0,7373x+25,306$

4.9 - Experiências visando aprimorar a representação das concentrações elevadas

Visando melhorar a qualidade das estimativas de concentrações elevadas de ozônio, foi ajustada uma rede neural utilizando apenas concentrações maiores ou iguais a $160 \mu\text{g.m}^{-3}$. Entretanto, foram poucos os casos em que esse nível de concentração foi atingido, portanto, o conjunto de aprendizado ficou com apenas 89 eventos, e o de teste com 11. Essa condição levou ao erro de sobreajuste da rede, em que, na comparação entre valores medidos e calculados, os dados do conjunto de aprendizado ficaram bem ajustados (R^2 de 0,91 e 0,95 para concentrações médias e máximas de ozônio, respectivamente), mas os de teste não (R^2 de 0,68 e 0,26), o que inviabiliza a aplicação dessa rede neural.

Outra experiência consistiu em selecionar todos os casos em que ocorreram concentrações de ozônio maior ou igual a $160 \mu\text{g.m}^{-3}$, repetindo-os dez vezes dentro da base de dados original. A idéia era tentar induzir a rede neural a dar maior ênfase a esses eventos durante o processo de ajuste. Porém, tais réplicas constituíram informação redundante ao ajuste de redes neurais. Isso já era esperado, de acordo com o que existe na literatura (HEO e KIM, 2004), mas ainda assim esse teste foi praticado, no intuito de entender melhor o comportamento das redes neurais. O resultado para as concentrações médias de ozônio ficou bom, chegando a 0,83 no coeficiente de determinação na comparação entre valores medidos e calculados, mas para as concentrações máximas o valor de R^2 caiu para 0,74.

A experiência foi repetida, com a adição de uma variação aleatória aos dados, a qual, por regra, não pôde ser superior ao desvio padrão para cada variável, em cada caso. Em outras

palavras, foi adicionado um "ruído" em cada medida, de forma que não desconfigurasse o padrão das variáveis, mas que não gerasse réplicas exatas dos dados originais. Como resultado foram criados diversos eventos muito semelhantes, o que levou a rede a se ajustar em função desses pontos, perdendo seu significado, por ter sido ajustada com um conjunto artificial de dados.

4.10 - Informação quanto ao dia da semana

Uma das abordagens empregadas no ajuste de redes neurais foi a diferenciação dos dados de acordo com os dias da semana. De acordo com a figura 54, o número de ultrapassagens do padrão de $160 \mu\text{g.m}^{-3}$ de ozônio é semelhante ao longo dos dias da semana, exceto para o sábado. Na figura 55 a média de concentração é quase a mesma durante os dias úteis, mas nos finais de semana ela é um pouco mais elevada. Tal perfil levou à realização de dois testes de ajuste: no primeiro, foi inserida como entrada uma informação quanto ao dia da semana (variável de 1 a 7); no segundo, foi usado um índice binário, indicando se era dia útil ou não (0- dias úteis, 1- finais de semana). A rede foi configurada tal como no item 4.7, ou seja, com 14 variáveis meteorológicas de entrada, mais a informação quanto ao dia da semana; obtendo-se como saída as concentrações médias e máximas de ozônio para a tarde.

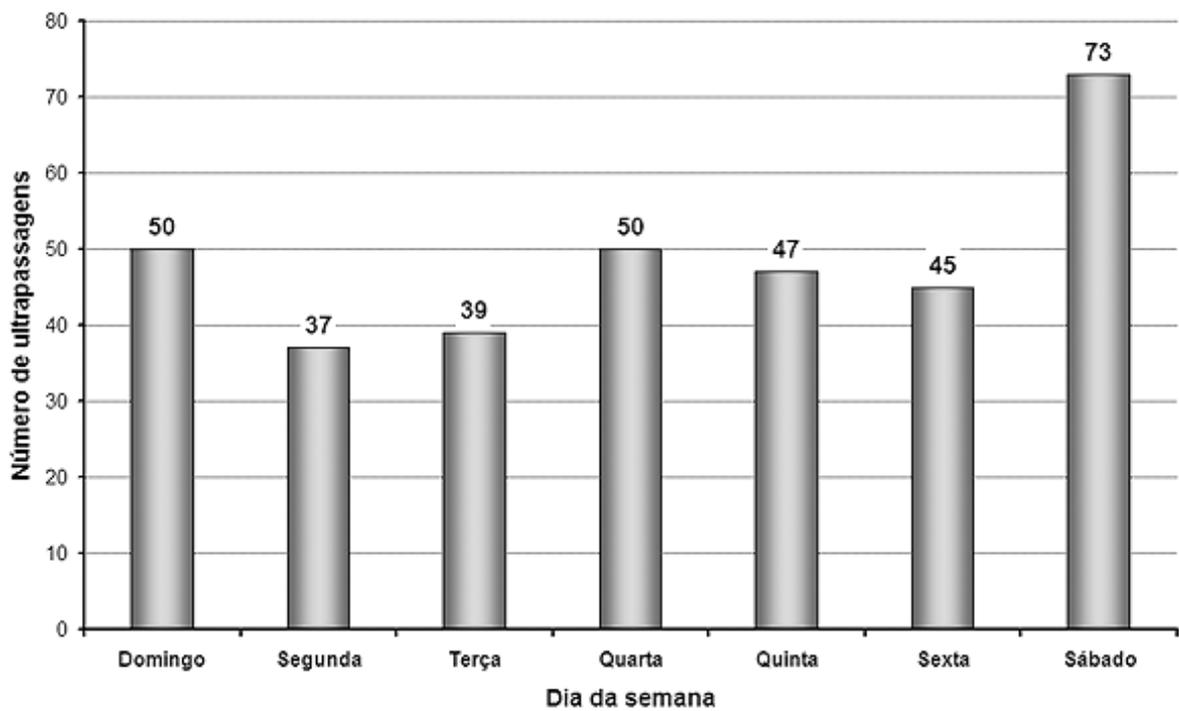


Figura 54: Número de ultrapassagens do padrão de qualidade do ar ($160 \mu\text{g.m}^{-3}$) para cada dia da semana de 1999 a 2003.

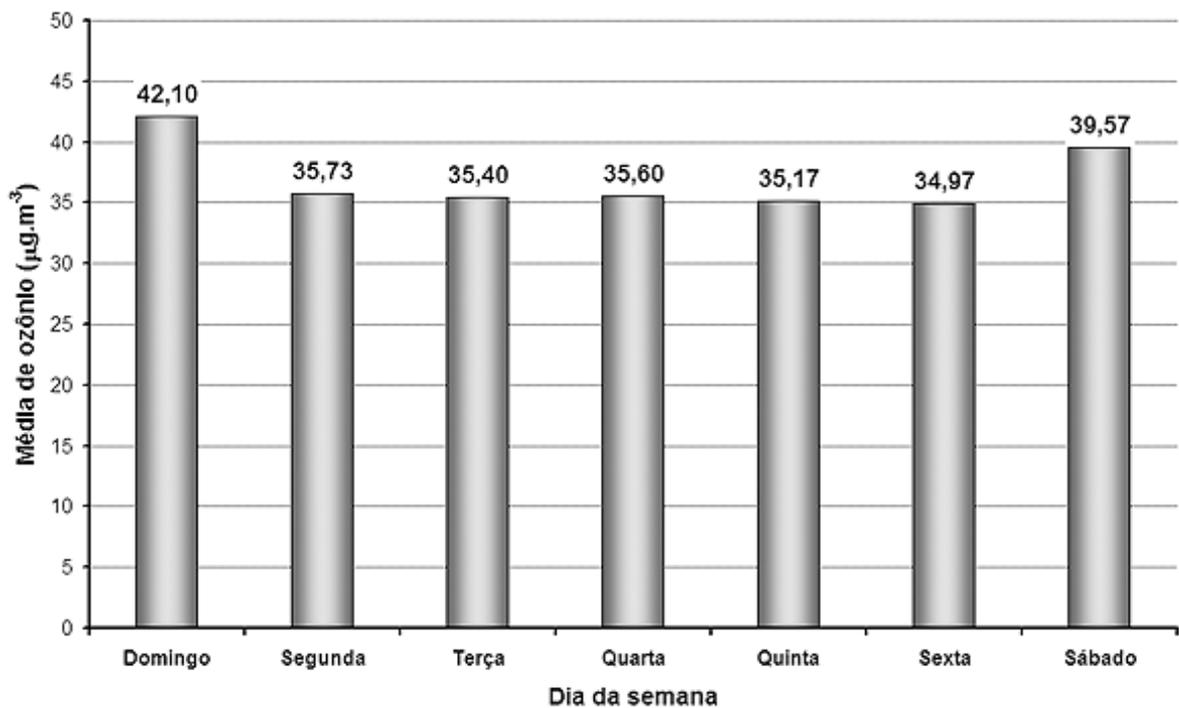


Figura 55: Concentração média de ozônio para cada dia da semana, de 1999 a 2003.

Na comparação entre concentrações médias de ozônio experimental e calculado (figura 56), nota-se uma melhoria em relação ao gráfico obtido no ajuste do item 4.7 (figura 50), pois os pontos estão menos dispersos, com o coeficiente R^2 de 0,83. Para as concentrações máximas de ozônio o resultado também foi melhor (figura 57); os pontos estão mais concisos e o valor de R^2 subiu para 0,78. Na soma absoluta dos pesos da figura 58 as componentes do vento médio de manhã tiveram os maiores pesos, seguidos pela temperatura e umidade relativa à tarde. Apesar das melhorias trazidas ao ajuste, o indicador do dia da semana foi a variável de menor peso, com valor próximo ao do termo independente (*bias*).

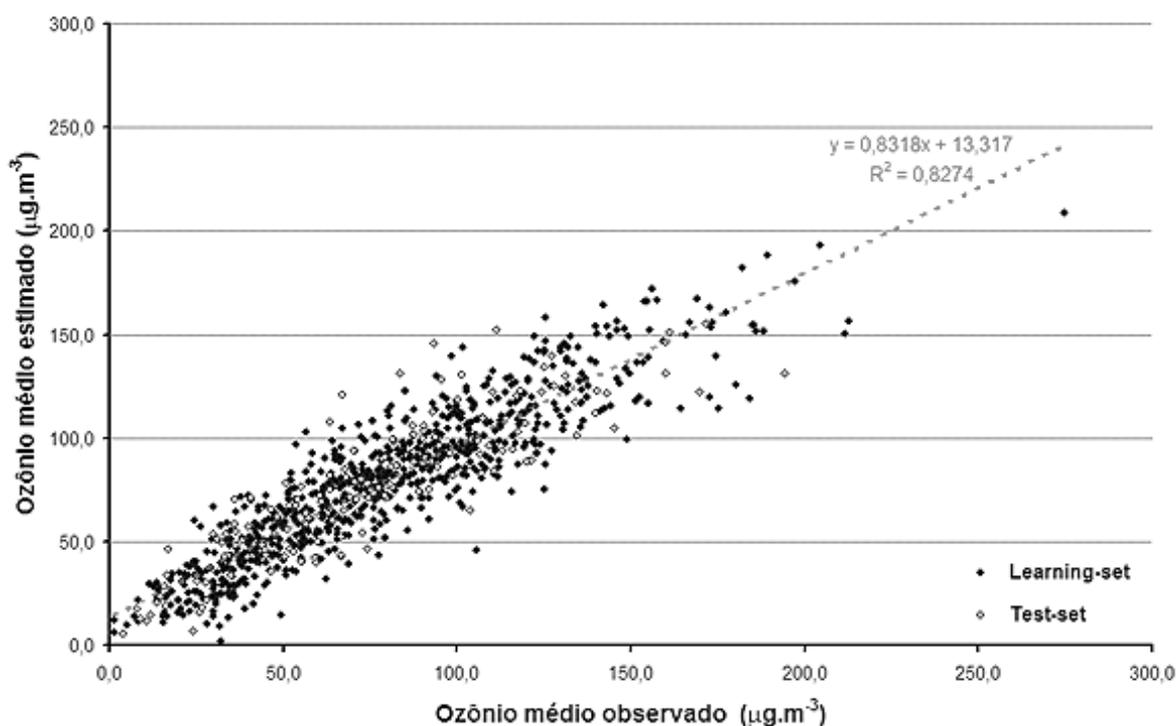


Figura 56: Relação entre valores calculados e medidos da concentração média de ozônio.

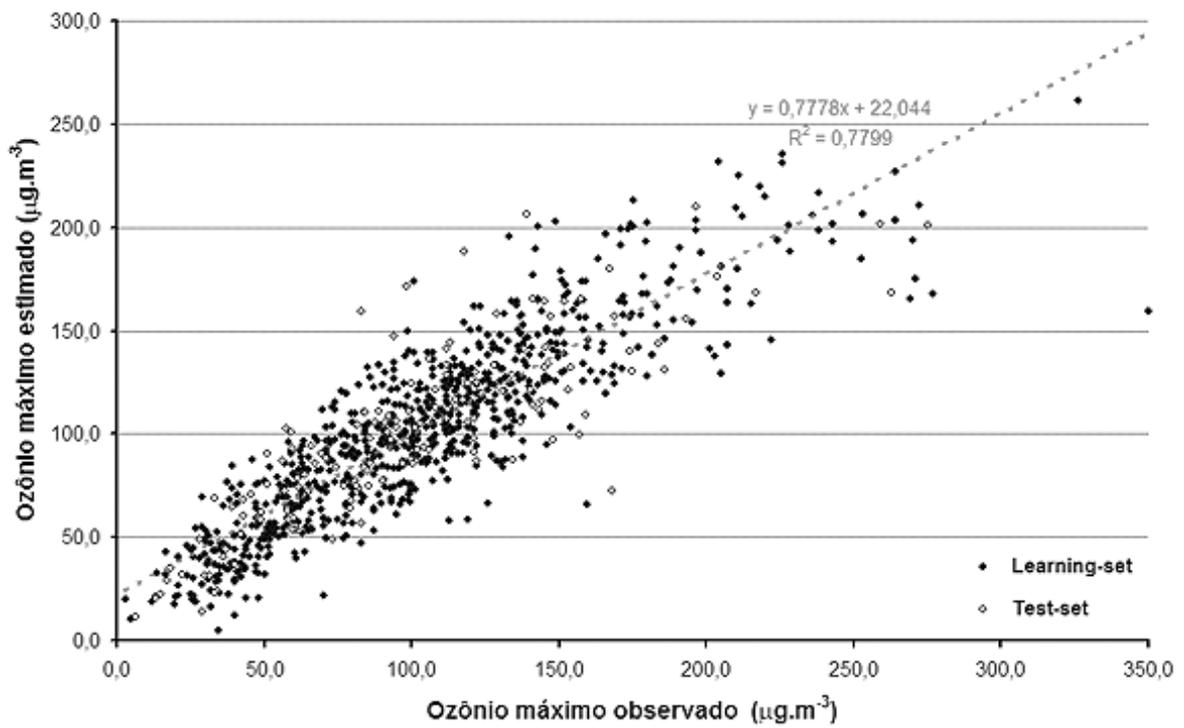


Figura 57: Relação entre valores calculados e medidos da concentração máxima de ozônio.

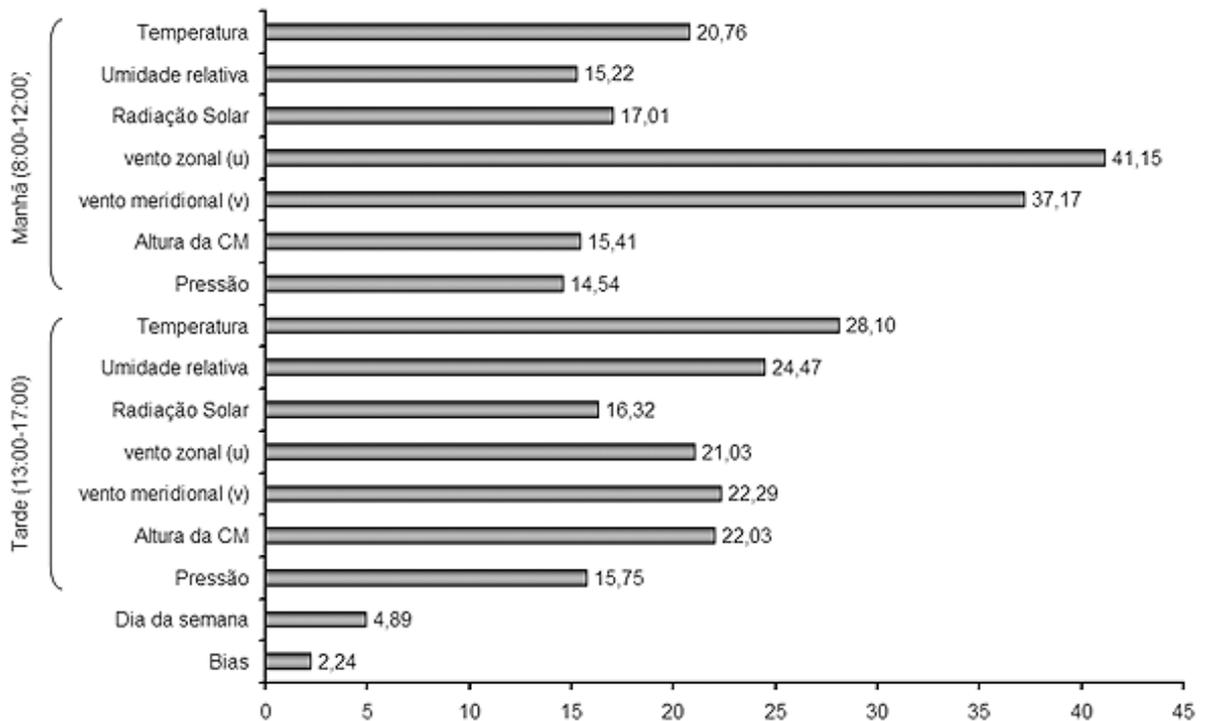


Figura 58: Soma dos valores absolutos dos pesos entre cada variável de entrada e os neurônios da camada oculta.

Comparando-se as concentrações médias de ozônio previsto com o medido (figura 59), observa-se uma pequena melhoria em relação ao teste anterior, identificado pelo coeficiente de determinação. Para as concentrações máximas de ozônio (figura 60), também houve melhora. É possível notar que alguns pontos ficaram mais próximos do eixo de 45°, e o valor de R^2 subiu um pouco. Pela soma absoluta dos pesos (figura 61), as componentes do vento de manhã também foram as variáveis mais significativas, seguidas pela componente meridional do vento e a temperatura média à tarde. A variável binária, indicadora de dia útil ou final de semana, aparece com o menor peso.

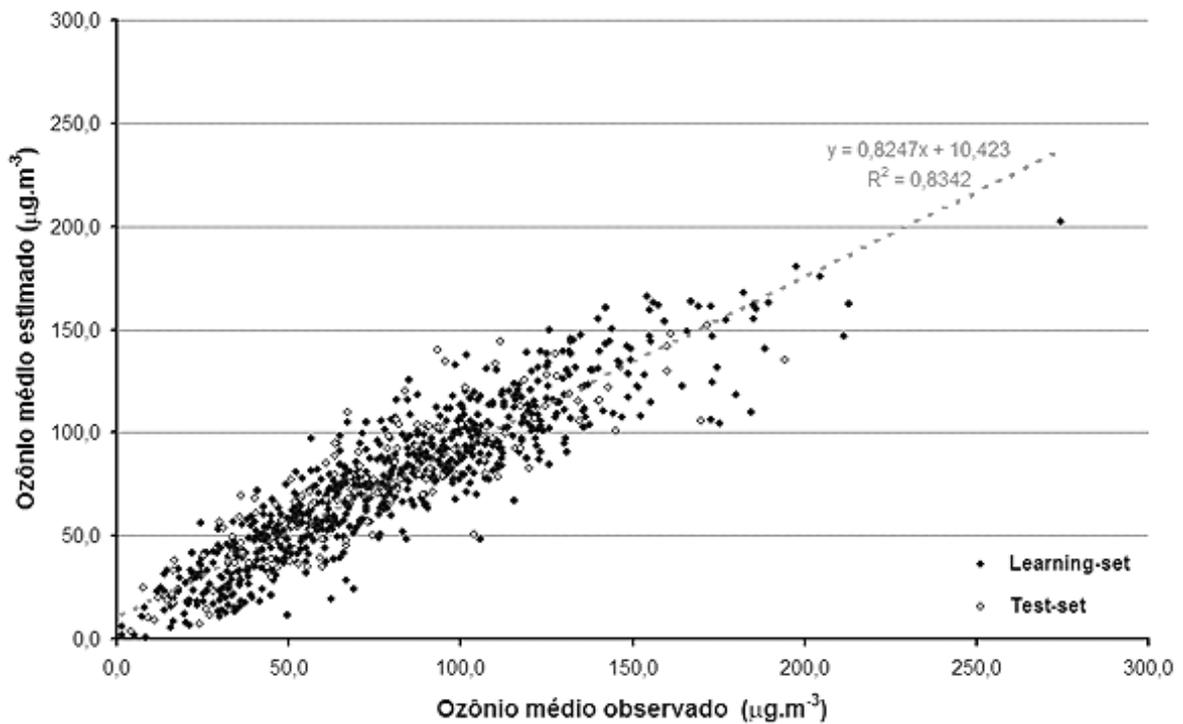


Figura 59: Relação entre valores calculados e medidos da concentração média de ozônio.

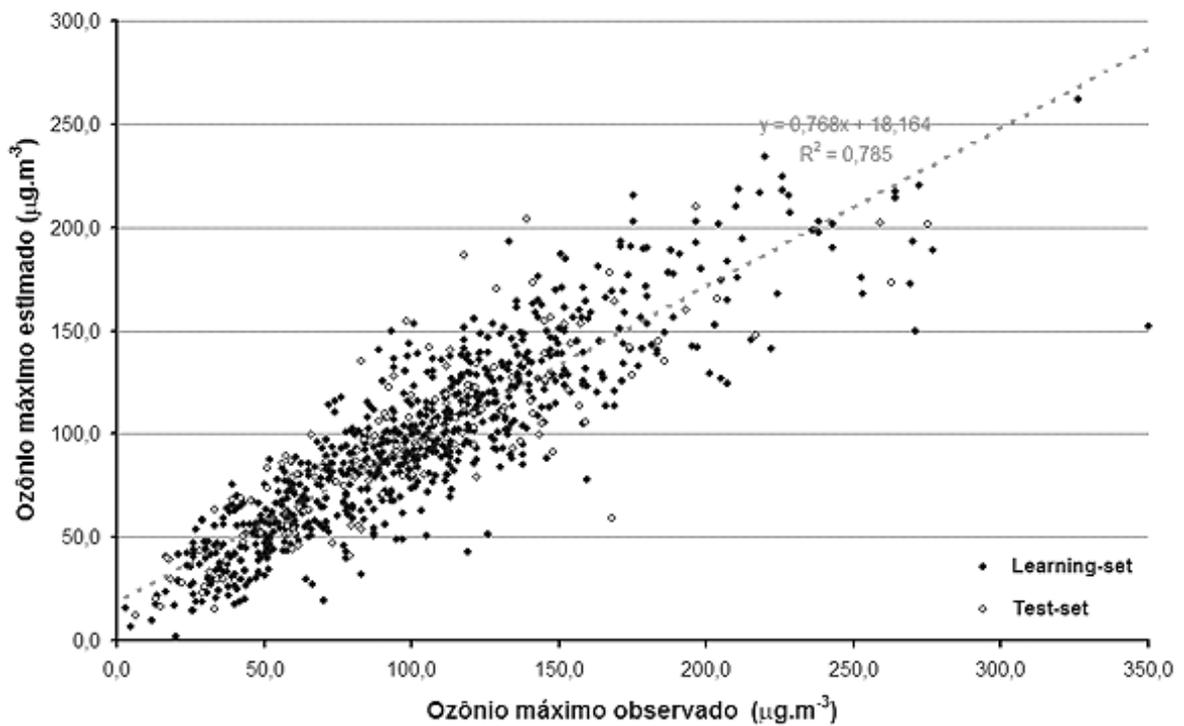


Figura 60: Relação entre valores calculados e medidos da concentração máxima de ozônio.

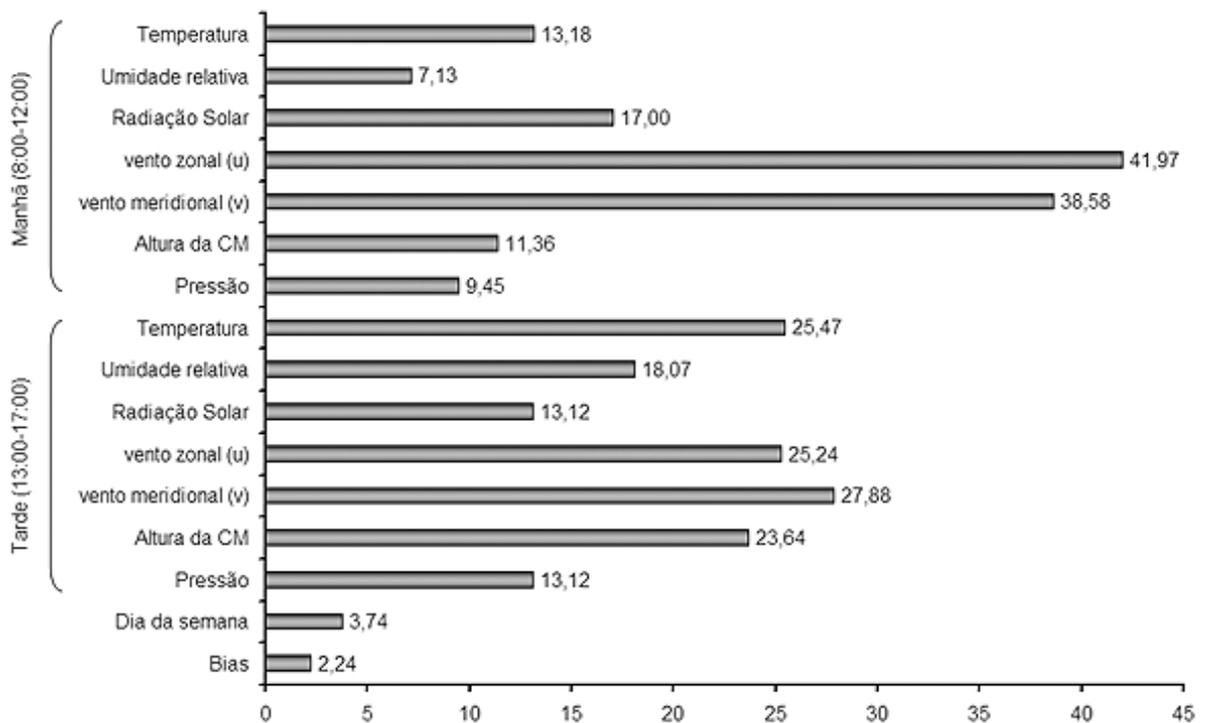


Figura 61: Soma dos valores absolutos dos pesos entre cada variável de entrada e os neurônios da camada oculta.

Dessa forma, a inclusão da variável indicadora de dias úteis e finais de semana contribuiu para a melhoria do ajuste da rede, sem adicionar um grau de incerteza (fonte de erros) à entrada da rede neural. Sua inclusão leva em consideração o fato de que o perfil temporal de geração de precursores de ozônio ser diferente nos finais de semana, nos quais o perfil temporal de densidade de tráfego de veículos na RMSF é diferente dos demais dias. Entretanto, a melhoria foi pequena; para as concentrações médias de ozônio foi de 1,0% e para as máximas foi de 1,4%. O peso indicado para essa variável na figura 61 mostra que ela contribui marginalmente para a qualidade dos resultados.

4.11- Ajustes finos

Após terem sido definidas as variáveis de entrada do modelo em estudo, foram feitos ajustes finos na configuração da rede neural artificial. Sendo assim, parâmetros da estrutura interna e do processo de ajuste (aprendizado) foram variados, no intuito de se obter a melhor qualidade de resultados, tais como: quantidade de neurônios na camada oculta, número de apresentações dos dados à rede, valores do fator de amortecimento (η) e limites de normalização da função de ativação (função sigmóide).

O primeiro parâmetro explorado foi o número de neurônios na camada oculta. Foram feitos diversos testes de ajuste, variando esse número de 3 a 16, em 20 mil apresentações dos dados à rede. Os resultados são apresentados nas figuras 62 (ozônio médio) e 63 (ozônio máximo), por meio da comparação entre os coeficientes de determinação (R^2) obtidos tanto para os casos no conjunto de aprendizado como no de treinamento. De acordo com os gráficos, a melhor configuração de rede neural é a que utiliza oito neurônios na camada oculta, pois apresenta o maior índice para o conjunto de treinamento sem prejuízo ao de teste. Abaixo de oito neurônios os valores de R^2 são inferiores para os dois conjuntos; acima, os de

treinamento apresentam valores maiores, porém, com perda de qualidade no de teste, indicando sobreajuste da rede.

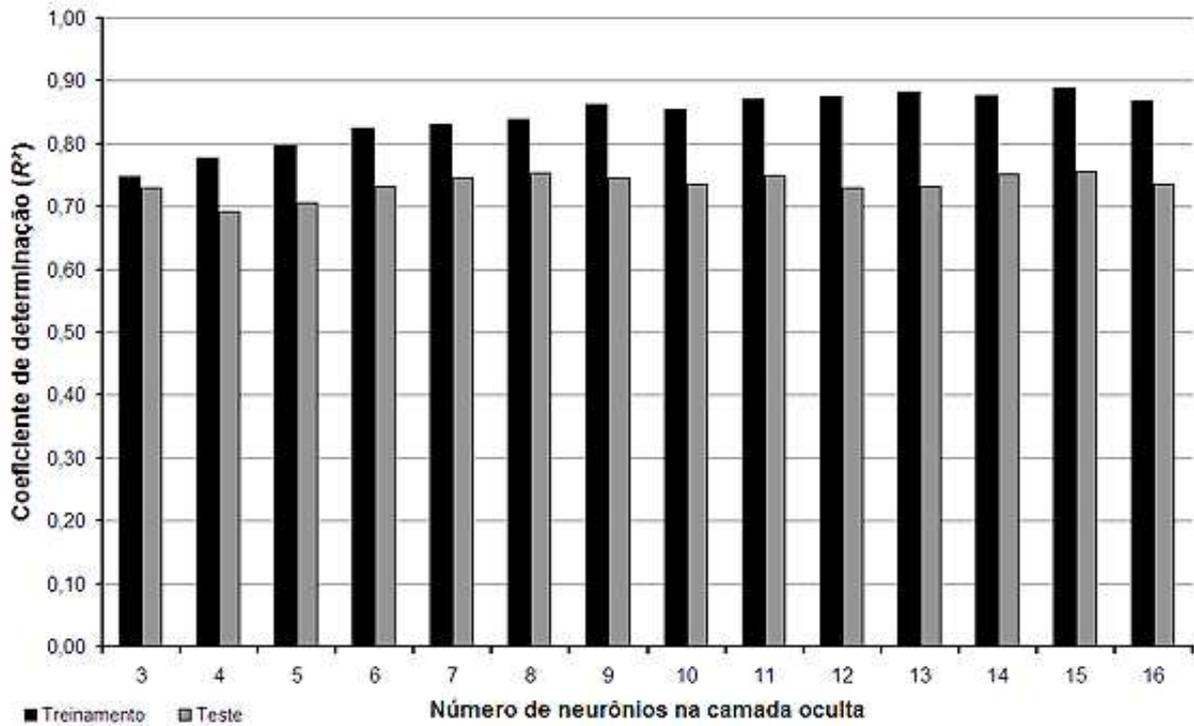


Figura 62: Coeficientes de determinação (R^2) obtidos nos conjuntos de aprendizado e de teste para diferentes configurações de redes neurais, variando o número de neurônios na camada oculta. Resultados referentes à concentração média de ozônio.

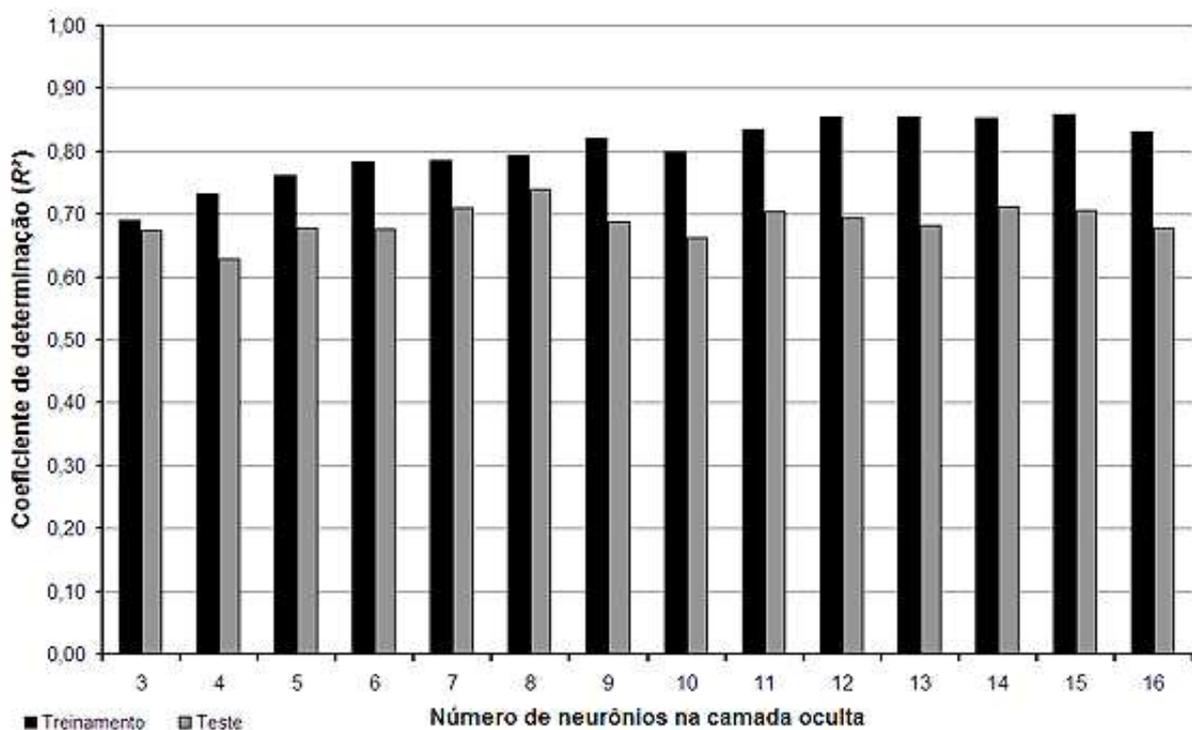


Figura 63: Coeficientes de determinação (R^2) obtidos nos conjuntos de aprendizado e de teste para diferentes configurações de redes neurais, variando o número de neurônios na camada oculta. Resultados referentes à concentração máxima de ozônio.

Em seguida foram avaliados os valores do fator η . Após alguns testes foi escolhida a configuração com η inicial e mínimo de 0,1 e máximo em 1,11, o que levou a coeficientes R^2 de 0,84 e 0,79, para concentrações médias e máximas de ozônio, respectivamente, em relação aos dados do conjunto de aprendizado. Também foram determinados os limites de normalização para a função de ativação, que no final ficaram entre 0,2 e 0,8. Foi observado também que os resultados ficaram melhores quando os dados eram apresentados 40 mil vezes à rede neural.

4.12 - Validação

Conforme descrito anteriormente (item 4.1), foram feitos diversos testes para se determinar quais variáveis são mais adequadas à entrada do modelo de rede neural artificial, com o objetivo de estudar a previsibilidade do ozônio troposférico. Dessa forma, as variáveis

meteorológicas escolhidas foram: temperatura do ar, umidade relativa, radiação solar, campo de vento em 50 m (expresso nas componentes zonal e meridional), altura da camada de mistura e pressão atmosférica. Para cada uma das variáveis foram obtidas médias representativas do período da manhã (08:00-12:00h) e tarde (13:00-17:00h). Além destas, foi inserida à entrada uma informação binária quanto ao dia da semana (0- Seg. a Sex.; 1- Sab. e Dom.). Como saída, o modelo obtém estimativas de concentração média e máxima de ozônio para o período da tarde. Em seguida foram feitos ajustes de parâmetros internos da rede, o que levou a uma RNA configurada com 8 neurônios na camada oculta e limites da função de ativação entre 0,2 e 0,8. Essa rede foi ajustada com η (fator de amortecimento) com valor mínimo e inicial de 0,1 e máximo de 1,11 e com o conjunto de dados de treinamento sendo apresentados 40 mil vezes à rede. No processo de treinamento foram considerados os dados de 1999 a 2003 da estação de monitoramento da Cetesb e seu SODAR, localizados no Parque do Ibirapuera.

Para a validação dessa rede foram utilizadas as medições feitas no ano de 2004, o qual não esteve presente no conjunto de treinamento e nem no de teste do modelo. Este ano foi o último em que o SODAR esteve em operação no Ibirapuera. Ao todo foram 231 eventos considerados.

Os resultados obtidos da aplicação desses dados ao modelo são apresentados a seguir. Nas figuras 64 e 65 são feitas comparações entre os valores estimados e previstos, respectivamente, para concentrações médias e máximas de ozônio no decorrer do ano. Observa-se que a curva das estimativas acompanha o perfil das medições, embora alguns eventos de baixas concentrações tenham sido superestimados e outros de alta subestimados. Na comparação entre dados experimentais e calculados de concentrações médias de ozônio (figura 66), foi obtido um coeficiente de determinação (R^2) igual a 0,71; para as concentrações máximas de ozônio, esse índice foi um pouco menor: 0,69 (figura 67). Os resultados estão

adequados, uma vez que os valores obtidos durante o treinamento da rede foram de 0,84 para concentrações médias de ozônio e 0,79 para as máximas; além disso, os dados aparecem de forma pouco dispersa em relação ao eixo de 45°.

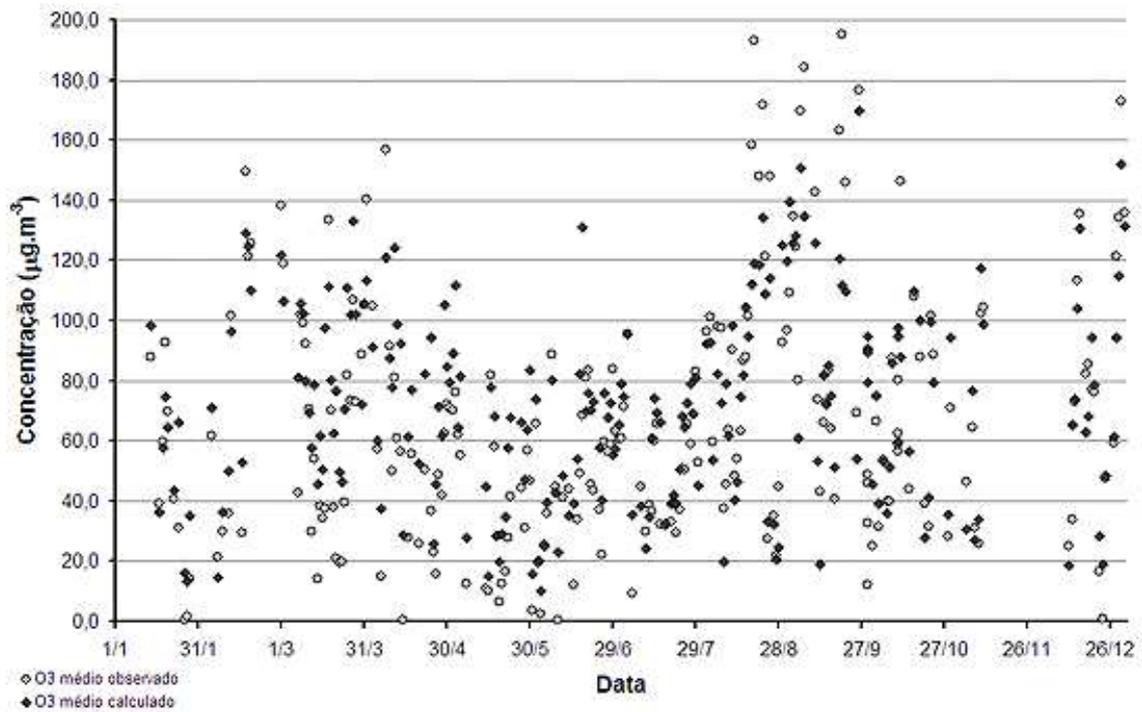


Figura 64: Evolução dos valores previstos e observados das concentrações médias de ozônio ao longo do ano de 2004.

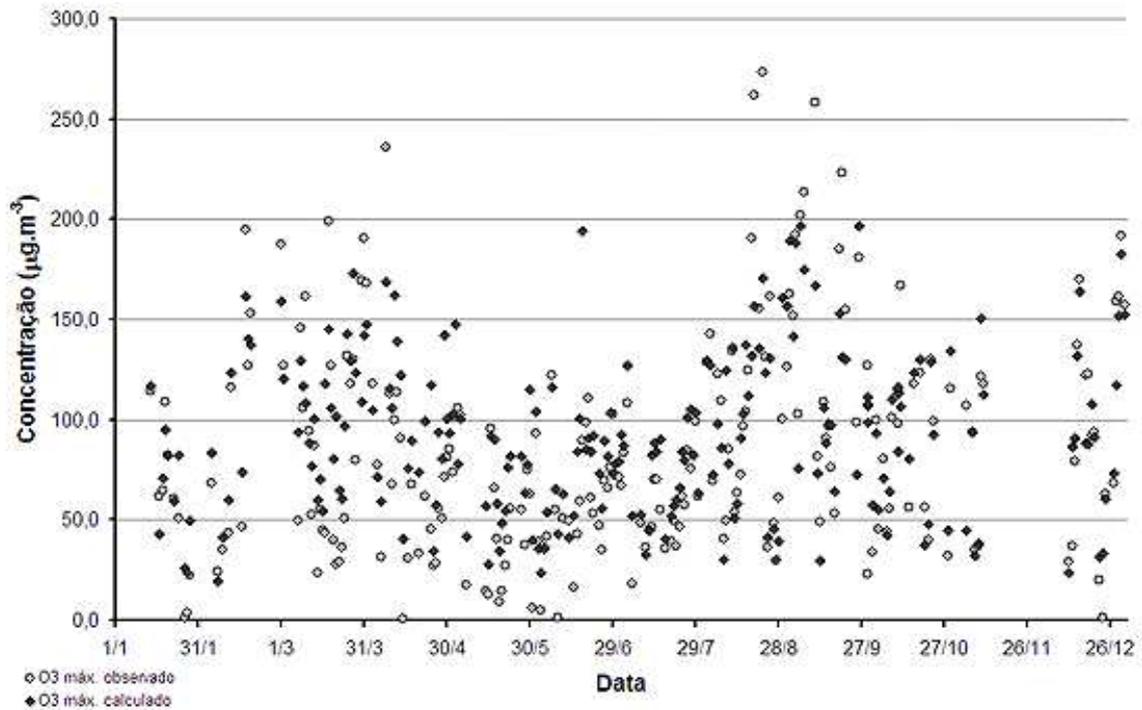


Figura 65: Evolução dos valores previstos e observados das concentrações máximas de ozônio ao longo do ano de 2004.

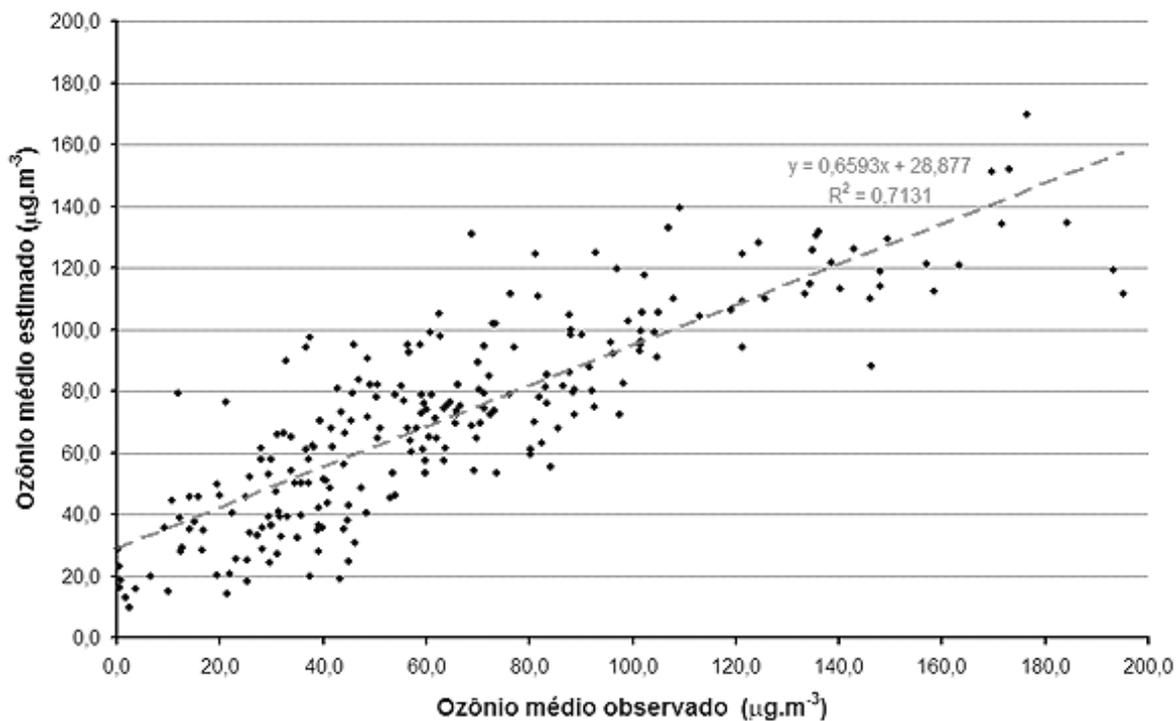


Figura 66: Relação entre valores calculados e medidos da concentração média de ozônio em 2004.

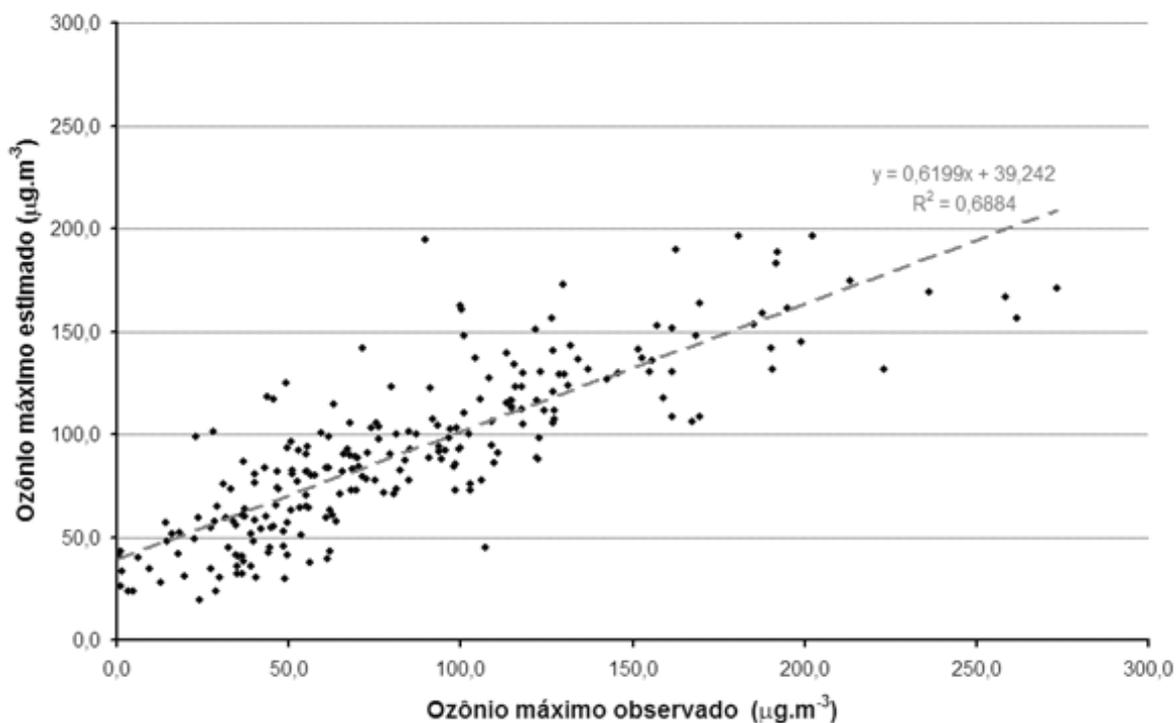


Figura 67: Relação entre valores calculados e medidos da concentração máxima de ozônio em 2004.

Para uma avaliação quantitativa mais abrangente foram adotados alguns parâmetros estatísticos, seguindo um procedimento similar ao aplicado por BORGES (2003) e por FREITAS (2003).

Inicialmente foi feito um cálculo das correlações entre observações e resultados das simulações para o grau de associação linear existente. O método utilizado foi a correlação linear de Pearson (equação 15), dada pela razão entre a covariância existente entre duas variáveis e o produto de seus desvios padrões.

$$r_{oy} = \frac{\text{cov}(o, y)}{\sigma_o \sigma_y} = \frac{\frac{1}{n-1} \sum_{i=1}^n [(o_i - \bar{o})(y_i - \bar{y})]}{\left\{ \left[\frac{1}{n-1} \sum_{i=1}^n (o_i - \bar{o})^2 \right]^{1/2} \right\} \left\{ \left[\frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 \right]^{1/2} \right\}} \quad (15)$$

Em que o é o valor medido e y o simulado.

Para avaliar os desvios que possam estar presentes nos resultados simulados e estimar a acurácia das simulações, foi calculado o erro médio, ME (equação 16), que é a diferença entre a média das simulações e a média das observações, expressando assim o desvio da simulação. Simulações que em média são maiores que as observações apresentam ME positivo, e simulações que em média são menores que as observações apresentam ME negativo.

$$ME = \frac{1}{n} \sum_{k=1}^n (y_k - o_k) = \bar{y} - \bar{o} \quad (16)$$

Como o erro médio é representativo de toda a série de simulações, este não fornece informações sobre erros individuais, não representando uma medida de acurácia. Para se ter uma medida de acurácia há o erro médio quadrático (MSE), que é a média quadrática das

diferenças entre pares de simulações e observações. Esse valor indica a magnitude típica dos erros das simulações; quanto maior o seu valor, maior a discrepância entre simulação e observação. Para os cálculos foi utilizada sua raiz quadrada (*RMSE*, conforme a equação 17), que tem as mesmas dimensões das previsões e observações, podendo ser considerada como a magnitude típica dos erros das simulações.

$$RMSE = \sqrt{\frac{1}{n} \sum_{k=1}^n (y_k - o_k)^2} \quad (17)$$

FREITAS (2003) apresenta um outro parâmetro (*RMSE_{UB}*) para avaliar a "habilidade" (*skill*) de uma simulação, representado pela raiz do erro médio quadrático após a remoção de um desvio médio (equação 18).

$$RMSE_{UB} = \left\{ \frac{1}{N} \sum_{k=1}^N \left[(y_k - \bar{y}) - (o_k - \bar{o}) \right]^2 \right\}^{1/2} \quad (18)$$

Na tabela 4.12.1 são apresentados os valores médios, máximos, mínimos e os desvios padrões referentes às medições da Cetesb e as estimativas da rede neural. Na tabela 4.12.2 estão os resultados dos cálculos de correlação, erro médio, raiz do erro médio quadrático e do parâmetro *RMSE_{ub}*. Observa-se que, em média, a simulação superestima as concentrações de ozônio em 6,7 µg.m⁻³, de acordo com *ME*. Pelos índices *RMSE* e *RMSE_{UB}*, que indicam a magnitude do erro e desempenho da simulação, a rede erra em média 22,8 µg.m⁻³ nas concentrações médias de ozônio e 30,4 µg.m⁻³ para as máximas.

Tabela 4.12.1: Caracterização das medidas e simulações para concentrações médias e máximas de ozônio (em $\mu\text{g.m}^{-3}$).

	O3 médio		O3 máximo	
	Obs.	Sim.	Obs.	Sim.
Média	65,41	72,00	85,37	92,16
σ (desvio)	41,54	32,43	53,32	39,84
Máximo	195,03	169,58	273,55	196,57
Mínimo	0,29	9,88	0,73	19,43

Tabela 4.12.2: Critérios de avaliação do desempenho da rede neural. *ME*: Erro médio ($\mu\text{g.m}^{-3}$); *RMSE*: Raiz do erro médio quadrático ($\mu\text{g.m}^{-3}$); *RMSE_{UB}*: Aptidão da simulação ($\mu\text{g.m}^{-3}$).

	O3 médio	O3 máximo
Correlação	0,84	0,83
ME	6,59	6,80
RMSE	23,31	30,78
RMSE_{ub}	22,36	30,02

Para cada evento foi calculado um termo de desvio, representado pela diferença entre o valor estimado pela rede neural e o medido pela estação, para as concentrações médias e máximas de ozônio. A figura 68 mostra como essa diferença varia em função das concentrações calculadas e observadas. Em relação às concentrações médias de ozônio obtidas pela RNA, é possível notar um aumento dos desvios por volta de $80 \mu\text{g.m}^{-3}$; isso também é observado com as concentrações máximas de ozônio, porém, com uma dispersão maior. Em relação às observações, os desvios assumem valores preferencialmente negativos a partir de $100 \mu\text{g.m}^{-3}$ para as médias de ozônio, e a partir de $150 \mu\text{g.m}^{-3}$ para as máximas, indicando uma tendência da rede neural em subestimar as concentrações mais elevadas do poluente. Como a incidência de eventos mais críticos de ozônio é bem menos freqüente que os de baixa concentração, o ajuste da rede neural não dispôs de muitos casos que permitissem uma melhor caracterização desses níveis mais elevados. Conforme descrito anteriormente (item 4.9), foram buscadas alternativas para se contornar o problema. Porém, a qualidade do ajuste depende da disponibilidade de mais dados nessa faixa de concentração.

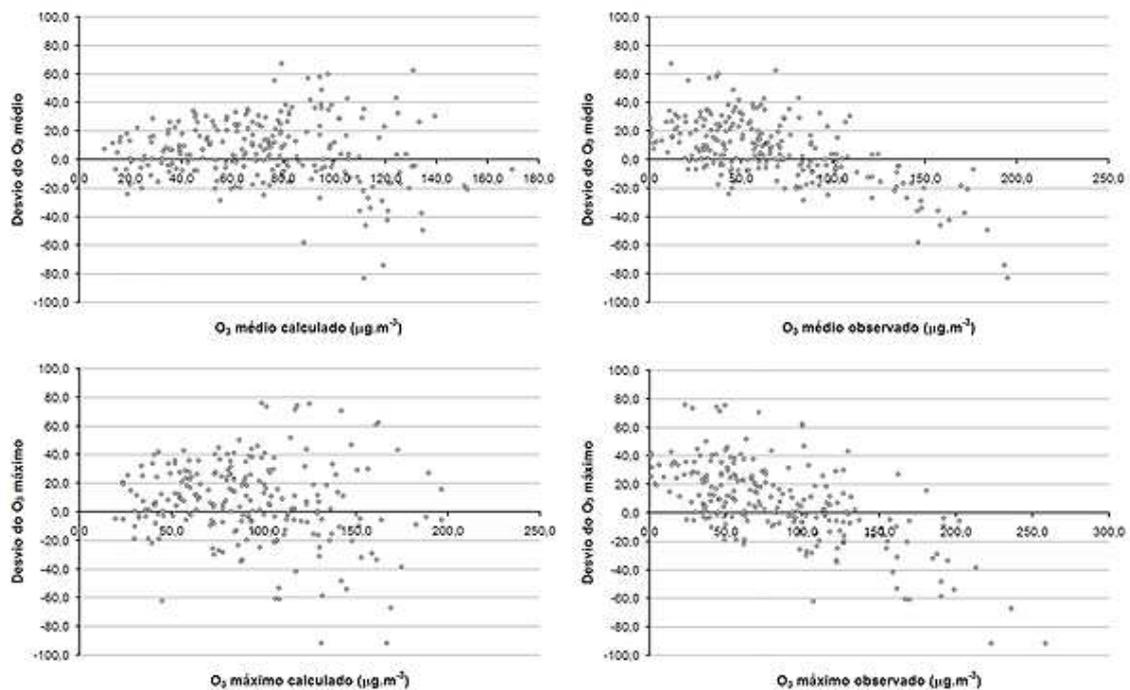


Figura 68: Relação entre os desvios (diferença entre valores calculados e observados) e as concentrações médias e máximas de ozônio.

Os desvios entre simulações e medições também foram comparados com as variáveis meteorológicas de entrada. Na relação entre temperatura e o desvio das concentrações médias de ozônio (figura 69) as diferenças tendem a aumentar com a elevação da temperatura, especialmente acima de 20 °C pela manhã, e 23 °C à tarde. No gráfico do desvio das concentrações máximas de ozônio (figura 70), as diferenças entre estimativa e observação são maiores de uma maneira geral, mas são mais significativas a partir de temperaturas em torno de 20 °C. No entanto, quando as temperaturas chegam próximas a 30°C, os desvios retornam a um patamar semelhante ao observado para temperaturas abaixo de 20°C: entre -40 a 0 $\mu\text{g}\cdot\text{m}^{-3}$ nas médias de ozônio e -30 a 20 $\mu\text{g}\cdot\text{m}^{-3}$ nas máximas.

Conforme visto anteriormente (item 3.2.2) as altas concentrações de ozônio estão associadas às temperaturas em torno de 22°C a 33°C (dados: 1999-2004), e o modelo não é capaz de prevê-las com a mesma exatidão com que faz para as mais baixas, logo são esperados erros maiores dentro dessa faixa de temperatura. Para temperaturas suficientemente

altas (próximas a 30°C) o modelo é melhor, pois nesse ponto já se configura uma condição claramente propícia à formação de ozônio.

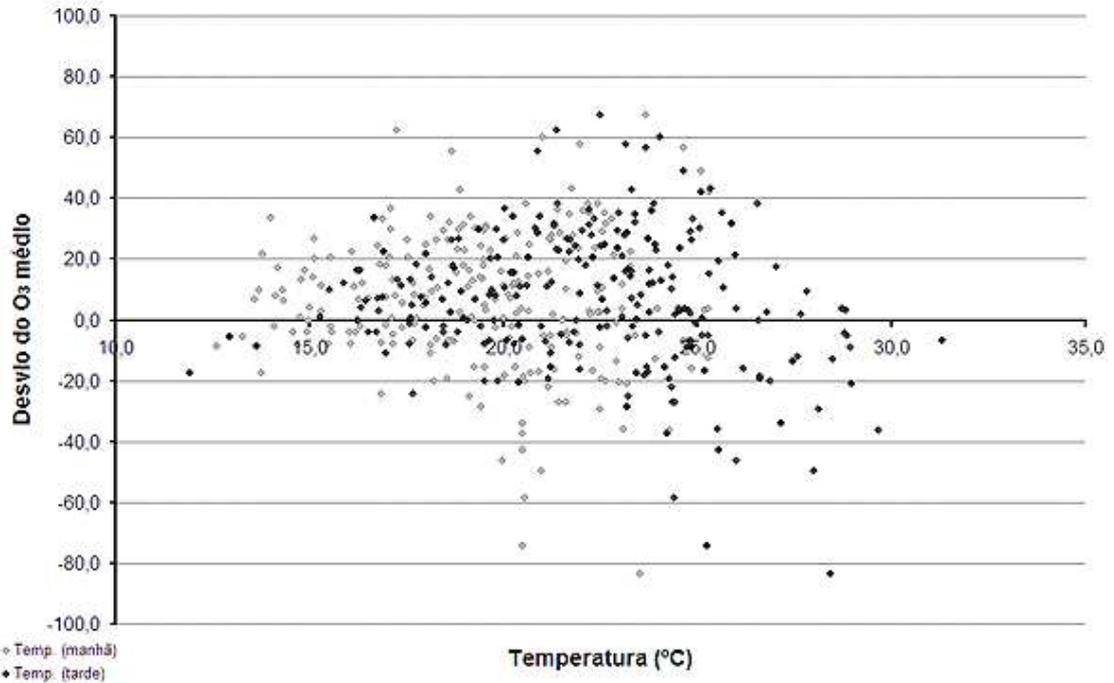


Figura 69: Relação entre os desvios (simulação-observação) das concentrações médias de ozônio e a temperatura do ar de manhã e de tarde.

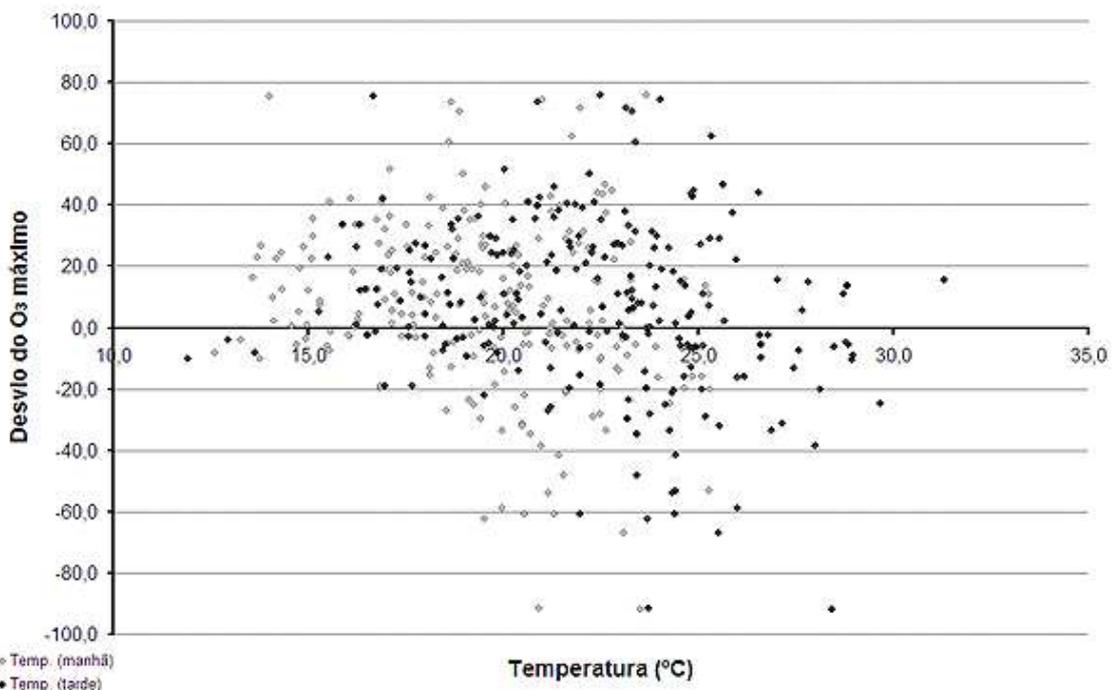


Figura 70: Relação entre os desvios (simulação-observação) das concentrações máximas de ozônio e a temperatura do ar de manhã e de tarde.

Para a umidade relativa os desvios observados nas concentrações médias de ozônio (figura 71) mostram-se mais expressivos abaixo de 50% de UR, especialmente para as médias da tarde; de manhã, índices abaixo de 60% de umidade já apresentam alguns desvios consideráveis (maior que $60 \mu\text{g}\cdot\text{m}^{-3}$). Nos desvios de concentrações máximas de ozônio (figura 72) segue-se um padrão semelhante, mas com desvios maiores.

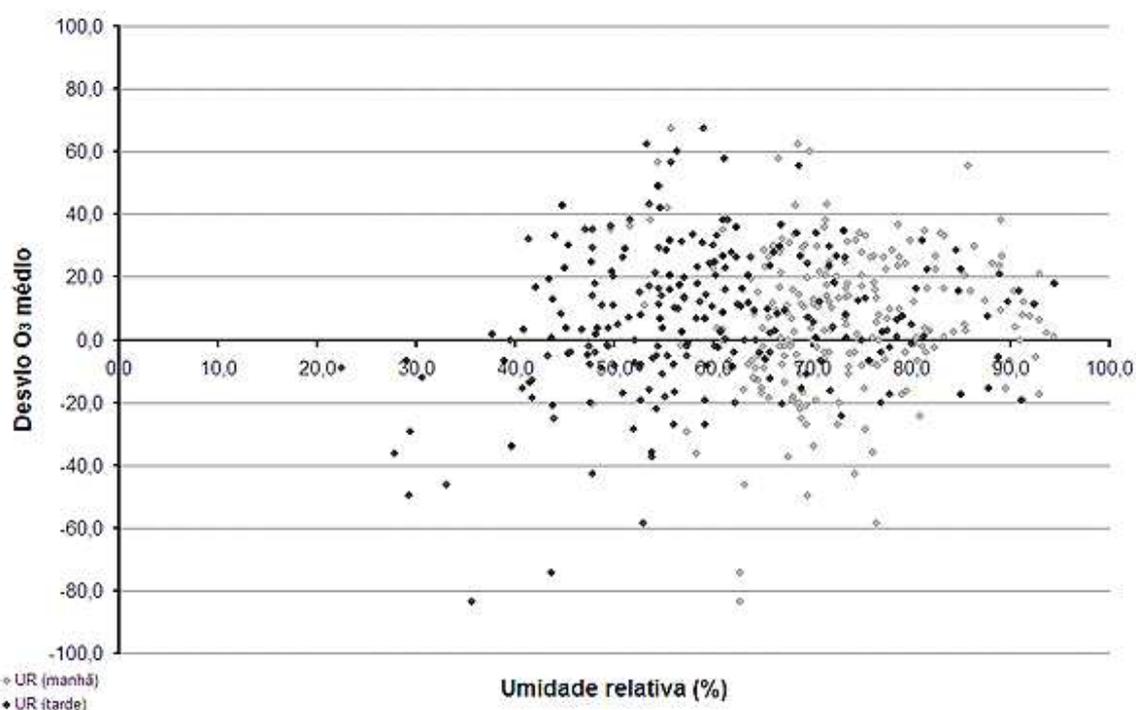


Figura 71: Relação entre os desvios (simulação-observação) das concentrações médias de ozônio e a umidade relativa de manhã e de tarde.

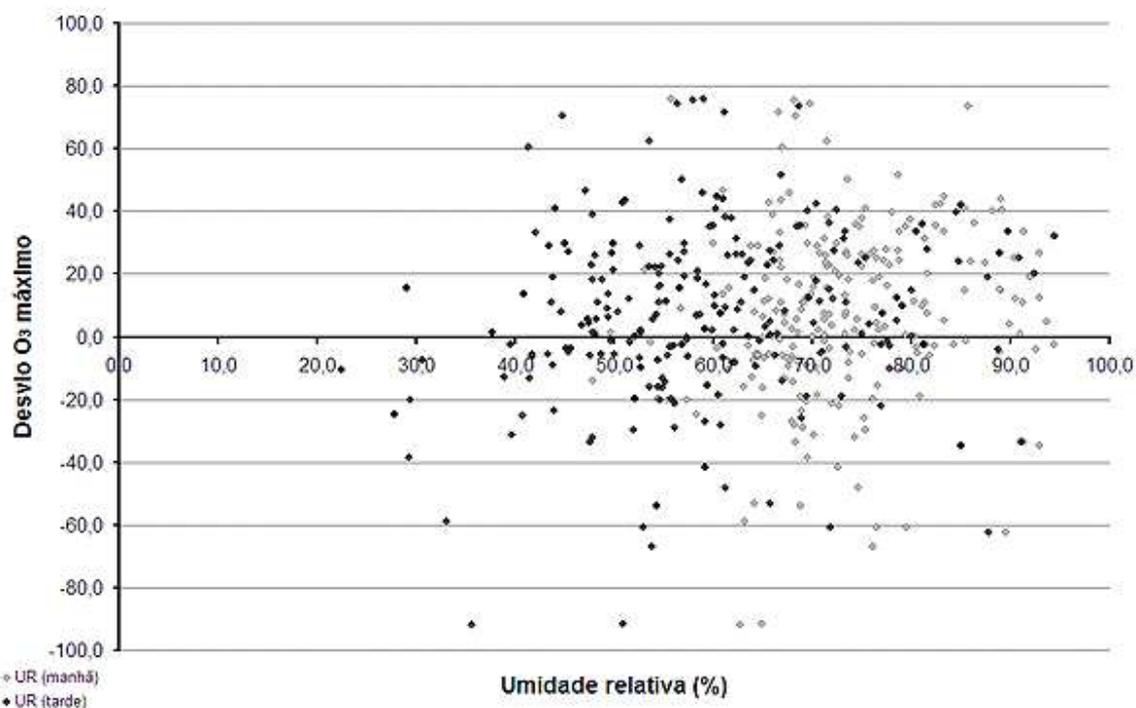


Figura 72: Relação entre os desvios (simulação-observação) das concentrações máximas de ozônio e a umidade relativa de manhã e de tarde.

Em relação à radiação solar, os desvios da concentração média do ozônio (figura 73) se mantêm em -20 a $40 \mu\text{g.m}^{-3}$ até por volta de 200 W.m^{-2} , quando então, eventos de maior discrepância (acima de $60 \mu\text{g.m}^{-3}$, em valores absolutos) começam a surgir, sendo mais significativos a partir de 500 W.m^{-2} . Entretanto, acima de 700 W.m^{-2} , o limite diminui para -25 a $15 \mu\text{g.m}^{-3}$. Para as concentrações máximas de ozônio (figura 74) os desvios são maiores, indo de -60 a $40 \mu\text{g.m}^{-3}$ até os 100 W.m^{-2} , passando então a -60 a $80 \mu\text{g.m}^{-3}$ a partir de 200 W.m^{-2} , aumentando ainda mais após 500 W.m^{-2} , mas caindo para a faixa entre -20 e $20 \mu\text{g.m}^{-3}$ após 700 W.m^{-2} .

Novamente se observa que os desvios são menores para os valores extremos das variáveis. Para níveis de radiação solar baixos (200 W.m^{-2} em relação as concentrações médias de ozônio, ou 100 W.m^{-2} para as máximas) os desvios são menores pois a condição é pouco propícia à formação de ozônio. Para valores mais altos de radiação solar (acima de 700

W.m^{-2}), caracteriza-se condição bem favorável à formação de ozônio, portanto os desvios diminuem.

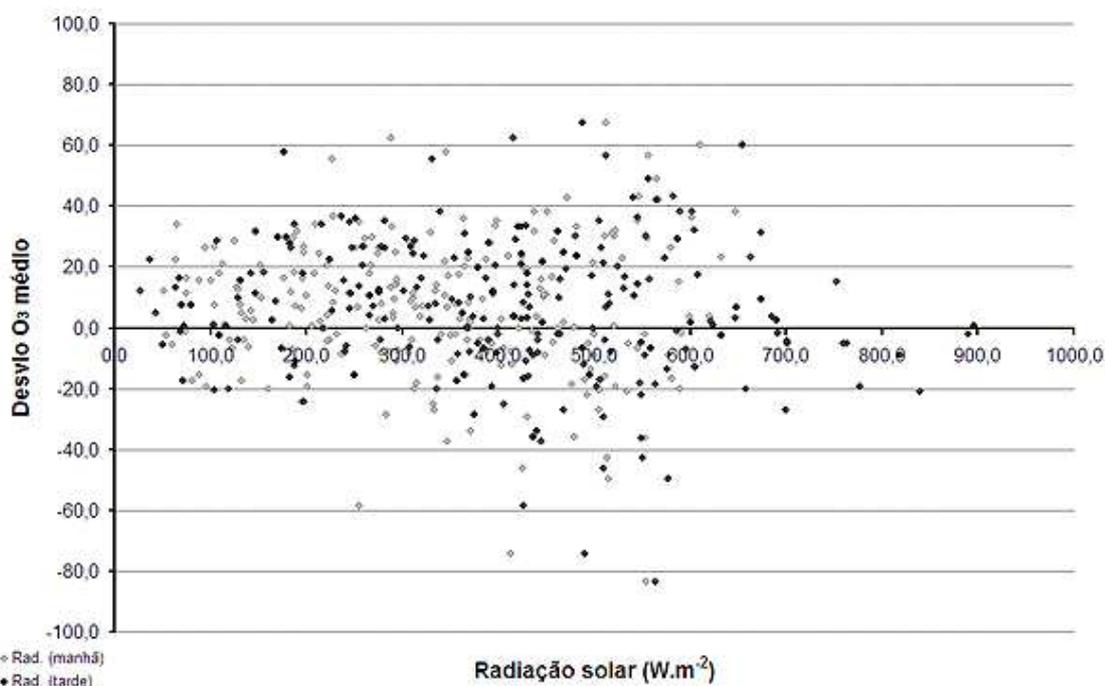


Figura 73: Relação entre os desvios (simulação-observação) das concentrações médias de ozônio e a radiação solar de manhã e de tarde.

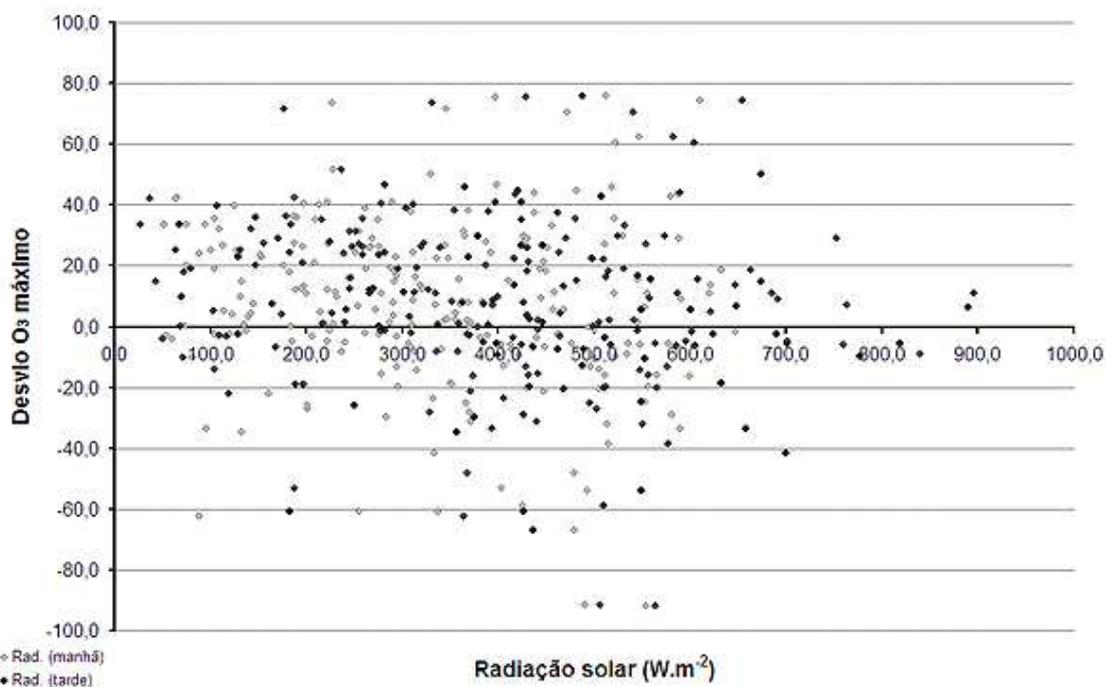


Figura 74: Relação entre os desvios (simulação-observação) das concentrações máximas de ozônio e a radiação solar de manhã e de tarde.

A relação entre o campo de vento de manhã (componentes u e v) e os desvios obtidos para as concentrações médias de ozônio à tarde (figura 75) não indica um padrão definido segundo o qual os desvios tendam a ser maiores ou menores para uma dada orientação e velocidade de vento. Observam-se apenas alguns pontos isolados denotando desvios de diferentes ordens de magnitude. Para a relação feita com os desvios de concentração máxima (figura 76) já é possível identificar um padrão, em que uma faixa circular, aproximadamente entre valores absolutos de 1 e 4 m.s^{-1} , apresenta a maior parte dos desvios representativos.

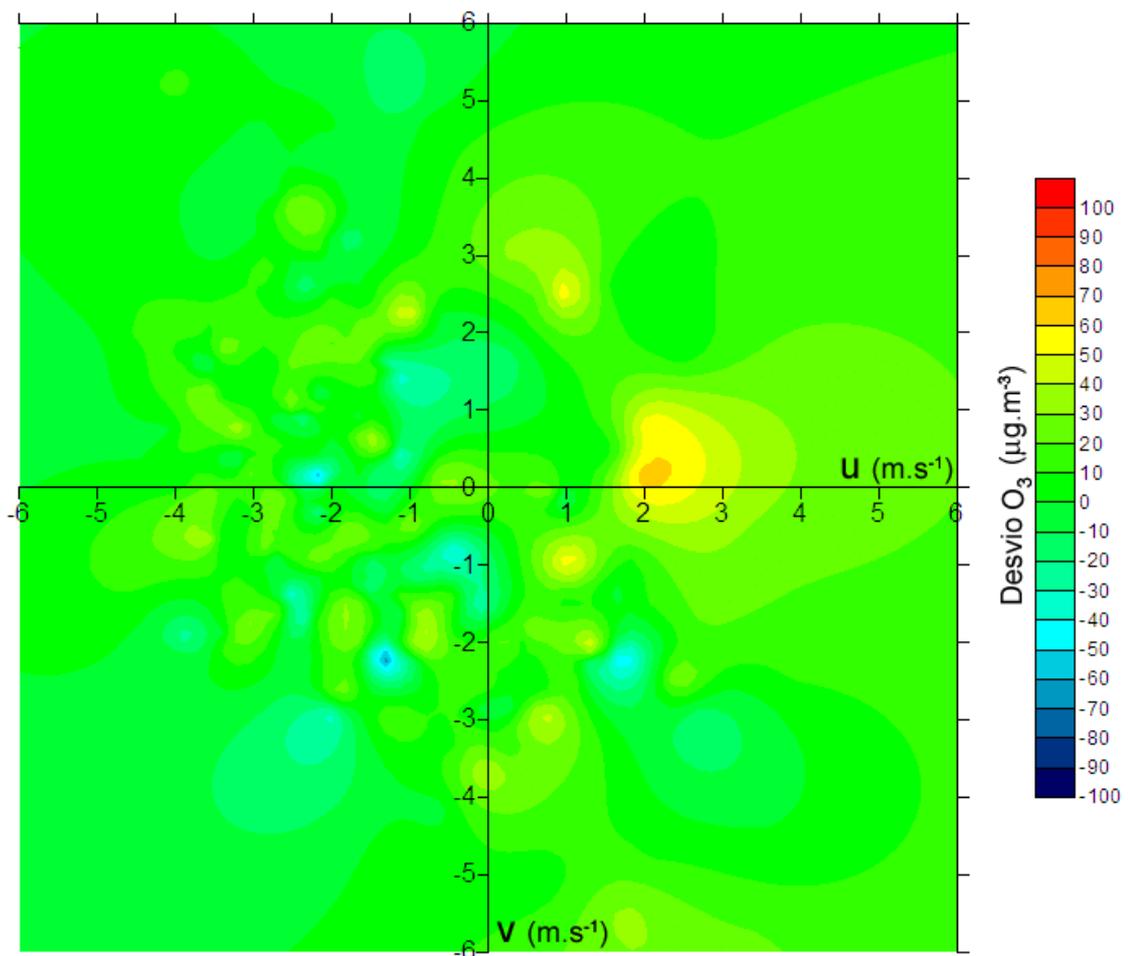


Figura 75: Distribuição dos desvios de concentrações médias de ozônio à tarde de acordo com o campo de vento de manhã (expresso por suas componentes u e v).

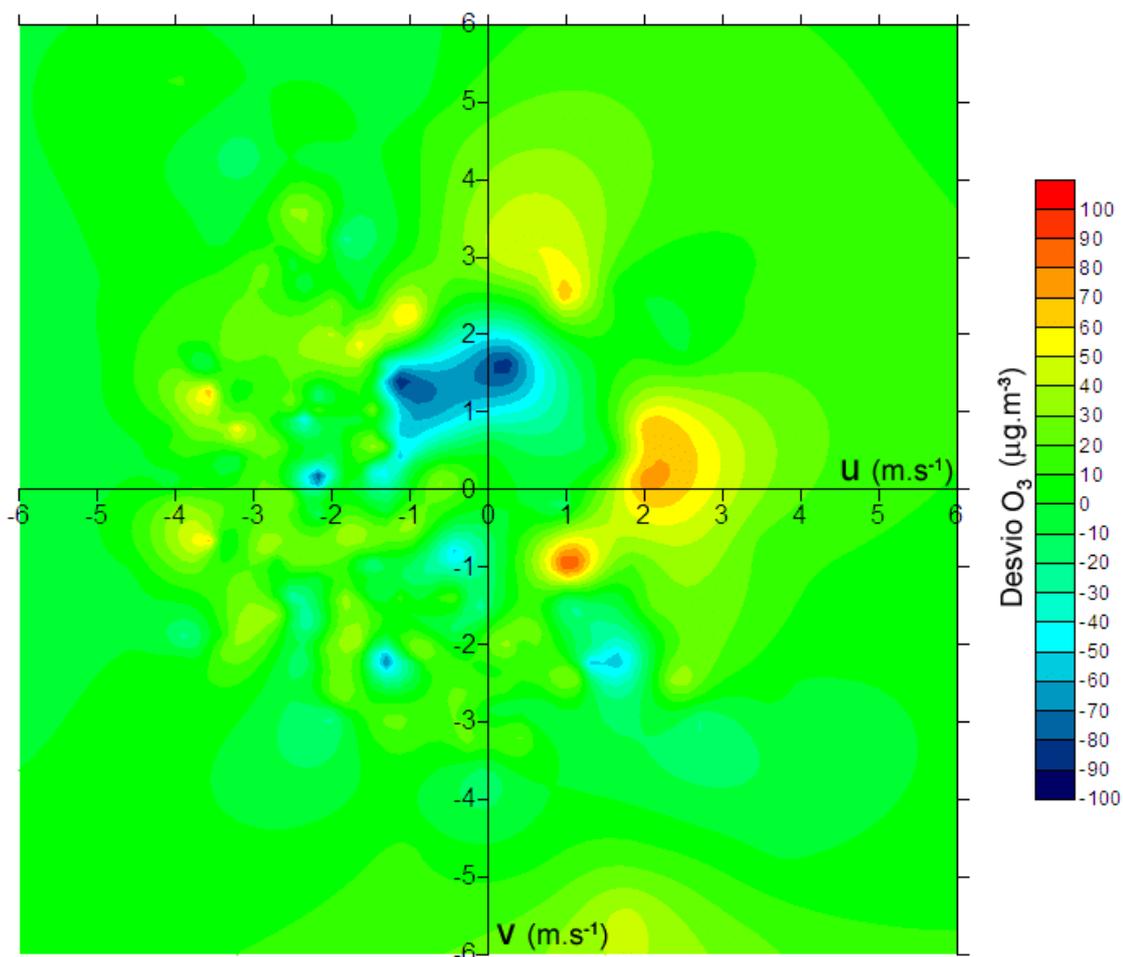


Figura 76: Distribuição dos desvios de concentrações máximas de ozônio à tarde de acordo com o campo de vento da manhã (expresso por suas componentes u e v).

Associando o desvio das concentrações médias de ozônio com o campo de vento à tarde (figura 77), observa-se que a maior parte dos desvios mais significativos ocorre preferencialmente para componentes do vento abaixo de 3 m.s^{-1} (em valor absoluto). Desvios marcantes para componentes acima dessa ordem ocorrem para ventos de sul e de noroeste. Esse perfil é mais notável no gráfico dos desvios das concentrações máximas de ozônio (figura 78), em que as diferenças são mais representativas.

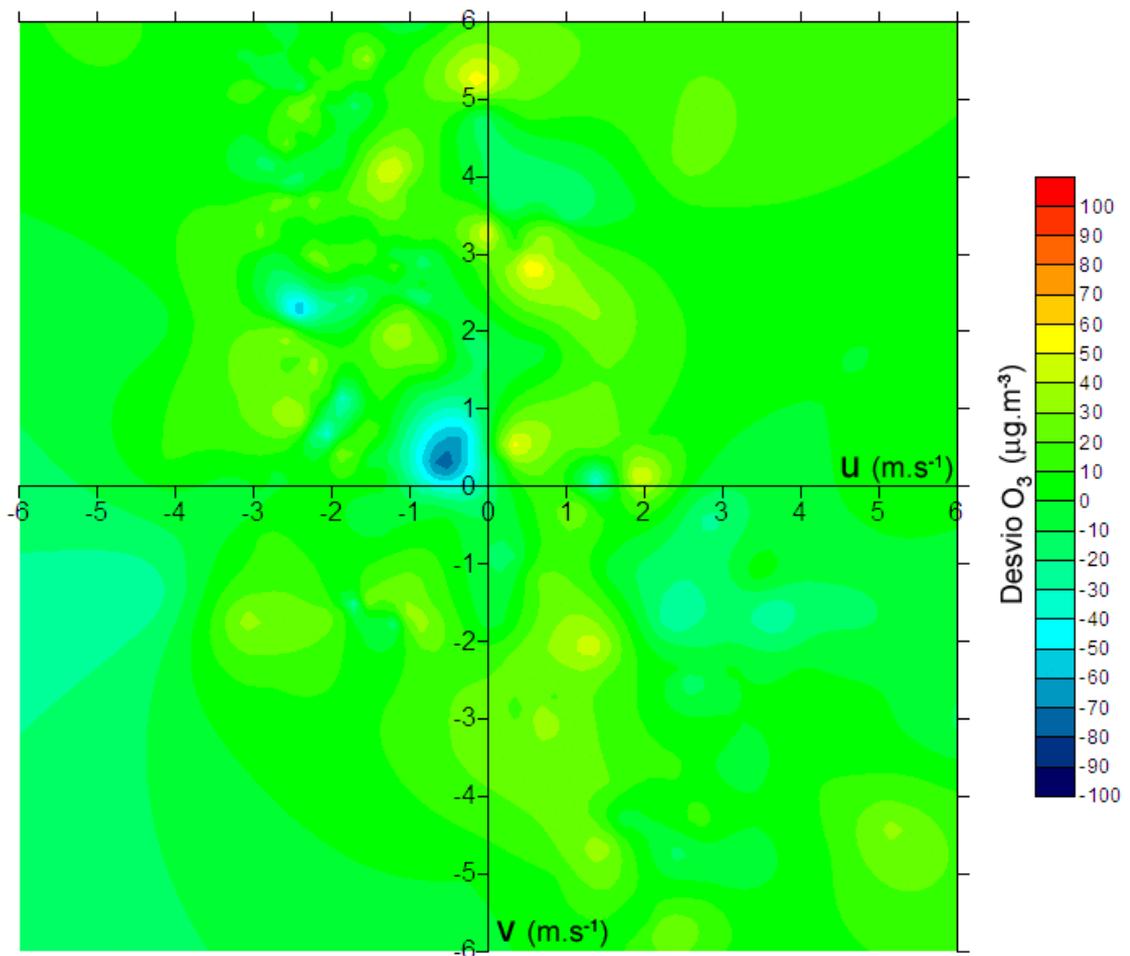


Figura 77: Distribuição dos desvios de concentrações médias de ozônio à tarde de acordo com o campo de vento à tarde (expresso por suas componentes u e v).

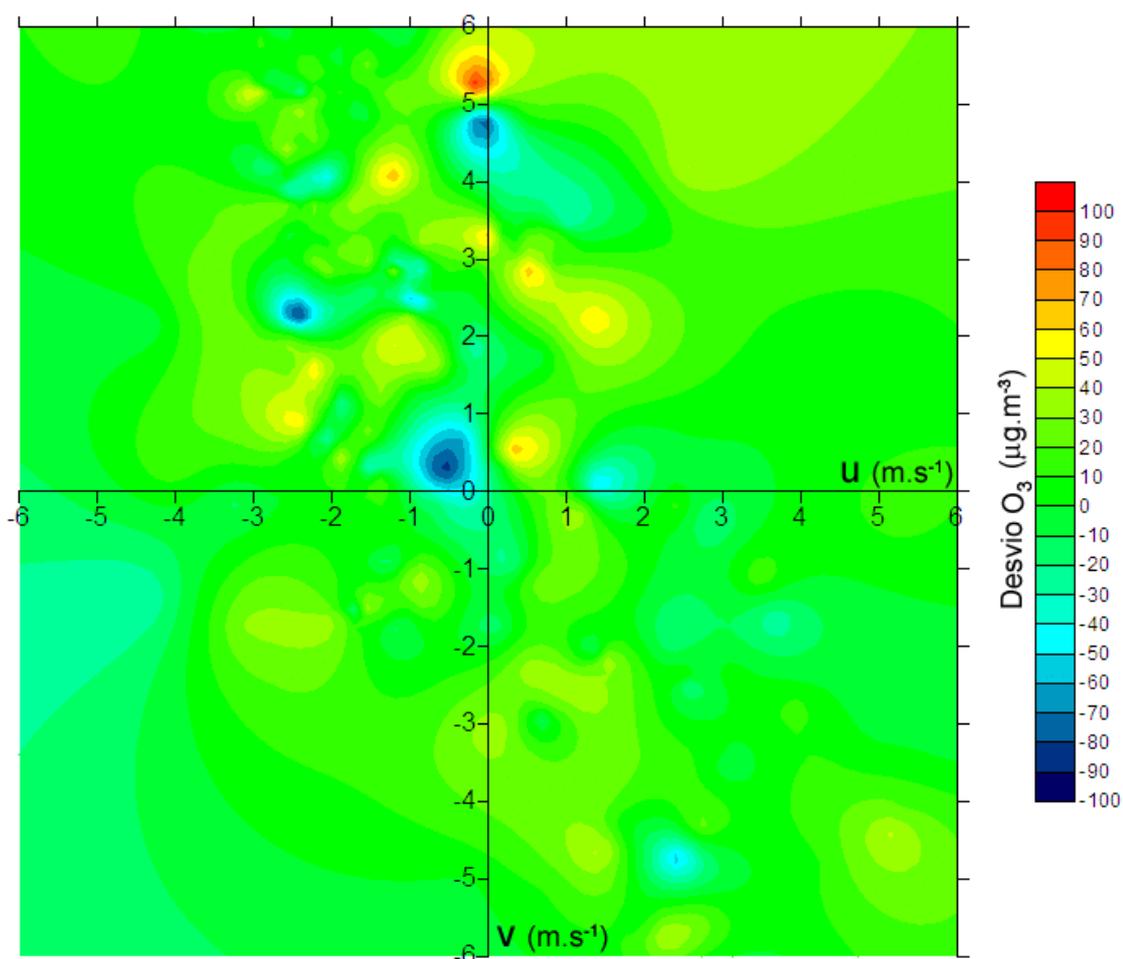


Figura 78: Distribuição dos desvios de concentrações máximas de ozônio à tarde de acordo com o campo de vento à tarde (expresso por suas componentes u e v).

As comparações entre os desvios das concentrações médias e máximas de ozônio também foram feitas para as medidas de altura da camada de mistura e pressão atmosférica. Entretanto, assim como nos itens 3.2.6 e 3.2.7, respectivamente, não foi possível determinar uma relação clara entre essas variáveis com as concentrações de ozônio, também não foi identificado um padrão de tendência dos desvios em relação às suas medidas.

4.13 - Teste de aplicação da rede neural

Para avaliar a qualidade de previsão de ozônio por redes neurais artificiais de uma forma operacional, foram inseridas como entradas as previsões meteorológicas fornecidas pelo Laboratório de Meteorologia Aplicada a Sistemas de Tempo Regionais (MASTER – DCA/IAG/USP). O período considerado foi de Agosto a Dezembro de 2006 e as variáveis são as mesmas das condições de melhor ajuste discutidas no item 4.11, ou seja, médias de manhã e de tarde para: temperatura do ar, umidade relativa, radiação solar, campo de vento (u,v), altura da camada de mistura e pressão atmosférica; além de uma informação binária quanto ao dia da semana (0-seg. a sex.,1-sab. e dom.). No entanto, foram usadas medidas de radiação solar em vez de sua previsão, por estas se mostrarem mais apropriadas.

Os resultados das simulações são comparados graficamente com as medições através das figuras 79 e 80, para as concentrações médias e máximas de ozônio, respectivamente. Em relação às concentrações médias de ozônio (figura 79), até outubro os resultados são mais concordantes com as medições, exceto por alguns eventos de baixa concentração que foram superestimados. Ao longo de outubro ocorreram as maiores discrepâncias, quando a rede exagerou as tendências de aumento ou de redução do poluente. Em novembro as simulações voltaram a apresentar um desempenho mais adequado e em dezembro houve poucos eventos que permitissem uma avaliação do período. Para as concentrações máximas de ozônio (figura 80), foi observado um padrão semelhante ao do gráfico de média de ozônio, porém, com desvios mais consideráveis entre as previsões e as medidas. De maneira geral, as estimativas acompanham as tendências de evolução das concentrações, mas extrapolando-as algumas vezes.

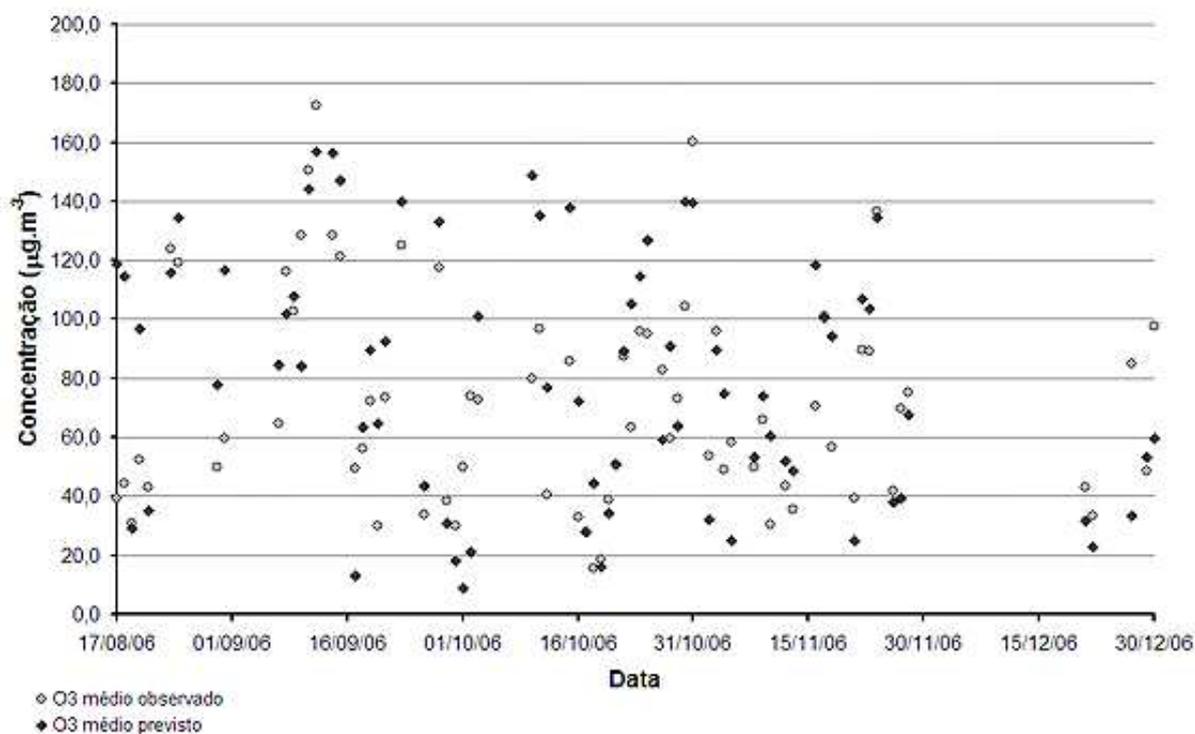


Figura 79: Evolução das concentrações médias de ozônio, nos valores previstos e observados, ao longo do segundo semestre de 2006.

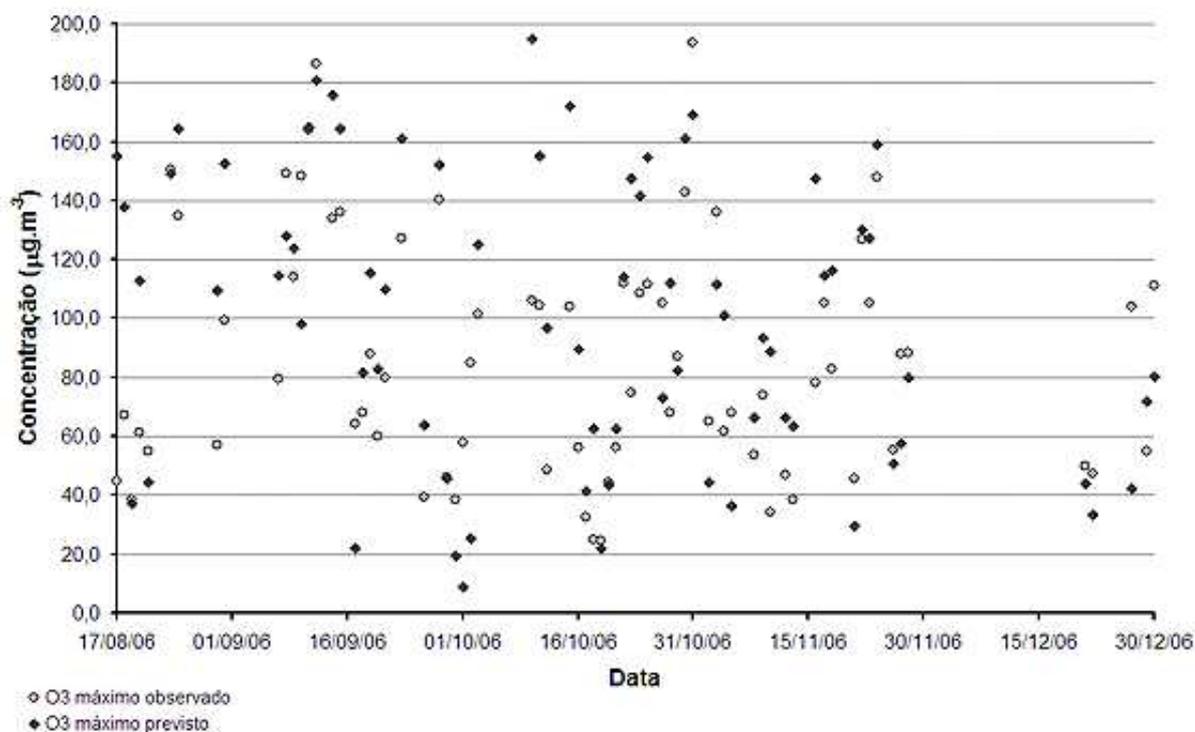


Figura 80: Evolução das concentrações máximas de ozônio, nos valores previstos e observados, ao longo do segundo semestre de 2006.

Em outra comparação, as previsões são relacionadas às observações pelas figuras 81 e 82, referentes às concentrações médias e máximas de ozônio, respectivamente. Por estas é possível observar graficamente o grau de concordância entre os resultados da rede neural e as medidas correspondentes. Em ambas as figuras há muita dispersão dos pontos, sendo ela menor para concentrações médias de ozônio, que apresenta um coeficiente de determinação (R^2) de 0,55, contra o 0,51 obtido para as concentrações máximas de ozônio.

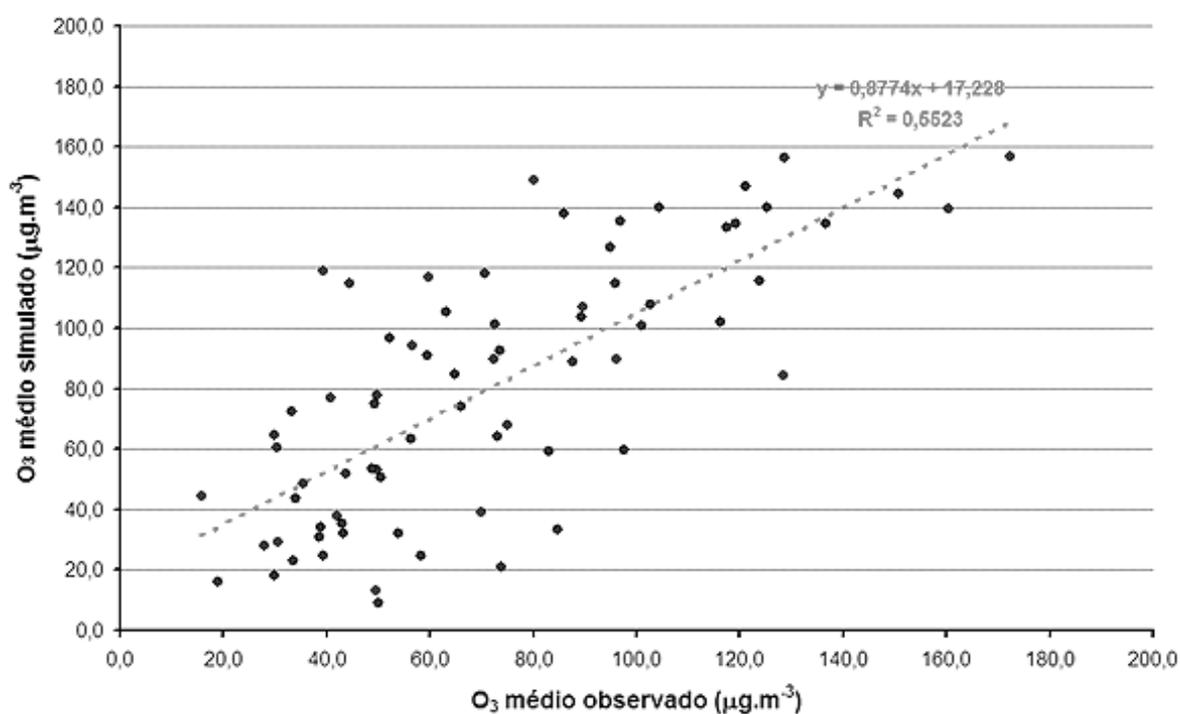


Figura 81: Relação entre valores calculados e medidos da concentração média de ozônio.

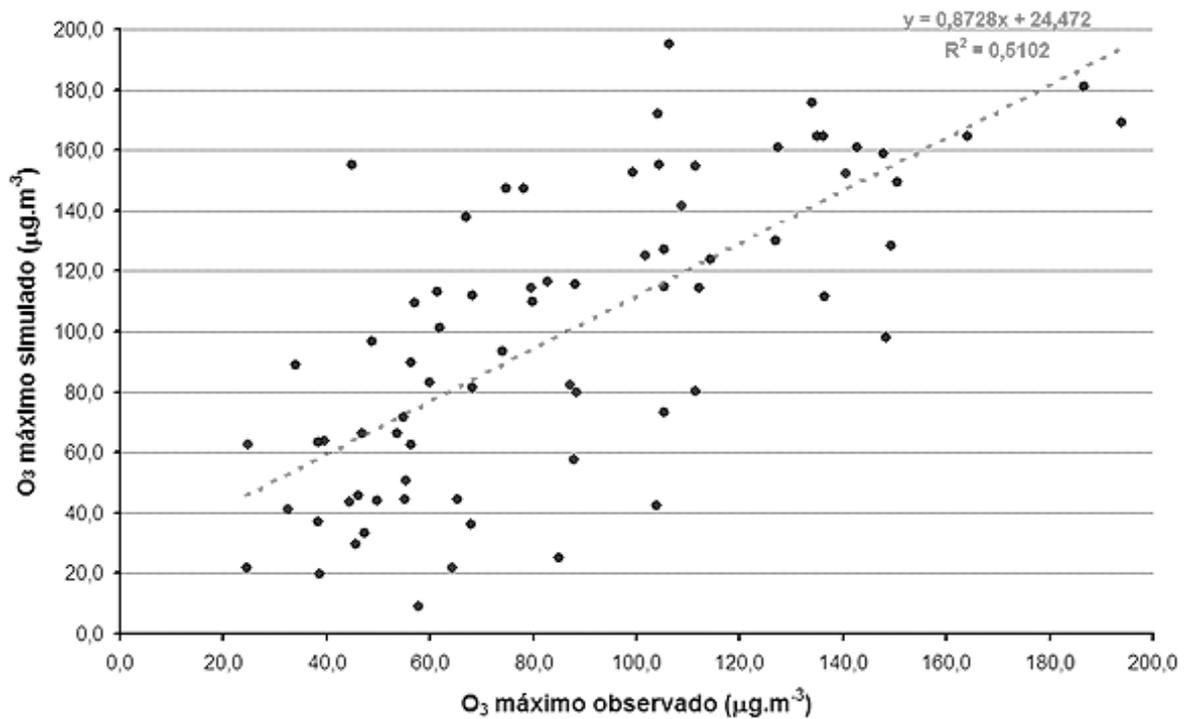


Figura 82: Relação entre valores calculados e medidos da concentração máxima de ozônio.

Na tabela 4.13.1 são apresentados os valores médios, máximos, mínimos e os desvios padrões referentes às medições da Cetesb e as estimativas da rede neural. Na tabela 4.13.2 estão os resultados dos cálculos de correlação, erro médio, raiz do erro médio quadrático e do parâmetro $RMSE_{ub}$. O erro médio indica que a simulação superestima as concentrações em $5,4 \mu\text{g.m}^{-3}$ para as médias de ozônio e $9,4 \mu\text{g.m}^{-3}$ para as máximas. Entretanto, de acordo com os parâmetros $RMSE$ e $RMSE_{ub}$, os erros são, em média, $34,4 \mu\text{g.m}^{-3}$ para concentrações médias de ozônio e $41,7 \mu\text{g.m}^{-3}$ para as máximas.

Tabela 4.13.1: Caracterização das medidas e simulações para concentrações médias e máximas de ozônio (em $\mu\text{g.m}^{-3}$).

	O3 médio		O3 máximo	
	Obs.	Sim.	Obs.	Sim.
Média	71,46	76,83	85,99	95,41
σ (desvio)	34,70	45,47	39,23	52,94
Máximo	172,43	156,74	193,83	195,00
Mínimo	15,84	-30,38	24,43	-36,27

Tabela 4.13.2: Critérios de avaliação do desempenho da rede neural. *ME*: Erro médio ($\mu\text{g.m}^{-3}$); *RMSE*: Raiz do erro médio quadrático ($\mu\text{g.m}^{-3}$); *RMSE_{UB}*: Aptidão da simulação ($\mu\text{g.m}^{-3}$).

	O3 médio	O3 máximo
Correlação	0,66	0,63
ME	5,37	9,42
RMSE	34,61	42,24
RMSE_{ub}	34,19	41,17

**Desenvolvimento de RNA para outras estações
de qualidade do ar selecionadas**

5.1 - Seleção das novas localidades

O estudo realizado para a estação Ibirapuera foi estendido a outras estações de monitoramento da Cetesb, a fim de avaliar a relação do ozônio troposférico com os parâmetros meteorológicos, no contexto de sua previsibilidade dentro do município de São Paulo. A seleção das estações foi baseada em diversos critérios, dentre eles, estão a disponibilidade de dados, localização espacial e representatividade estatística.

Quanto aos dados, foram selecionadas as estações que monitorassem concentrações de ozônio e tivessem registros para o período entre 1999 a 2003. Dentre as estações da cidade de São Paulo, e municípios vizinhos, poucas satisfizeram esse critério. Diadema, Mauá, Moóca, P.D.Pedro II, Pinheiros, Santana, São Caetano do Sul e São Miguel Paulista foram as que atenderam o requisito.

Do ponto de vista estatístico, foram feitas algumas análises para classificar as estações mais relevantes ao estudo. Inicialmente foi feita uma inspeção dos dados, através de indicadores básicos, como valores médios e máximos e mais uma comparação gráfica. Observou-se que Mauá, Santana e São Caetano do Sul apresentam as maiores concentrações de ozônio, portanto, sendo interessantes para investigação, dada a preocupação com o impacto à saúde devido às altas concentrações de ozônio.

Pela análise de correlação (tabela 5.1.1), as estações de Ibirapuera, Moóca e P.D.Pedro II estão muito bem associadas entre si, sendo que Ibirapuera é a que mais se correlaciona com as demais estações. São Caetano do Sul e São Miguel Paulista também estão bem relacionadas com as demais. Já Mauá é a que se apresenta com comportamento mais distinto.

Tabela 5.1.1: Correlação entre as concentrações de ozônio medidas em diferentes estações da Cetesb.

	Diadema	Ibirapuera	Mauá	Moóca	P. D. Pedro II	Pinheiros	Santana	S. C. do Sul	S. M. Paulista
Diadema	1,00	0,90	0,86	0,85	0,82	0,82	0,85	0,91	0,82
Ibirapuera	0,90	1,00	0,81	0,94	0,91	0,90	0,90	0,89	0,87
Mauá	0,86	0,81	1,00	0,76	0,71	0,71	0,80	0,87	0,80
Moóca	0,85	0,94	0,76	1,00	0,94	0,88	0,89	0,86	0,89
P. D. Pedro II	0,82	0,91	0,71	0,94	1,00	0,88	0,87	0,82	0,84
Pinheiros	0,82	0,90	0,71	0,88	0,88	1,00	0,83	0,81	0,79
Santana	0,85	0,90	0,80	0,89	0,87	0,83	1,00	0,86	0,88
S. C. do Sul	0,91	0,89	0,87	0,86	0,82	0,81	0,86	1,00	0,86
S. M. Paulista	0,82	0,87	0,80	0,89	0,84	0,79	0,88	0,86	1,00

A análise de componentes principais (tabela 5.1.2) mostra que a variabilidade dos dados é explicada por todas as estações e que estas estão correlacionadas entre si e agrupadas no primeiro fator. A segunda componente que explica parte da variabilidade dos dados tem a estação Mauá como principal variável. A terceira componente apresenta uma associação negativa entre São Miguel Paulista e as estações Pinheiros e Diadema, mas com baixa representatividade na variabilidade dos dados. As demais componentes possuem relações pouco expressivas para serem consideradas.

Tabela 5.1.2: Análise de componentes principais para as diferentes estações de monitoramento de ozônio.

	Comp. 1	Comp. 2	Comp. 3	Comp. 4	Comp. 5	Comp. 6	Comp. 7	Comp. 8	Comp. 9
Diadema	-0,93	-0,18	0,19	0,11	-0,14	0,05	0,14	0,09	0,03
Ibirapuera	-0,97	0,07	0,06	0,03	-0,02	-0,01	0,08	-0,16	-0,14
Mauá	-0,87	-0,43	0,01	-0,13	0,10	-0,16	0,01	0,00	0,00
Moóca	-0,96	0,18	-0,05	0,09	0,01	-0,10	0,03	-0,11	0,14
P. D. Pedro II	-0,93	0,28	-0,03	0,05	-0,04	-0,17	-0,08	0,13	-0,06
Pinheiros	-0,91	0,25	0,22	-0,15	0,16	0,10	0,00	0,03	0,02
Santana	-0,94	0,02	-0,14	-0,20	-0,22	0,07	-0,05	-0,02	0,02
S. C. do Sul	-0,94	-0,20	0,06	0,14	0,02	0,10	-0,21	-0,03	0,00
S. M. Paulista	-0,93	-0,02	-0,32	0,05	0,13	0,11	0,08	0,06	-0,02
Autovalores	7,80	0,44	0,21	0,13	0,12	0,10	0,09	0,07	0,04
% Total	86,67	4,85	2,37	1,42	1,34	1,13	0,97	0,75	0,50
% cumulativa	86,67	91,52	93,88	95,31	96,65	97,78	98,75	99,50	100,00

Adicionalmente foi feita uma análise de agrupamento (*cluster*), com diferentes procedimentos hierárquicos e medidas de distância. O dendograma da figura 83 (*Single Linkage*, com medida de distância Euclidiana) foi o mais representativo. Observa-se que

P.D.Pedro II e Pinheiros formam um grupo muito próximo, assim como São Caetano do Sul e Diadema, portanto, o ajuste de rede neural pode ser feito para apenas uma integrante de cada par, pois o outro estará razoavelmente representado. As demais estações formam grupos mais distintos entre si, com exceção de Moóca, que está bem associada com P. D. Pedro II e Pinheiros.

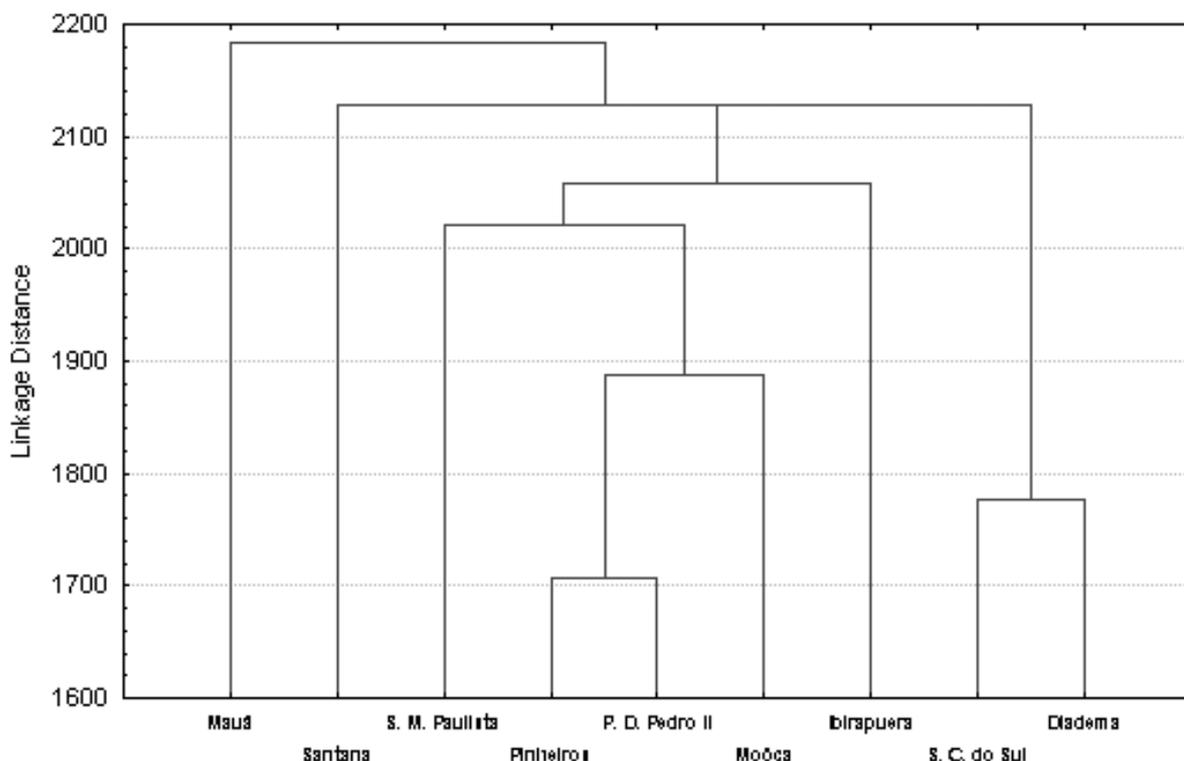


Figura 83: Dendrograma de associação entre as medições de ozônio em diferentes estações da Cetesb.

Ao final das análises as estações selecionadas para ajuste de rede neural foram: Mauá, São Miguel Paulista, P. D. Pedro II, São Caetano do Sul e Santana. Além de serem estatisticamente representativas, elas cobrem diferentes regiões de São Paulo.

Em todas as redes foram usados os mesmos tipos de variáveis de entrada e saída. Ao todo foram oito entradas, com valores médios representativos do período da manhã (08:00 – 12:00h) e da tarde (13:00 – 17:00h) de: temperatura do ar, umidade relativa e campo de vento horizontal (u,v) em 50 m. As medidas de temperatura e umidade foram obtidas de cada uma das estações selecionadas, salvo quando foi preciso recorrer a alguma adaptação (como foi o

caso da estação Mauá e Santana). A princípio foram feitos ajustes considerando as observações de vento em superfície de cada estação, ou de suas vizinhas quando na ausência desses registros. No entanto, foi constatado que alguns desses campos sofreram problema de canalização de vento, ou seja, obstáculos ao redor da estação comprometeram a qualidade das medidas. Por isso foram utilizadas as medições do SODAR no Ibirapuera como referencial.

Como saída, as redes neurais obtêm estimativas de concentração média e máxima de ozônio para o período da tarde.

Todas as redes foram configuradas de modo a ter oito neurônios na camada oculta, e com os dados sendo apresentados 40 mil vezes durante o processo de ajuste. A base de dados usada no treinamento corresponde ao período de 1999 a 2003. Para validação, foram empregadas as medições de 2004.

Para cada rede neural foi aplicado um teste de previsão, inserindo na rede os prognósticos meteorológicos obtidos pelo Laboratório de Meteorologia Aplicada a Sistemas de Tempo Regionais (MASTER – DCA/IAG/USP), considerando o período de Agosto a Dezembro de 2006.

5.2 - Estação de monitoramento de qualidade do ar em Mauá

A estação Mauá não dispõe de medidas de temperatura e umidade. Para se estimarem esses valores, a opção foi adotar os dados da estação vizinha São Caetano do Sul. Nenhuma das demais estações na região dispunha de medidas dessas variáveis. Para o ajuste da rede neural foram considerados 896 eventos para o conjunto de aprendizado e 177 para o de teste. Os resultados são apresentados graficamente nas figuras 84 e 85, comparando os resultados das simulações com os valores observados, respectivamente para as concentrações médias e máximas de ozônio.

As estimativas de concentração média de ozônio correspondem bem às observações, levando a um coeficiente de determinação de 0,80. Para as máximas, alguns pontos de elevadas concentrações (acima de $160 \mu\text{g.m}^{-3}$) foram subestimados pela rede neural; o valor de R^2 foi de 0,75. De acordo com a soma absoluta dos pesos (figura 86), as médias de temperatura, para manhã e tarde, respectivamente, foram as variáveis de maior importância nessa rede neural, seguidas pelo vento meridional, de tarde e de manhã.

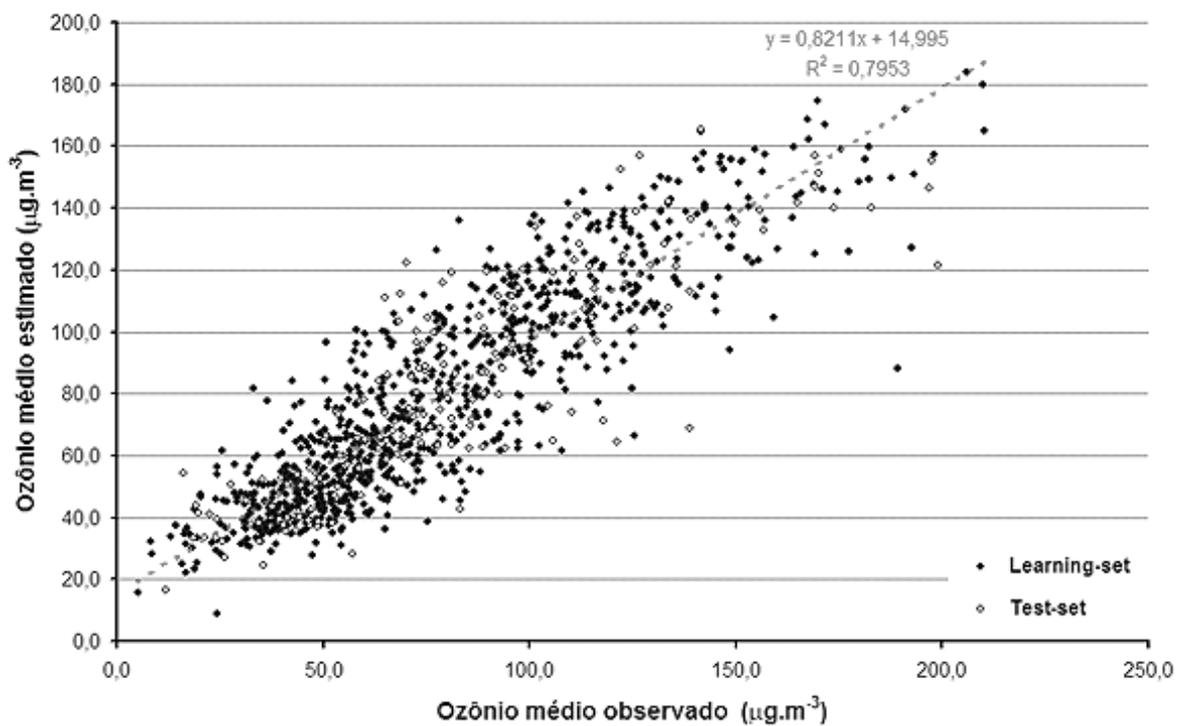


Figura 84: Relação entre valores calculados e medidos da concentração média de ozônio para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em Mauá.

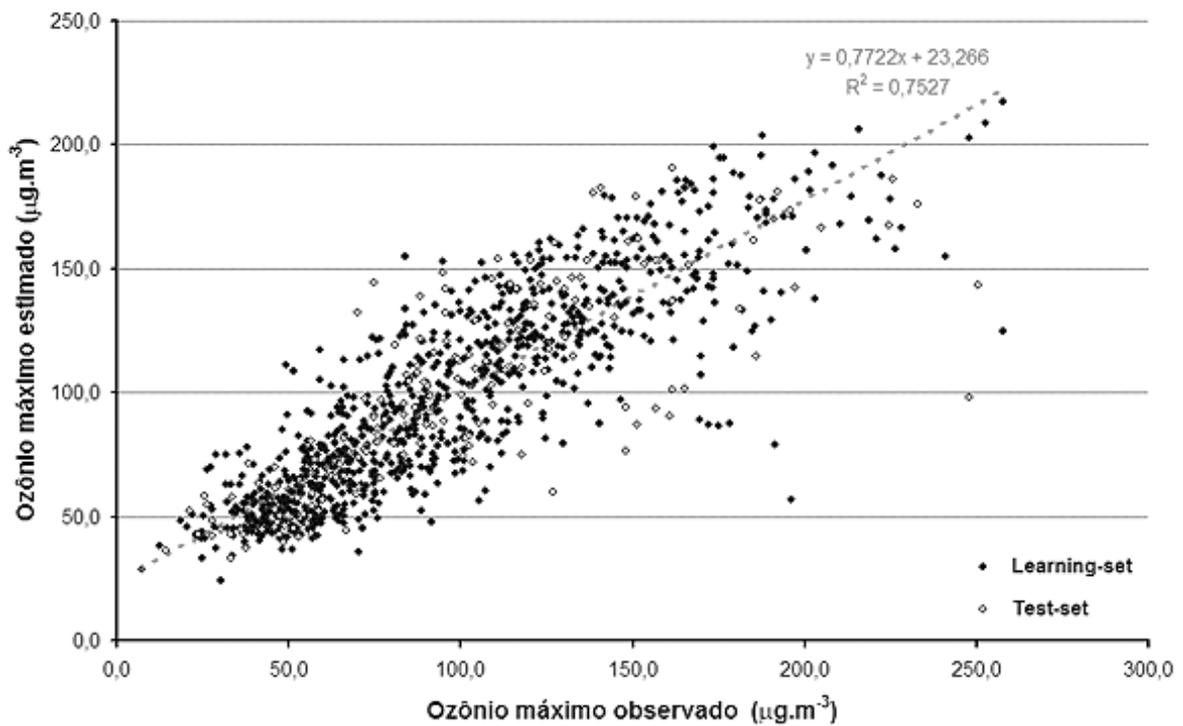


Figura 85: Relação entre valores calculados e medidos da concentração máxima de ozônio para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em Mauá.

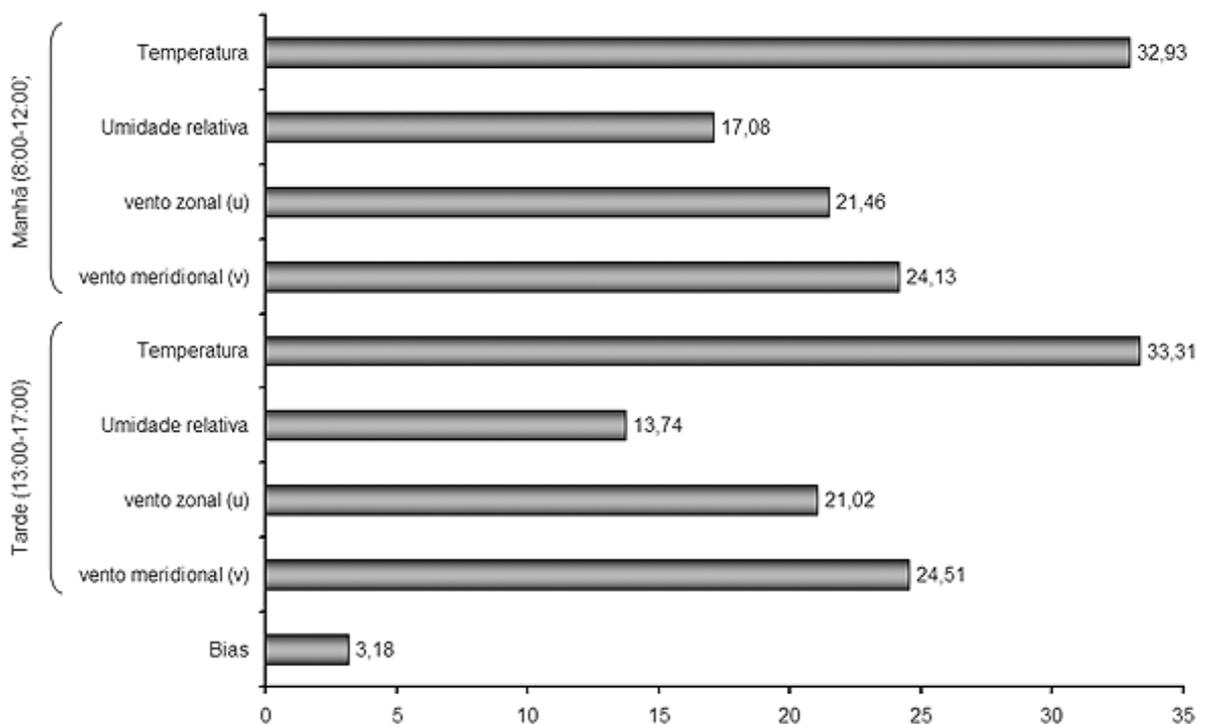


Figura 86: Soma dos valores absolutos dos pesos entre cada variável de entrada e os neurônios da camada oculta para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em Mauá.

Com a validação dessa rede, utilizando 196 eventos de 2004, foram obtidos os gráficos das figuras 87 e 88, nas quais são feitas comparações entre os valores estimados e previstos, respectivamente para concentrações médias e máximas de ozônio, com o decorrer do ano. Em ambas as curvas, as simulações acompanham bem o perfil das observações, exceto para alguns eventos que foram superestimados pela rede neural, mais notadamente durante o final de junho e começo de julho.

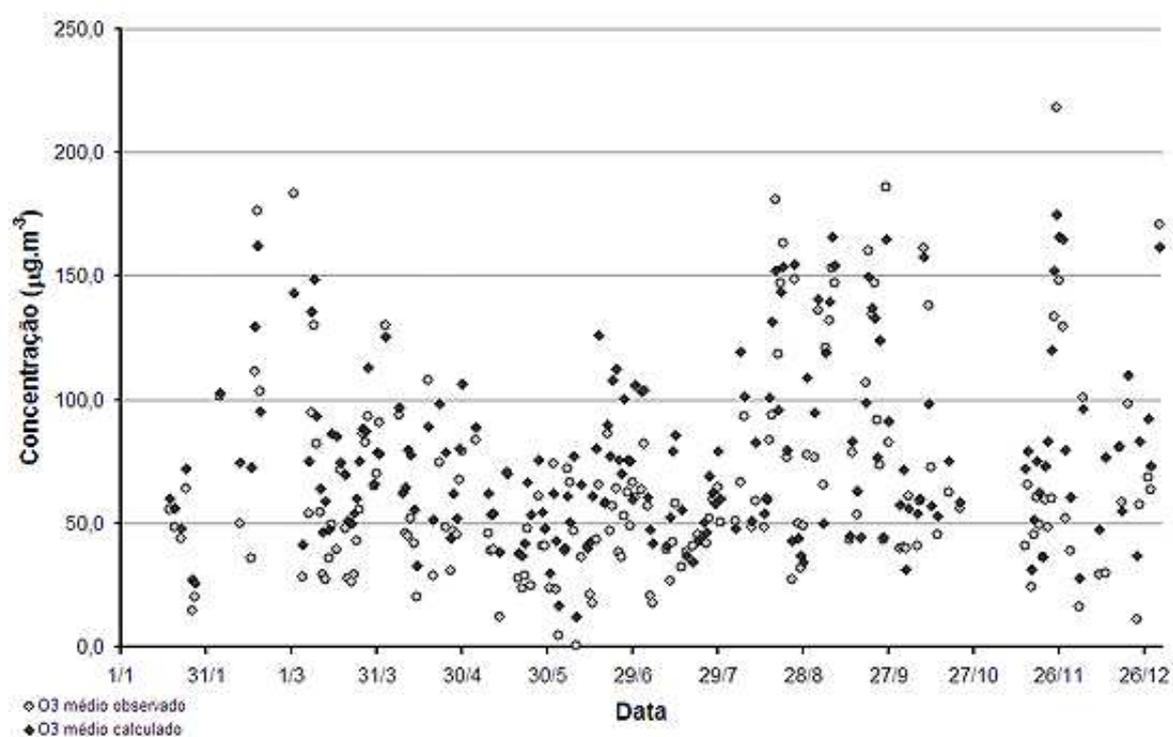


Figura 87: Evolução das concentrações médias de ozônio, valores previstos e observados, ao longo do ano de 2004 para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em Mauá.

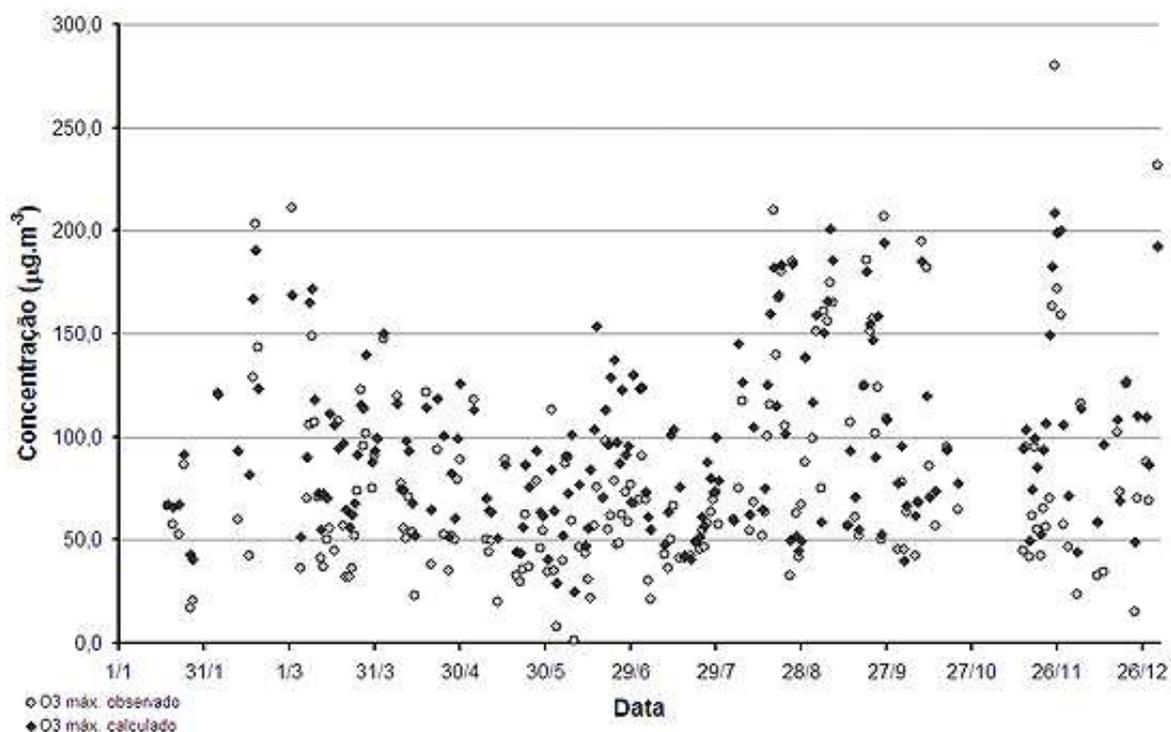


Figura 88: Evolução das concentrações máximas de ozônio, valores previstos e observados, ao longo do ano de 2004 para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em Mauá.

Em outra comparação, os valores simulados são relacionados com as observações nas figuras 89 e 90, para as concentrações médias e máximas de ozônio, respectivamente. Os resultados foram bons, considerando que os coeficientes de determinação foram um pouco superiores ao encontrado para os dados de treinamento da rede. No gráfico de concentrações médias de ozônio, o valor de R^2 foi de 0,80, enquanto no de máximas foi de 0,76. Em ambas as figuras não se observam grandes dispersões entre os valores calculados e os medidos.

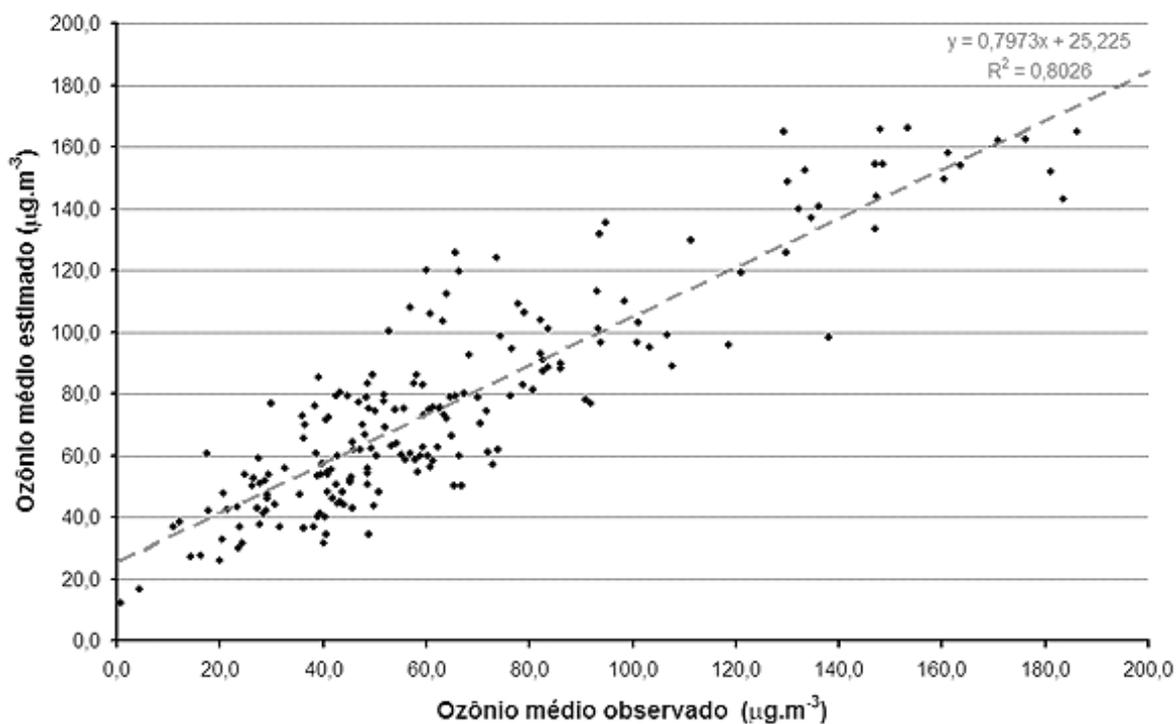


Figura 89: Relação entre valores calculados e medidos da concentração média de ozônio para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em Mauá.

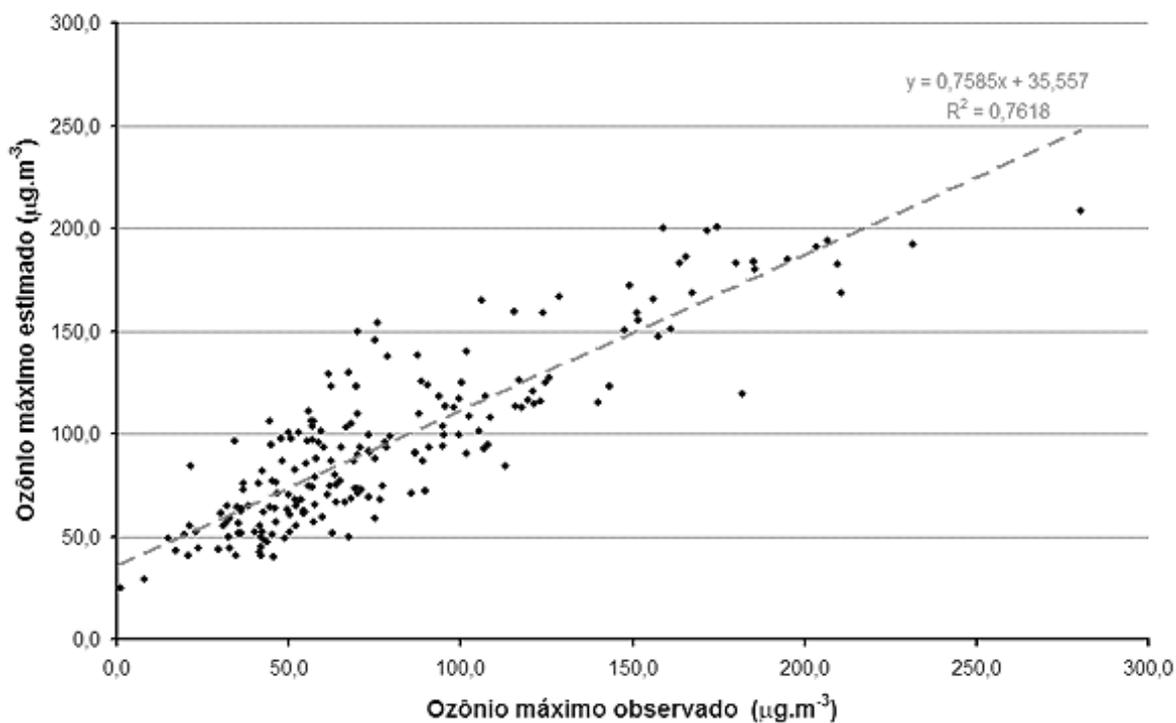


Figura 90: Relação entre valores calculados e medidos da concentração máxima de ozônio para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em Mauá.

Na tabela 5.2.1 são apresentados os valores médios, máximos, mínimos e os desvios padrões referentes às medições da Cetesb e as estimativas da rede neural. Na tabela 5.2.2 estão os resultados dos cálculos de correlação, erro médio, raiz do erro médio quadrático e do parâmetro $RMSE_{ub}$. Em média a simulação superestima as observações em $12 \mu\text{g.m}^{-3}$ para as concentrações médias de ozônio e em $16,4 \mu\text{g.m}^{-3}$ para as máximas; valores superiores aos obtidos para a estação Ibirapuera. Entretanto, nos índices de desempenho ($RMSE$ e $RMSE_{ub}$), os valores de erro foram menores. Em média foram de $19,6 \mu\text{g.m}^{-3}$ para as concentrações médias de ozônio, e $25,9 \mu\text{g.m}^{-3}$ para as máximas.

Tabela 5.2.1: Caracterização das medidas (obs.) e simulações (sim.) para concentrações médias e máximas de ozônio (em $\mu\text{g.m}^{-3}$).

	O3 médio		O3 máximo	
	Obs.	Sim.	Obs.	Sim.
MÉDIA	65,41	77,38	79,26	95,68
σ (desvio)	40,07	35,66	47,83	41,56
MÁXIMO	218,43	175,14	280,16	208,77
MÍNIMO	0,71	11,96	0,95	25,05

Tabela 5.2.2: Critérios de avaliação do desempenho da rede neural. ME : Erro médio ($\mu\text{g.m}^{-3}$); $RMSE$: Raiz do erro médio quadrático ($\mu\text{g.m}^{-3}$); $RMSE_{UB}$: Aptidão da simulação ($\mu\text{g.m}^{-3}$).

	O3 médio	O3 máximo
Correlação	0,90	0,87
ME	11,97	16,42
RMSE	21,42	28,49
$RMSE_{ub}$	17,76	23,28

Na figura 91 a diferença entre simulação e observação (desvio) para cada evento de concentrações médias e máximas de ozônio é comparada com os valores calculados e os medidos. Semelhante ao obtido para a estação Ibirapuera, observa-se uma tendência de aumento dos desvios em relação à elevação das concentrações, e a subestimativa da rede neural para com esses níveis mais altos do poluente (acima de $150 \mu\text{g.m}^{-3}$, aproximadamente).

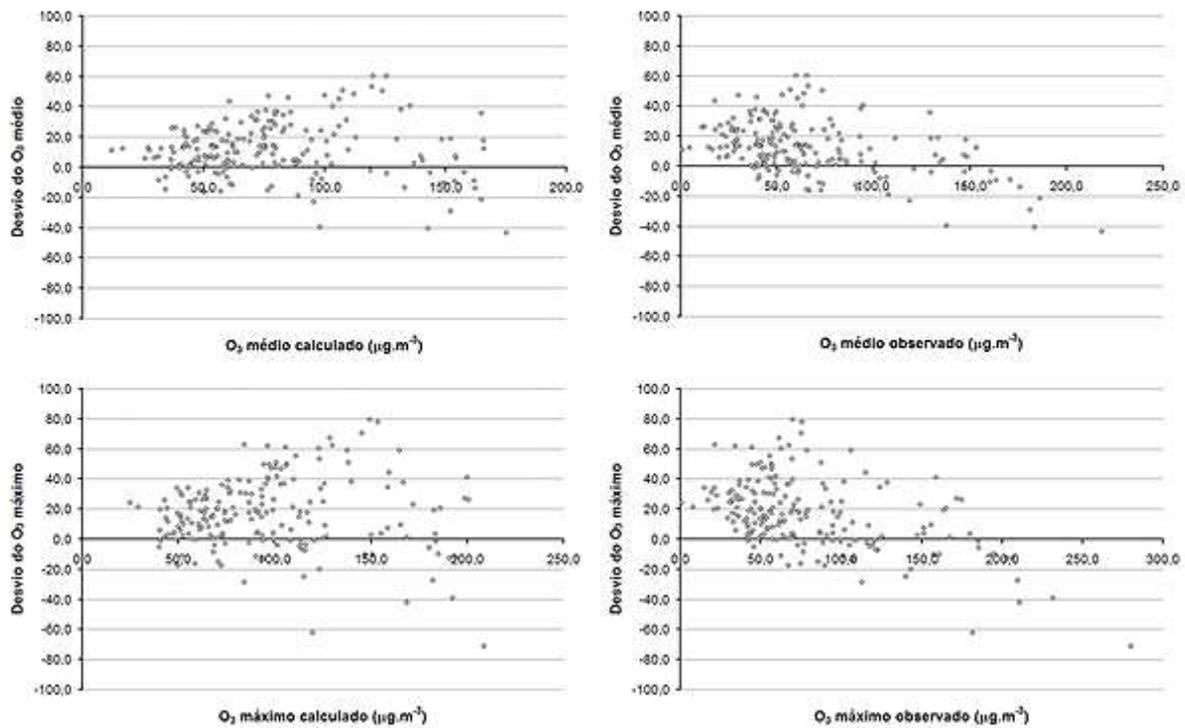


Figura 91: Relação entre os desvios (diferença entre valores calculados e observados) e as concentrações médias e máximas de ozônio para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em Mauá.

As comparações entre os desvios e as medidas de temperatura estão presentes na figura 92. A distribuição dos desvios de acordo com a temperatura possui um caráter simétrico em desvios positivos e negativos, exceto que, acima de 20°C, há uma maior frequência de eventos subestimados (desvios negativos). Em relação à umidade relativa (figura 93), não há um padrão característico na distribuição dos desvios, apenas os desvios das concentrações máximas de ozônio são maiores que os das concentrações médias de ozônio. Quanto ao campo de vento horizontal, os maiores desvios, seja para médias ou máximas de concentração de ozônio (figuras 94 e 95, respectivamente), ocorrem para ventos de baixa intensidade vindos de norte; particularmente para a componente zonal entre -2 e 2 m.s^{-1} e meridional em 0 e -2 m.s^{-1} . Há alguns eventos de superestimativa devido a vento sul, que são melhor identificados no gráfico de desvios de concentrações máximas de ozônio.

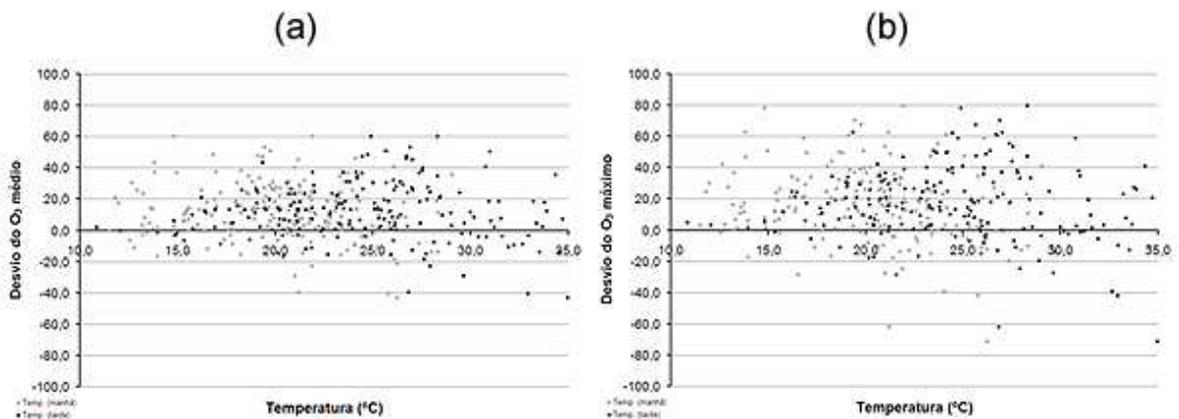


Figura 92: Relação entre os desvios das saídas de concentrações médias (a) e máximas (b) de ozônio e as médias de temperatura do ar (manhã e tarde) para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em Mauá.

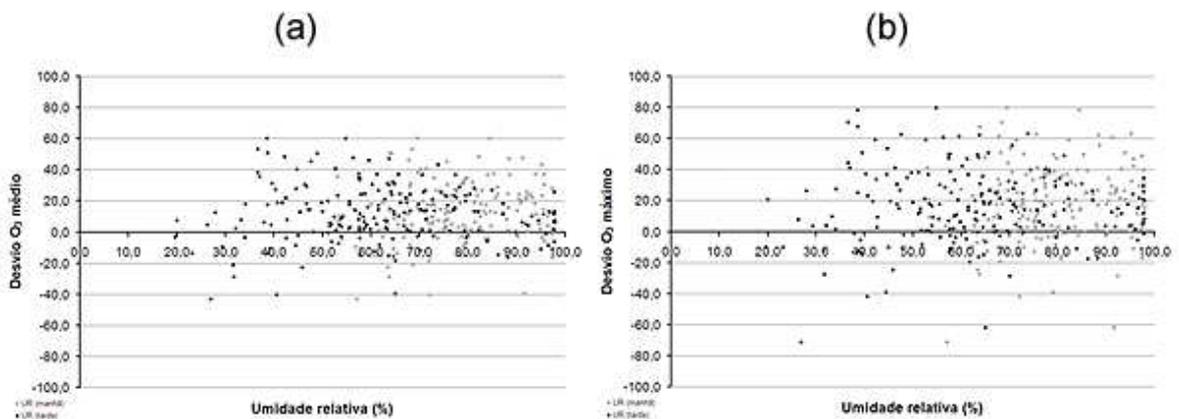


Figura 93: Relação entre os desvios das saídas de concentrações médias (a) e máximas (b) de ozônio e as médias de umidade relativa (manhã e tarde) para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em Mauá.

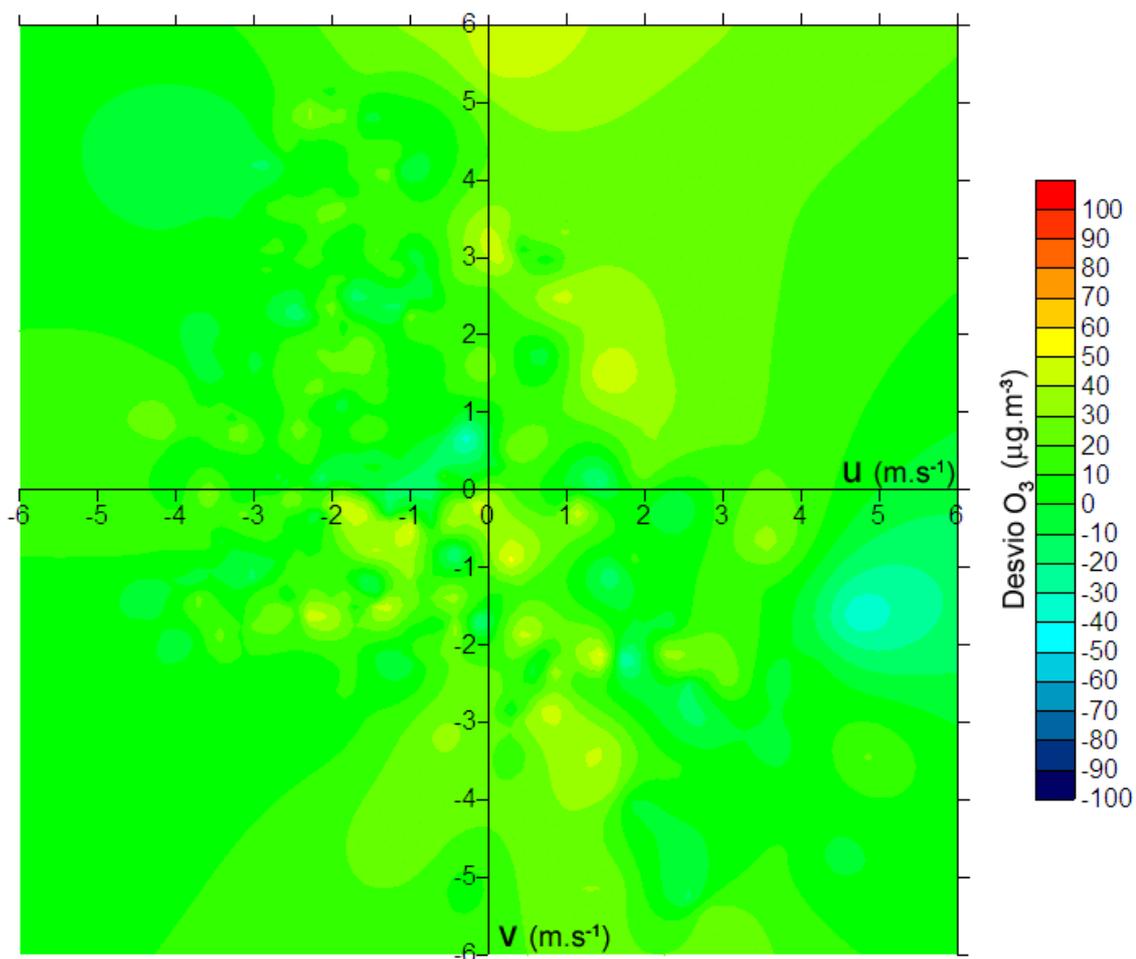


Figura 94: Distribuição dos desvios de concentração média de ozônio à tarde de acordo com o campo de vento (expresso por suas componentes u e v , de manhã e de tarde) para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em Mauá.

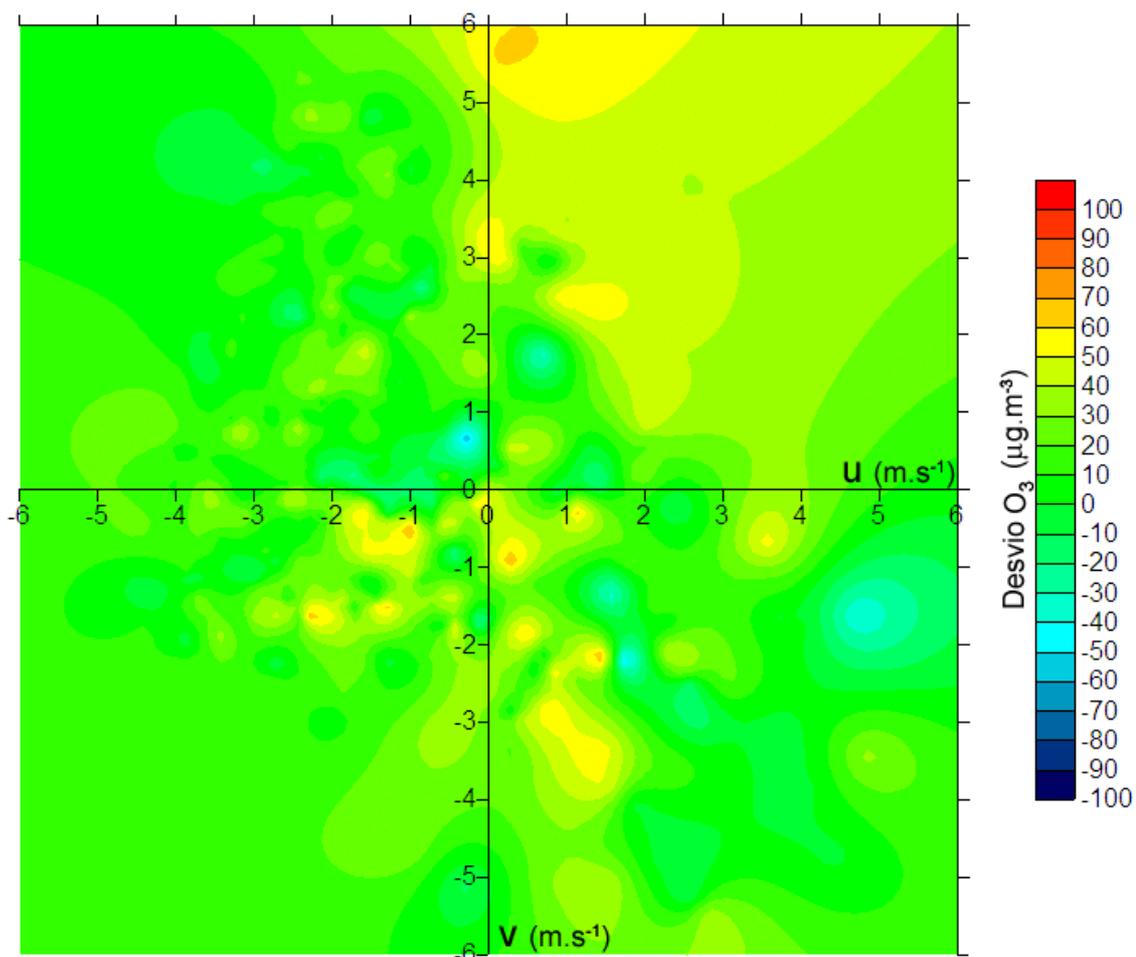


Figura 95: Distribuição dos desvios de concentrações máximas de ozônio à tarde de acordo com o campo de vento (expresso por suas componentes u e v, de manhã e de tarde) para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em Mauá.

Os resultados do teste de aplicação, inserindo as previsões meteorológicas como entradas da rede, foram comparados com os valores observados na estação Mauá, de acordo com as figuras 96 a 99, para concentrações médias e máximas de ozônio, respectivamente. Nas figuras 96 e 97 os dados são dispostos ao longo do tempo, enquanto em 98 e 99 são relacionados diretamente. De maneira geral a simulação acompanhou bem o perfil das observações, resultando em coeficientes R^2 de 0,63 e 0,65, respectivos às concentrações médias e máximas de ozônio.

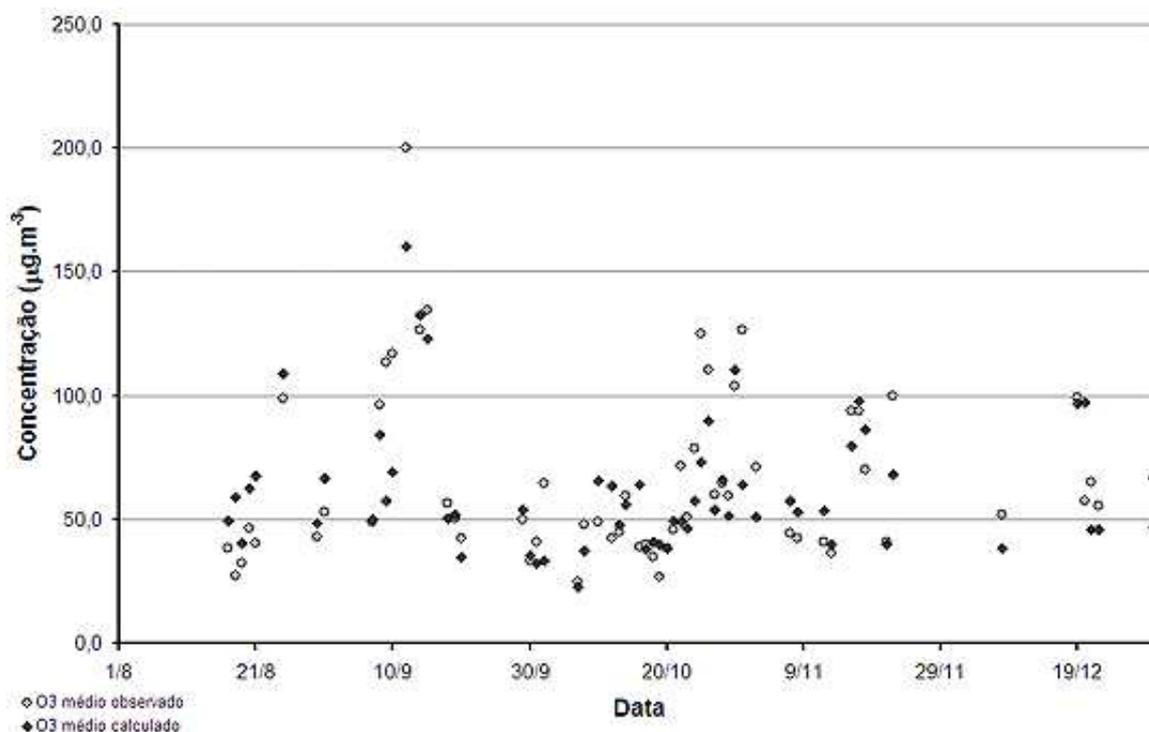


Figura 96: Evolução das concentrações médias de ozônio, nos valores previstos e observados, ao longo do segundo semestre de 2006 para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em Mauá.

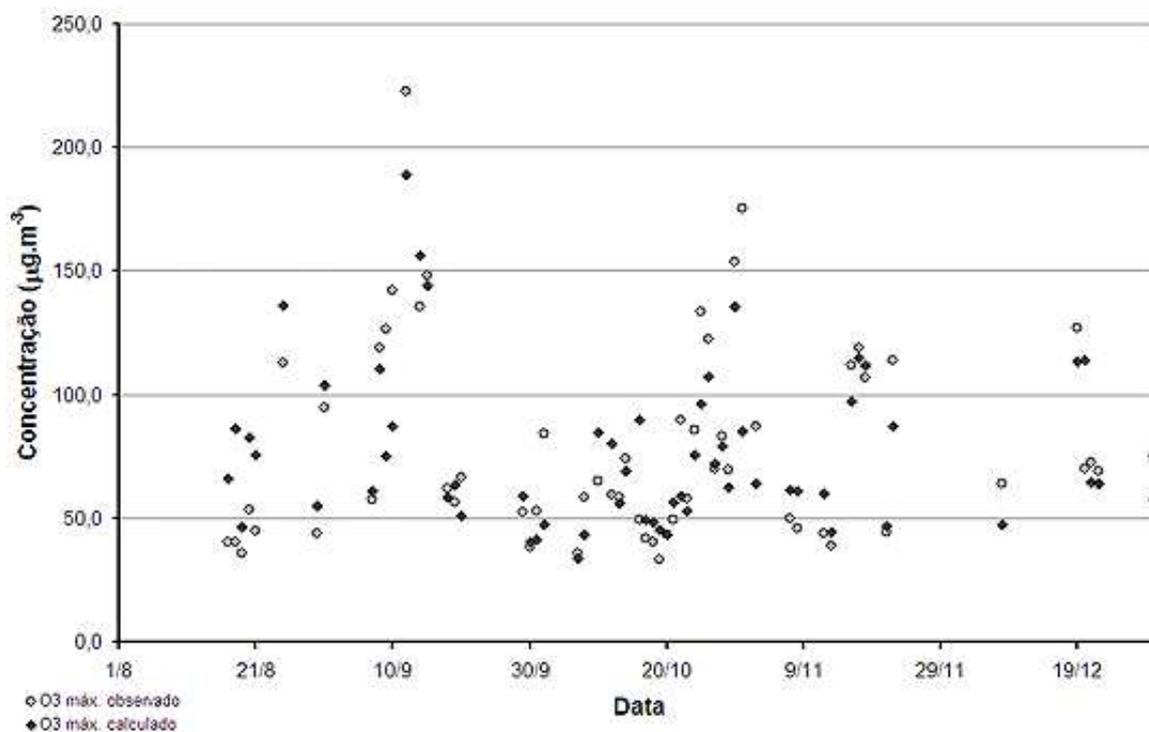


Figura 97: Evolução das concentrações máximas de ozônio, nos valores previstos e observados, ao longo do segundo semestre de 2006 para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em Mauá.

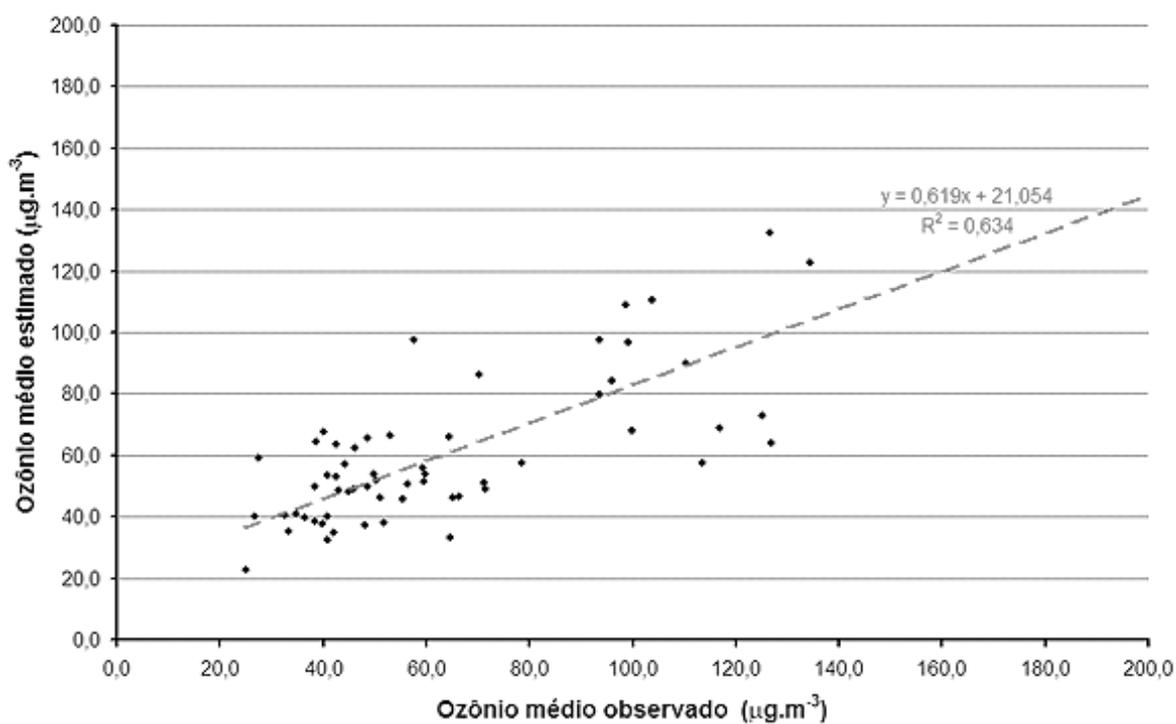


Figura 98: Relação entre valores calculados e medidos da concentração média de ozônio para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em Mauá.

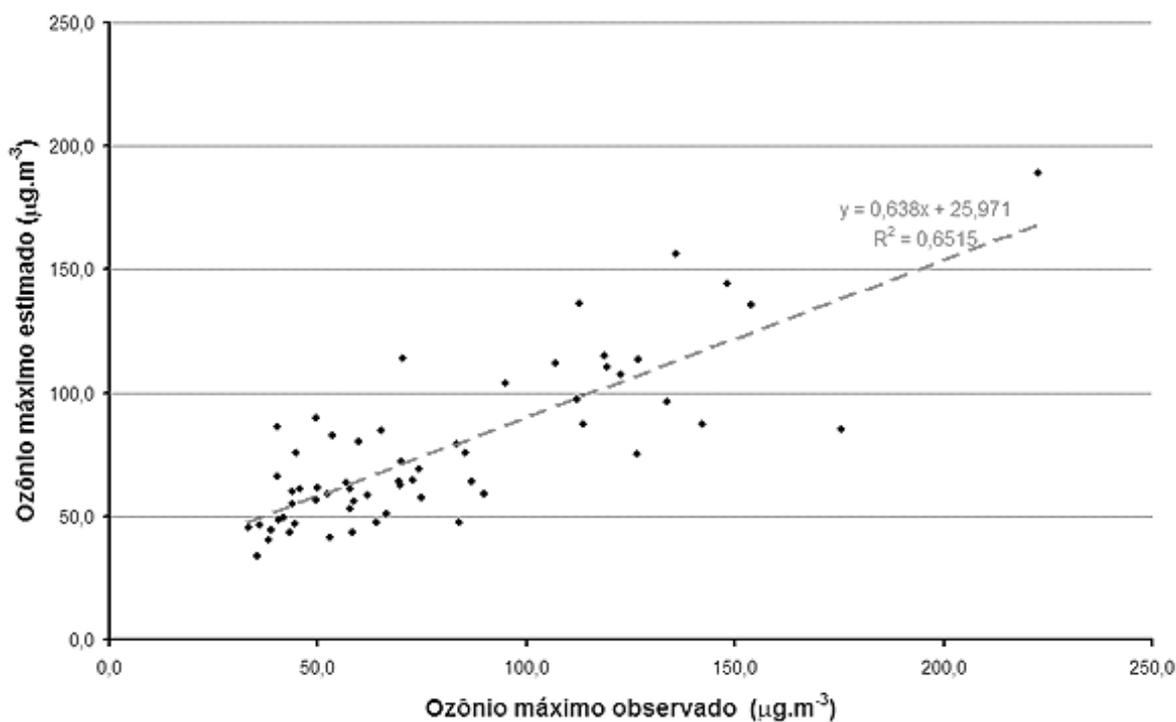


Figura 99: Relação entre valores calculados e medidos da concentração máxima de ozônio para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em Mauá.

De acordo com as tabelas 5.2.3 e 5.2.4, a rede neural tendeu a subestimar os valores observados de concentração média de ozônio em $3,9 \mu\text{g.m}^{-3}$ e as concentrações máximas de ozônio em $2,5 \mu\text{g.m}^{-3}$. De maneira geral, considerando a média dos índices $RMSE$ e $RMSE_{ub}$, o erro das concentrações médias de ozônio foi de $20,6 \mu\text{g.m}^{-3}$ e o das máximas foi $23,6 \mu\text{g.m}^{-3}$. Em comparação ao obtido para a rede de Ibirapuera, o resultado foi melhor.

Tabela 5.2.3: Caracterização das medidas e simulações para concentrações médias e máximas de ozônio (em $\mu\text{g.m}^{-3}$).

	O3 médio		O3 máximo	
	Obs.	Sim.	Obs.	Sim.
MÉDIA	65,40	61,54	78,68	76,17
σ (desvio)	34,05	26,47	40,26	31,82
MÁXIMO	200,22	160,26	222,80	188,79
MÍNIMO	24,92	22,62	33,30	33,58

Tabela 5.2.4: Critérios de avaliação do desempenho da rede neural. ME : Erro médio ($\mu\text{g.m}^{-3}$); $RMSE$: Raiz do erro médio quadrático ($\mu\text{g.m}^{-3}$); $RMSE_{UB}$: Aptidão da simulação ($\mu\text{g.m}^{-3}$).

	O3 médio	O3 máximo
Correlação	0,80	0,81
ME	-3,9	-2,5
RMSE	20,80	23,71
RMSE_{ub}	20,44	23,58

5.3 - Estação de monitoramento de qualidade do ar em São Miguel Paulista

Para o ajuste da rede neural de São Miguel Paulista foram considerados 815 eventos para o conjunto de aprendizado e 160 para teste. Os resultados se encontram graficamente nas figuras 100 e 101, comparando as simulações com os valores observados, respectivamente para as concentrações médias e máximas de ozônio. Observa-se uma boa concordância entre estimativas e medições em ambos os gráficos, embora o coeficiente de determinação tenha sido um pouco inferior ao encontrado em outras redes. Para as concentrações médias de ozônio o valor de R^2 foi de 0,75 e para as máximas foi de 0,71.

Na figura 102 são identificadas as variáveis mais importantes para a rede neural, segundo a soma absoluta dos pesos. A média do vento meridional de manhã e a temperatura à tarde foram as que apresentaram os maiores pesos.

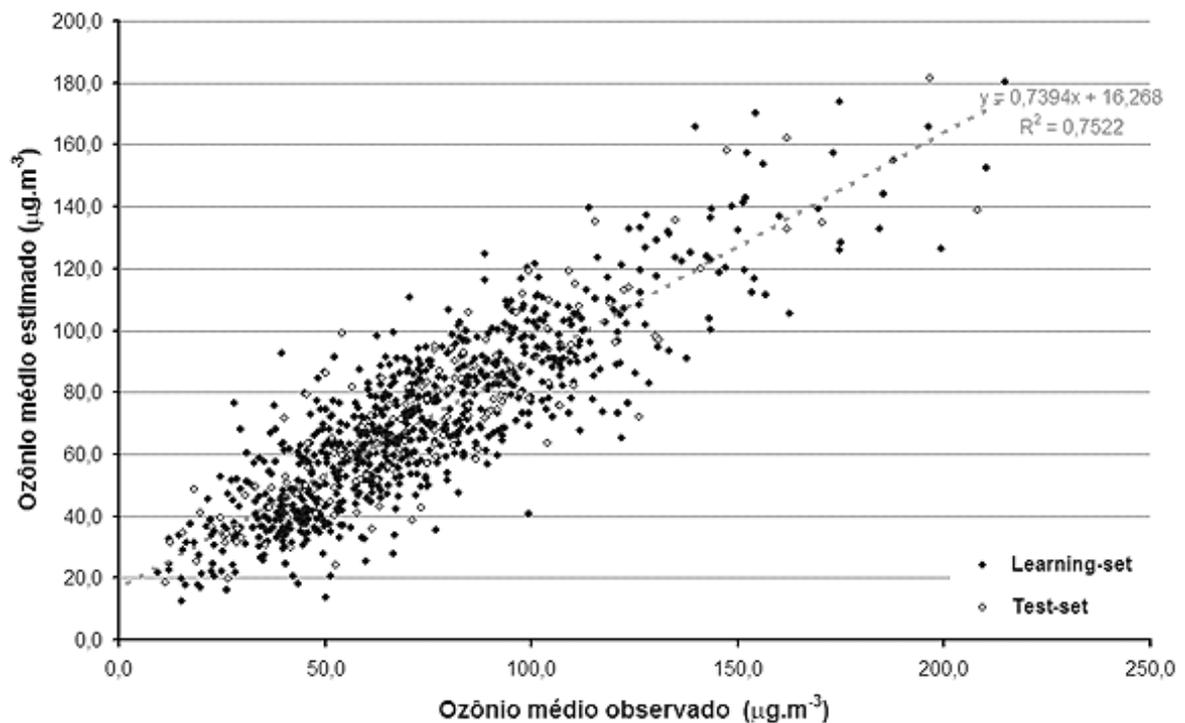


Figura 100: Relação entre valores calculados e medidos da concentração média de ozônio para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em São Miguel Paulista.

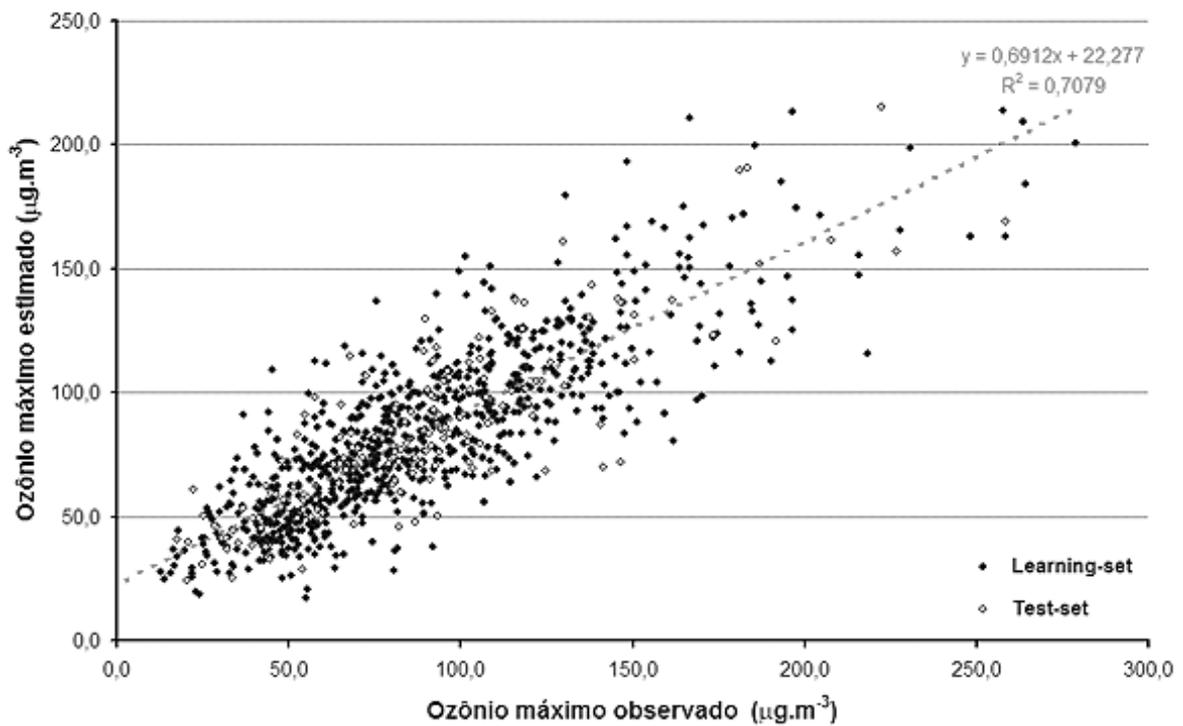


Figura 101: Relação entre valores calculados e medidos da concentração máxima de ozônio para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em São Miguel Paulista.

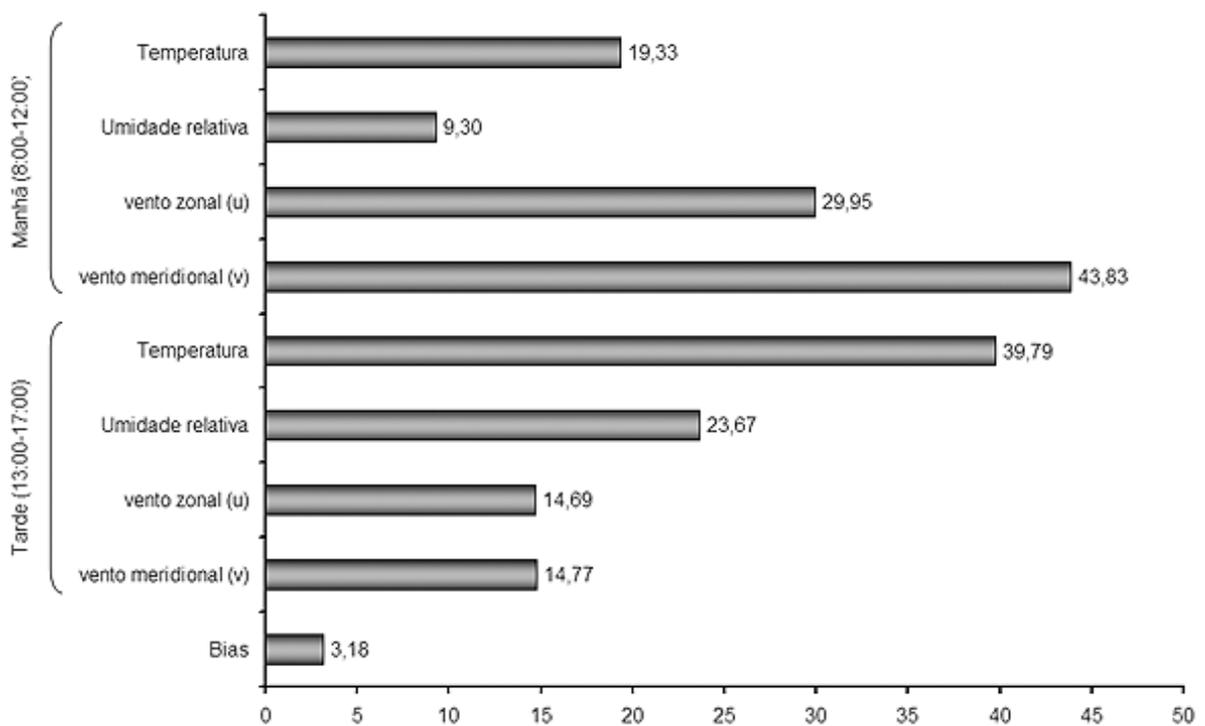


Figura 102: Soma dos valores absolutos dos pesos entre cada variável de entrada e os neurônios da camada oculta para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em São Miguel Paulista.

Na figura 103 são comparados os valores experimentais e simulados de concentração média de ozônio ao longo de 2004, no processo de validação da rede neural. A figura 104 corresponde ao resultado obtido para as concentrações máximas de ozônio. O desempenho em ambos os gráficos foi semelhante, com uma boa correspondência entre medições e estimativas, embora alguns eventos não tenham sido bem representados. Isso é mais evidente nos meses de janeiro a março e outubro a dezembro, em que a rede superestimou algumas concentrações baixas ou não foi capaz de acompanhar a intensidade de alguns picos.

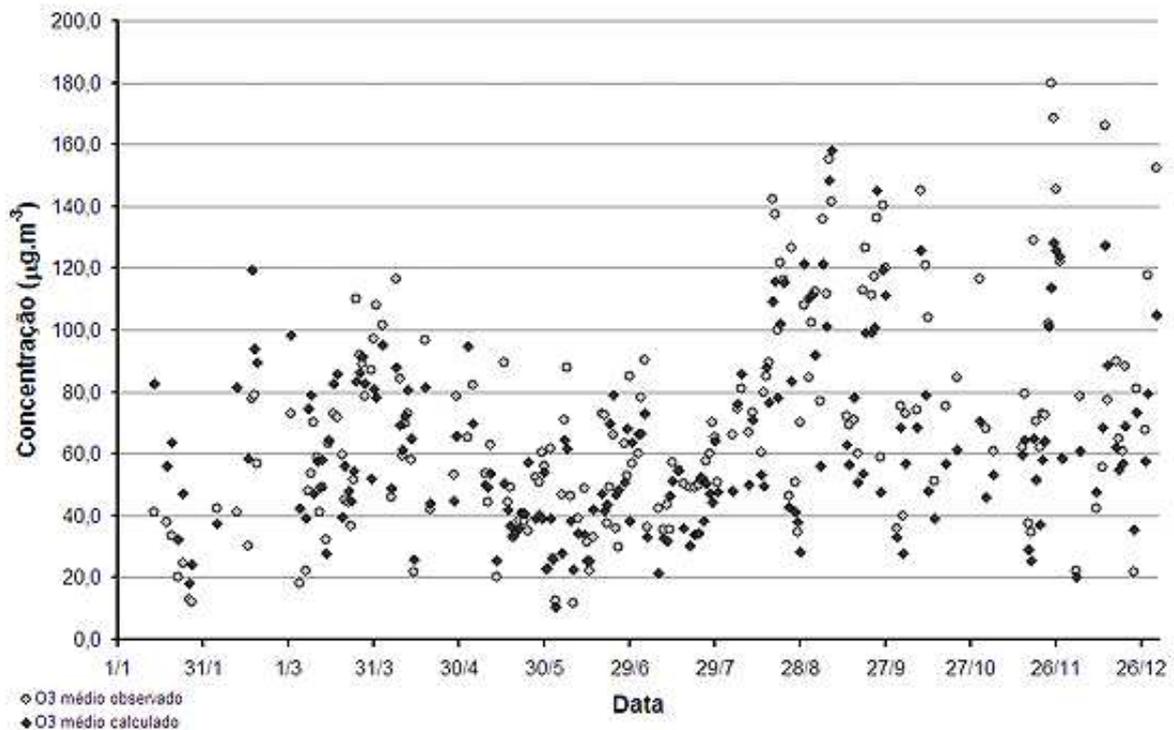


Figura 103: Evolução das concentrações médias de ozônio, valores previstos e observados, ao longo do ano de 2004 para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em São Miguel Paulista.

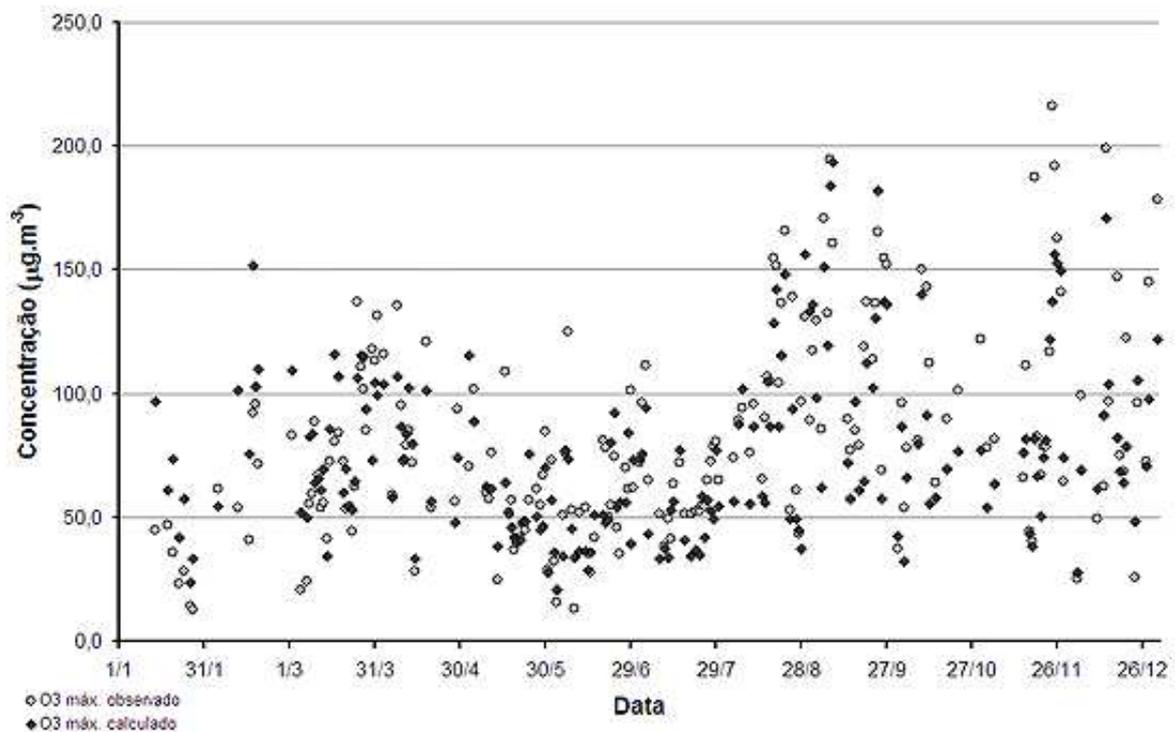


Figura 104: Evolução das concentrações máximas de ozônio, valores previstos e observados, ao longo do ano de 2004 para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em São Miguel Paulista.

Nas figuras 105 e 106 os resultados são relacionados com as medições, para concentrações médias e máximas de ozônio, respectivamente. Observa-se uma dispersão de pontos um pouco maior para as concentrações máximas de ozônio, mas, em geral, ambos os gráficos apresentam a mesma qualidade de desempenho. O coeficiente de determinação das concentrações médias de ozônio foi de 0,70, enquanto o de concentrações máximas de ozônio foi ligeiramente inferior, com R^2 igual a 0,66.

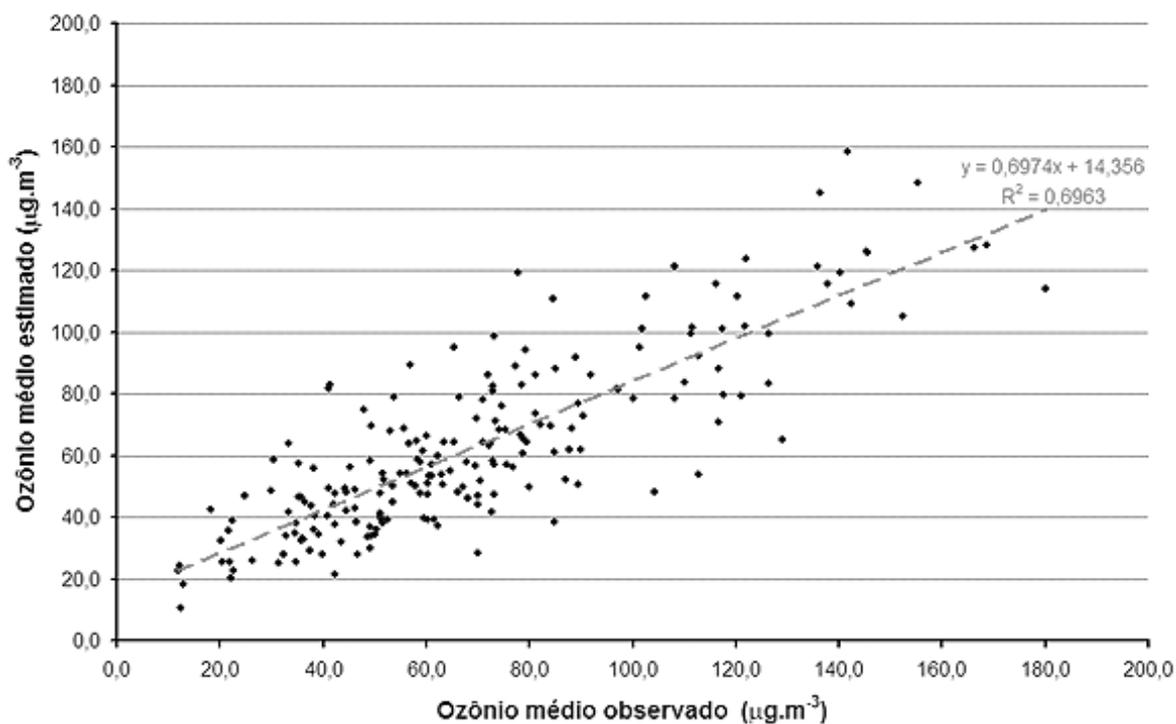


Figura 105: Relação entre valores calculados e medidos da concentração média de ozônio para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em São Miguel Paulista.

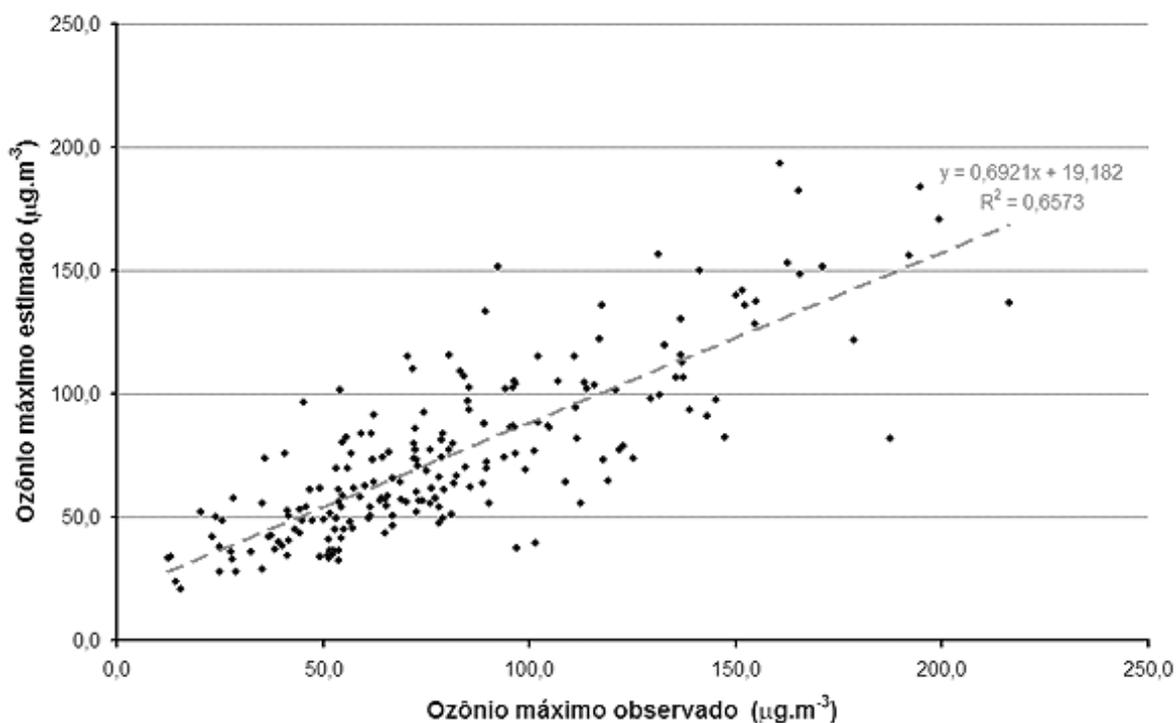


Figura 106: Relação entre valores calculados e medidos da concentração máxima de ozônio para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em São Miguel Paulista.

Parâmetros de validação são apresentados nas tabelas 5.3.1 e 5.3.2. O erro médio de concentrações médias de ozônio foi de $-6,5 \mu\text{g.m}^{-3}$, e o das máximas foi $-6,0 \mu\text{g.m}^{-3}$. Pelos indicadores de desempenho da simulação, o erro foi em torno de $19,3 \mu\text{g.m}^{-3}$ para as concentrações médias de ozônio e $24,1 \mu\text{g.m}^{-3}$ para as máximas. Em comparação a Mauá, a qualidade foi ligeiramente melhor. A rede de São Miguel Paulista também se saiu melhor no teste de validação do que a de Ibirapuera.

Tabela 5.3.1: Caracterização das medidas e simulações para concentrações médias e máximas de ozônio (em $\mu\text{g.m}^{-3}$).

	O3 médio		O3 máximo	
	Obs.	Sim.	Obs.	Sim.
MÉDIA	69,10	62,55	81,66	75,70
σ (desvio)	34,06	28,47	40,54	34,61
MÁXIMO	179,96	158,25	216,33	193,57
MÍNIMO	11,83	10,62	12,48	20,90

Tabela 5.3.2: Critérios de avaliação do desempenho da rede neural. *ME*: Erro médio ($\mu\text{g.m}^{-3}$); *RMSE*: Raiz do erro médio quadrático ($\mu\text{g.m}^{-3}$); *RMSE_{UB}*: Aptidão da simulação ($\mu\text{g.m}^{-3}$).

	O3 médio	O3 máximo
Correlação	0,83	0,81
ME	-6,5	-6,0
RMSE	19,84	24,47
RMSEub	18,72	23,73

Comparando as diferenças entre simulação e observação (figura 107), tanto para concentrações médias como máximas de ozônio observa-se um aumento dos desvios conforme o aumento das concentrações, que se torna mais expressivo acima de $100 \mu\text{g.m}^{-3}$, com a rede subestimando esses valores.

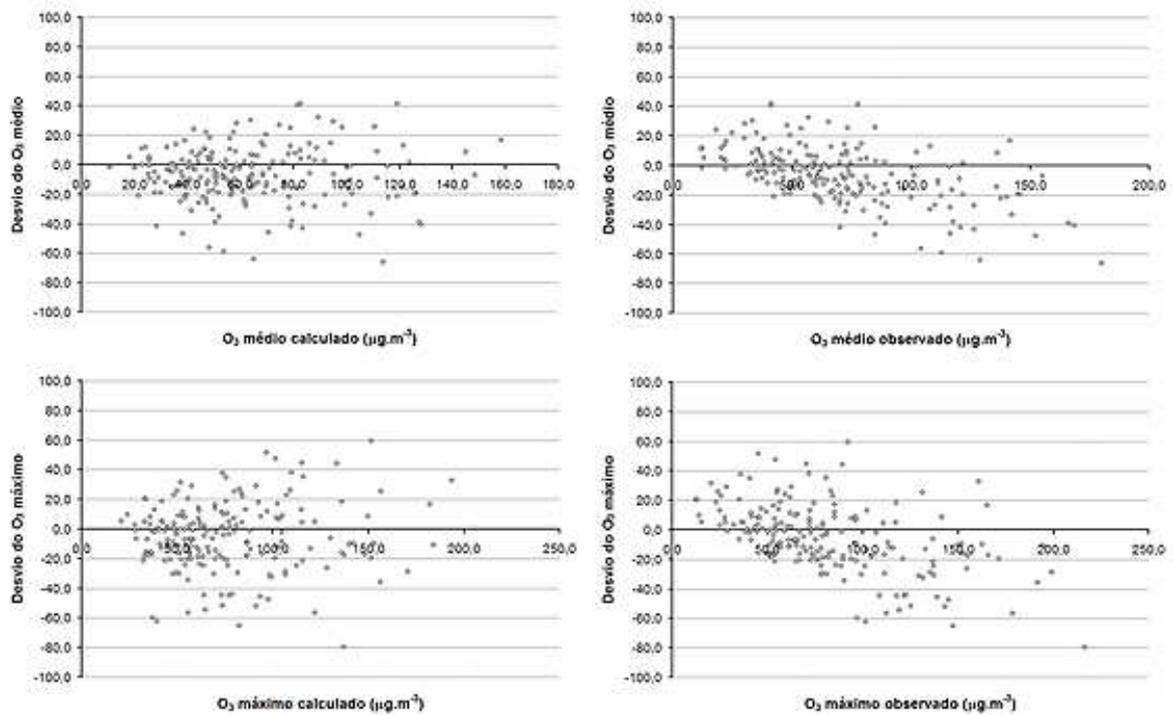


Figura 107: Relação entre os desvios (diferença entre valores calculados e observados) e as concentrações médias e máximas de ozônio para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em São Miguel Paulista.

Em comparação às medidas de temperatura (figura 108), os desvios das concentrações médias e máximas de ozônio aumentam por volta de 20°C, sendo que os erros por subestimativa são mais expressivos. Em relação à umidade (figura 109), não há um padrão de variação dos desvios, exceto pelos erros das concentrações máximas de ozônio terem sido maiores que os das concentrações médias. Os desvios de acordo com o campo de vento (figuras 110 e 111) mostram-se mais representativos nos sentidos sudeste e noroeste, em diferentes velocidades, com maior tendência a subestimativa das concentrações (desvios negativos).

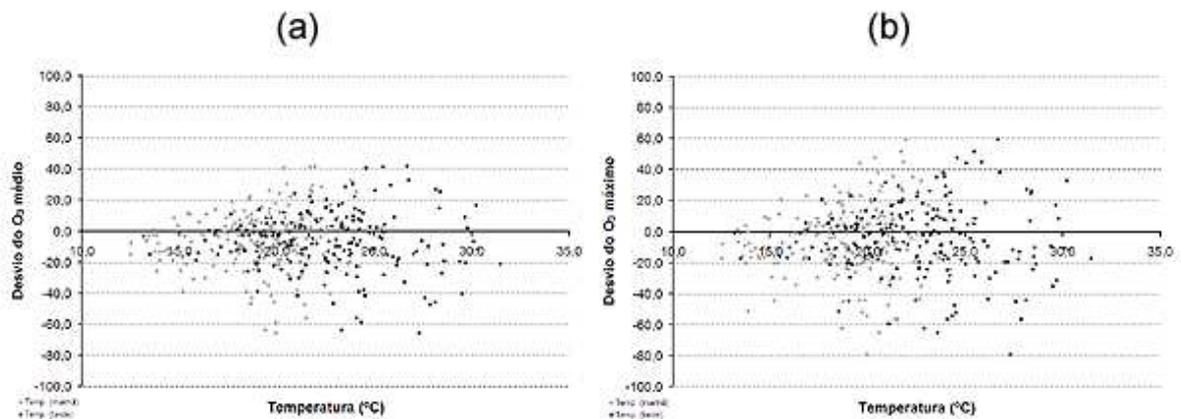


Figura 108: Relação entre os desvios das saídas de concentrações médias (a) e máximas (b) de ozônio e as médias de temperatura do ar (manhã e tarde) para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em São Miguel Paulista.

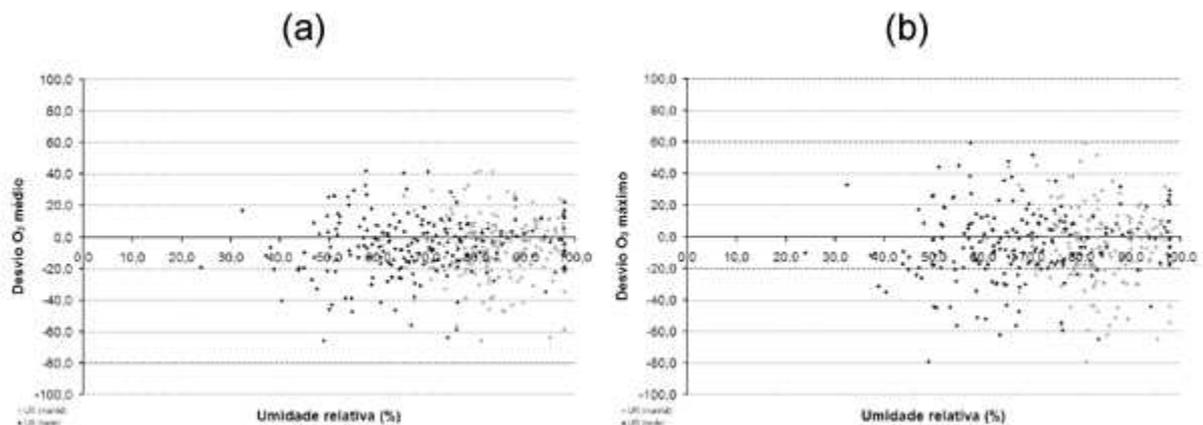


Figura 109: Relação entre os desvios das saídas de concentrações médias (a) e máximas (b) de ozônio e as médias de umidade relativa (manhã e tarde) para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em São Miguel Paulista.

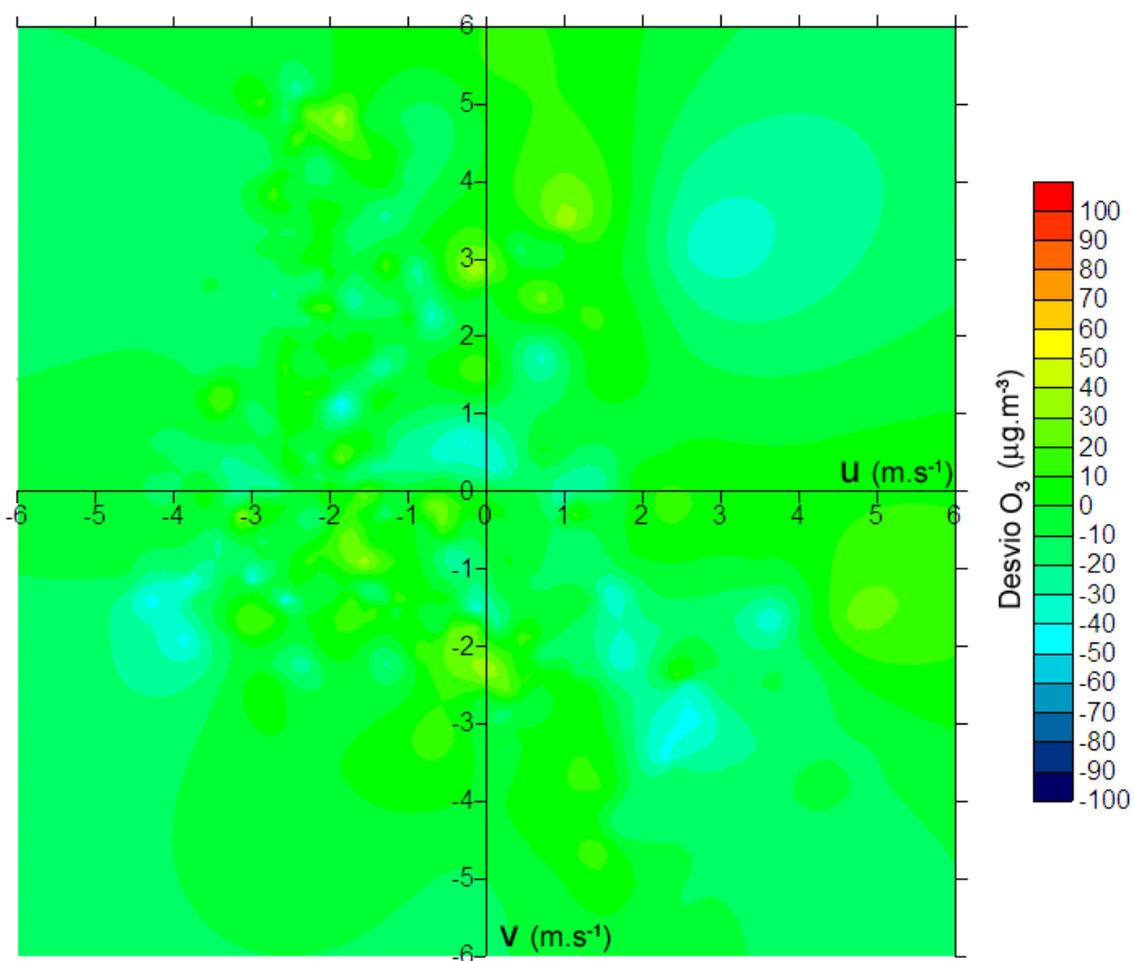


Figura 110: Distribuição dos desvios de concentrações médias de ozônio à tarde de acordo com o campo de vento (expresso por suas componentes u e v, de manhã e de tarde) para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em São Miguel Paulista.

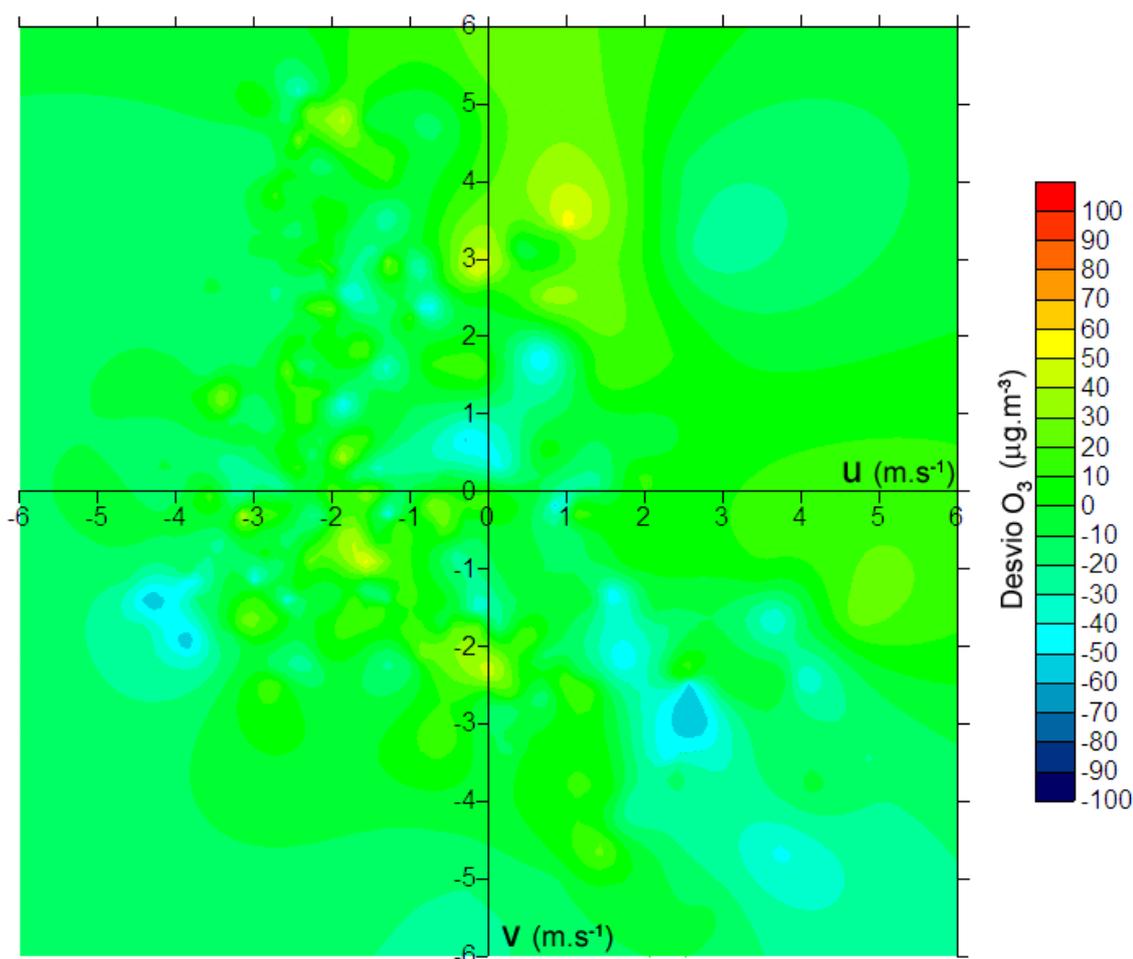


Figura 111: Distribuição dos desvios de concentrações máximas de ozônio à tarde de acordo com o campo de vento (expresso por suas componentes u e v , de manhã e de tarde) para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em São Miguel Paulista.

Foi feito um teste de aplicação dessa rede como ferramenta de previsão, utilizando os prognósticos de variáveis meteorológicas como entrada. Entretanto, não foi possível comparar os resultados das simulações com os valores medidos pela estação, pois a quantidade de dados disponíveis foi muito escassa para se fazer uma avaliação estatística. Como forma de compensação, as estimativas da rede neural foram comparadas com as medições de ozônio da estação Ibirapuera, por esta ser a referencial deste estudo.

Pelos gráficos na figura 112 os resultados são comparados ao longo do tempo, dentro do segundo semestre de 2006. Na figura 113 as simulações e medidas são relacionadas

diretamente. Observa-se que as simulações acompanham o perfil de evolução de ozônio, embora tenham ocorrido muitas superestimativas.

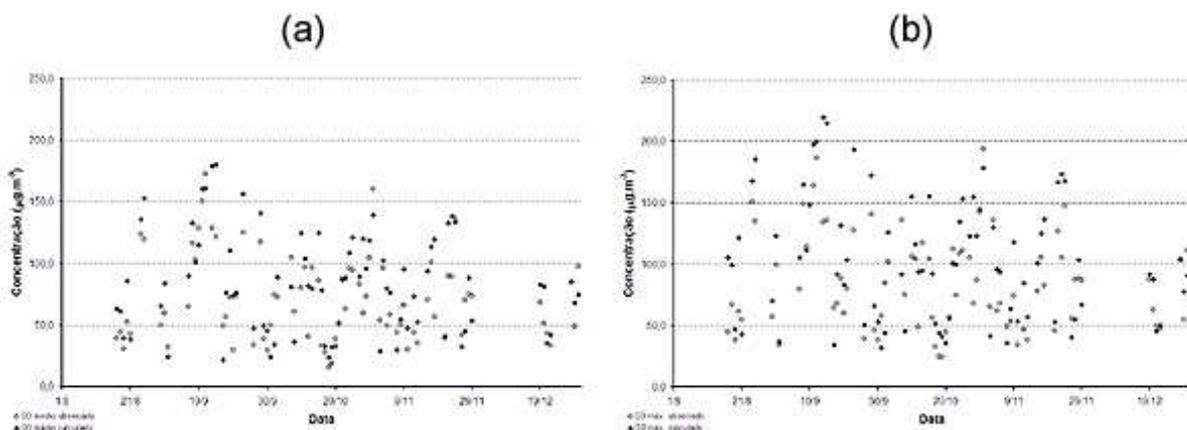


Figura 112: Evolução das concentrações médias de ozônio (a) e máximas (b), valores previstos e observados, ao longo do segundo semestre de 2006 para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em São Miguel Paulista.

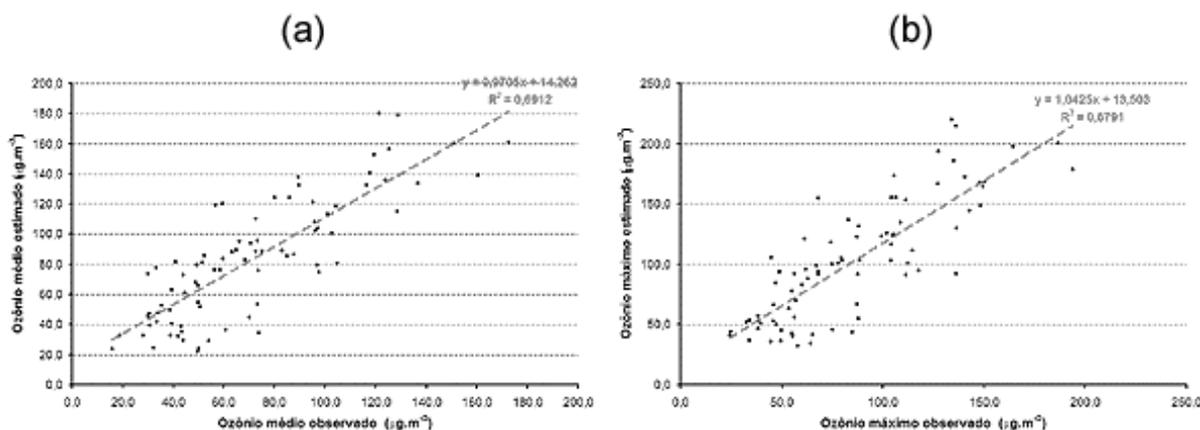


Figura 113: Relação entre valores calculados e medidos da concentração média (a) e máxima (b) de ozônio para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em São Miguel Paulista.

5.4 - Estação de monitoramento de qualidade do ar em Parque Dom Pedro II

No treinamento da rede neural de Parque D. Pedro II foram utilizados 855 dados no conjunto de aprendizado e 164 no de teste. Na figura 114, comparando as estimativas de concentrações médias de ozônio com as observações, os pontos se encontram de forma pouco dispersa, resultando em um coeficiente de determinação de 0,73. Entretanto, na mesma comparação feita para as concentrações máximas de ozônio (figura 115), houve diversas

falhas de subestimativa do poluente para concentrações observadas acima de $200 \mu\text{g.m}^{-3}$. Isso não se deve à presença de anomalias nos dados, uma vez que o mesmo não foi observado para as concentrações médias de ozônio; esse resultado se deve possivelmente à falta de habilidade da rede em estimar corretamente estes eventos. A maior parte dos dados são representativos de concentrações abaixo de $150 \mu\text{g.m}^{-3}$, sendo assim, a rede neural foi ajustada com ênfase nesses eventos, não podendo, portanto, caracterizar adequadamente as concentrações mais elevadas; apesar disso, o valor de R^2 foi de 0,65. Pela figura 116 são indicadas as variáveis de maior importância para a rede neural, segundo a soma absoluta dos pesos. De maneira geral, as variáveis mais significativas foram as do período da tarde. Porém, a média da componente meridional do vento de manhã foi a mais representativa.

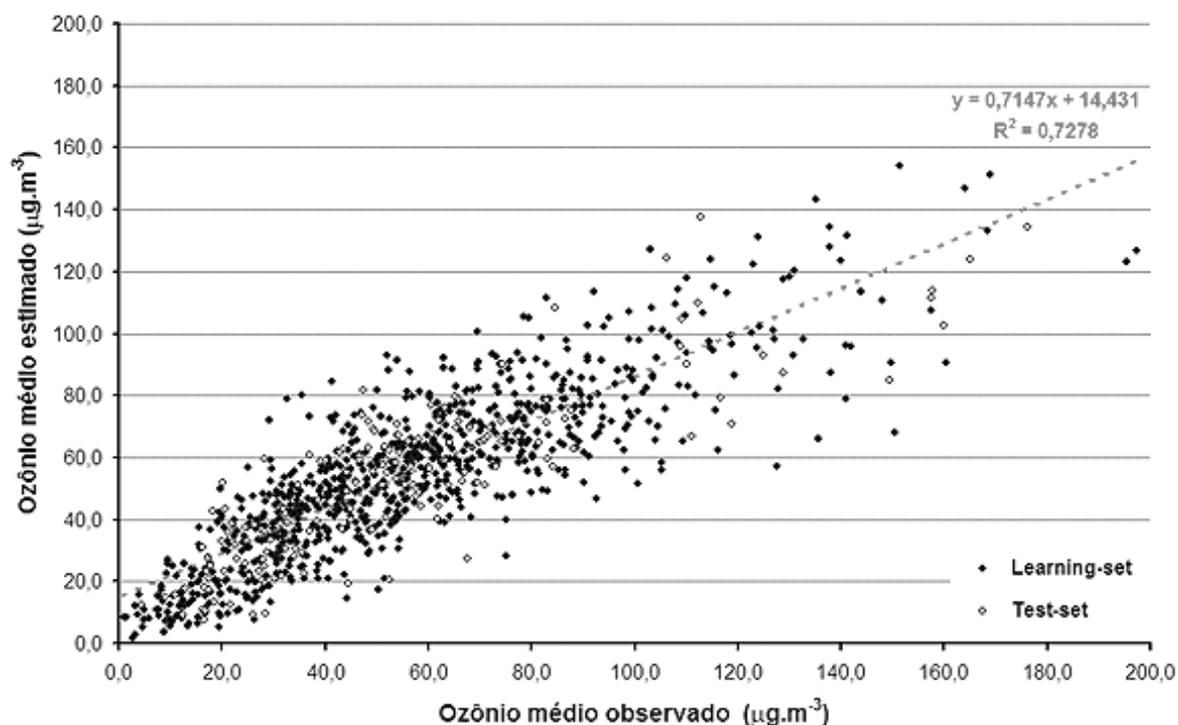


Figura 114: Relação entre valores calculados e medidos da concentração média de ozônio para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em P. D. Pedro II.

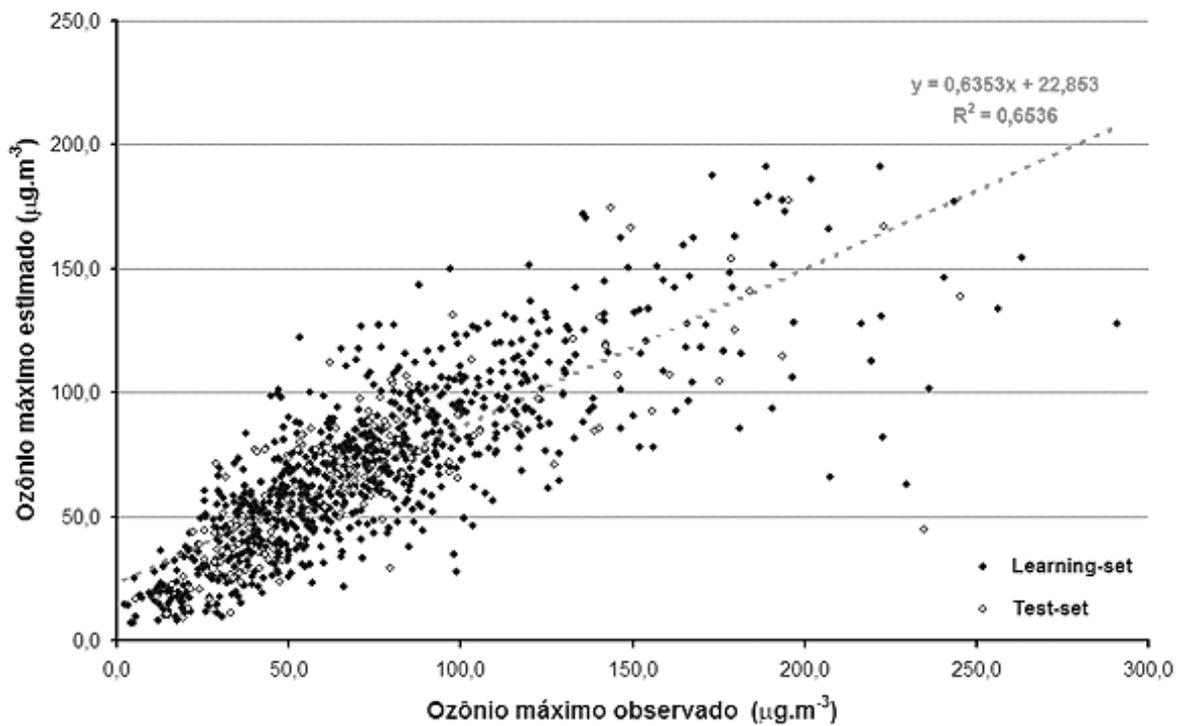


Figura 115: Relação entre valores calculados e medidos da concentração máxima de ozônio para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em P. D. Pedro II.

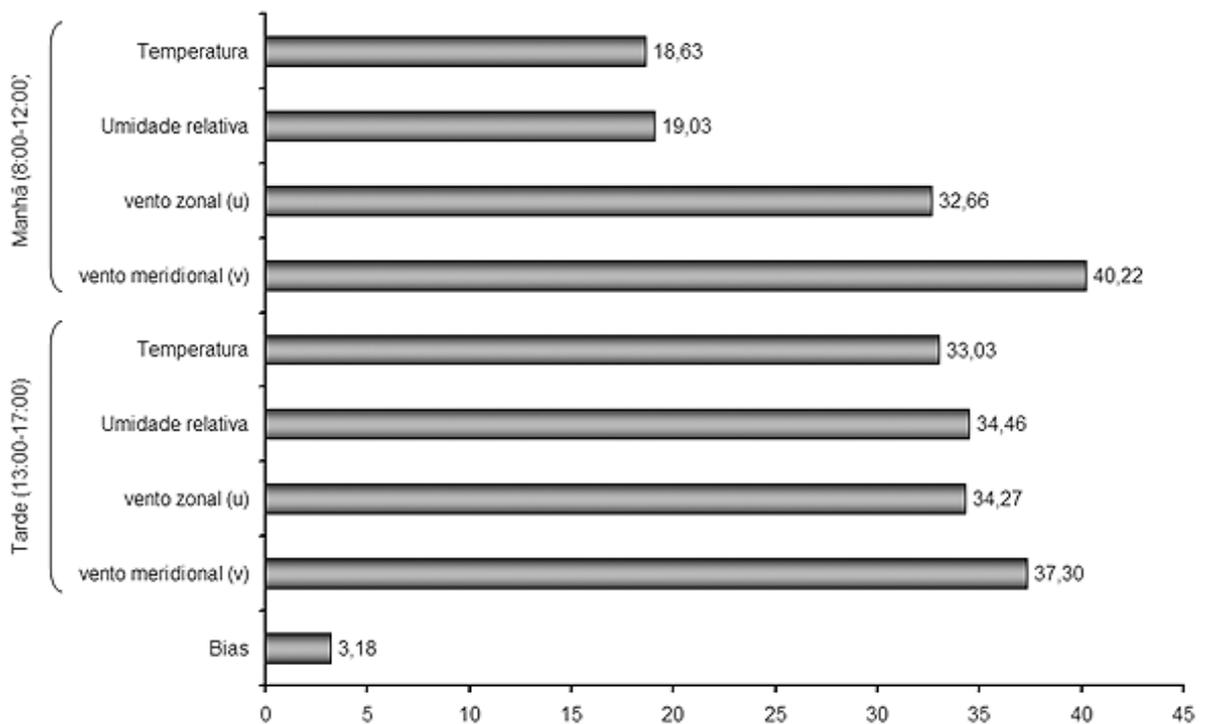


Figura 116: Soma dos valores absolutos dos pesos entre cada variável de entrada e os neurônios da camada oculta para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em P. D. Pedro II.

Para essa rede neural não foi possível realizar a validação, pois não houve dados para esse fim. Em 2004 as medições foram feitas apenas em meados de janeiro a começo de fevereiro, em que somente sete eventos puderam ser selecionados. Em 2005 e 2006 houve apenas medidas de ozônio, mas nenhuma de temperatura e umidade. Além disso, a estação foi transferida para outra localidade.

A falta de dados também afeta o teste de aplicação da rede neural. Embora as previsões meteorológicas estivessem disponíveis, não foi possível fazer uma comparação adequada das simulações com as observações, devido à mudança de local da estação, o que afeta a natureza dos dados. Como compensação, as previsões de ozônio foram comparadas com as observações de Ibirapuera (figuras 117 e 118). Entretanto, P.D. Pedro II é característica por apresentar baixas concentrações por estar localizada próxima à via de tráfego, enquanto Ibirapuera, dentro do parque, registra níveis mais elevados. Sendo assim, diversos eventos aparecem subestimados nessa comparação. Apesar disso, a simulação procura acompanhar bem as tendências de aumento e redução do poluente (figura 117). Os coeficientes de determinação obtidos (figura 118) ficaram um pouco baixos em comparação às demais redes neurais, valendo 0,51 e 0,49, respectivamente para concentrações médias e máximas de ozônio. Esta estação, no entanto, tem pouco interesse em relação à poluição por ozônio na RMSP, devido às características de sua localização, que resultam em baixos valores de ozônio medido.

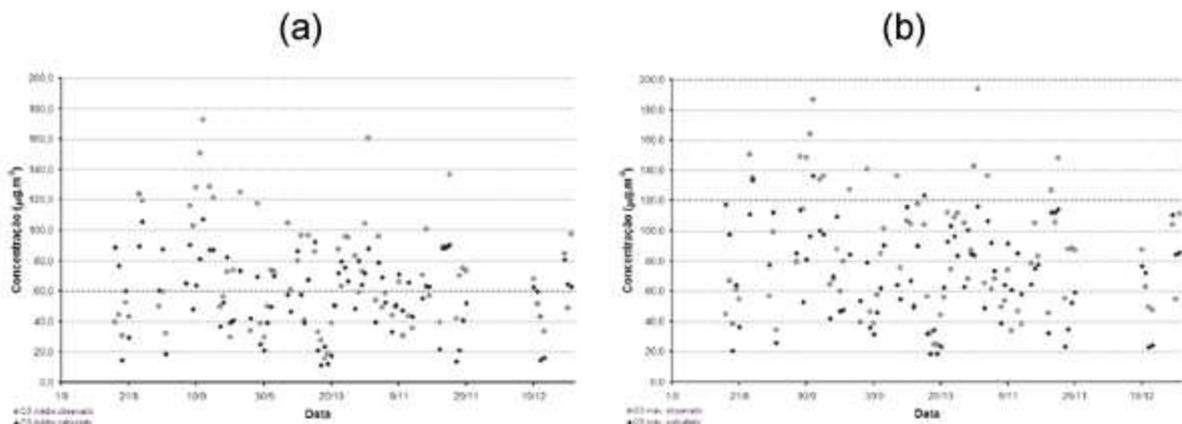


Figura 117: Evolução das concentrações médias de ozônio (a) e máximas (b), valores previstos e observados, ao longo do segundo semestre de 2006 para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em P. D. Pedro II.

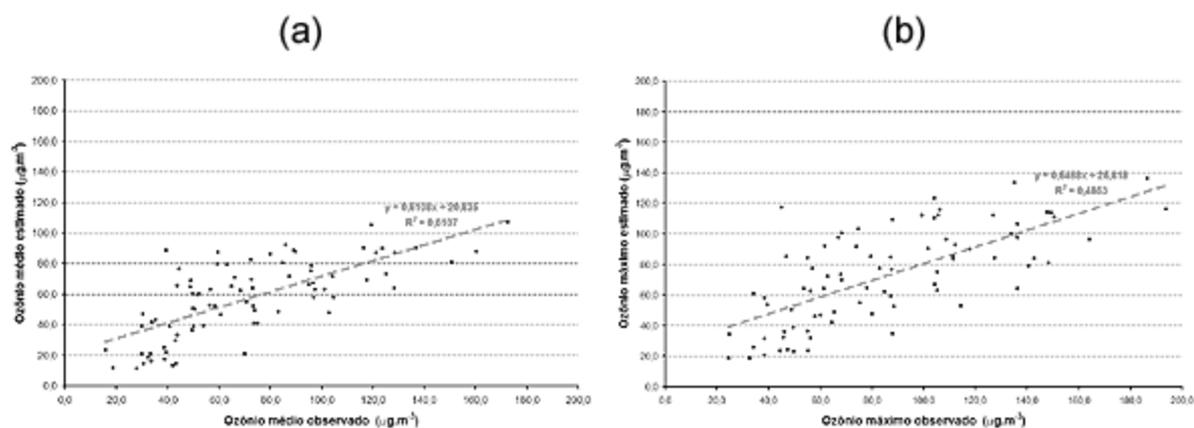


Figura 118: Relação entre valores calculados e medidos da concentração média (a) e máxima (b) de ozônio para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em P. D. Pedro II.

5.5 - Estação de monitoramento de qualidade do ar em São Caetano do Sul

A rede neural para São Caetano do Sul foi treinada com 889 eventos no conjunto de dados de aprendizado e 175 casos para teste. Nas comparações entre os valores medidos e os simulados (figuras 119 e 120) observa-se uma menor dispersão nos resultados de concentrações médias de ozônio do que nas de máximas. Os coeficientes de determinação foram de, respectivamente, 0,75 e 0,69. A soma absoluta dos pesos (figura 121) aponta a média do vento zonal de manhã como a variável mais representativa, seguida pela média da componente meridional à tarde.

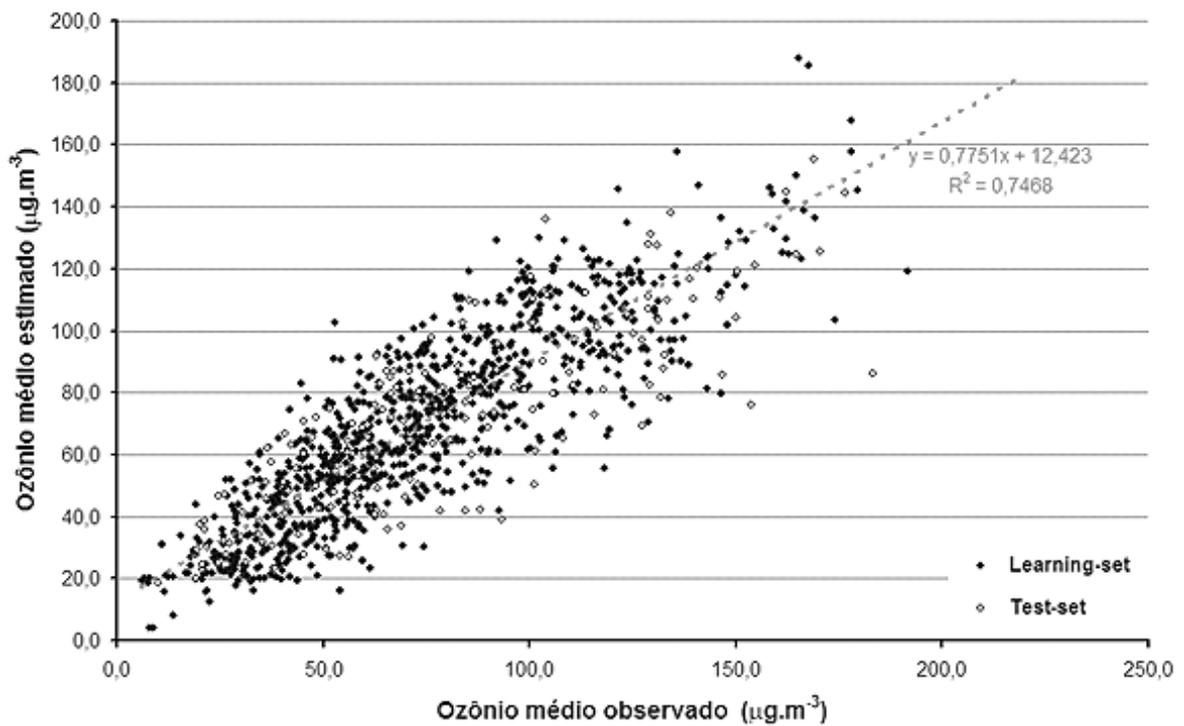


Figura 119: Relação entre valores calculados e medidos da concentração média de ozônio para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em São Caetano do Sul.

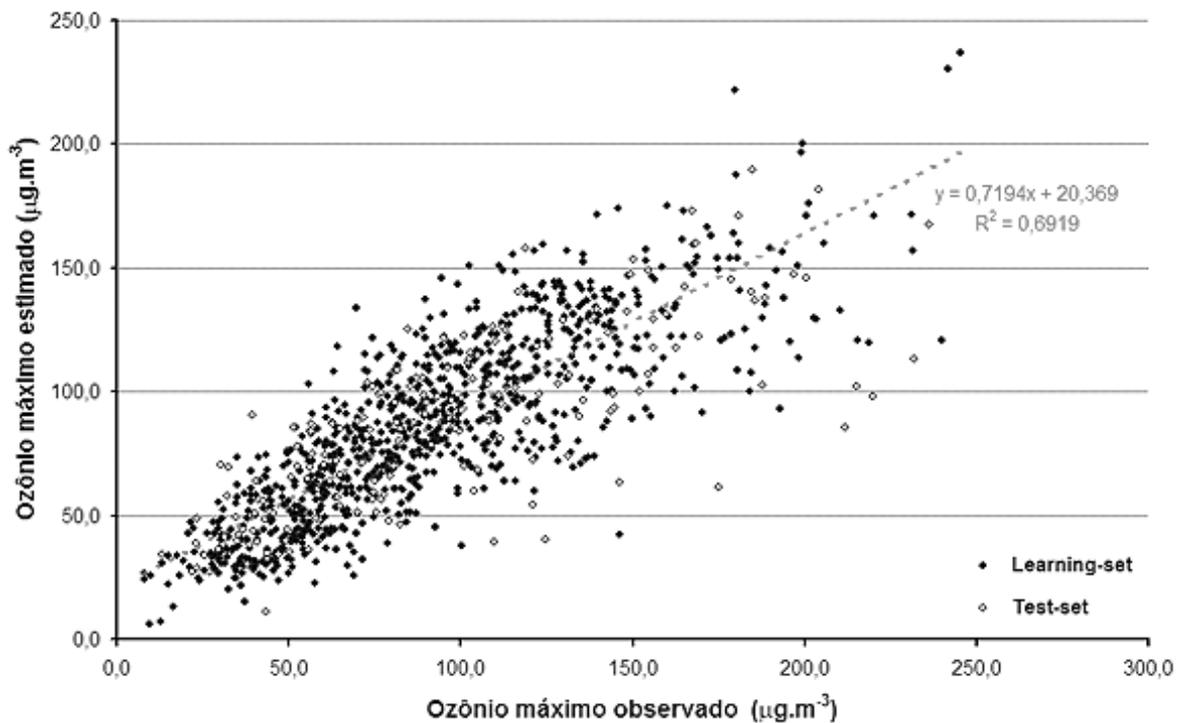


Figura 121: Relação entre valores calculados e medidos da concentração máxima de ozônio para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em São Caetano do Sul.

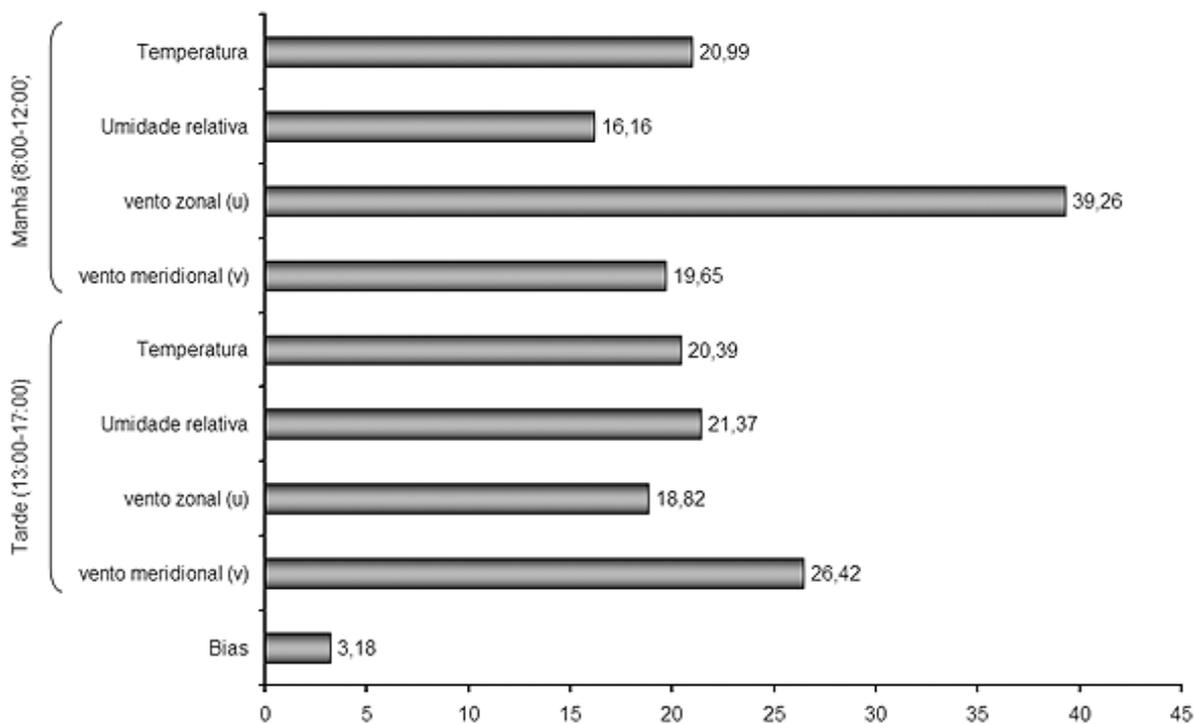


Figura 122: Soma dos valores absolutos dos pesos entre cada variável de entrada e os neurônios da camada oculta para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em São Caetano do Sul.

A validação dessa rede neural consistiu da aplicação de 202 casos ao longo de 2004, resultando nos gráficos das figuras 123 a 126. Tanto pelas figuras 123 e 124, comparando simulações e medições de concentrações médias e máximas de ozônio ao longo do tempo, observa-se uma boa concordância entre os dados, sem muitas ocorrências de diferenças marcantes. Nas figuras 125 e 126, relacionando observações com os valores calculados por evento, a dispersão entre os pontos é pequena, e os coeficientes de determinação foram de 0,69 e 0,75, respectivamente para concentrações médias e máximas de ozônio. Tal resultado é bastante próximo ao obtido no processo de treinamento da rede.

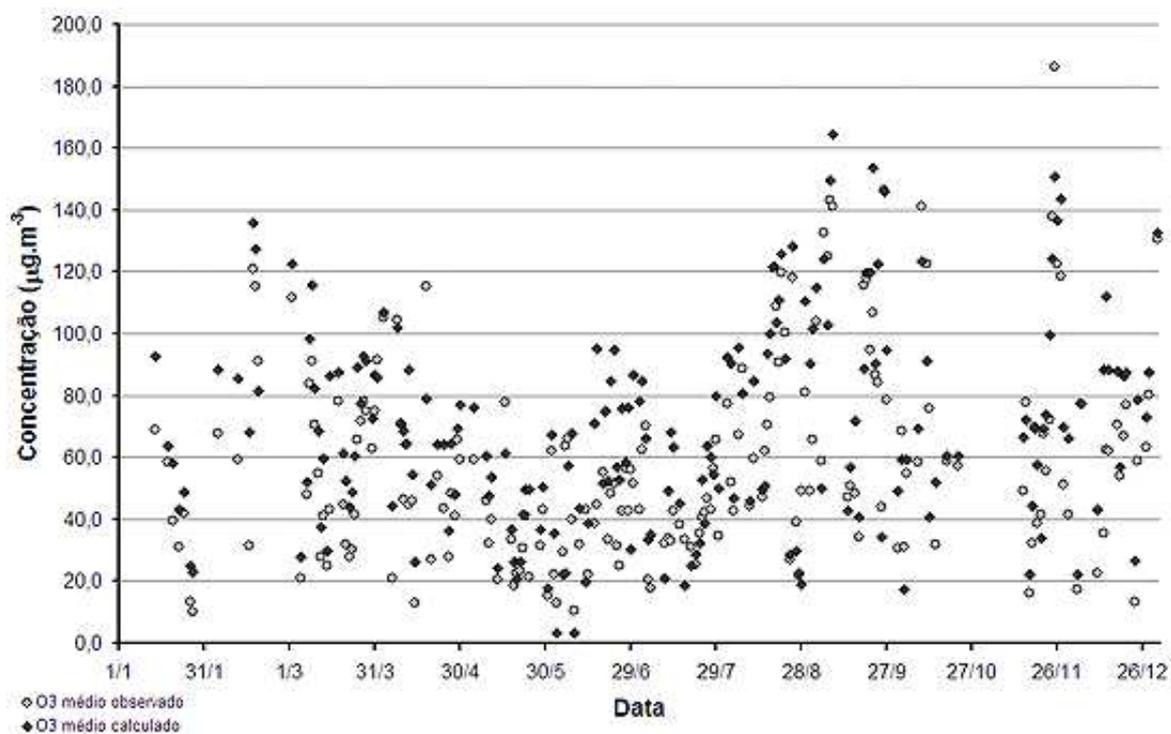


Figura 123: Evolução das concentrações médias de ozônio, valores previstos e observados, ao longo do ano de 2004 para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em São Caetano do Sul.

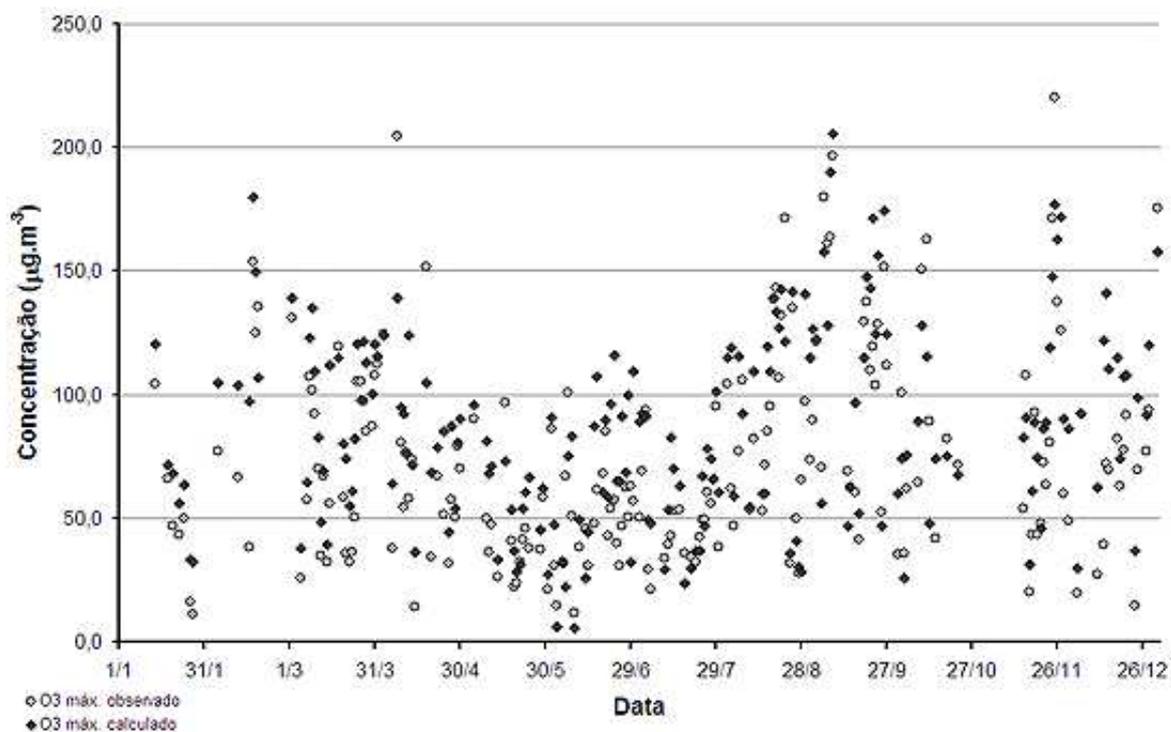


Figura 124: Evolução das concentrações máximas de ozônio, valores previstos e observados, ao longo do ano de 2004 para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em São Caetano do Sul.

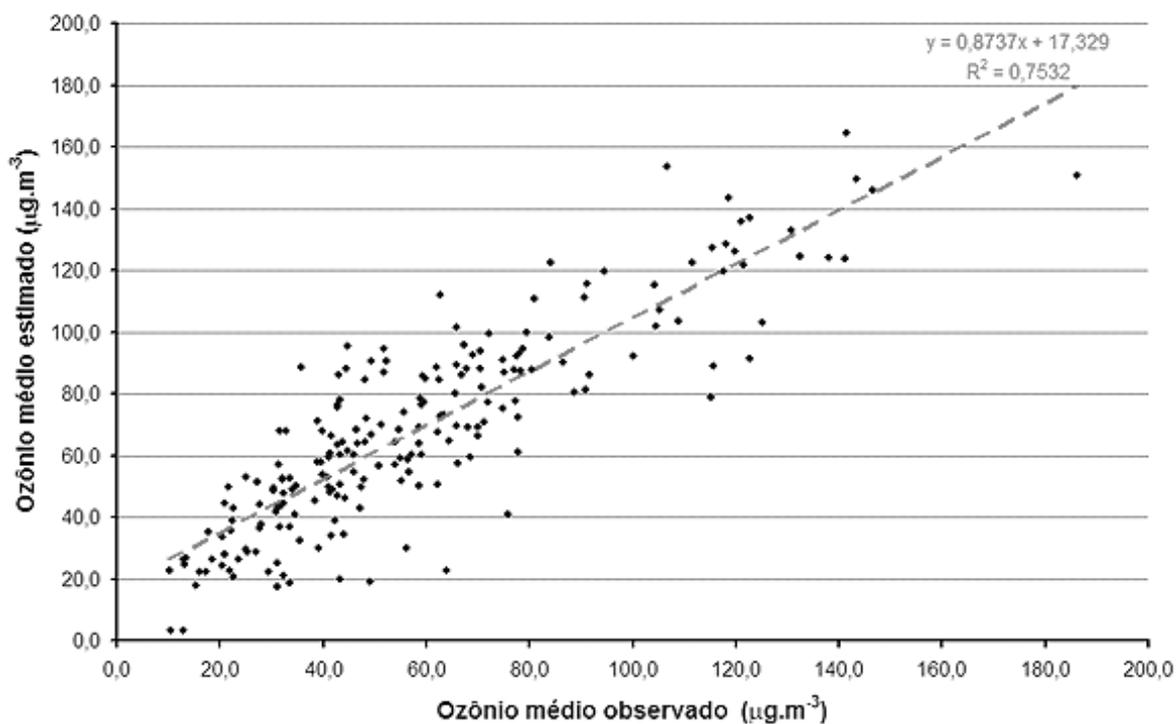


Figura 125: Relação entre valores calculados e medidos da concentração média de ozônio para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em São Caetano do Sul.

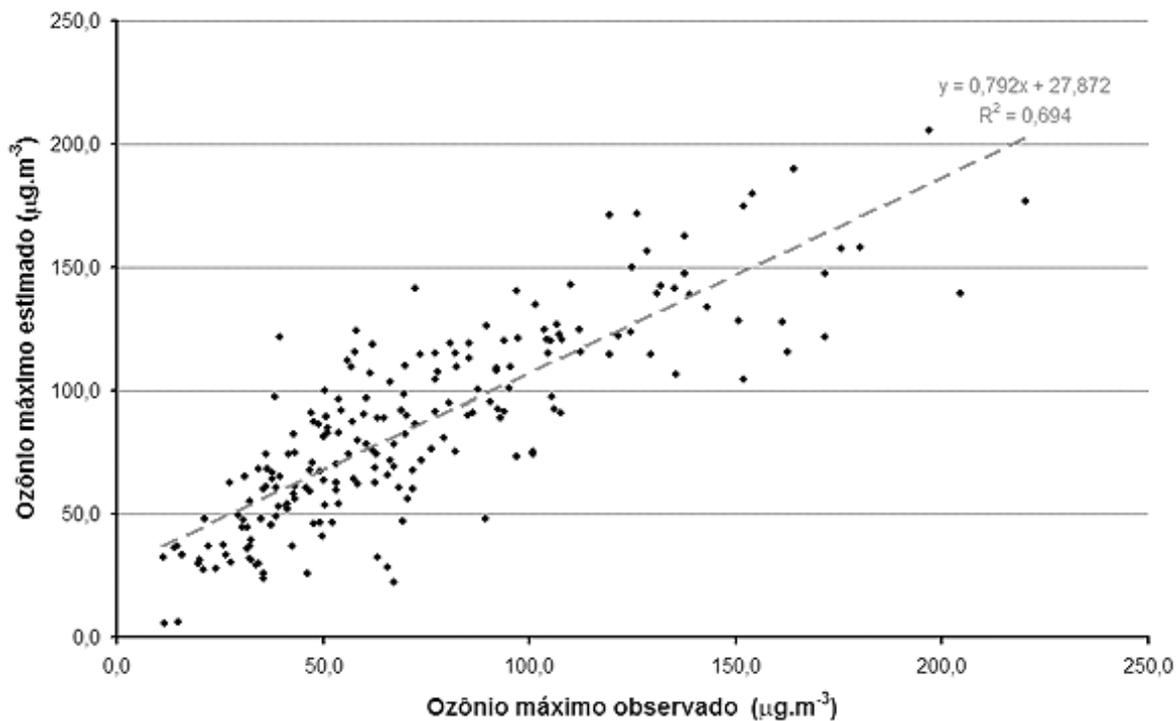


Figura 126: Relação entre valores calculados e medidos da concentração máxima de ozônio para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em São Caetano do Sul.

De acordo com os resultados das tabelas 5.5.1 e 5.5.2, há uma pequena tendência de superestimativa pela rede neural da ordem de $10 \mu\text{g.m}^{-3}$, sendo que os parâmetros de desempenho indicam uma média de erro de $18 \mu\text{g.m}^{-3}$ e $25 \mu\text{g.m}^{-3}$, respectivamente para concentrações médias e máximas de ozônio.

Tabela 5.5.1: Caracterização das medidas (obs.) e simulações (sim.) para concentrações médias e máximas de ozônio (em $\mu\text{g.m}^{-3}$).

	O3 médio		O3 máximo	
	Obs.	Sim.	Obs.	Sim.
MÉDIA	58,67	68,59	72,84	85,57
σ (desvio)	32,39	32,61	41,18	39,15
MÁXIMO	186,17	164,44	220,33	205,50
MÍNIMO	10,23	3,36	11,23	5,52

Tabela 5.5.2: Critérios de avaliação do desempenho da rede neural. *ME*: Erro médio ($\mu\text{g.m}^{-3}$); *RMSE*: Raiz do erro médio quadrático ($\mu\text{g.m}^{-3}$); *RMSE_{UB}*: Aptidão da simulação ($\mu\text{g.m}^{-3}$).

	O3 médio	O3 máximo
Correlação	0,87	0,83
ME	9,9	12,7
RMSE	19,39	26,48
RMSEub	16,67	23,23

Os desvios entre medidas e simulações são representados na figura 127, tanto para comparações de concentrações médias de ozônio quanto para máximas. Para as concentrações médias de ozônio não há um padrão marcante, pois as diferenças ficam entre -40 a $60 \mu\text{g.m}^{-3}$, sendo que para as concentrações máximas de ozônio alguns pontos ultrapassam essa faixa, estando entre -60 e $80 \mu\text{g.m}^{-3}$. Para as concentrações máximas de ozônio a tendência é de aumento dos desvios com a elevação das concentrações; em comparação aos valores observados, essa propensão é de subestimativa.

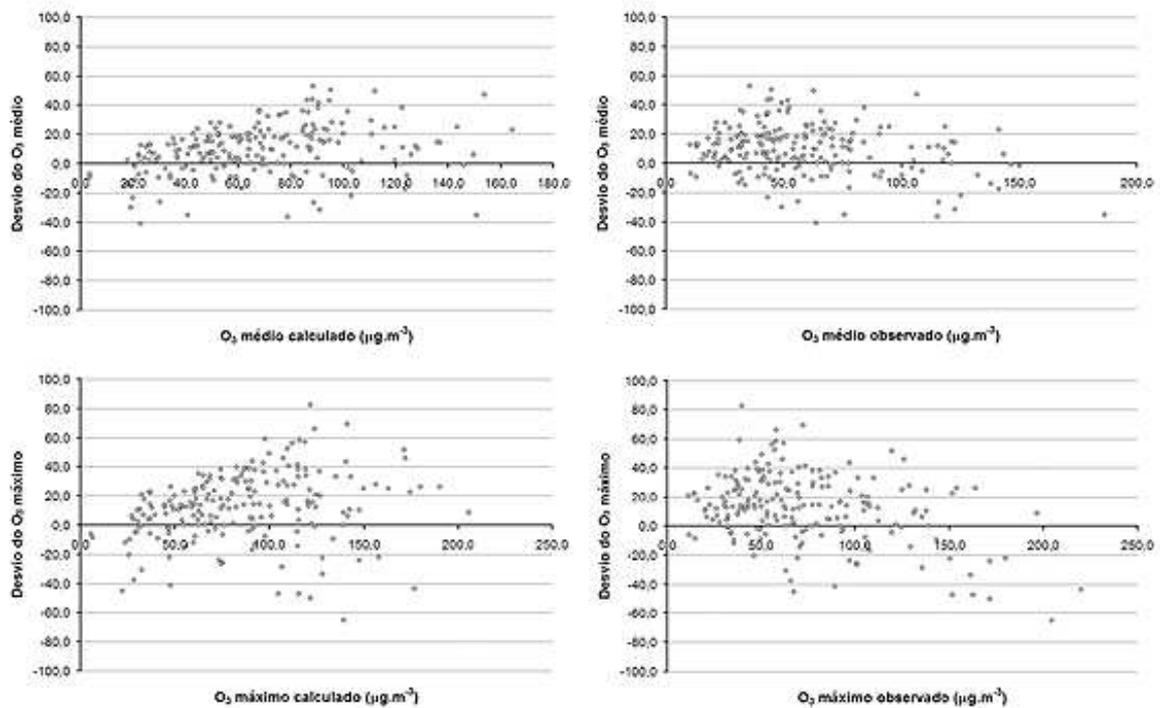


Figura 127: Relação entre os desvios (diferença entre valores calculados e observados) e as concentrações médias e máximas de ozônio para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em São Caetano do Sul.

As diferenças entre simulações e observações são comparadas com valores de temperatura (figura 128), umidade (figura 129) e vento (figuras 130 e 131), tanto para desvios de concentrações médias de ozônio como máximas. Na figura 128 os desvios nas estimativas de concentrações máximas de ozônio são maiores que para as concentrações médias, mas não há um padrão característico de variação de acordo com a temperatura. Na figura 129 observa-se o mesmo com relação à umidade; os desvios não variam muito com a evolução dessa variável. A variação dos desvios de acordo com o campo de vento (figuras 130 e 131) indica uma maior ocorrência de erros significativos para ventos de sul e sudoeste, e com velocidades acima de 2 m.s^{-1} .

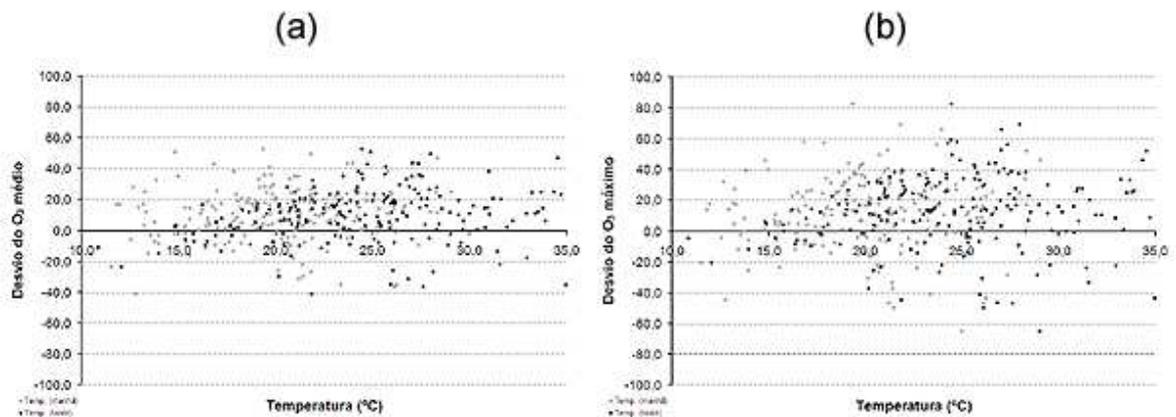


Figura 128: Relação entre os desvios das saídas de concentrações médias (a) e máximas (b) de ozônio e as médias de temperatura do ar (manhã e tarde) para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em São Caetano do Sul.

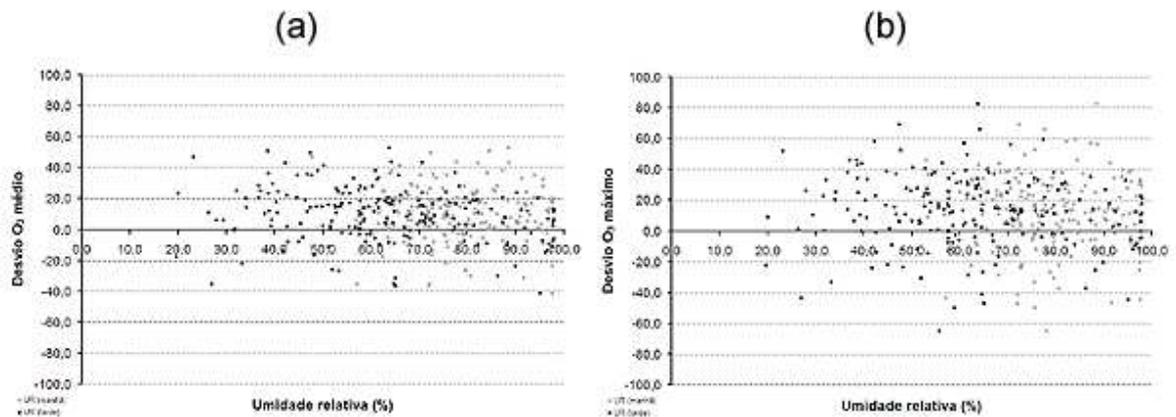


Figura 129: Relação entre os desvios das saídas de concentrações médias (a) e máximas (b) de ozônio e as médias de umidade relativa (manhã e tarde) para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em São Caetano do Sul.

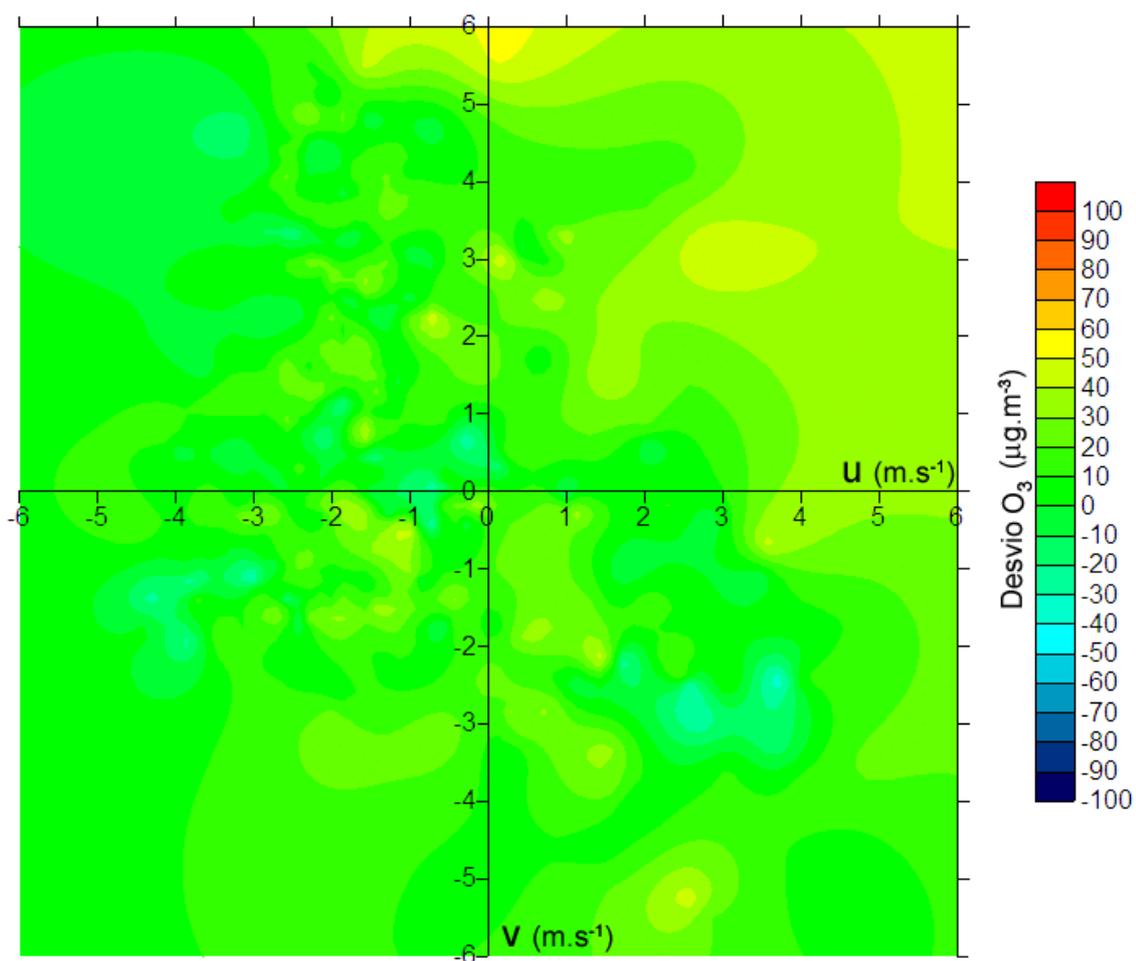


Figura 130: Distribuição dos desvios de concentrações médias de ozônio à tarde de acordo com o campo de vento (expresso por suas componentes u e v , de manhã e de tarde) para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em São Caetano do Sul.

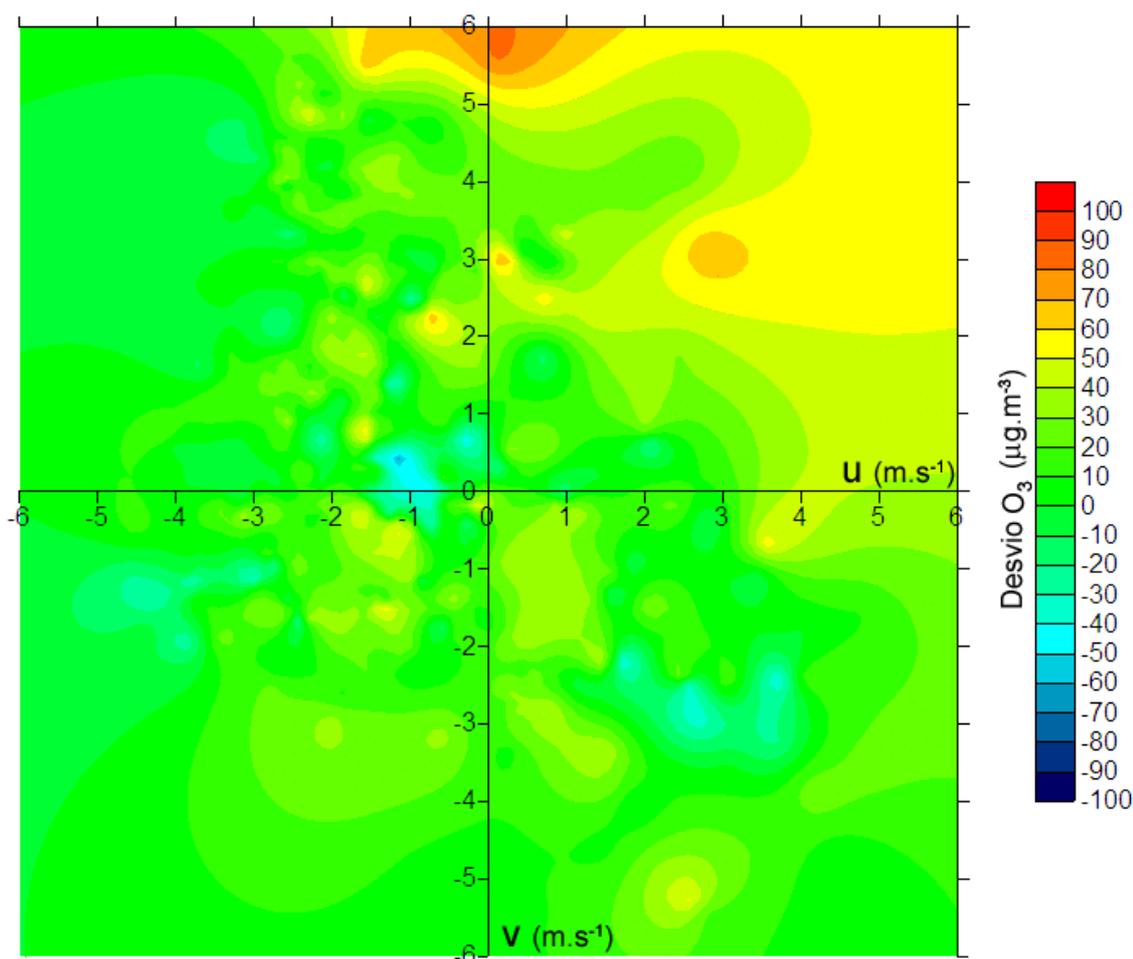


Figura 131: Distribuição dos desvios de concentrações máximas de ozônio à tarde de acordo com o campo de vento (expresso por suas componentes u e v , de manhã e de tarde) para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em São Caetano do Sul.

Foi feito um teste de aplicação dessa rede neural, utilizando as previsões meteorológicas do laboratório Master para São Caetano do Sul. Entretanto, os resultados não puderam ser comparados com as observações dessa estação, pois apenas 10 casos estiveram disponíveis dentro do período de Agosto a Dezembro de 2006. Portanto, a comparação foi feita em relação a outras duas estações. A primeira foi Ibirapuera, por ser a estação de referência neste estudo; a segunda foi Diadema, pois na avaliação estatística (item 5.1) esta se mostrou associada com São Caetano do Sul.

Os resultados são apresentados graficamente nas figuras 132 a 135. Com o decorrer do semestre, tanto para concentrações médias de ozônio (figura 132) como para máximas (figura

133), a comparação entre ambas as estações não apresenta grandes discrepâncias. Em geral, as observações de Ibirapuera apresentam concentrações mais elevadas que Diadema, portanto alguns resultados da rede neural aparecem subestimados com relação a primeira e superestimados em comparação a segunda. Na figura 134, a associação entre a previsão e os dados das duas estações não apresenta grandes diferenças. Para Ibirapuera foi obtido um coeficiente de determinação de 0,59 enquanto para Diadema foi de 0,58. Já para a figura 135, em relação às concentrações máximas de ozônio, a comparação com Ibirapuera foi um pouco melhor, apresentando R^2 de 0,58, contra o 0,53 para Diadema.

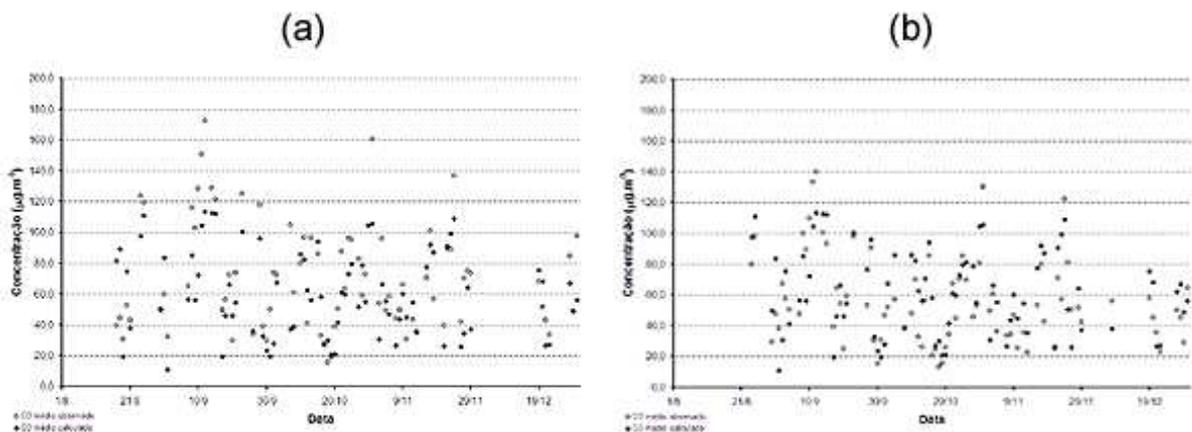


Figura 132: Evolução das concentrações médias de ozônio ao longo do segundo semestre de 2006, comparando os valores previstos pela rede neural e os observados pelas estações (a) Ibirapuera e (b) Diadema.

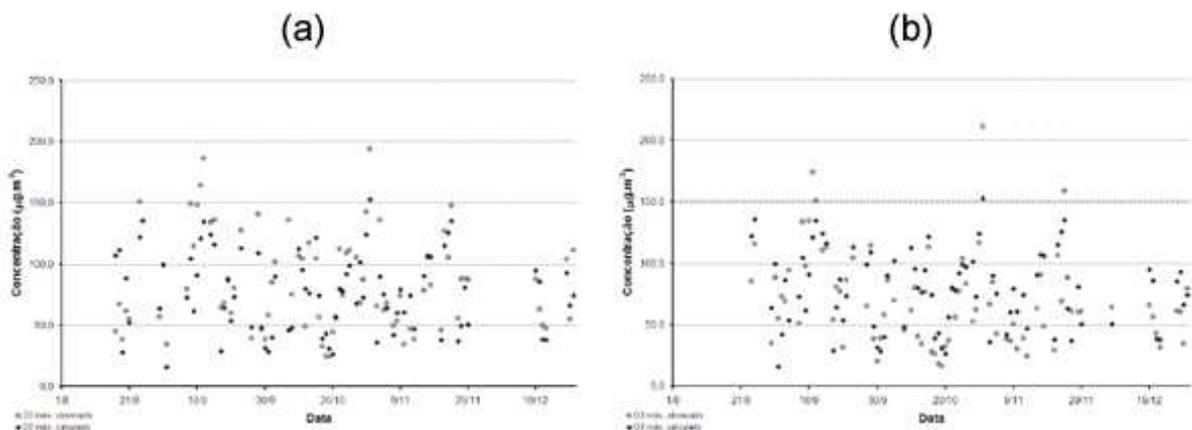


Figura 133: Evolução das concentrações máximas de ozônio ao longo do segundo semestre de 2006, comparando os valores previstos pela rede neural e os observados pelas estações (a) Ibirapuera e (b) Diadema.

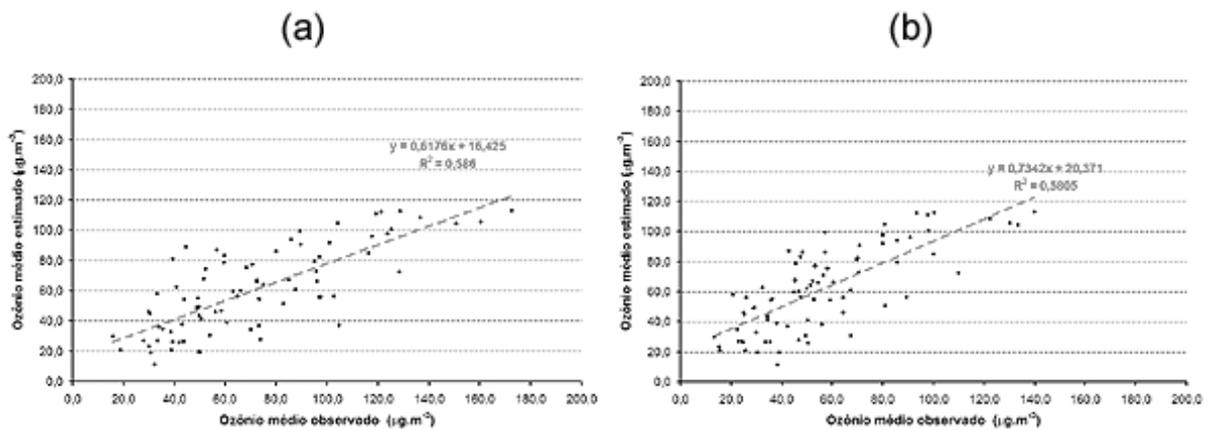


Figura 134: Relação entre previsão de rede neural e medidas de (a) Ibirapuera e (b) Diadema, para concentrações médias de ozônio à tarde.

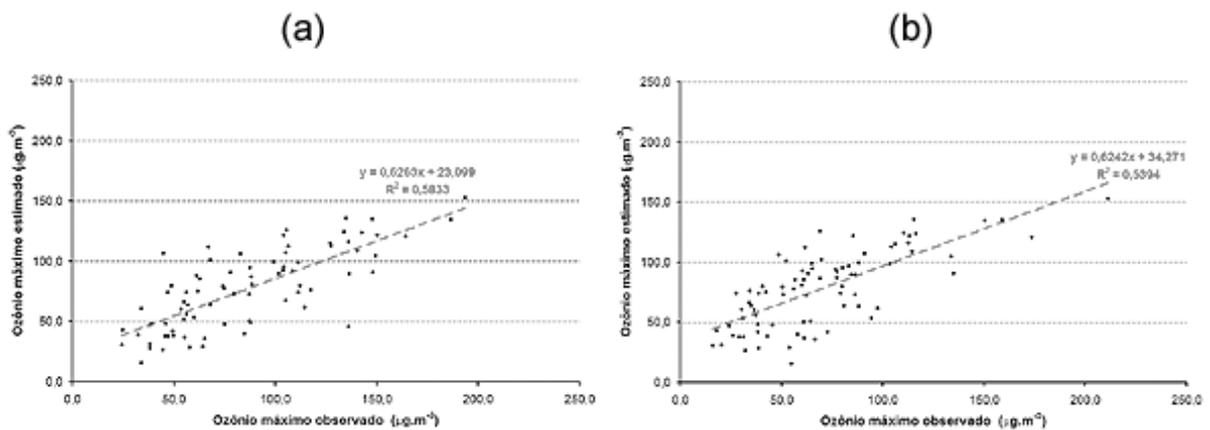


Figura 135: Relação entre previsão de rede neural e medidas de (a) Ibirapuera e (b) Diadema, para concentrações máximas de ozônio à tarde.

Na tabela 5.5.3 são apresentados alguns resultados de indicadores de desempenho. Pelo erro médio (*ME*) a comparação com Ibirapuera denota tendência da rede em subestimar os resultados ($-9,9 \mu\text{g.m}^{-3}$, em média), porém, com relação a Diadema, ocorre uma superestimativa ($6,6 \mu\text{g.m}^{-3}$, em média). Os erros, segundo os parâmetros *RMSE* e *RMSE_{ub}* são em média 8,1% maiores para Ibirapuera do que Diadema.

Tabela 5.5.3: Critérios de avaliação do desempenho da rede neural. *ME*: Erro médio ($\mu\text{g.m}^{-3}$); *RMSE*: Raiz do erro médio quadrático ($\mu\text{g.m}^{-3}$); *RMSE_{UB}*: Aptidão da simulação ($\mu\text{g.m}^{-3}$).

Comparação com Ibirapuera			Comparação com Diadema		
	O3 médio	O3 máximo		O3 médio	O3 máximo
Correlação	0,77	0,76	Correlação	0,76	0,73
ME	-10,8	-8,9	ME	5,4	7,8
RMSE	24,71	26,86	RMSE	19,94	26,58
RMSEub	22,21	25,33	RMSEub	19,20	25,40

5.6 - Estação de monitoramento de qualidade do ar em Santana

Para a estação de Santana foi ajustada uma rede neural utilizando 861 eventos no conjunto de treinamento e 172 no de teste, gerando-se os gráficos das figuras 136 a 138. Entretanto, como essa estação não dispõe de medidas de temperatura do ar e umidade relativa, foi necessário recorrer aos dados de P. D. Pedro II, por ser a mais próxima com essas medições.

Na comparação entre medidas e simulações de concentrações médias de ozônio (figura 136) a dispersão dos pontos é pequena e o coeficiente de determinação obtido foi de 0,81. Em relação às concentrações máximas de ozônio (figura 137), alguns eventos, especialmente em concentrações acima de $200 \mu\text{g}\cdot\text{m}^{-3}$, não ficaram bem correlacionados, mas de maneira geral os demais eventos apresentaram boa concordância entre observação e simulação, resultando em um coeficiente R^2 de 0,73. A soma absoluta dos pesos, indicada pela figura 138, coloca as médias do vento zonal de manhã e da temperatura à tarde como as variáveis mais importantes no ajuste, seguidas pela umidade relativa e vento zonal no período da tarde. A umidade relativa de manhã foi a variável de menor representatividade.

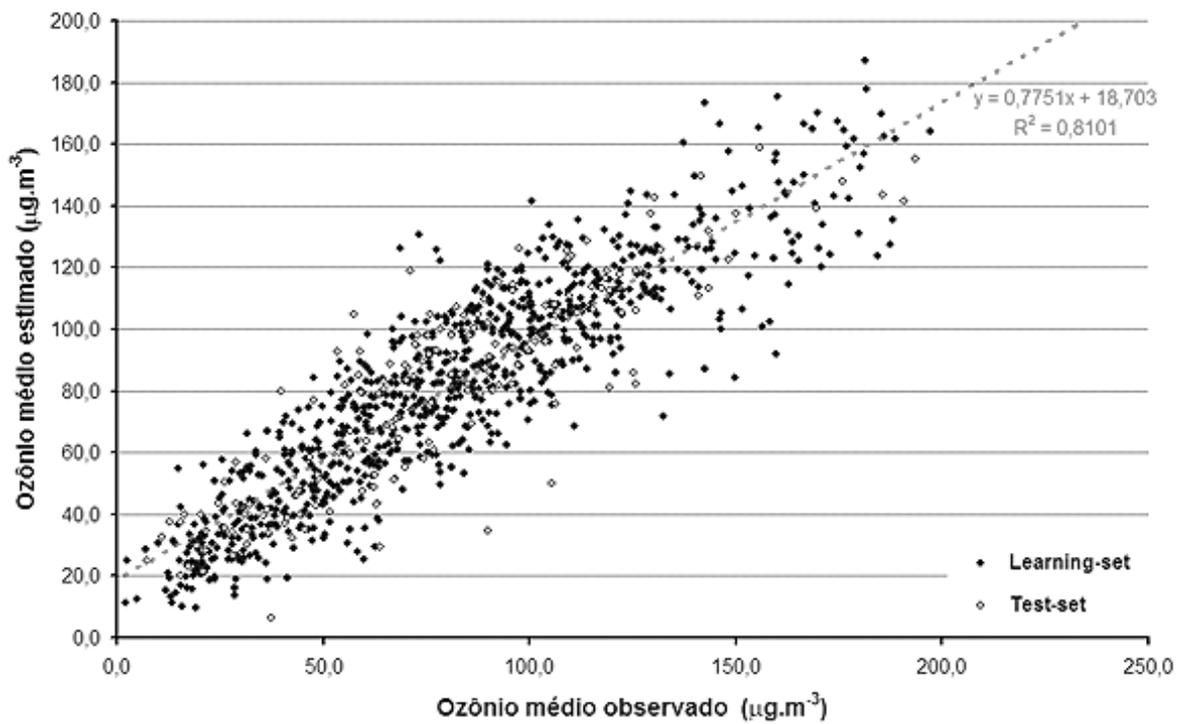


Figura 136: Relação entre valores calculados e medidos da concentração média de ozônio para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em Santana.

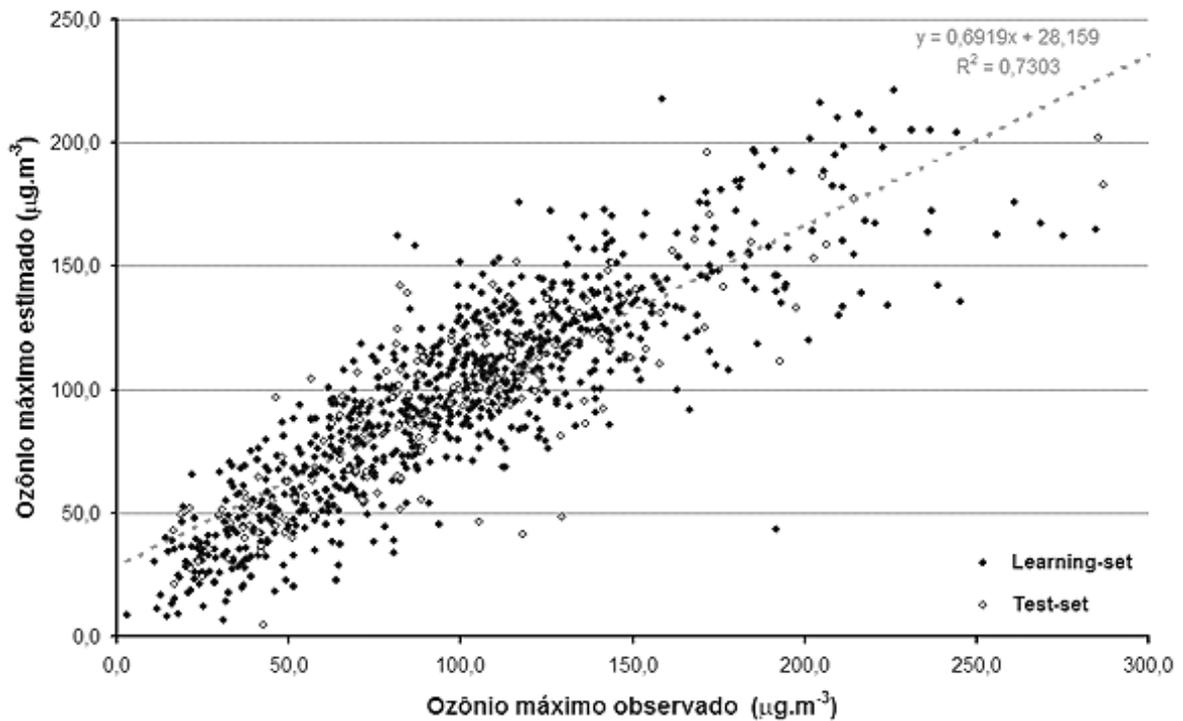


Figura 137: Relação entre valores calculados e medidos da concentração máxima de ozônio para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em Santana.

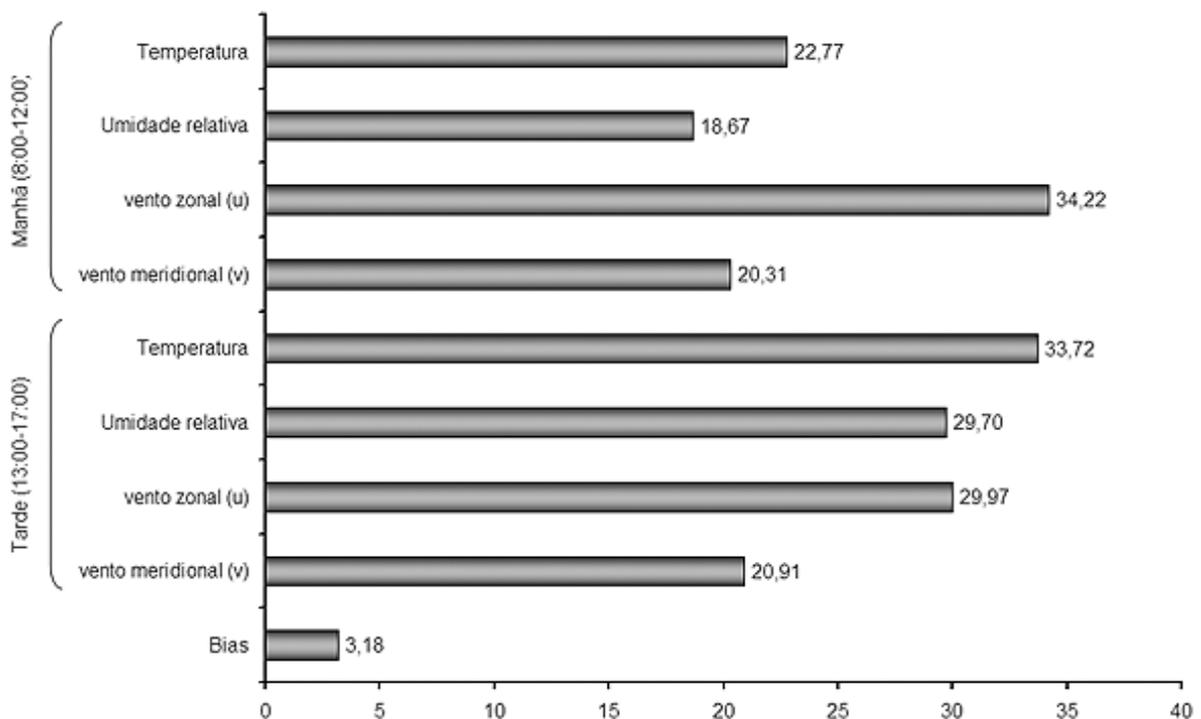


Figura 138: Soma dos valores absolutos dos pesos entre cada variável de entrada e os neurônios da camada oculta para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em Santana.

Por depender da temperatura e umidade de P. D. Pedro II, também não foi possível validar a rede neural de Santana, devido à falta de dados para estas variáveis. Já o teste de aplicação foi feito inserindo as previsões meteorológicas provindas do laboratório MASTER como entradas e comparando os resultados com as medidas de ozônio pela estação. Ao todo foram 57 eventos considerados.

As evoluções das concentrações médias e máximas de ozônio (figuras 139 e 140) apresentam boa concordância entre as estimativas da rede neural e as medições. O mesmo pode ser observado nas figuras 141 e 142, comparando as simulações com os valores medidos. Exceto por alguns casos isolados, os pontos não se dispersam muito, resultando em coeficientes de determinação de 0,62 e 0,57, respectivos às concentrações médias e máximas de ozônio.

Alternativamente, devido ao reduzido número de casos considerado nesse teste, foi feita uma comparação dos resultados da rede neural com as observações da estação Ibirapuera

(81 eventos). As simulações estiveram de acordo com os valores medidos. Porém, os desvios foram um pouco maiores, resultando em coeficientes de determinação de 0,46 e 0,45, em relação às concentrações médias e máximas de ozônio.

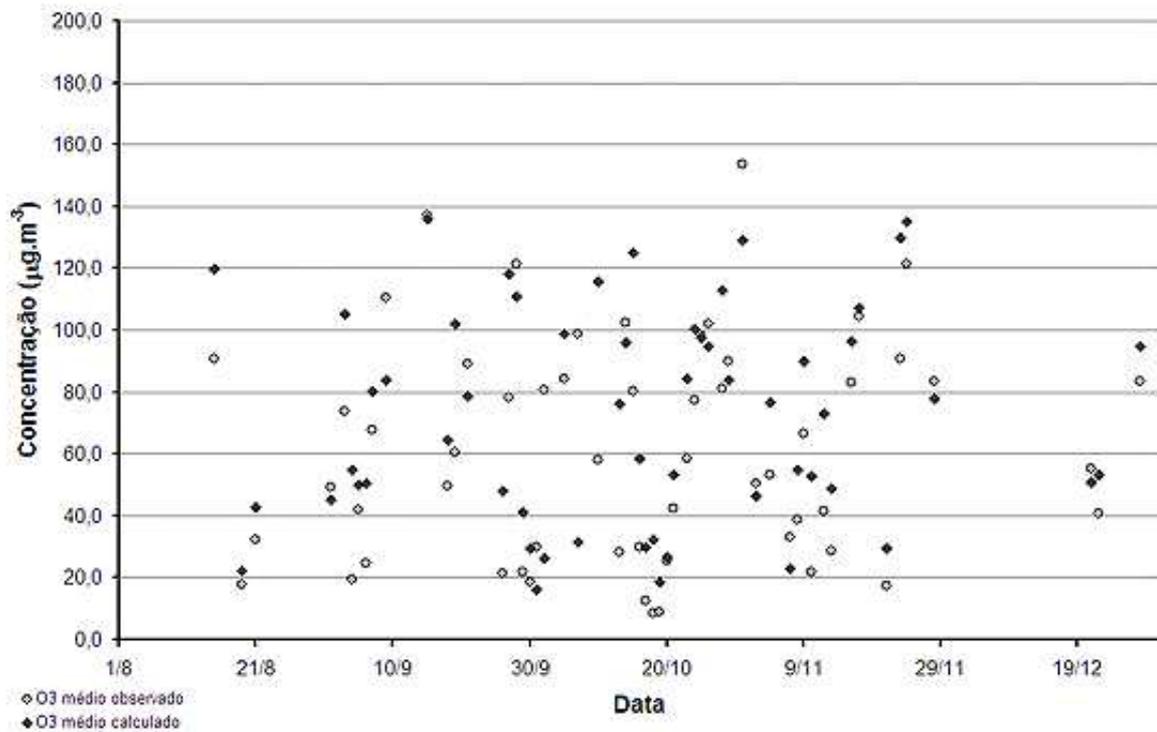


Figura 139: Evolução das concentrações médias de ozônio, valores previstos e observados, ao longo do segundo semestre de 2006 para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em Santana.

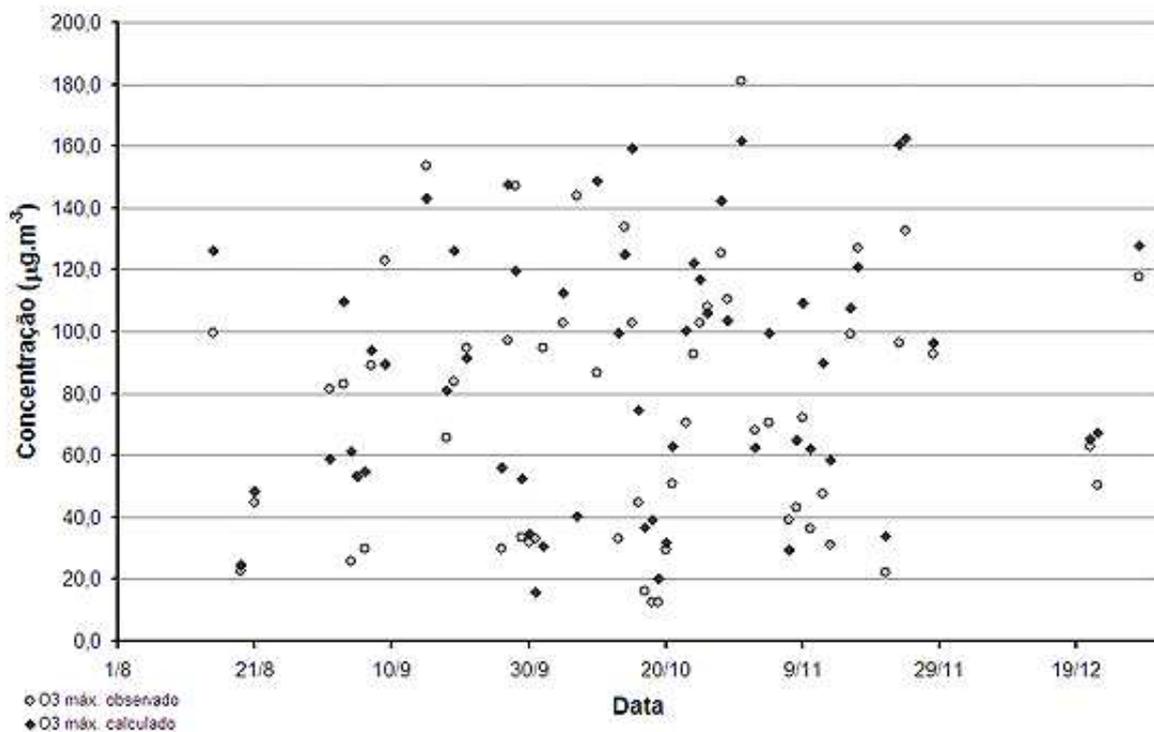


Figura 140: Evolução das concentrações máximas de ozônio, valores previstos e observados, ao longo do segundo semestre de 2006 para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em Santana.

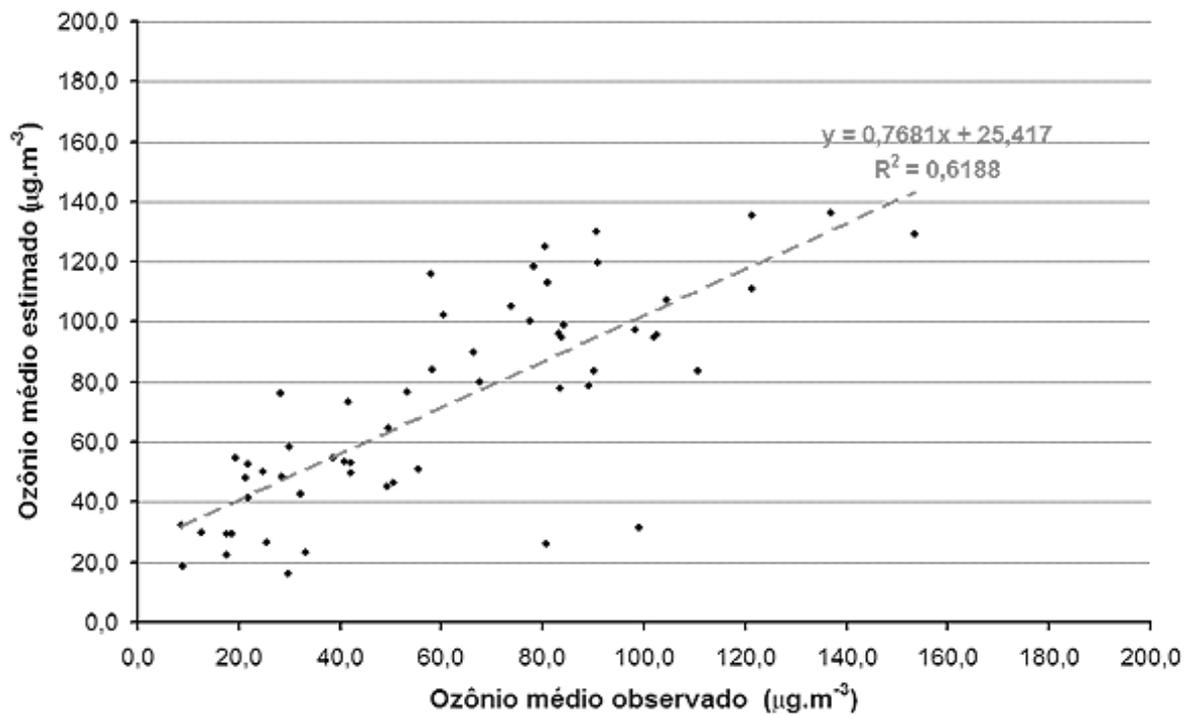


Figura 141: Relação entre valores calculados e medidos da concentração média de ozônio para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em Santana.

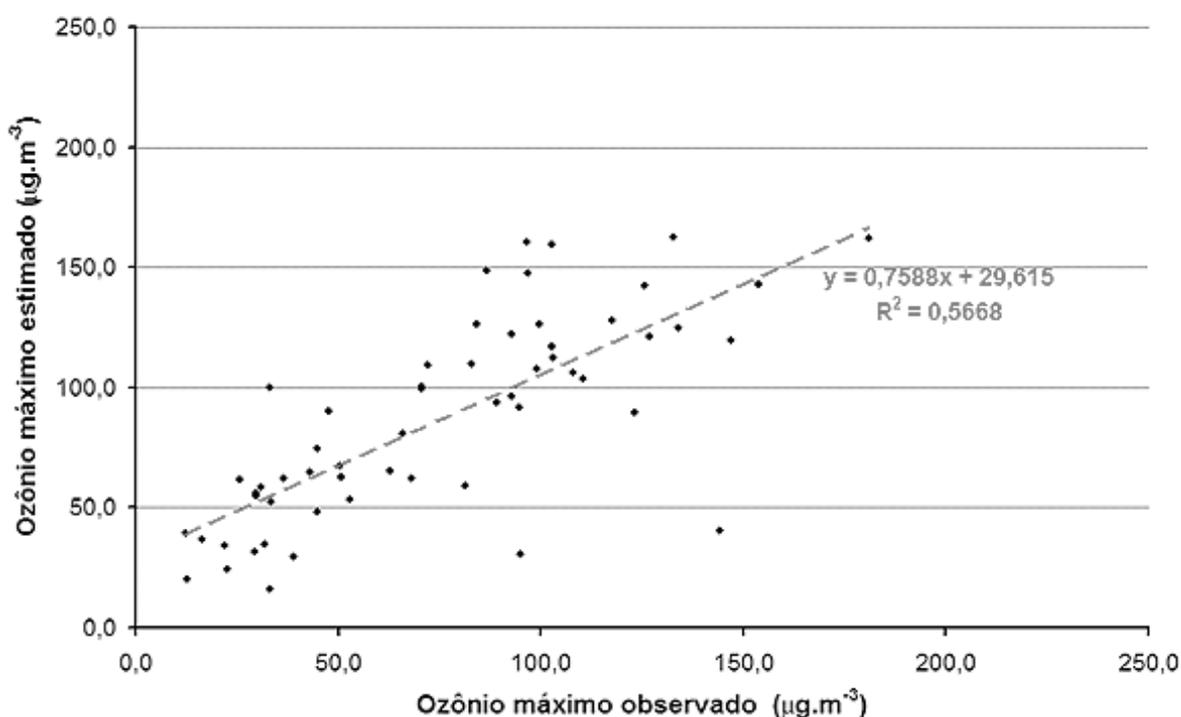


Figura 142: Relação entre valores calculados e medidos da concentração máxima de ozônio para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em Santana.

Os parâmetros de desempenho, descritos pelas tabelas 5.6.1 e 5.6.2, mostram que em média a rede erra em torno de $24 \mu\text{g.m}^{-3}$ nas concentrações médias de ozônio e $30 \mu\text{g.m}^{-3}$ para as máximas, sendo que a tendência é de superestimativa dos resultados em $11,3 \mu\text{g.m}^{-3}$. Paralelamente, as simulações de Santana foram comparadas com as medidas de ozônio em Ibirapuera. O erro médio foi um pouco menor, com $8,2 \mu\text{g.m}^{-3}$, em média, porém, segundo os parâmetros $RMSE$ e $RMSE_{ub}$, os erros foram de $29,0 \mu\text{g.m}^{-3}$ e $34,1 \mu\text{g.m}^{-3}$, em média, para as concentrações médias e máximas de ozônio, respectivamente.

Tabela 5.6.1: Caracterização das medidas (obs.) e simulações (sim.) para concentrações médias e máximas de ozônio (em $\mu\text{g.m}^{-3}$).

	O3 médio		O3 máximo	
	Obs.	Sim.	Obs.	Sim.
MÉDIA	61,23	72,44	75,21	86,69
σ (desvio)	35,39	34,56	41,20	41,53
MÁXIMO	153,54	136,06	180,89	162,57
MÍNIMO	8,54	16,04	12,46	15,83

Tabela 5.6.2: Critérios de avaliação do desempenho da rede neural. *ME*: Erro médio ($\mu\text{g.m}^{-3}$); *RMSE*: Raiz do erro médio quadrático ($\mu\text{g.m}^{-3}$); *RMSE_{UB}*: Aptidão da simulação ($\mu\text{g.m}^{-3}$).

	O3 médio	O3 máximo
Correlação	0,79	0,75
ME	11,2	11,5
RMSE	25,28	31,03
RMSEub	22,66	28,83

5.7 - Análise geral

Como critério de avaliação, foi adotado o coeficiente de determinação (R^2) obtido do ajuste de reta na comparação de valores medidos e calculados. De uma maneira geral, segundo este índice, as redes neurais apresentaram um bom desempenho de qualidade no ajuste, em especial para simulações de concentrações médias de ozônio, do que para as máximas do poluente (figura 143). As concentrações máximas de ozônio estão associadas aos eventos extremos de concentração do poluente, portanto, de ocorrência menos freqüente se comparados aos valores médios. Sendo assim, a rede neural é ajustada com maior ênfase para concentrações mais baixas, por estas serem mais freqüentes no conjunto de dados. Os eventos de concentração mais elevada acabam não sendo bem caracterizados, afetando assim os valores de R^2 .

Os menores coeficientes de determinação obtidos foram os da rede de P. D. Pedro II (figura 143). O fato se deve à localização física da estação, próxima de uma via de tráfego. Devido a isso, há uma maior quantidade de registros de baixa concentração do que das mais elevadas. Portanto, os poucos eventos mais intensos não foram suficientes para que a rede neural identificasse seu padrão de ocorrência. Assim, foram enfatizadas as concentrações mais baixas (menores que $150 \mu\text{g.m}^{-3}$). Os valores de R^2 mais baixos não significam que o ajuste tenha sido pior, mas sim que essa rede em particular não representa bem os eventos mais altos. De acordo com a figura 114 do item 5.4, a maioria dos pontos simulados corresponde adequadamente aos dados observados.

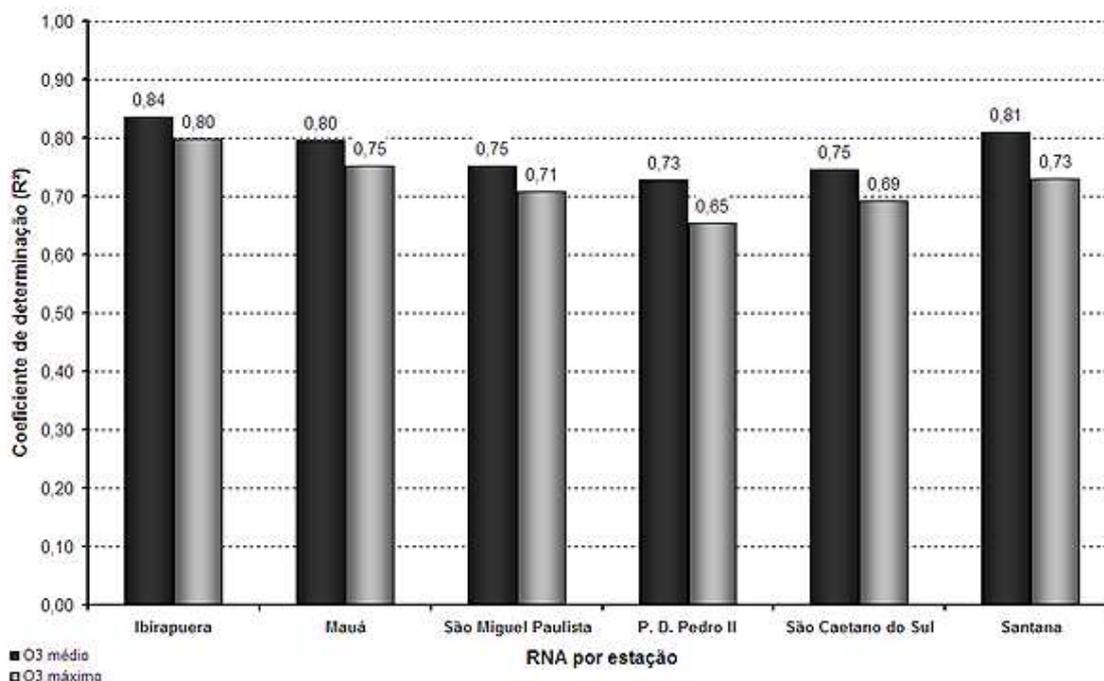


Figura 143: Coeficientes de determinação para concentrações médias (preto) e máximas (cinza) de ozônio em cada rede neural artificial de cada estação de qualidade do ar.

Outro parâmetro de avaliação adotado foi o cálculo do erro médio (*ME*) que, conforme descrito anteriormente, trata-se da diferença entre as médias das simulações e medições (equação 16 em 4.12). Em geral os desvios encontrados foram pequenos, na maioria das vezes com tendência das redes neurais em subestimar os valores das observações (figura 144). Em termos médios, as redes subestimam em $1,52 \mu\text{g.m}^{-3}$ as concentrações médias de ozônio, e em $2,89 \mu\text{g.m}^{-3}$ as máximas. Os maiores desvios foram os da RNA de São Caetano do Sul, com $-4,03 \mu\text{g.m}^{-3}$ e $-5,18 \mu\text{g.m}^{-3}$ para as concentrações médias e máximas de ozônio, respectivamente.

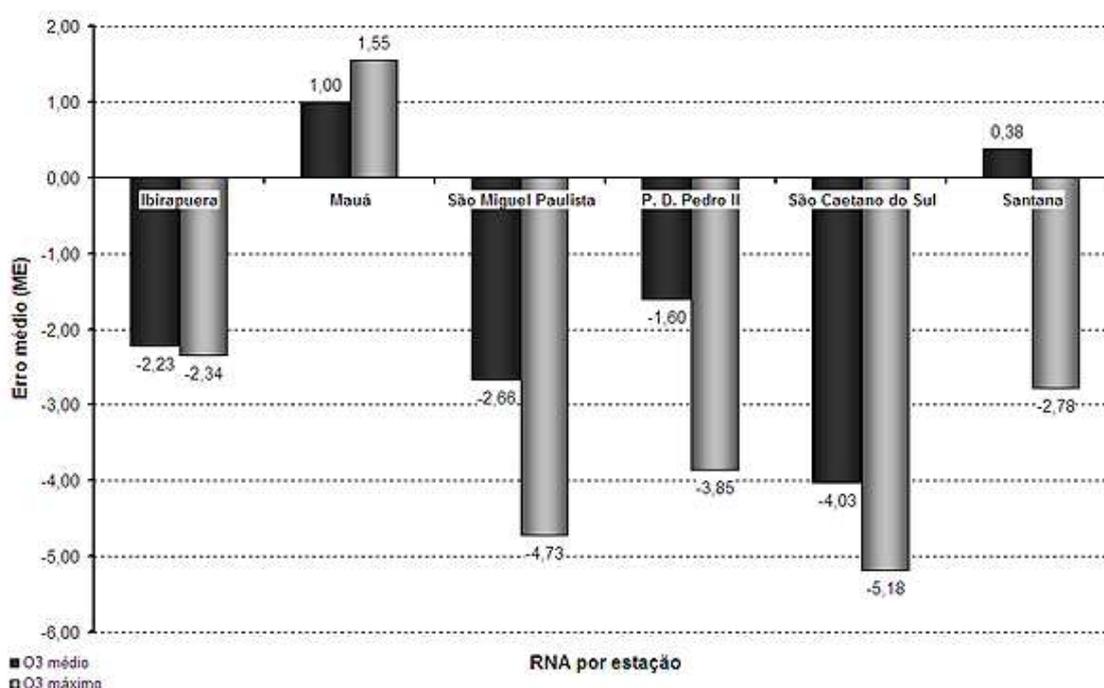


Figura 144: Erro médio (*ME*) para concentrações médias (preto) e máximas (cinza) de ozônio em cada rede neural artificial de cada estação de qualidade do ar.

Para uma medida da acurácia, foi calculada a raiz quadrada do erro médio quadrático (*RMSE*), que é a raiz da média quadrática das diferenças entre pares de simulações e medidas (equação 17 em 4.12). Tais erros estiveram dentro de um mesmo patamar para todas as redes neurais, ou seja, sem nenhuma se sobressaindo demais às outras (figura 145). Em média o *RMSE* foi de $17,01 \mu\text{g.m}^{-3}$ para as concentrações médias de ozônio e $24,29 \mu\text{g.m}^{-3}$ para as máximas. A rede que apresentou maior magnitude de erro foi Santana, com $17,99 \mu\text{g.m}^{-3}$ e $27,02 \mu\text{g.m}^{-3}$ para as concentrações médias e máximas de ozônio, respectivamente.

Outro indicador considerado foi o $RMSE_{ub}$, para avaliar a habilidade da simulação, representado pela raiz do erro médio quadrático após a remoção de um desvio médio (equação 18 em 4.12). Entretanto, no que se refere à comparação entre as redes neurais artificiais, os resultados não diferem muito do obtido para o índice *RMSE*. Em média o $RMSE_{ub}$ encontrado foi de $16,85 \mu\text{g.m}^{-3}$ e $24,02 \mu\text{g.m}^{-3}$, para concentrações médias e máximas de ozônio, respectivamente.

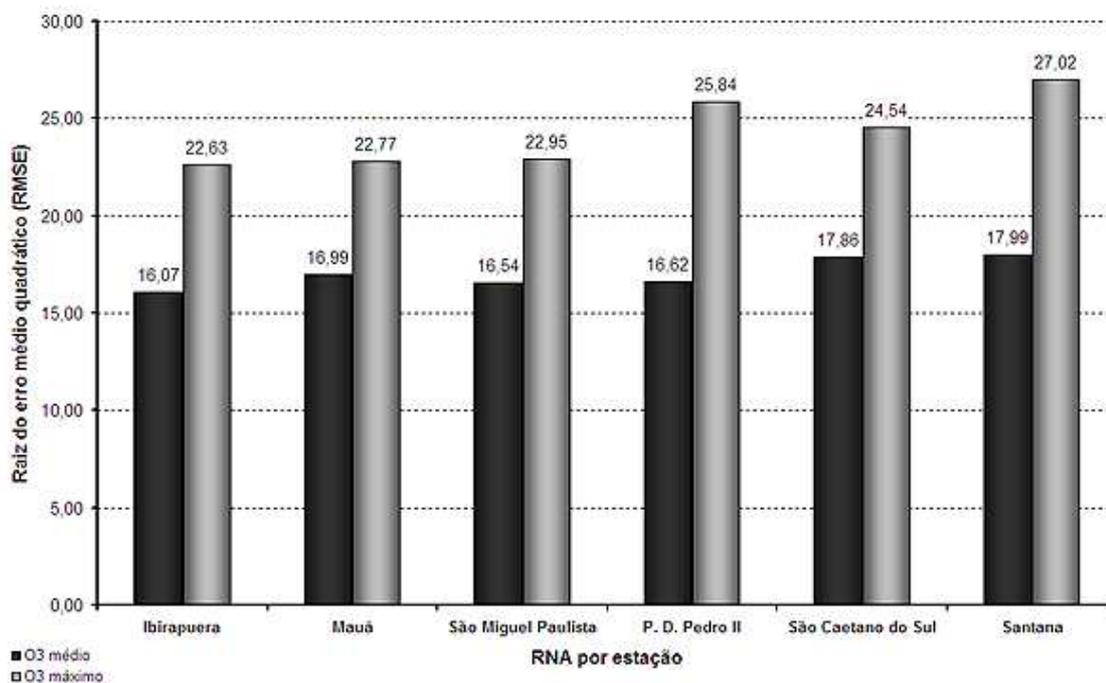


Figura 145: Raiz do erro médio quadrático (*RMSE*) para concentrações médias (preto) e máximas (cinza) de ozônio em cada rede neural artificial de cada estação de qualidade do ar.

Foram calculadas as somas absolutas dos pesos de cada variável de entrada, para avaliar sua importância na simulação de ozônio troposférico. Para fins comparativos, os resultados foram expressos em termos percentuais, de forma a indicar a magnitude de importância de cada variável nas diferentes redes neurais preparadas.

A temperatura do ar no período da tarde é mais importante na caracterização do ozônio de São Miguel Paulista do que para as demais localidades, representando 20,0% da soma absoluta dos pesos de sua RNA (figura 146). A temperatura também é importante para Mauá, tanto no período da manhã como da tarde, sendo respectivamente 17,2% e 17,4% da soma absoluta dos pesos. Santana (15,8%) e Ibirapuera (15,4%) são outras regiões onde a temperatura da tarde exerce uma influência significativa.

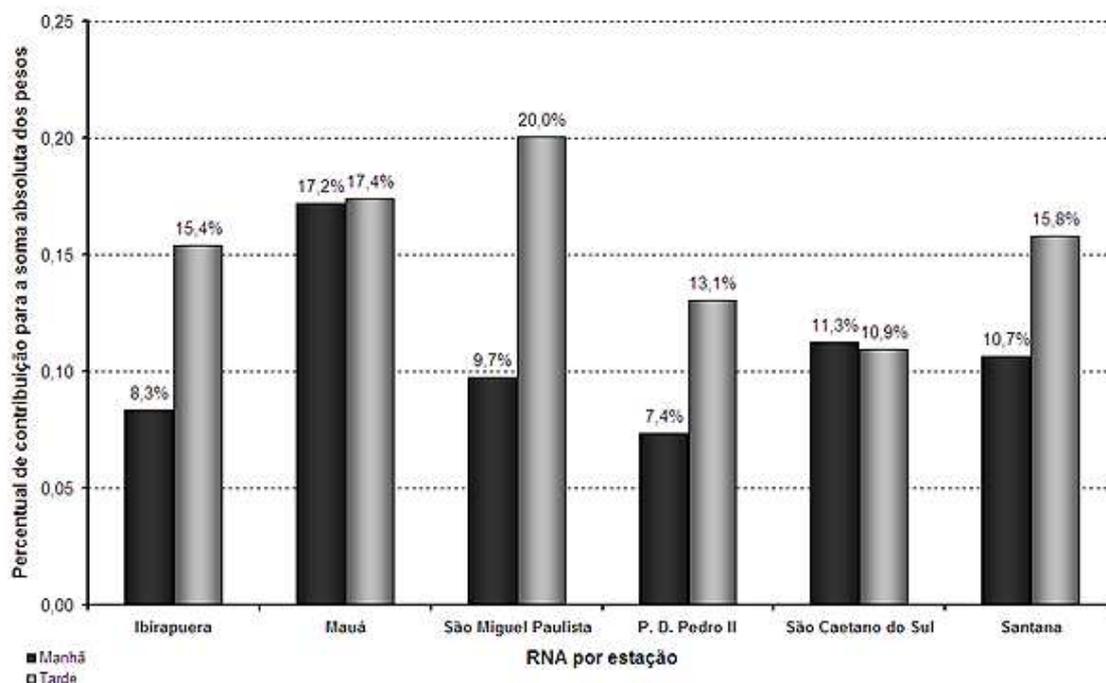


Figura 146: Percentual de contribuição da temperatura do ar para a soma absoluta dos pesos em cada rede neural artificial de cada estação de qualidade do ar.

De maneira geral a umidade relativa não influencia muito na simulação do ozônio pela rede neural (figura 147). Essa variável é mais representativa no período da tarde para Santana e P. D. Pedro II, se comparada às demais redes. No entanto, a contribuição é de apenas 13,9% e 13,6%, respectivamente. Em São Miguel Paulista e São Caetano esse percentual é de 11,9% e 11,5%, para cada estação à tarde.

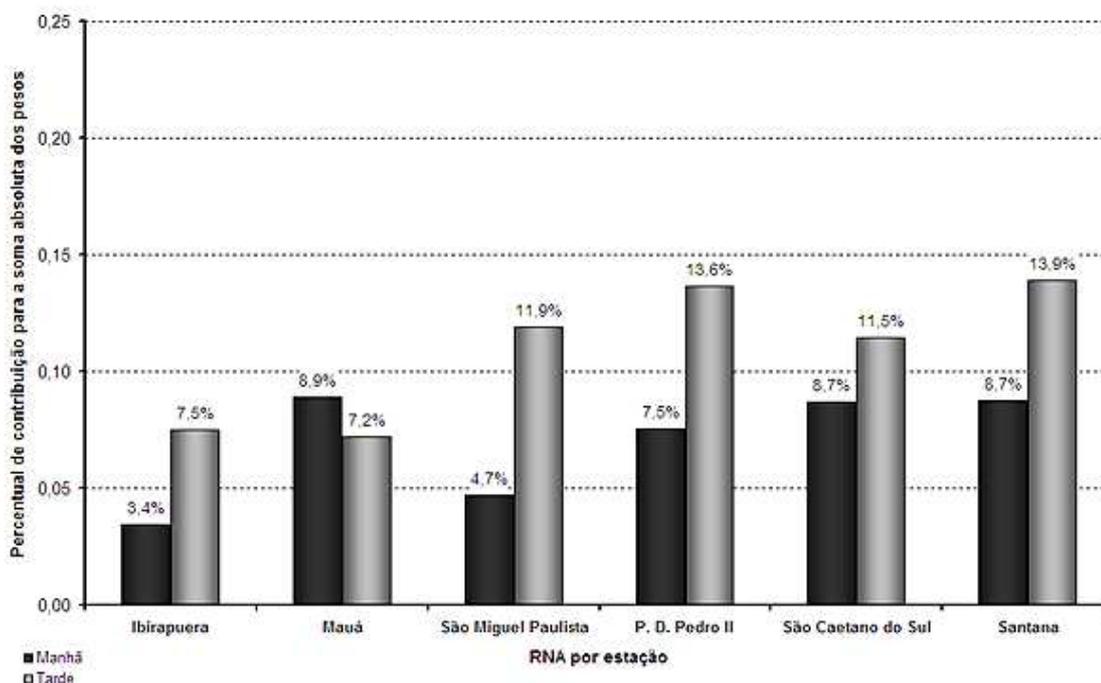


Figura 147: Percentual de contribuição da umidade relativa para a soma absoluta dos pesos em cada rede neural artificial de cada estação de qualidade do ar.

Em Ibirapuera e São Caetano do Sul a componente zonal do vento de manhã tem um papel importante na caracterização do ozônio, respondendo respectivamente por 21,2% e 21,1% da soma absoluta dos pesos (figura 148). Em Santana e São Miguel Paulista o vento zonal de manhã também é importante, com respectivas proporções de 16,0% e 15,1%. Em geral as entradas do período da tarde não são tão representativas.

A componente meridional do vento de manhã é mais importante para São Miguel Paulista e Ibirapuera do que para as demais redes neurais, com 22,1% e 19,5% da soma absoluta dos pesos (figura 149). Em P. D. Pedro II essa variável também é significativa, em proporção de 15,9% para as médias de manhã e 14,8% para tarde.

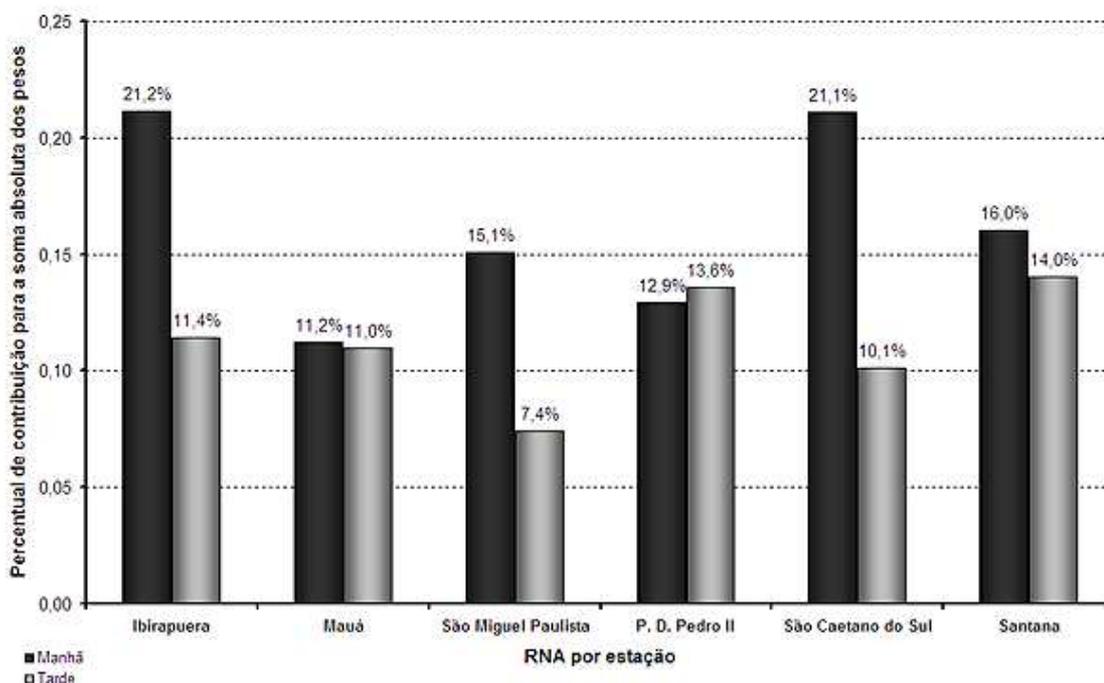


Figura 148: Percentual de contribuição do vento zonal para a soma absoluta dos pesos em cada rede neural artificial de cada estação de qualidade do ar.

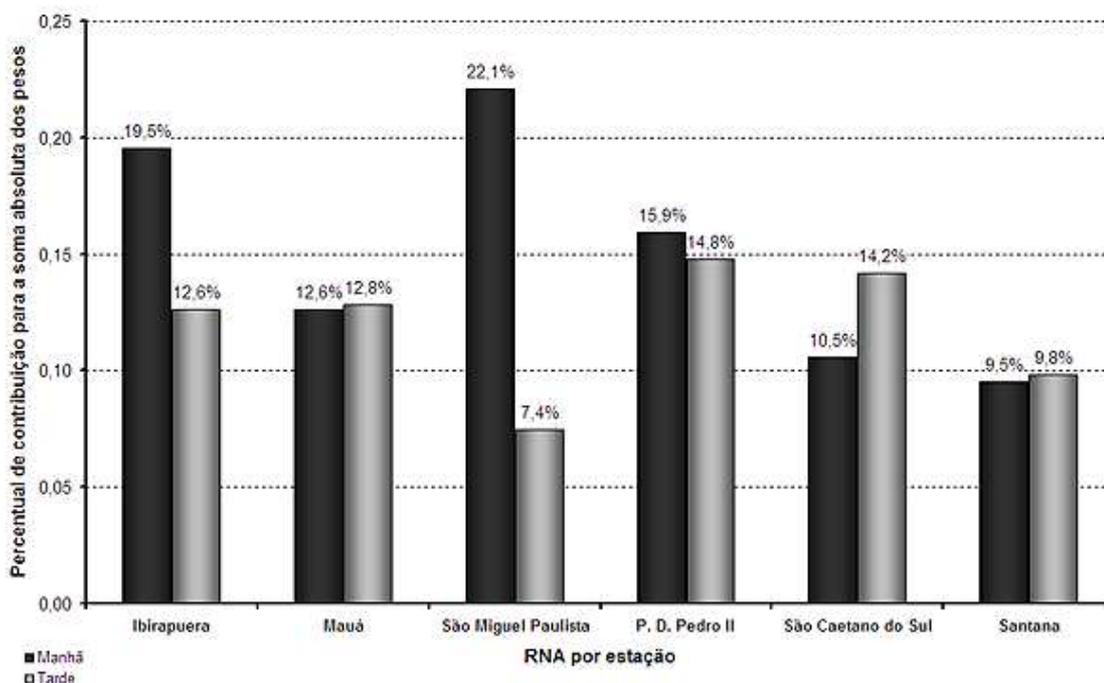


Figura 149: Percentual de contribuição do vento meridional para a soma absoluta dos pesos em cada rede neural artificial de cada estação de qualidade do ar.

Em valores médios, uma das variáveis mais importantes na estimativa de ozônio pelas redes neurais foi o campo médio de vento de manhã, com uma proporção da soma absoluta dos pesos de 16,25% na componente zonal e 15,03% na meridional. A média de temperatura à

tarde também se configura entre as mais representativas, com 15,44%. Em seguida está o campo médio de vento à tarde, com 11,94% para a componente meridional e 11,25% para a zonal. A média de umidade relativa de tarde aparece com 10,94%; a média de temperatura de manhã com 10,77%; e, por fim, está a média de manhã para umidade relativa, com 7,00%.

**Ajustes de redes neurais artificiais incluindo
como entrada as concentrações de poluentes
atmosféricos**

6.1 Descrição

A exemplo do que foi feito no capítulo 5, novas redes neurais artificiais foram desenvolvidas para cada uma das estações selecionadas, porém, introduzindo como variáveis de entrada as concentrações de diferentes poluentes atmosféricos, além dos parâmetros meteorológicos. A escolha dos poluentes dependeu da disponibilidade de dados em cada estação, tornando cada ajuste de rede neural um caso particular.

Tais redes neurais não são apropriadas para aplicação prática como ferramenta de previsão, uma vez que dependem da inserção de prognósticos das concentrações de poluentes usadas nas entradas do modelo. Não havendo uma metodologia confiável para se obter tais estimativas, as redes neurais artificiais apresentadas a seguir servem apenas para fins de estudo da relação entre as variáveis de entrada para a previsibilidade do ozônio troposférico.

Todas as redes foram configuradas de modo a ter oito neurônios na camada oculta e com os dados sendo apresentados 40 mil vezes durante o processo de ajuste. A base de dados usada no treinamento corresponde ao período de 1999 a 2003. Para validação foram empregadas as medições de 2004. As variáveis de entrada correspondem a valores médios representativos do período da manhã (08:00 – 12:00h) e da tarde (13:00 – 17:00h) para parâmetros meteorológicos (tal como aplicado no capítulo 5) e concentrações de poluente (variáveis em cada caso, de acordo com sua disponibilidade).

6.2 Estação de monitoramento de qualidade do ar em Mauá

Para essa rede neural foram considerados 971 eventos para ajuste e 192 para teste, representados por 16 entradas, com médias de manhã e tarde de: temperatura do ar, umidade relativa, campo de vento horizontal em 50 m e concentrações de NO, NO₂, NO_x e particulado (PM₁₀). Os parâmetros químicos foram obtidos diretamente da estação, já os meteorológicos

seguem a metodologia descrita em 5.2, ou seja, a temperatura e umidade são da estação São Caetano do Sul e o vento do SODAR do Ibirapuera.

Os resultados são apresentados graficamente nas figuras 150 e 151, onde é feita a comparação dos resultados das simulações com os valores observados, respectivamente para concentrações médias e máximas de ozônio. A inclusão das concentrações de poluentes à entrada da rede neural contribuiu para a qualidade do ajuste da rede neural. Os coeficientes de determinação passaram de 0,80 e 0,75 para 0,85 e 0,82, para concentrações médias e máximas de ozônio, respectivamente. Pela soma absoluta dos pesos (figura 152), a média de NO no período da tarde foi a variável mais representativa, seguido pelo NO_x e Particulado inalável (PI).

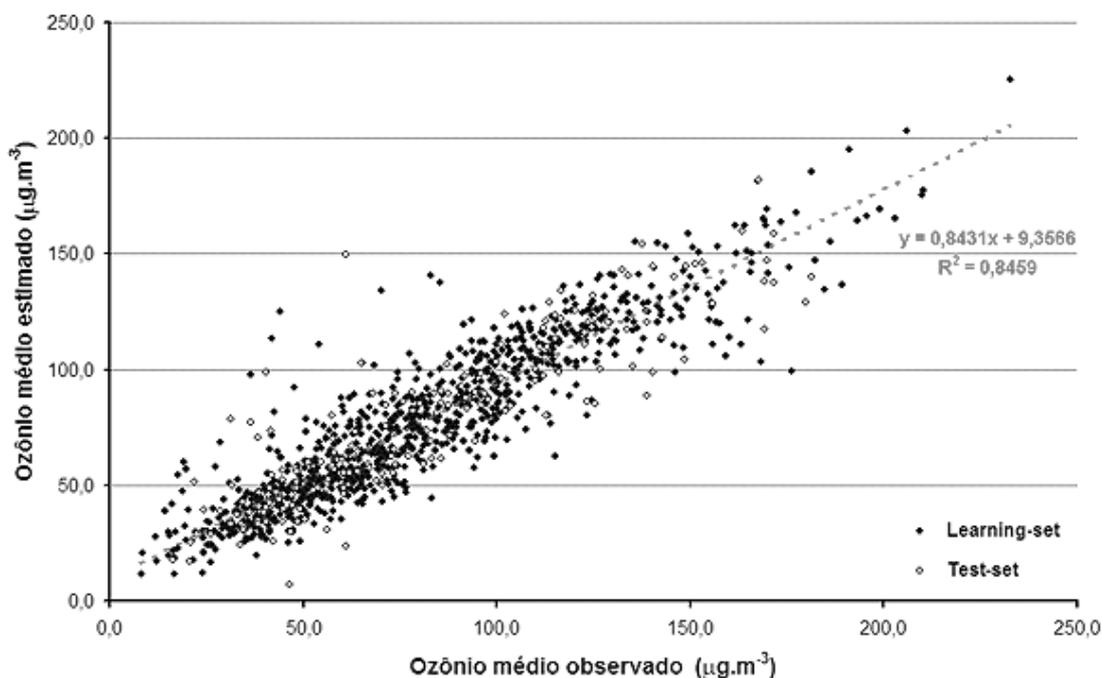


Figura 150: Relação entre valores calculados e medidos da concentração média de ozônio para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em Mauá.

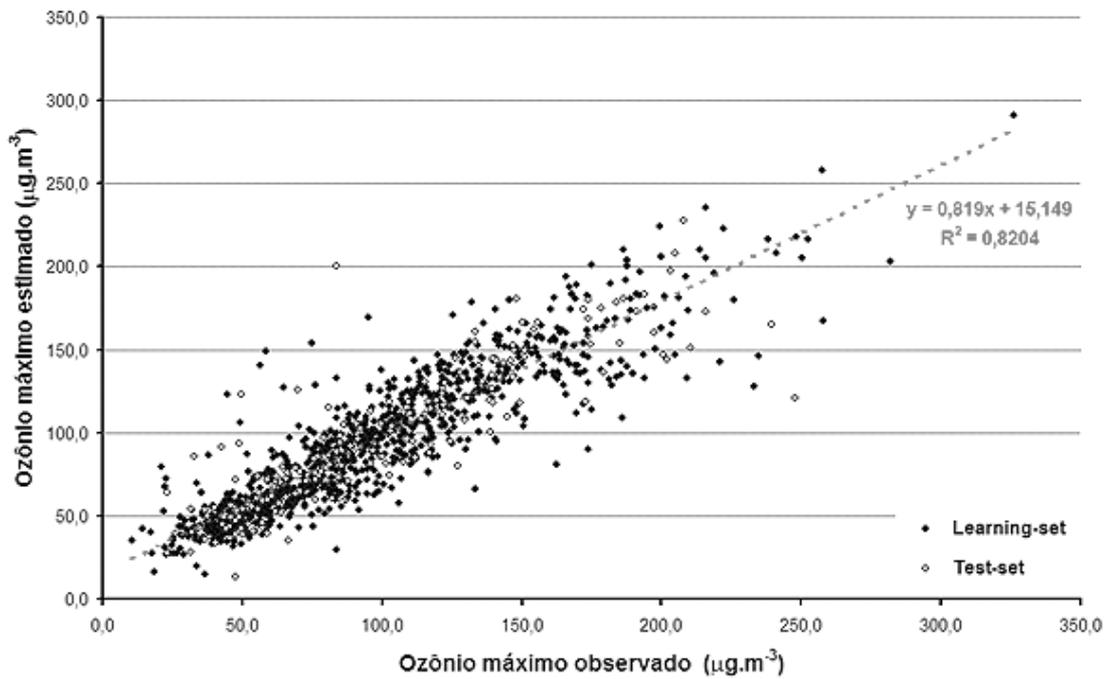


Figura 151: Relação entre valores calculados e medidos da concentração máxima de ozônio para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em Mauá.

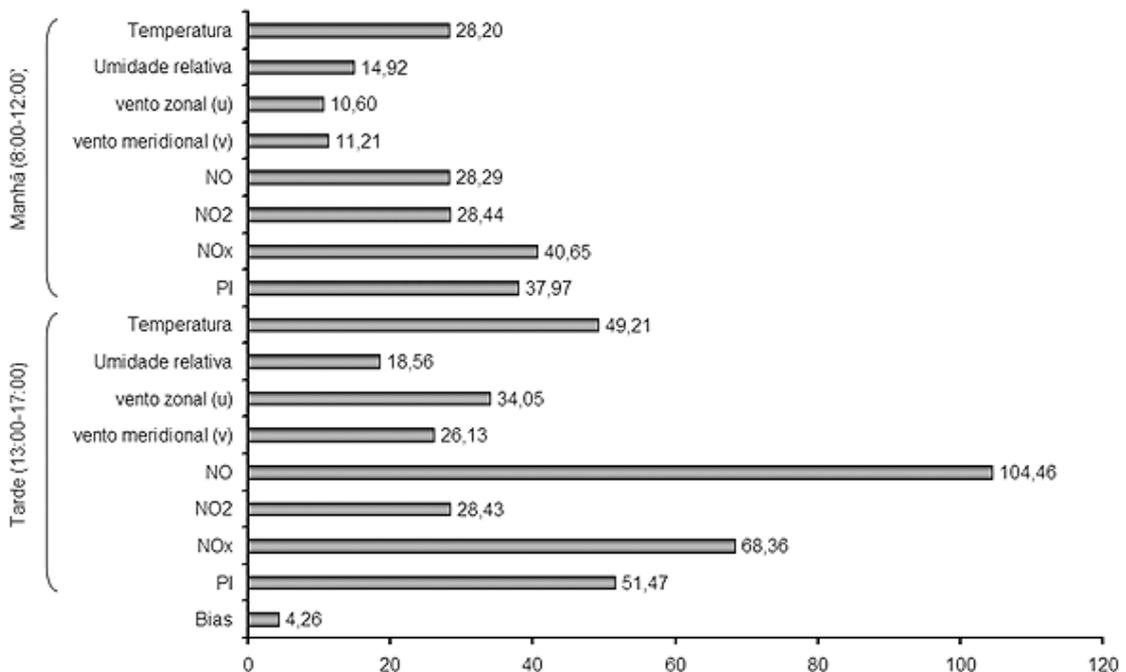


Figura 152: Soma dos valores absolutos dos pesos entre cada variável de entrada e os neurônios da camada oculta para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em Mauá.

Os resultados da validação aplicando 285 eventos de 2004 à entrada do modelo, estão representados pelas figuras 153 e 154, de onde se observa a relação entre as concentrações medidas de ozônio e suas respectivas previsões obtidas pela RNA. Assim como ocorreu na rede que usa apenas entradas meteorológicas (item 5.2), os resultados de validação se mostraram melhores que o do próprio ajuste, resultando em coeficientes de determinação de 0,90 e 0,89, respectivos as concentrações médias e máximas.

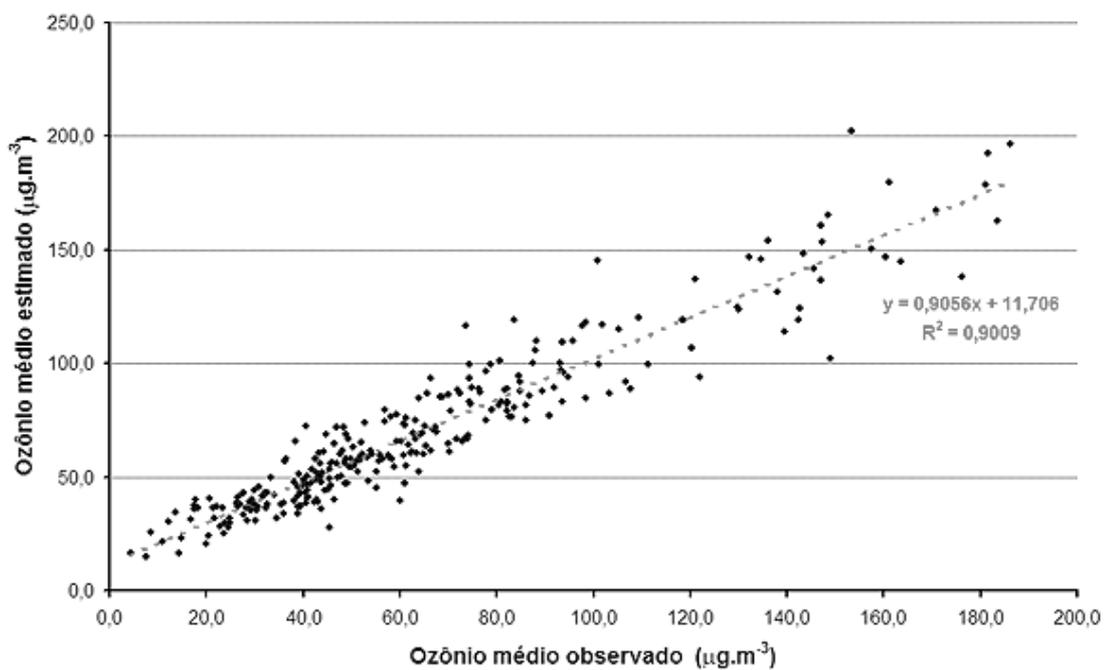


Figura 153: Relação entre valores calculados e medidos da concentração média de ozônio para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em Mauá.

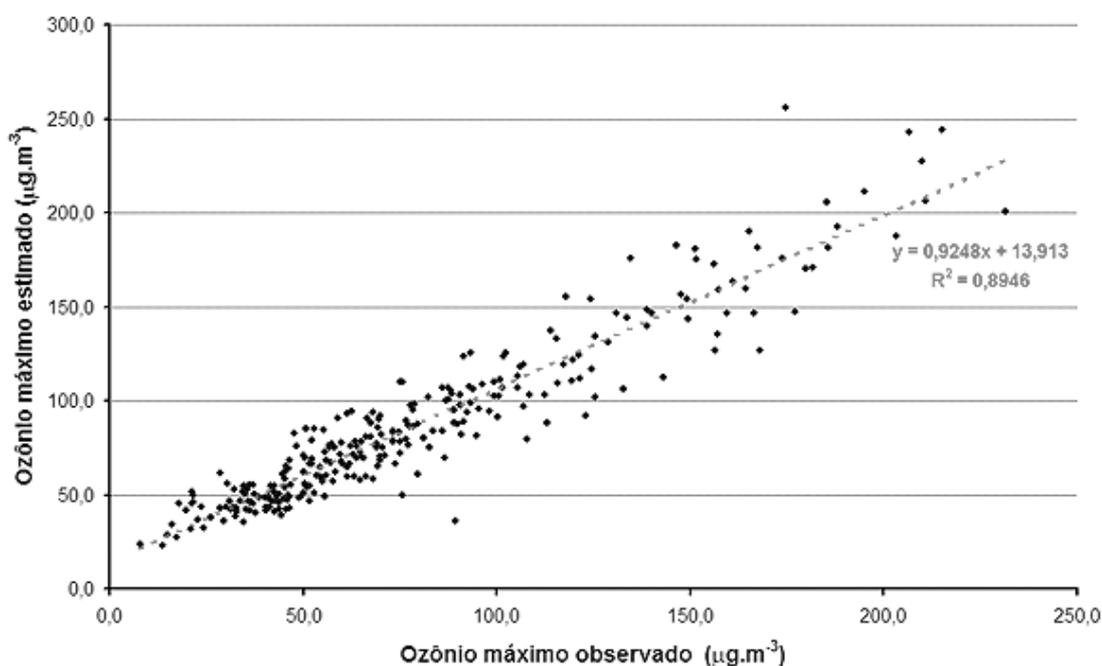


Figura 154: Relação entre valores calculados e medidos da concentração máxima de ozônio para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em Mauá.

Na tabela 6.2.1 são apresentadas estatísticas características das medições de concentração de ozônio e as estimativas da rede neural. Na tabela 6.2.2 encontram-se os resultados de cálculos de correlação, erro médio, raiz do erro médio quadrático e do parâmetro $RMSE_{ub}$, usados para avaliar o desempenho da rede. Em média a simulação difere das observações em $10,2 \mu\text{g.m}^{-3}$ para as concentrações médias de ozônio, e $13,1 \mu\text{g.m}^{-3}$ para as máximas, não diferindo muito do obtido para rede do item 5.2.

Tabela 6.2.1: Caracterização das medidas (obs.) e simulações (sim.) para O_3 médio e máximo (em $\mu\text{g.m}^{-3}$).

	O3 médio		O3 máximo	
	Obs.	Sim.	Obs.	Sim.
MÉDIA	64,42	70,04	78,12	86,16
σ (desvio)	37,55	35,82	45,02	44,02
MÁXIMO	186,21	202,09	231,44	256,10
MÍNIMO	4,35	14,65	8,00	22,82

Tabela 6.2.2: Critérios de avaliação do desempenho da rede neural. *ME*: Erro médio ($\mu\text{g.m}^{-3}$); *RMSE*: Raiz do erro médio quadrático ($\mu\text{g.m}^{-3}$); *RMSE_{UB}*: Aptidão da simulação ($\mu\text{g.m}^{-3}$).

	O3 médio	O3 máximo
Correlação	0,95	0,95
ME	5,6	8,0
RMSE	13,07	16,72
RMSEub	11,80	14,66

6.3 Estação de monitoramento de qualidade do ar em São Miguel Paulista

No ajuste dessa RNA foram apresentados 1034 eventos do conjunto de aprendizado e 198 para o de teste. Ao todo foram 10 entradas, com médias de: temperatura do ar, umidade relativa, campo de vento em 50m (do SODAR em Ibirapuera) e concentração de particulado inalável (PI ou PM_{10}). A representação dos resultados é mostrada nas figuras 155 e 156, respectivas às concentrações de ozônio em termos médios e máximos, fazendo a comparação entre os valores simulados e as medidas da estação. Na figura 157 se encontra a soma absoluta dos pesos da RNA, indicando as variáveis de entrada mais representativas no processo de treinamento da rede.

Com relação às concentrações médias, o novo ajuste foi semelhante ao obtido em 5.3, onde a diferença só pode ser identificada pelos coeficientes de determinação, que passou de 0,75 para 0,74. Já para as concentrações máximas observa-se uma pequena melhora. Os pontos acima de $150 \mu\text{g.m}^{-3}$ se encontram menos dispersos do que fora visto em 5.3, embora o R^2 tenha subido apenas de 0,71 para 0,72. Pela soma absoluta dos pesos (fig.157) observa-se que quatro variáveis tiveram maior relevância no ajuste da rede. Sendo estas a temperatura média no período da manhã, as concentrações médias de material particulado inalável à tarde e o campo de vento médio também do período da tarde.

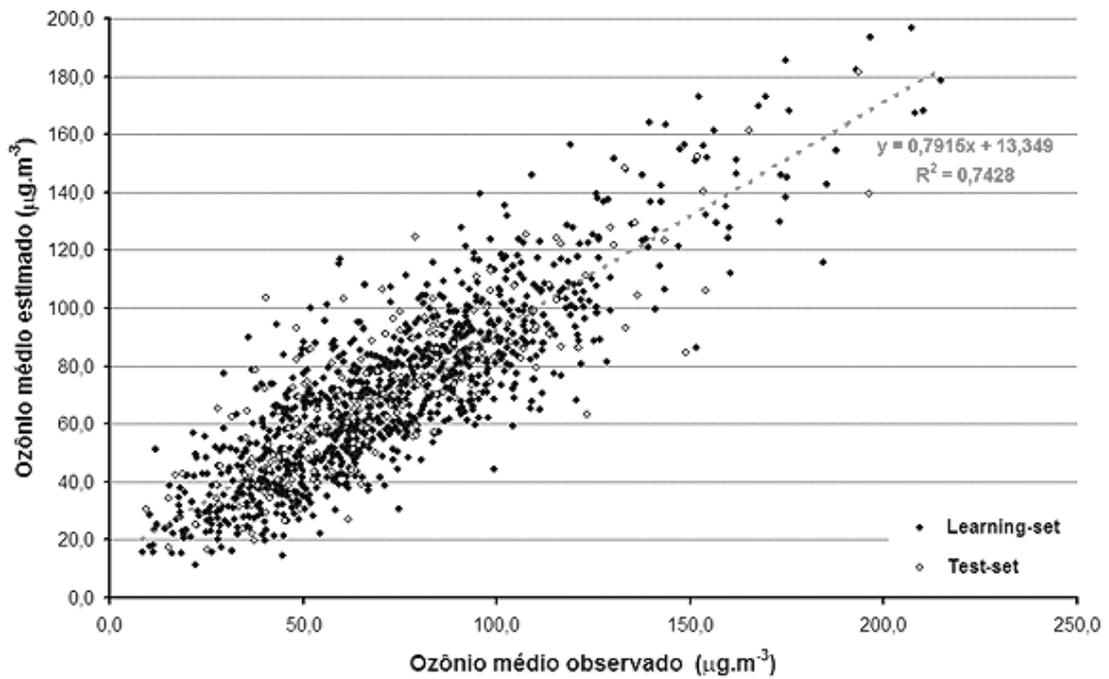


Figura 155: Relação entre valores calculados e medidos da concentração média de ozônio para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em São Miguel Paulista.

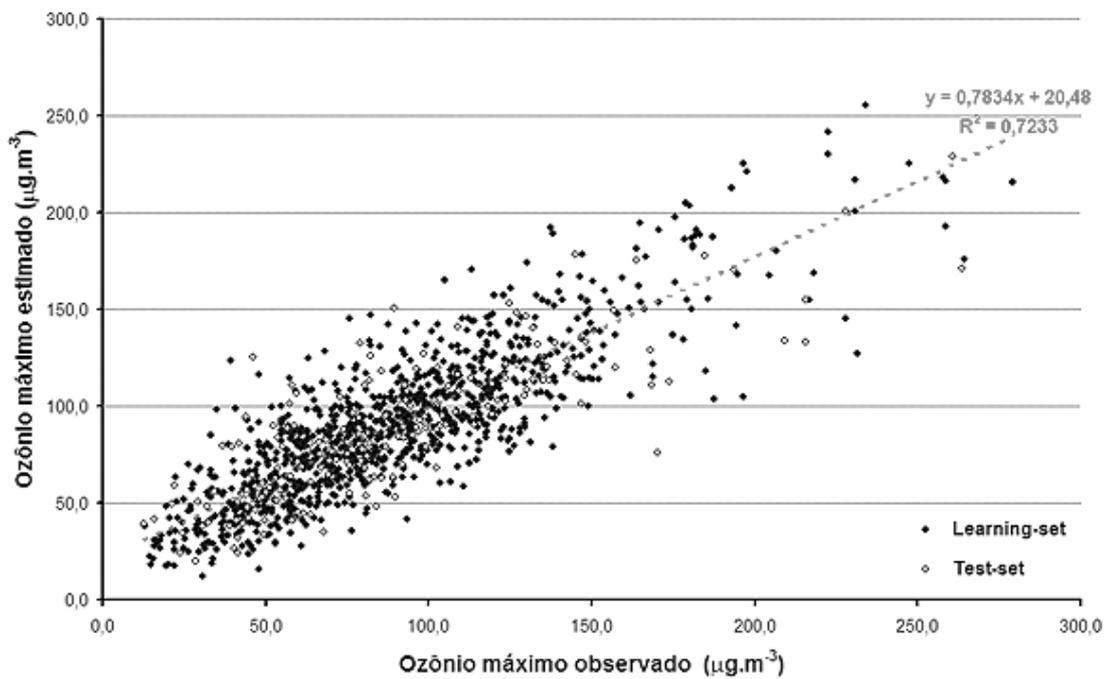


Figura 156: Relação entre valores calculados e medidos da concentração máxima de ozônio para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em São Miguel Paulista.

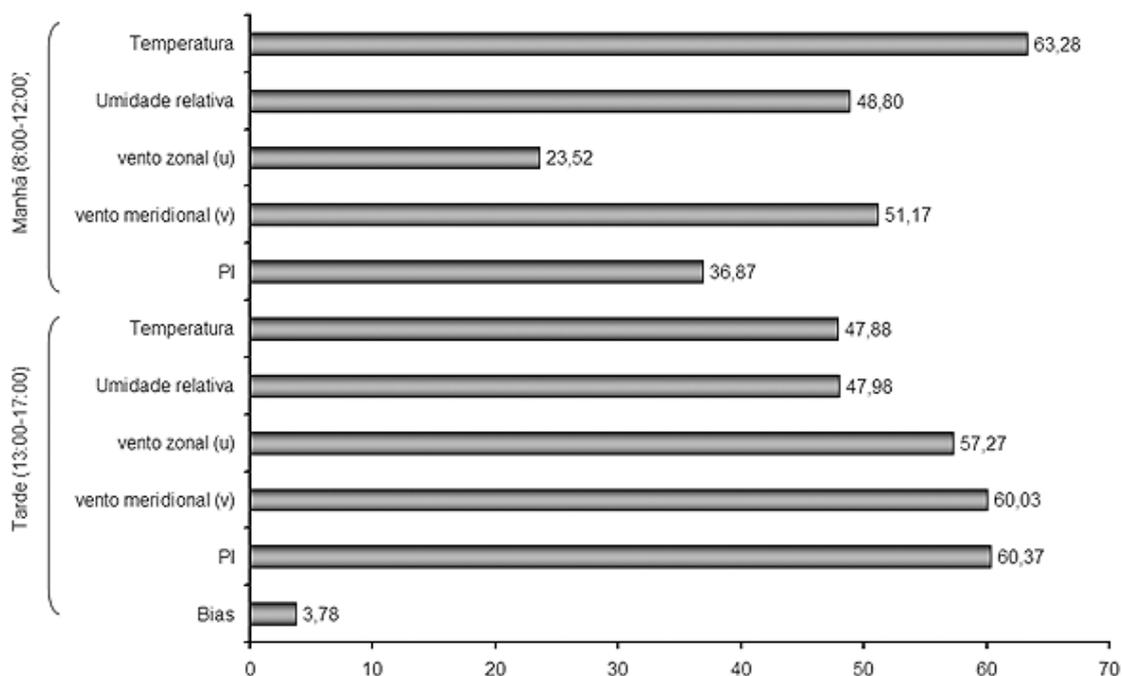


Figura 157: Soma dos valores absolutos dos pesos entre cada variável de entrada e os neurônios da camada oculta para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em São Miguel Paulista.

A validação dessa rede neural contou com 168 casos do ano de 2004, obtidos da própria estação. Os resultados são representados pelas figuras 158 e 159. Com relação à validação da RNA descrita em 5.3, o desempenho foi inferior. Os coeficientes de determinação que eram de 0,70 e 0,66, respectivos às concentrações médias e máximas de ozônio, passaram para 0,52 e 0,50. Observa-se também uma maior dispersão dos pontos. Isso se deve a menor ocorrência de eventos de baixa concentração para essa validação, os quais normalmente são bem previstos pela rede. Embora se refira ao mesmo ano de 2004, a inclusão do PI implicou na redução da base de dados quando houve falha de medição desta variável, coincidindo com alguns eventos de baixa concentração de ozônio.

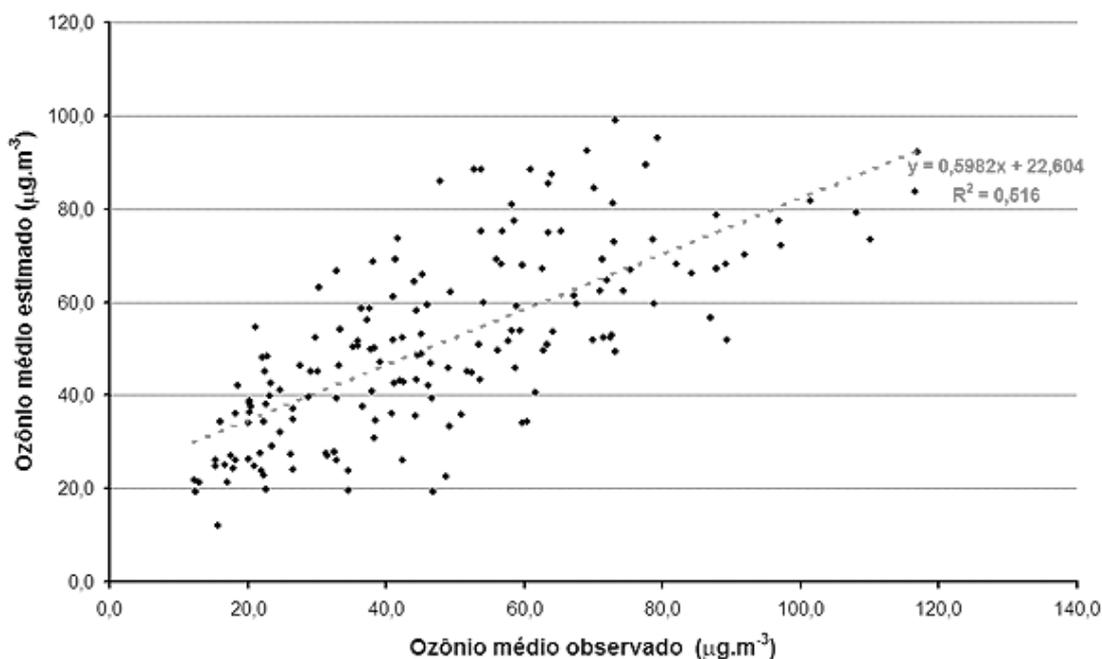


Figura 158: Relação entre valores calculados e medidos da concentração média de ozônio para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em São Miguel Paulista.

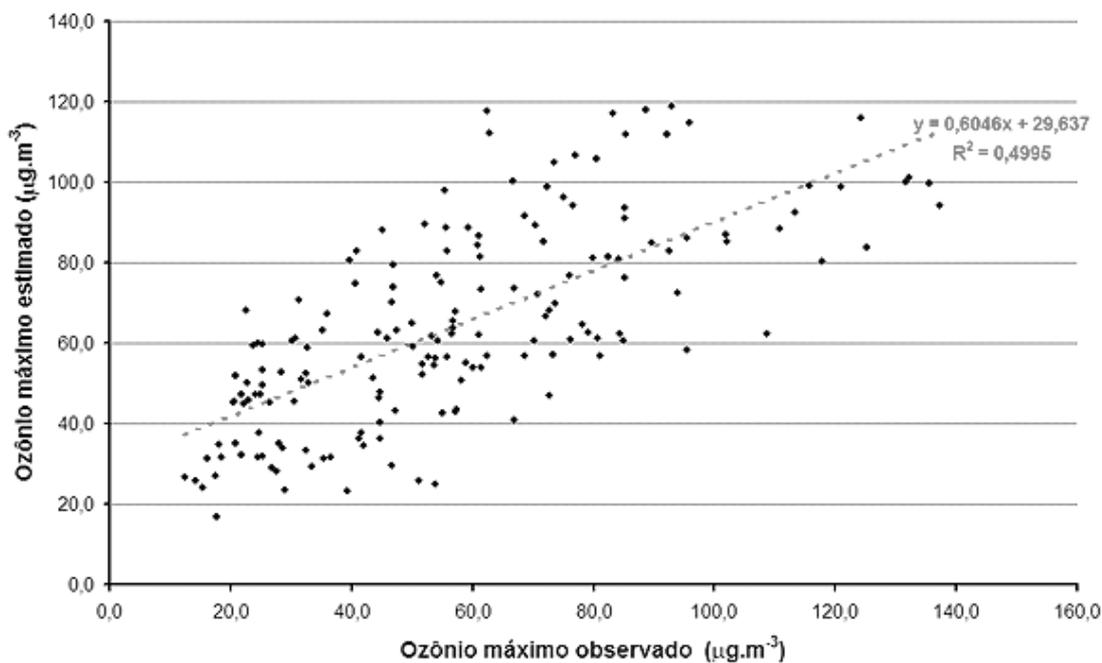


Figura 159: Relação entre valores calculados e medidos da concentração máxima de ozônio para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em São Miguel Paulista.

A caracterização das medidas de concentração de ozônio e as simulações da RNA é apresentada na tabela 6.3.1 por meio de parâmetros estatísticos. Já na tabela 6.3.2 estão os resultados de alguns indicadores de desempenho. De acordo com o erro médio, a rede tende a superestimar o valor das observações em $3,4 \mu\text{g.m}^{-3}$ para concentrações médias de ozônio e $7,0 \mu\text{g.m}^{-3}$ para as máximas. Já o $RMSE$ e $RMSE_{ub}$ indicam que, em média, o erro é de $16,7 \mu\text{g.m}^{-3}$ para o ozônio médio e $21,3 \mu\text{g.m}^{-3}$ para o máximo.

Tabela 6.3.1: Caracterização das medidas (obs.) e simulações (sim.) para O₃ médio e máximo(em $\mu\text{g.m}^{-3}$).

	O3 médio		O3 máximo	
	Obs.	Sim.	Obs.	Sim.
MÉDIA	47,68	51,13	57,31	64,28
σ (desvio)	23,56	19,62	28,78	24,62
MÁXIMO	116,80	99,02	137,26	118,71
MÍNIMO	12,21	11,94	12,48	16,91

Tabela 6.3.2: Critérios de avaliação do desempenho da rede neural. ME : Erro médio ($\mu\text{g.m}^{-3}$); $RMSE$: Raiz do erro médio quadrático ($\mu\text{g.m}^{-3}$); $RMSE_{UB}$: Aptidão da simulação ($\mu\text{g.m}^{-3}$).

	O3 médio	O3 máximo
Correlação	0,72	0,71
ME	3,4	7,0
RMSE	16,92	21,88
RMSEub	16,56	20,74

6.4 Estação de monitoramento de qualidade do ar em Parque Dom Pedro II

Para a RNA de P. D. Pedro II foram usados 1058 casos no conjunto de aprendizado e 205 para o de teste. Ao todo foram 20 variáveis de entrada, correspondentes à médias de manhã e de tarde para: temperatura do ar, umidade relativa, campo de vento horizontal (obtido do SODAR em Ibirapuera), e concentrações de NO, NO₂, NO_x, PI, CO e SO₂. Os resultados encontram-se representados pelas figuras 160 e 161, comparando observações e estimativas da rede neural, respectivamente para concentrações médias e máximas de ozônio no período da tarde. As somas absolutas dos pesos das variáveis na RNA são comparadas na figura 162.

Houve uma melhora considerável na qualidade do ajuste com a inclusão das concentrações de poluentes à entrada do modelo. Além dos pontos se mostrarem bem menos dispersos, o índice de determinação passou de 0,73 para 0,87 com relação às concentrações médias de ozônio e de 0,65 para 0,85 com relação às máximas. Nessa RNA as variáveis mais importantes foram os óxidos de nitrogênio e a componente meridional do vento. Esta última já aparecia como a entrada mais importante da rede descrita em 5.4, mas os óxidos de nitrogênio se mostraram ainda mais relevantes. Isso sugere que, por P. D. Pedro II ser uma estação próxima de via de tráfego veicular, a previsibilidade do ozônio seja mais dependente da alta emissão de NOx pelos veículos, em especial os veículos à diesel. A estação P.D. Pedro II fica próxima a terminais de ônibus e a química de formação de ozônio é determinada pela razão entre NOx e os COVs.

Assim como ocorreu no item 5.4, não houve dados suficientes do ano de 2004 para se fazer uma validação do modelo. Ao todo foram reunidos apenas 14 eventos que estiveram disponíveis, o que não é suficiente para uma avaliação.

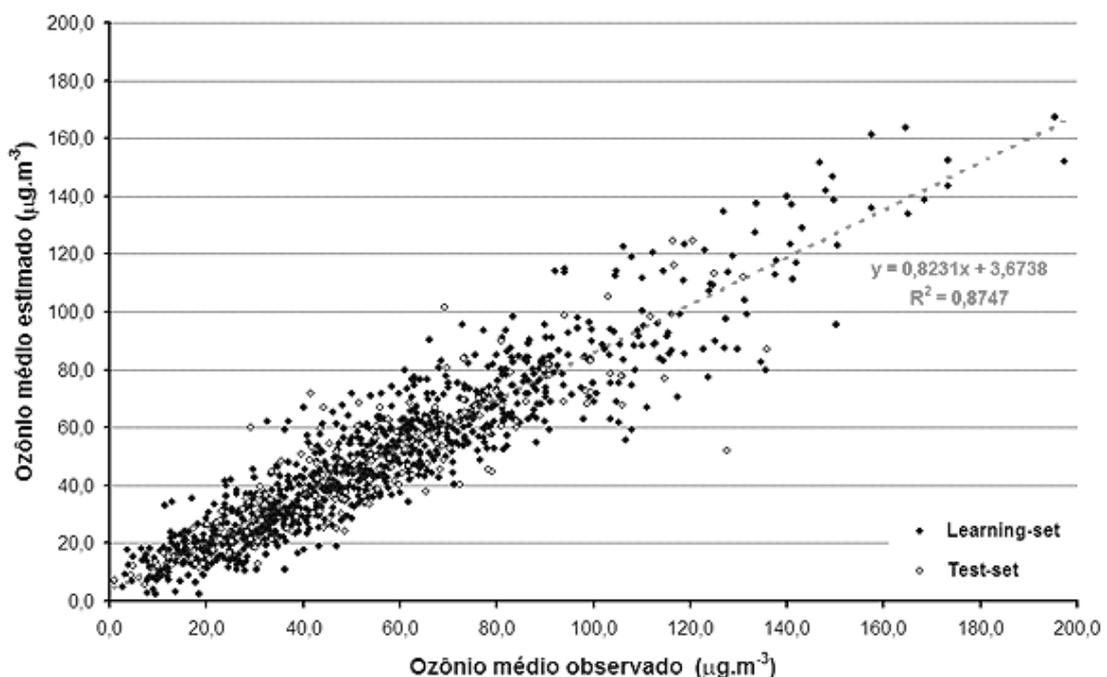


Figura 160: Relação entre valores calculados e medidos da concentração média de ozônio para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em P. D. Pedro II.

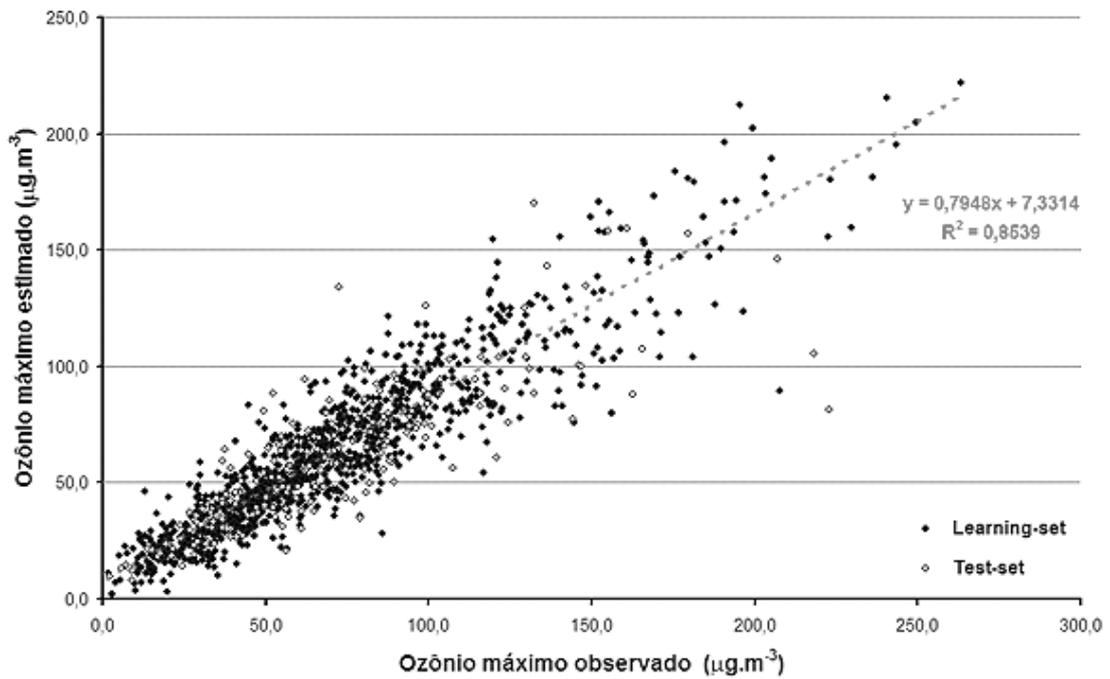


Figura 161: Relação entre valores calculados e medidos da concentração máxima de ozônio para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em P. D. Pedro II.

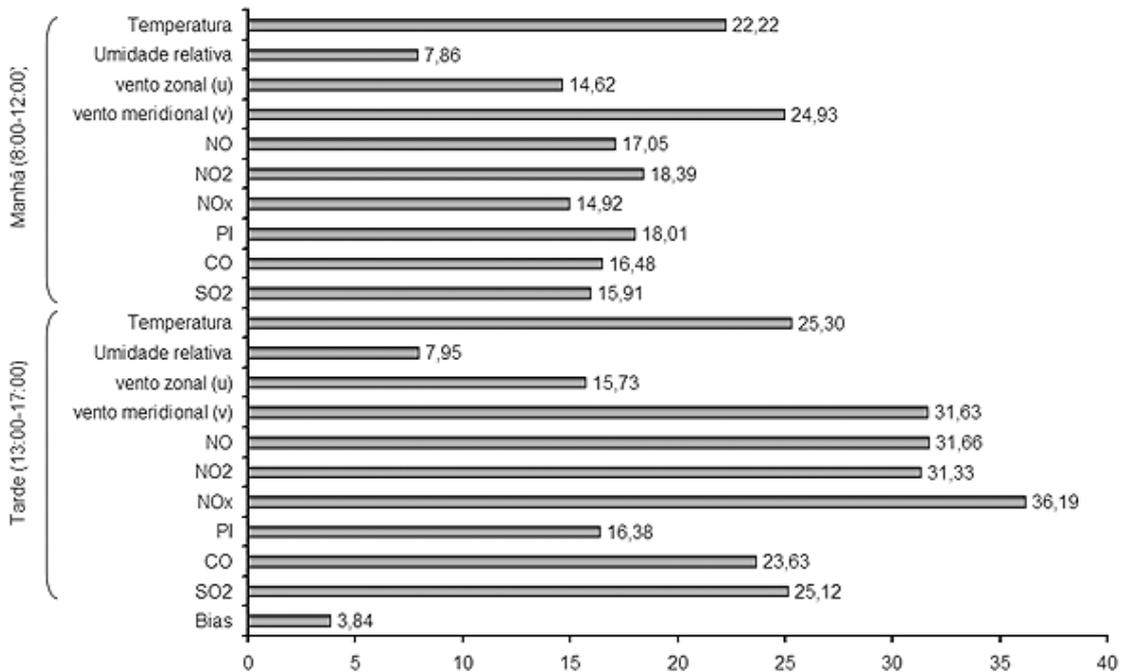


Figura 162: Soma dos valores absolutos dos pesos entre cada variável de entrada e os neurônios da camada oculta para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em P. D. Pedro II.

6.5 Estação de monitoramento de qualidade do ar em São Caetano do Sul

Foram considerados nesse ajuste 775 casos para o conjunto de treinamento da RNA, além de 147 eventos no conjunto de teste. No total foram 24 variáveis, correspondendo as médias de manhã e de tarde para: temperatura do ar, umidade relativa, campo de vento (SODAR Ibirapuera) e concentrações de NO, NO₂, NO_x, PI, CO, SO₂, CH₄ e NMHC (hidrocarbonetos, exceto metano). Os resultados são ilustrados nas figuras 163 e 164, nas quais são feitas comparações entre valores medidos e calculados de concentrações médias e máximas de ozônio, respectivamente. As somas absolutas dos pesos das variáveis de entrada são comparadas na figura 165.

Tal como observado com a RNA de P. D. Pedro II (item 6.4), a inclusão das concentrações de espécies químicas trouxe melhoria na qualidade das estimativas de ozônio pelo modelo. Nos gráficos 163 e 164 os pontos se mostram mais concisos do que no observado para a rede descrita em 5.5. Além disso, com relação a essa RNA que usa apenas fatores meteorológicos como entrada, os coeficientes de determinação aumentaram com esse novo ajuste. Para concentrações médias de ozônio, o R^2 subiu de 0,75 para 0,86, enquanto para as estimativas de máximas do poluente foi de 0,69 para 0,84.

Entretanto, não foi possível confirmar essa melhoria através da validação do modelo, uma vez que houve poucos eventos disponíveis do ano de 2004. A inclusão das novas variáveis implicou na redução da base de dados devido à ocorrência de falhas de medição. Sendo assim, se no item 5.5 havia 202 casos disponíveis, com as novas variáveis passa para apenas 45 eventos.

De acordo com a figura 165, as variáveis mais relevantes no modelo foram as médias no período da tarde para concentrações de óxidos de nitrogênio (NO₂, principalmente) e particulado inalável. Observa-se que essa foi outra rede neural onde a química atmosférica teve um papel importante na previsibilidade do ozônio troposférico.

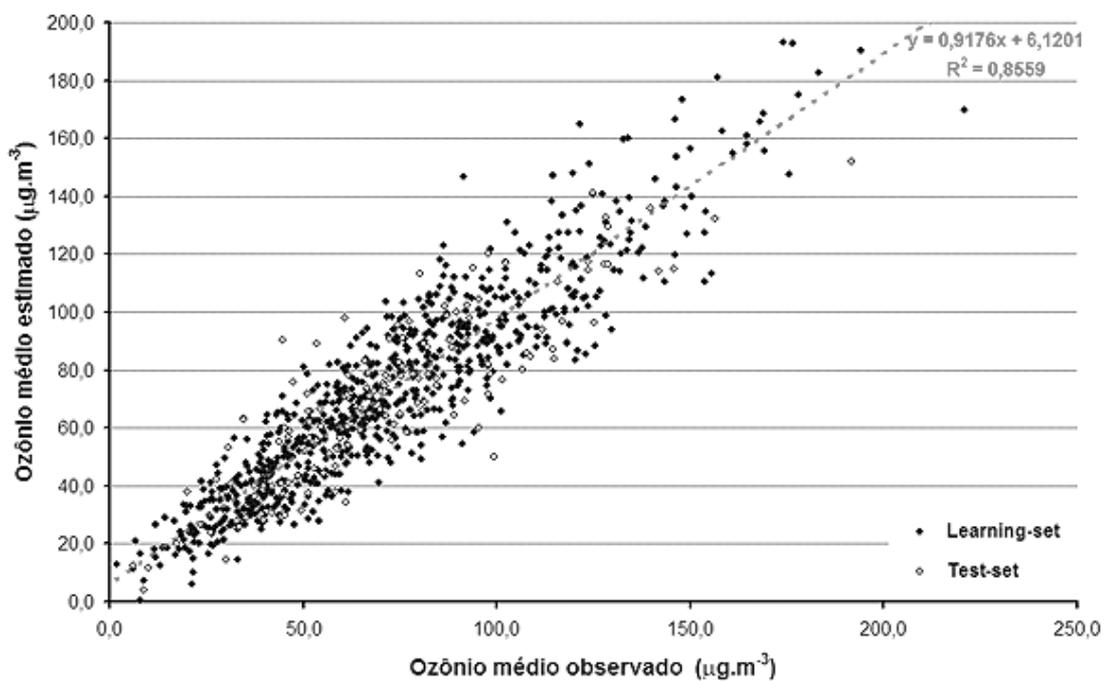


Figura 163: Relação entre valores calculados e medidos da concentração média de ozônio para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em São Caetano do Sul.

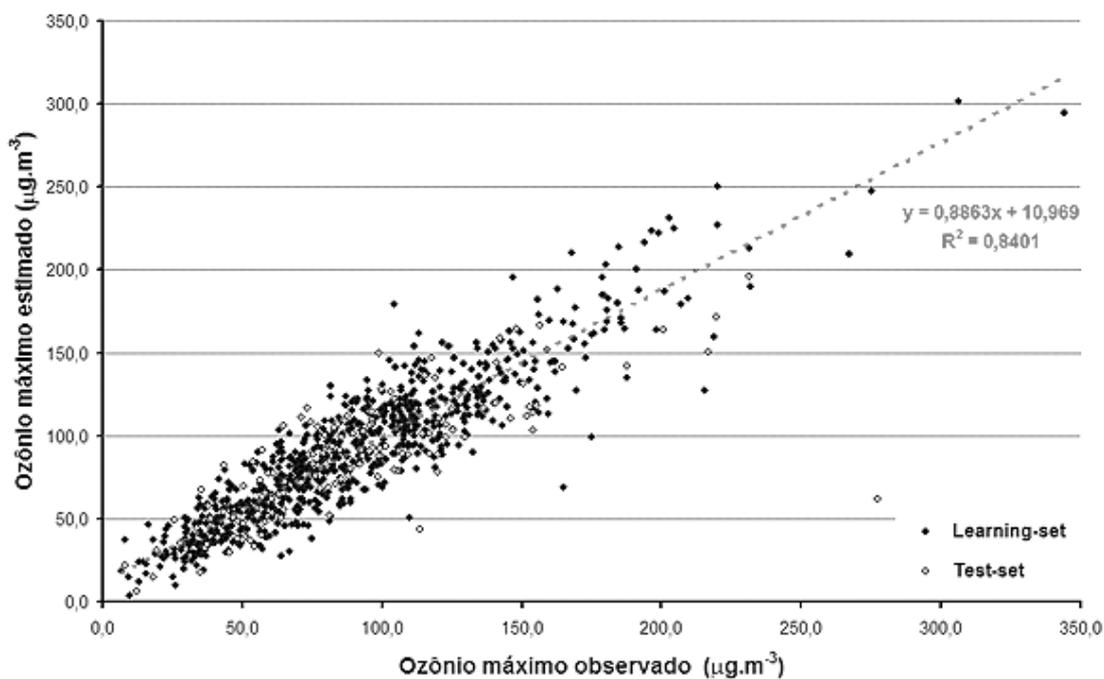


Figura 164: Relação entre valores calculados e medidos da concentração máxima de ozônio para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em São Caetano do Sul.

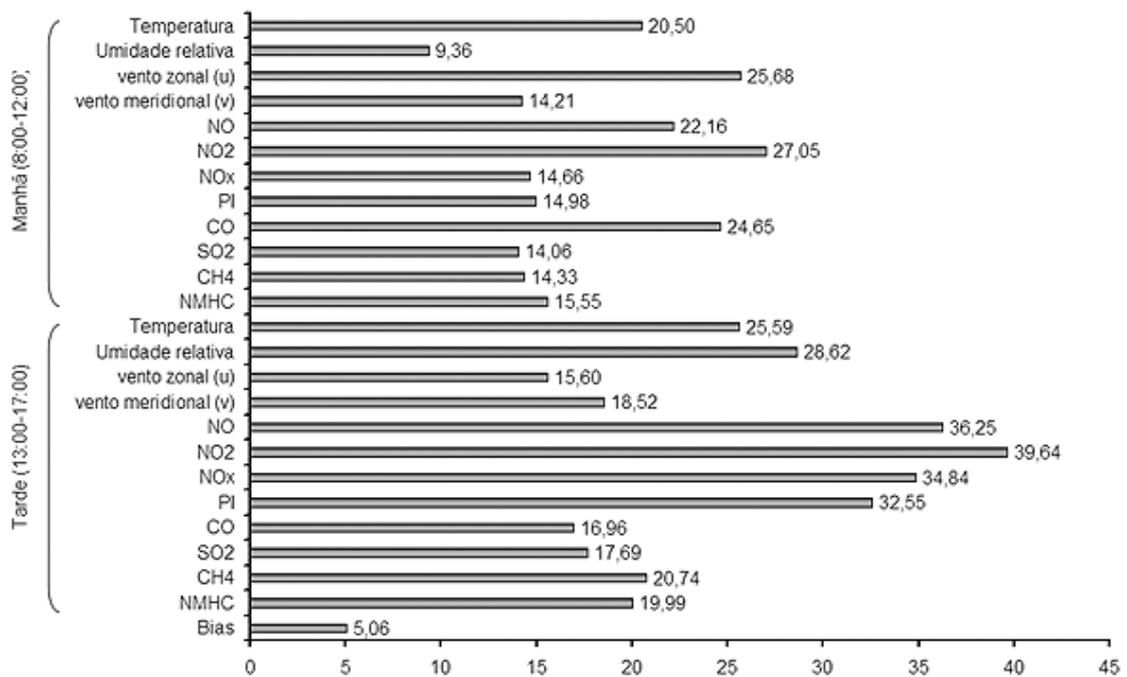


Figura 165: Soma dos valores absolutos dos pesos entre cada variável de entrada e os neurônios da camada oculta para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em São Caetano do Sul.

6.6 Estação de monitoramento de qualidade do ar em Santana

O ajuste da rede neural artificial para a estação Santana contou com 409 eventos no conjunto de treinamento e 77 no de teste. No total foram usadas 10 variáveis de entrada, correspondentes a médias de manhã e de tarde para: temperatura do ar, umidade relativa, campo de vento em 50 m (componentes zonal e meridional) e concentrações de material particulado inalável (PI). O número reduzido de casos para treinamento da RNA deve-se principalmente a ausência de medições de PI em diversos períodos, notadamente no ano de 2002. Os dados de temperatura e umidade foram obtidos da estação P. D. Pedro II, seguindo o critério descrito em 5.6, e o campo de vento foi do SODAR em Ibirapuera.

Os resultados são apresentados nas figuras 166 a 167. Na comparação entre valores medidos e simulados, a qualidade se mostrou melhor do que obtido para a rede descrita em 5.6. Com relação às concentrações médias de ozônio (fig. 166), o índice R^2 passou de 0,81 para 0,84; com as máximas (fig. 167) aumentou de 0,73 para 0,82. Porém, não foi possível

validar essa RNA devido à dependência de dados da estação P. D. Pedro II, a qual apresentou problemas de medição após 2003 (item 5.4).

As somas absolutas dos pesos, representadas na figura 168, indicam que a variável mais importante dentro dessa rede neural foi a média do campo de vento pela manhã, expresso em suas componentes zonal e meridional. Outros parâmetros meteorológicos também foram relevantes, deixando o particulado inalável como a entrada menos significativa, seja pelas médias de manhã ou de tarde. Isso poderia sugerir uma menor dependência da química atmosférica para a previsibilidade do ozônio na região, entretanto, é preciso considerar que o uso de dados de estações vizinhas (P. D. Pedro II e Ibirapuera) implica em uma menor acurácia nos resultados.

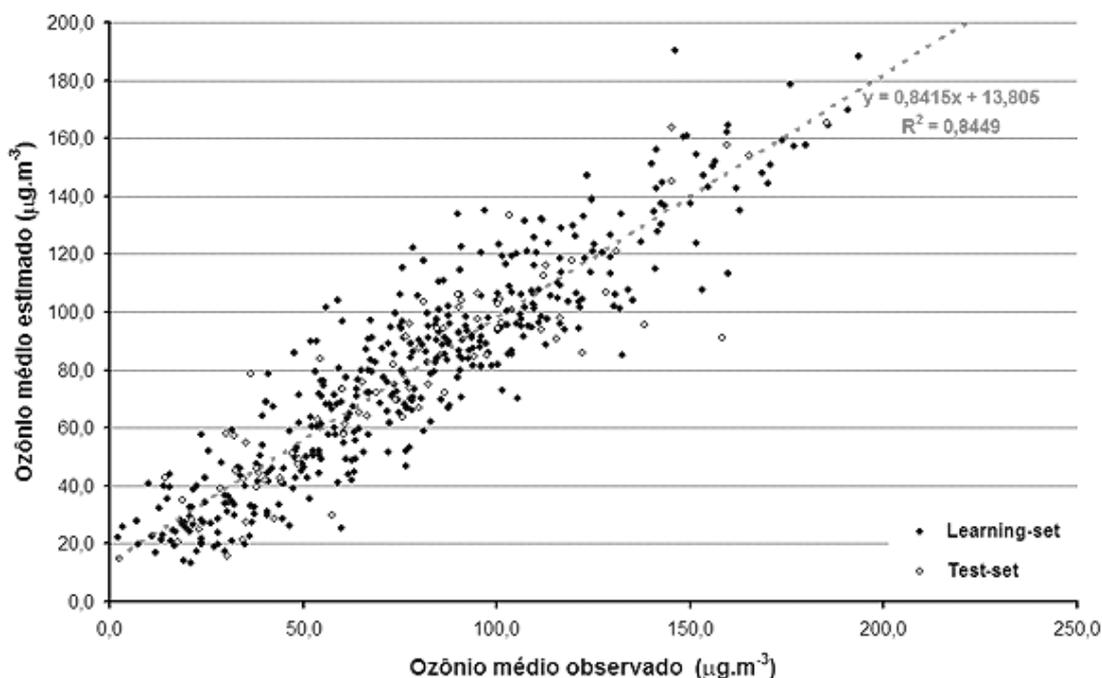


Figura 166: Relação entre valores calculados e medidos da concentração média de ozônio para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em Santana.

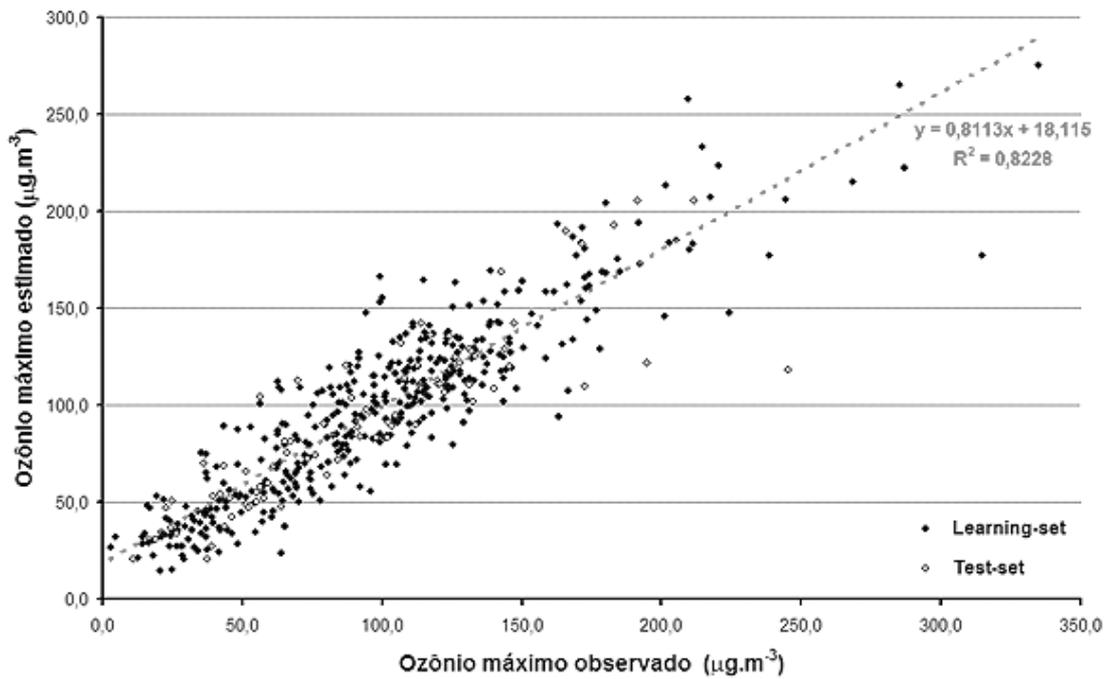


Figura 167: Relação entre valores calculados e medidos da concentração máxima de ozônio para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em Santana.

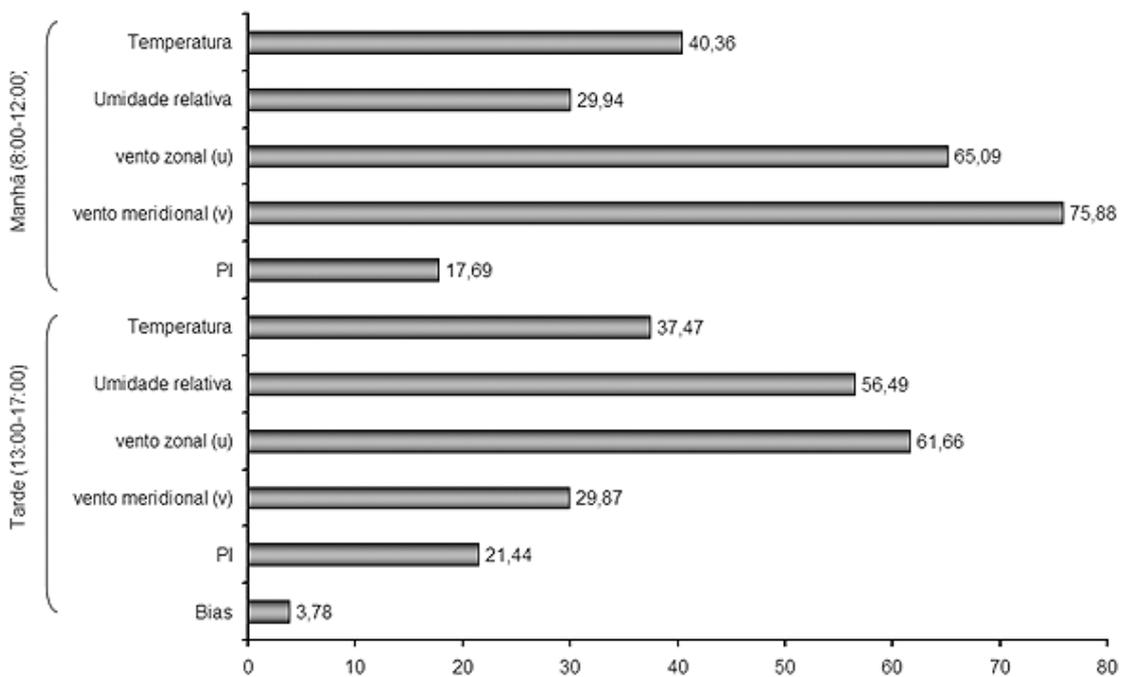


Figura 168: Soma dos valores absolutos dos pesos entre cada variável de entrada e os neurônios da camada oculta para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em Santana.

6.7 Estação de monitoramento de qualidade do ar em Ibirapuera

Essa rede neural artificial visa complementar o estudo descrito no capítulo 4, demonstrando a influência da química atmosférica para a previsibilidade do ozônio troposférico, uma vez que no referido capítulo a ênfase foi dada sobre os fatores meteorológicos. Para o ajuste desta RNA foram usados 614 eventos no conjunto de aprendizado e 116 no de teste. No total foram introduzidas 24 entradas, que se referem às médias de manhã e de tarde para: temperatura do ar, umidade relativa, pressão atmosférica, radiação solar, campo de vento horizontal em 50m (obtido do SODAR), além de concentrações de NO, NO₂, NO_x, PI, CO e SO₂. Os resultados são apresentados graficamente pelas figuras 169 a 171.

No item 4.11, a melhor configuração de RNA levou à coeficientes de determinação iguais a 0,84 e 0,79, respectivamente para concentrações médias e máximas de ozônio. Com a inclusão de concentrações químicas à entrada do modelo, esse índice passou a valer 0,89 para as médias de ozônio (fig. 169) e 0,86 para as máximas (fig. 170). Essa melhora na qualidade deve-se principalmente às médias de óxido de nitrogênio no período da tarde que, conforme mostrado na figura 171, foi a variável mais representativa dentro dessa rede neural.

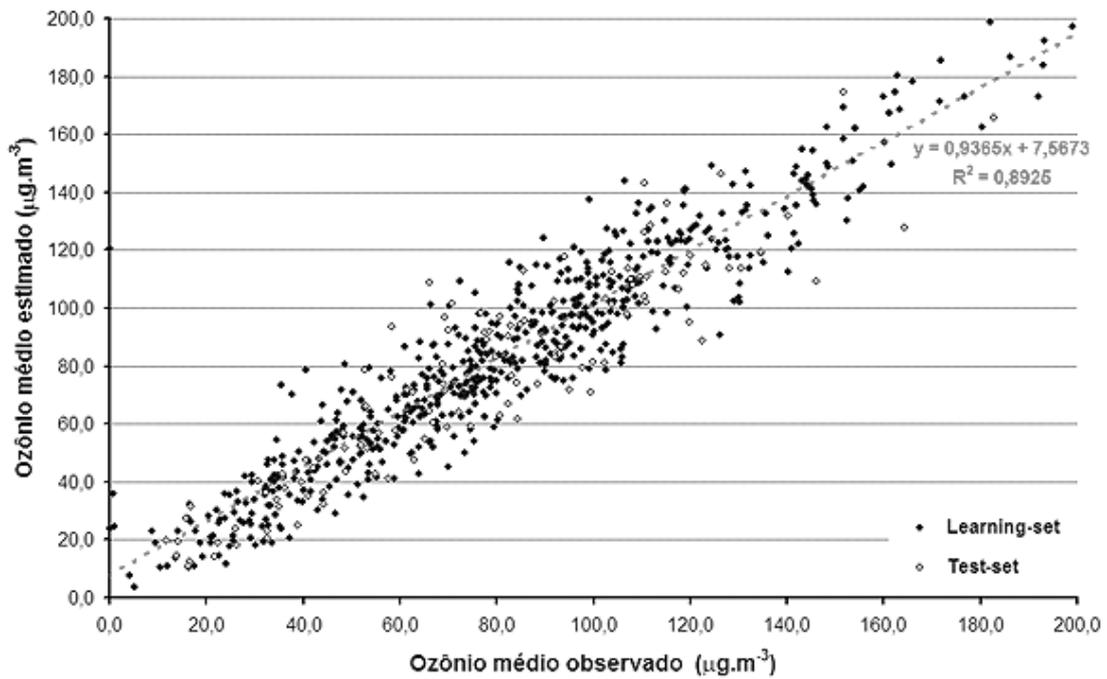


Figura 169: Relação entre valores calculados e medidos da concentração média de ozônio para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em Ibirapuera.

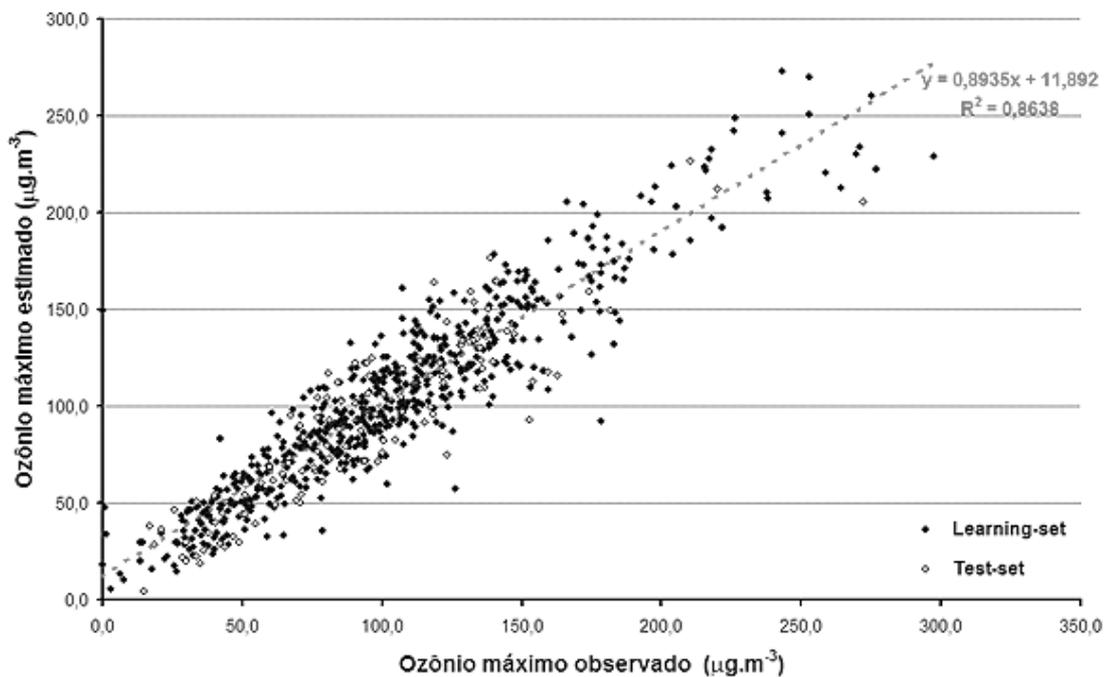


Figura 170: Relação entre valores calculados e medidos da concentração máxima de ozônio para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em Ibirapuera.

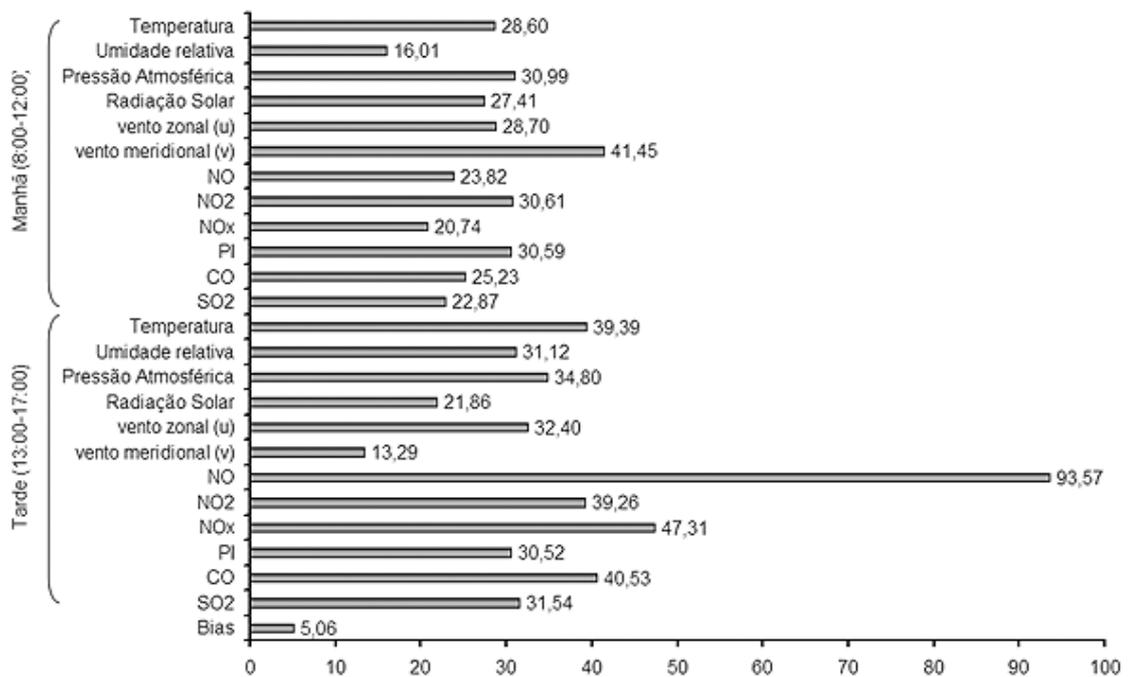


Figura 171: Soma dos valores absolutos dos pesos entre cada variável de entrada e os neurônios da camada oculta para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em Ibirapuera.

Para a validação dessa rede neural artificial foram usados 82 casos referentes ao ano de 2004. Os resultados são apresentados nas figuras 172 e 173, comparando simulações e medidas de concentrações médias e máximas de ozônio, respectivamente. Apesar da melhoria na qualidade do ajuste (figuras 169 e 170), a validação apresentou desempenho semelhante ao obtido em 4.12, resultando em índices R^2 de 0,71 e 0,65 respectivos às concentrações médias e máximas de ozônio.

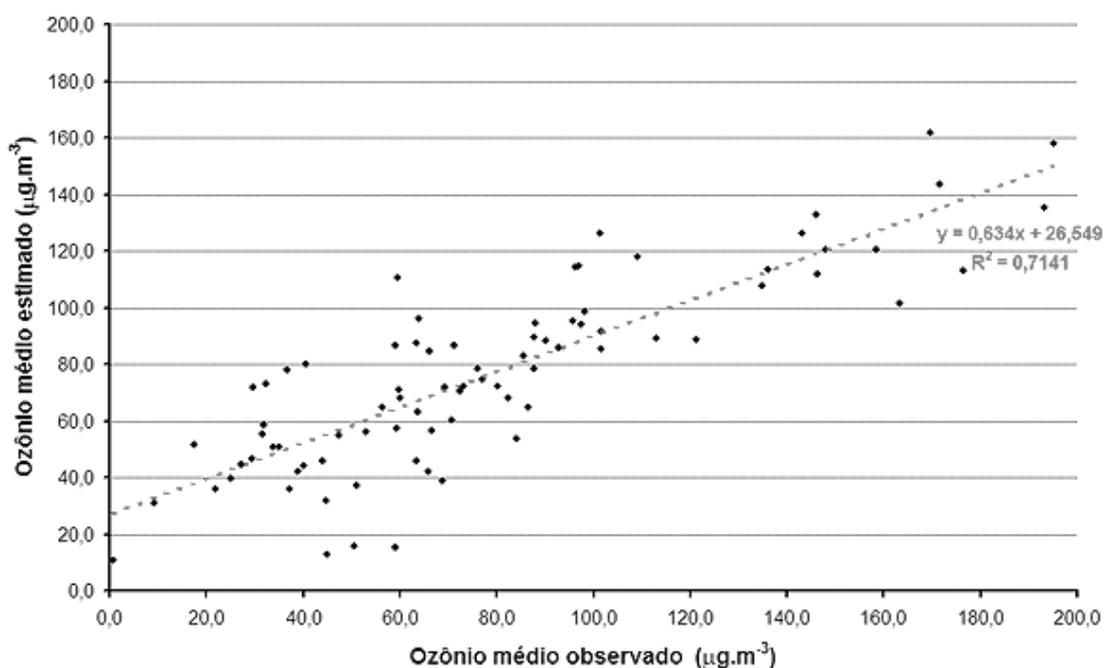


Figura 172: Relação entre valores calculados e medidos da concentração média de ozônio para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em Ibirapuera.

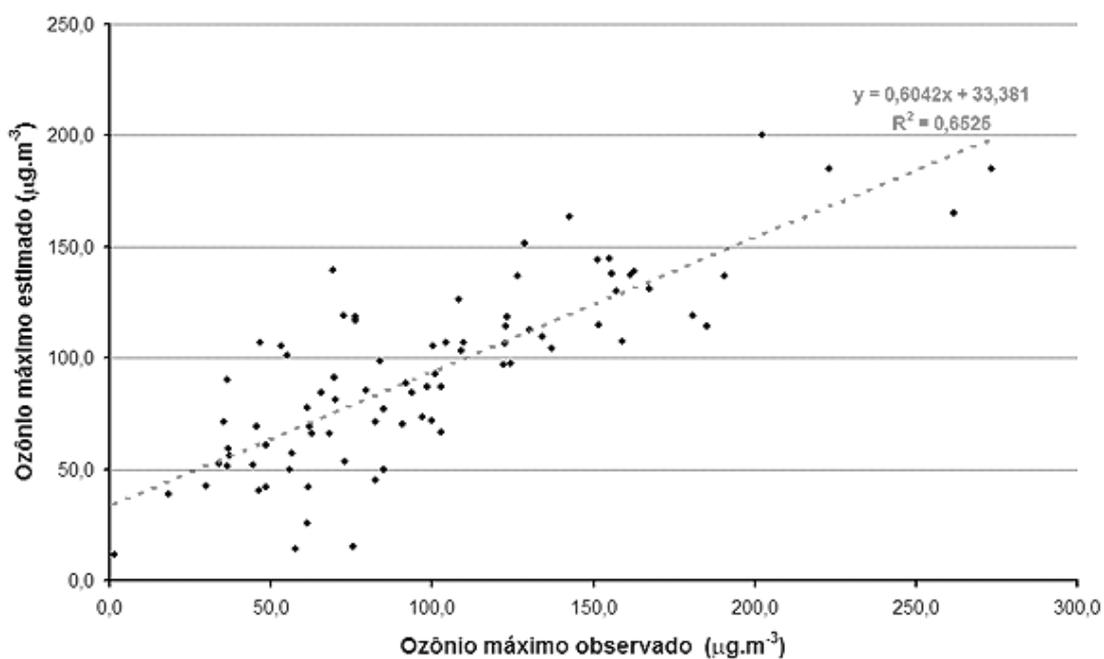


Figura 173: Relação entre valores calculados e medidos da concentração máxima de ozônio para a RNA ajustada com dados da estação de monitoramento de qualidade do ar em Ibirapuera.

Na tabela 6.7.1 as medidas da Cetesb e resultados da simulação são caracterizados estatisticamente, enquanto que na tabela 6.7.2 estão os resultados de cálculos de parâmetros de desempenho. Segundo o erro médio, a RNA tende a subestimar as medidas em $2,6 \mu\text{g.m}^{-3}$ para as concentrações médias de ozônio e $5,8 \mu\text{g.m}^{-3}$ para as máximas. Em geral, pelos indicadores $RMSE$ e $RMSE_{ub}$, a RNA erra em $24,2 \mu\text{g.m}^{-3}$ nas médias de ozônio e $31,9 \mu\text{g.m}^{-3}$ nas máximas.

Tabela 6.7.1: Caracterização das medidas (obs.) e simulações (sim.) para O₃ médio e máximo (em $\mu\text{g.m}^{-3}$).

	O3 médio		O3 máximo	
	Obs.	Sim.	Obs.	Sim.
MÉDIA	79,58	77,01	98,89	93,14
σ (desvio)	44,68	33,52	53,77	40,22
MÁXIMO	195,03	161,77	273,55	200,38
MÍNIMO	0,82	10,68	1,48	11,29

Tabela 6.7.2: Critérios de avaliação do desempenho da rede neural. ME : Erro médio ($\mu\text{g.m}^{-3}$); $RMSE$: Raiz do erro médio quadrático ($\mu\text{g.m}^{-3}$); $RMSE_{UB}$: Aptidão da simulação ($\mu\text{g.m}^{-3}$).

	O3 médio	O3 máximo
Correlação	0,85	0,81
ME	-2,6	-5,8
RMSE	24,25	32,18
RMSEub	24,11	31,66

Conclusões e perspectivas para trabalhos futuros

7.1 Conclusões

Os principais resultados obtidos através desta pesquisa foram a obtenção de modelos estatísticos de redes neurais artificiais dedicados à previsão de ozônio troposférico em diferentes localidades da Região Metropolitana de São Paulo e um melhor entendimento da influência dos fatores meteorológicos sobre os níveis de concentração desse poluente.

Esse estudo foi baseado na aplicação da técnica de redes neurais artificiais (RNA), envolvendo a aplicação de diferentes parâmetros meteorológicos e concentração de poluentes obtidos de estações de monitoramento de qualidade do ar mantidas pela CETESB (Companhia de Tecnologia e Saneamento Ambiental). Os dados foram tratados e analisados (conforme descrito no capítulo 3) para servirem ao treinamento das RNA. Numa primeira etapa foi considerada apenas a estação localizada no Parque do Ibirapuera, por esta ser a mais representativa na RMSP na caracterização do ozônio em superfície. Em seguida o trabalho foi estendido para outras estações de monitoramento desse poluente, selecionadas de acordo com as necessidades da pesquisa. As redes neurais artificiais desenvolvidas nos capítulos 4 e 5 são modelos estatísticos que estimam concentrações médias e máximas de ozônio troposférico para o período da tarde, de acordo com as variáveis meteorológicas fornecidas como entradas. No capítulo 6 essa estimativa é feita através da inserção de concentrações de diferentes poluentes atmosféricos, além dos fatores meteorológicos previamente considerados.

Por este estudo foram obtidos modelos estatísticos que podem vir a ser implementados por instituições ambientais, de pesquisa ou de saúde, além de centros de previsão do tempo, sejam estes públicos ou privados. A CETESB vem utilizando em caráter experimental um modelo de previsão de ozônio por Redes Neurais Artificiais. Este trabalho pretende subsidiar esse sistema de previsão.

Ao contrário dos modelos determinísticos, que dependem de um elevado custo de processamento computacional para sua operação, as redes neurais artificiais podem ser

implementadas mesmo onde não haja muita disponibilidade de recursos, uma vez que a estrutura do modelo já tenha sido desenvolvida. A aplicação do método depende da inserção de previsões meteorológicas à entrada do modelo que possuam um grau de antecedência, em dias, respectivo ao desejado para as concentrações de ozônio, a exemplo do que foi feito nos testes de aplicação nesta tese. Dessa forma, os modelos de RNA representam uma ferramenta útil na estimativa de concentrações de ozônio.

As incertezas associadas à aplicação do método estão dentro de um nível aceitável, considerando que a média dos erros se encontra em $19,9 \pm 1,5 \mu\text{g.m}^{-3}$ para concentrações médias de ozônio e $26,3 \pm 1,8 \mu\text{g.m}^{-3}$ para as máximas, de acordo com os valores dos parâmetros de desempenho (RMSE e RMSE_{ub}) encontrados em cada RNA dedicada a previsão. Conforme observado, a RNA é mais apta para prever concentrações mais baixas (menores que $150 \mu\text{g.m}^{-3}$, aproximadamente) do que as mais elevadas, por isso o desempenho foi melhor para a saída de concentrações médias do que para a de concentrações máximas. Essa é uma limitação decorrente do método em si, que não é bastante eficaz na estimativa de eventos extremos, uma vez que são de menor ocorrência e, portanto, possuem um padrão que não é bem caracterizado pela rede neural artificial, a qual enfatiza os casos mais frequentes.

Nos testes de aplicação expostos nos capítulos 4 e 5 as diferenças entre previsão e medição foram maiores, porém, além das incertezas do próprio método estatístico, há de se considerar aquelas associadas à previsão meteorológica. Portanto, o bom desempenho na previsão de ozônio troposférico é dependente da qualidade da previsão meteorológica fornecida como entrada do modelo. Outro fato é que São Paulo se caracteriza por ser uma região sujeita às instabilidades climáticas, o que dificulta a previsão meteorológica e, subsequentemente, a previsão de concentrações de ozônio. Entretanto, se os resultados forem interpretados de uma forma qualitativa, ou seja, classificando-os de acordo com faixas de

qualidade do ar (por exemplo: boa, regular e inadequada), as diferenças entre simulação e observação podem ser aceitáveis no intuito de previsão de índice de qualidade do ar.

Por esta pesquisa foi possível identificar as variáveis mais importantes, quantificando a importância relativa de cada uma para a previsão de ozônio troposférico. A temperatura média durante a tarde foi uma das variáveis mais importantes na caracterização do ozônio pela RNA, especialmente para Mauá e Santana. Outra variável significativa foi o campo de vento, expresso nas suas componentes zonais e meridionais. Em Ibirapuera o vento médio pela manhã desempenha uma grande importância na estimativa de ozônio, mais até que a temperatura do ar a tarde. Em São Miguel Paulista observa-se algo semelhante, mas apenas a média da componente zonal de manhã é que foi a mais relevante, sendo um pouco mais representativa que a temperatura média à tarde. Em São Caetano do Sul é a média da componente meridional do vento de manhã que aparece em destaque, deixando as demais variáveis aproximadamente em um mesmo nível de significância. Em P. D. Pedro II o campo de vento médio também foi mais importante que as demais variáveis, mas a diferença não foi muito significativa quando comparada com as RNAs das outras estações. Já a umidade relativa foi, de uma maneira geral, pouco importante na caracterização das concentrações de ozônio por RNA, em comparação as demais variáveis.

Com a inclusão de concentrações de poluentes precursores de ozônio à análise, observa-se, como esperado, a importância da química atmosférica para a previsão de ozônio. Para Ibirapuera é marcante a influência das médias de concentrações de NO no período da tarde. A RNA de São Caetano do Sul também aponta as concentrações médias de óxidos de nitrogênio à tarde como as majoritárias (NO_2 , NO e NO_x) na entrada do modelo; das variáveis meteorológicas, a umidade relativa média à tarde foi a mais importante, seguida das médias de vento zonal de manhã e temperatura à tarde. Em P. D. Pedro II as médias dos óxidos de nitrogênio à tarde também são importantes, mas a média da componente meridional do vento

à tarde continua sendo relevante. Para a RNA de Mauá as médias de concentrações de NO, NO_x e particulado inalável foram as variáveis de maior peso; dentre as variáveis meteorológicas, a temperatura ainda desempenha um papel significativo, especialmente nos valores do período da tarde. Já em São Miguel Paulista e Santana, onde não foi possível incluir medidas de óxidos de nitrogênio à RNA, a meteorologia prevalece como a mais importante. Em São Miguel Paulista a média de temperatura à tarde foi a mais representativa, seguida das médias à tarde de PI e vento meridional. Para a rede de Santana observou-se que o campo de vento de manhã foi importante na caracterização do padrão de ozônio pela rede neural.

A adição da química atmosférica à entrada da RNA trouxe melhorias para a qualidade dos resultados, com relação às redes que utilizam apenas fatores meteorológicos como entrada. Isso ocorre especialmente para as redes neurais das estações P. D. Pedro II e São Caetano do Sul. A primeira apresentou melhorias no ajuste de 19% e 31%, para estimativas de concentrações médias e máximas de ozônio, respectivamente. Enquanto que para São Caetano do Sul esses valores foram de 15% e 22%. Entretanto, em ambos os casos não foi possível realizar a validação do modelo, devido a falta dos dados necessários. Para a RNA de Mauá a melhora foi pequena, de 6% e 9%, respectivas às concentrações médias e máximas de ozônio, mas a validação apresentou bons resultados. Já para São Miguel Paulista, houve poucas diferenças com relação a RNA anterior (pouco mais de 1%), inclusive, a validação teve um desempenho inferior. A RNA de Santana também apresentou pequenas melhoras, sendo estas de 4% e 12%, para concentrações médias e máximas de ozônio, mas não foi possível validar o modelo para uma melhor verificação. Em Ibirapuera a qualidade do ajuste da rede neural subiu em 6% e 9%, para concentrações médias e máximas de ozônio, o que pode ser comprovado pela validação, embora os resultados não tenham diferenciado muito do obtido para a rede que utiliza apenas entradas meteorológicas.

Dessa forma, apesar da importância da química atmosférica para a formação do ozônio, é perfeitamente possível estimar as concentrações deste poluente com base apenas nas características meteorológicas, utilizando a técnica de redes neurais artificiais. Com isso é possível implementar os modelos estatísticos de forma operacional, na previsão de concentrações médias e máximas de ozônio para o período da tarde, quando acontecem suas concentrações mais elevadas.

7.2 Perspectivas para trabalhos futuros

Outras metodologias de previsão podem ser associadas à técnica de redes neurais artificiais para uma melhor caracterização das estimativas de concentração de ozônio, como o emprego de modelos tridimensionais de qualidade do ar, por exemplo. Sendo possível, dessa forma, estabelecer um sistema de previsão de concentração de poluentes através de diferentes ferramentas e análises.

Futuramente esta pesquisa precisará ser atualizada com novos dados obtidos das estações de monitoramento de qualidade do ar da Cetesb, de forma que reflitam as mudanças no clima e no perfil de emissões que eventualmente ocorrerão. O estudo da previsibilidade do ozônio troposférico pode ser feito por meio de técnicas estatísticas mais simples, como a regressão linear e árvores de classificação e regressão (CART). Embora não sejam esperados resultados com a mesma qualidade da aplicação das redes neurais artificiais, essa atividade pode ser interessante em um estudo de iniciação científica, por proporcionar um primeiro contato com o tratamento e análise de dados, além da compreensão dos processos químicos na atmosfera sob influência da meteorologia.

Outra recomendação é a utilização operacional da RNA com a validação dos resultados obtidos de previsão de máximo horário de ozônio. Nesse contexto uma análise considerando a variação espacial das concentrações deve ser incluída em análises futuras.

Referências

ABDUL-WAHAB, S.A.; BAKHEIT, C.S.B.; AL-ALAWI, S.M. (2004). Principal component and multiple regression analysis in modelling of ground-level ozone and factors affecting its concentrations. *Environmental Modelling & Software* (2004)

ANDRADE, M.F.; YNOUE, R.Y.; HARLEY, R.; MIGUEL, A.H. (2004). Air-quality model simulating photochemical formation of pollutants: the São Paulo Metropolitan Area, Brazil. *International Journal of Environment and Pollution* 22(4), 460-475.

BORGES, A.S. (2003). Impacto do refinamento da grade horizontal do modelo fotoquímico CIT nas estimativas de concentrações de poluentes atmosféricos. Dissertação de Mestrado. Instituto de Astronomia, Geofísica e de Ciências Atmosféricas da Universidade de São Paulo.

CARDOSO, M.F. (1997). Produção e Transporte de Ozônio Troposférico em Cubatão, São Paulo. Dissertação de Mestrado. Instituto Astronômico e Geofísico da Universidade de São Paulo.

CHALOULAKOU, A.; SAISANA, M.; SPYRELLIS, N. (2003). Comparative assessment of neural networks and regression models for forecasting summertime ozone in Athens. *The Science of the Total Environment* 313, 1-13

CETESB (2007). Relatório de Qualidade do Ar no Estado de São Paulo 2006. Secretaria do Meio Ambiente, Série Relatórios. São Paulo. Disponível em <<http://www.cetesb.sp.gov.br>>

DAYHOFF, J.E. (1990). *Neural Network Architectures - An Introduction*. Van Nostrand Reinhold, New York, 1990.

EPA - OAQPS (1999). Guideline for developing an ozone forecasting program. United States Environmental Protection - Office of Air Quality Planning and Standards; EPA-454/R-99-009.

FARHAT, S.C.L.; PAULO, R.L.P.; SHIMODA, T.M.; CONCEIÇÃO, G.M.S.; LIN, C.A.; BRAGA, A.L.F.; WARTH, M.P.N.; SALDIVA, P.H.N. (2005). Effect of air pollution on pediatric respiratory emergency room visits and hospital admissions. *Brazilian Journal of Medical and Biological Research* (2005) 38: 227-235

FINLAYSON-PITTS, B.J.; PITTS, J.N. (1997). Tropospheric Air Pollution: Ozone, Airborne Toxics, Polycyclic Aromatic Hydrocarbons, and Particles. *Science*, 276, 1045 - 1052.

FREITAS, E.D. (2003), Circulações locais em São Paulo e sua influência sobre a dispersão de poluentes. Tese de Doutorado, Instituto de Astronomia, Geofísica e Ciências Atmosféricas – Universidade de São Paulo.

FREITAS, E.D.; MARTINS, L.D.; DIAS, P.L.S.; ANDRADE, M.F. (2005). A simple photochemical module implemented in RAMS for tropospheric ozone concentration forecast in the Metropolitan Area of Sao Paulo-Brazil: Coupling and validation. *Atmospheric Environment* 39(36), 6352-6361.

GARDNER, M.W.; DORLING, S.R. (2000). Statistical surface ozone models: an improved methodology to account for non-linear behaviour. *Atmospheric Environment*, Volume 34, Issue 1, January 2000, Pages 21-34

GRIVAS, G.; CHALOULAKOU, A. (2006). Artificial neural network models for prediction of PM10 hourly concentrations, in the Greater Area of Athens, Greece. *Atmospheric Environment* 40, 1216-1229

GUARDANI, R.; NASCIMENTO, C.A.O.; GUARDANI, M.L.G.; MARTINS, M.H.R.B.; ROMANO, J. (1999). Study of Atmospheric Ozone Formation by Means of a Neural Network-Based Model. *Journal of the Air & Waste Management Association* 1999; 49: 316-323.

GUARDANI, R.; AGUIAR, J.L.; NASCIMENTO, C.A.O.; LACAVA, C.I.V.; YANAGI, Y. (2003). Ground-Level Ozone Mapping in Large Urban Areas Using Multivariate Statistical Analysis: Application to the São Paulo Metropolitan Area. *Journal of the Air & Waste Management Association* 2003; 53: 553-559. ISSN 1047-3289

HEO, J.; KIM, D. (2004). A new method of ozone forecasting using fuzzy expert and neural network systems. *Science of the Total Environment* 325, 221-237

JORQUERA, H.; PÉREZ, R.; CIPRIANO, A.; ESPEJO, A.; LETELIER, M.V.; ACUÑA, G. (1998). Forecasting ozone daily maximum levels at Santiago, Chile. *Atmospheric Environment* 32(20), 3415-3424

KOÇAK, K.; SAYLAN, L.; SEN, O. (2000). Nonlinear time series prediction of O₃ concentration in Istanbul. *Atmospheric Environment* 34, 1267-1271

KUKKONEN, J.; PARTANEN, L.; KARPPINEN, A.; RUUSKANEN, J.; JUNNINEN, H.; KOLEHMAINEN, M.; NISKA, H.; DORLING, S.; CHATTERTON, T.; FOXALL, R.; CAWLEY, G. (2003). Extensive evaluation of neural network models for the prediction of NO₂ and PM₁₀ concentrations, compared with a deterministic modelling system and measurements in central Helsinki. *Atmospheric Environment* 37, 4539-4550

LANDIM, P.M.B (2000). Introdução aos métodos de estimação espacial para confecção de mapas. DGA, IGCE, UNESP/Rio Claro, Lab. Geomatemática, Texto Didático 02, 20 pp. 2000. Disponível em <<http://www.rc.unesp.br/igce/aplicada/textodi.html>>

LIN, C.A. (1997). Efeitos da poluição atmosférica sobre a morbidade respiratória aguda na população infantil de São Paulo, 1997. Tese de Doutorado. Faculdade de Medicina – Universidade de São Paulo.

MAFFEIS, G. (1999). Prediction of carbon monoxide acute air pollution episodes. Model formulation and first application in Lombardy. *Atmospheric Environment* 33, 3859-3872

MARTINS, L.D. (2002). Estimativas do impacto das emissões de hidrocarbonetos pela vegetação na formação de oxidantes fotoquímicos em São Paulo. Dissertação de Mestrado. Instituto Astronômico e Geofísico - Universidade de São Paulo.

MARTINS, L.D.; VASCONCELLOS, P.C.; CARVALHO, L.R.F.; ANDRADE, M.F. (2006a). Estimated impact of biogenic hydrocarbon emissions on photochemical oxidant formation in São Paulo during two periods of the winters of 1999-2000. *Revista Brasileira de Meteorologia* 21(2), 2006.

MARTINS, L. D.; ANDRADE, M. F.; FREITAS, E. D.; PRETTO, A. G.; LUCIANA, A. E. L.; TOMAZ E.; GUARDANI, M. L.; MARTINS, M. H. R. B.; JUNIOR, O. M. A. (2006b). Emission Factors for Gas-powered Vehicles traveling through Road Tunnels in São Paulo City, Brazil. *Environmental Science and Technology Library*. , v.40, p.6722 - 6729, 2006.

MCRAE, G.; GOODIN, W.; SEINFELD, J. (1982a). Mathematical modeling of photochemical air pollution. Final report to the State of California Air Resources Board (Contract nos. A5-046-9-87 e A7-187-30).

MCRAE, G.; GOODIN, W.; SEINFELD, J. (1982b). Development of a second-generation mathematical model for urban air pollution - I. model formulation. *Atmospheric Environment* Vol 16, No 4, pp. 679-696.

MILFORD, J.; RUSSELL, A.; MCRAE, G. (1989). A new approach to photochemical pollution control: implications of spatial patterns in pollutant responses to reductions in nitrogen oxides and reactive organic gas emissions. *Environment Science Technology*, 23, 1290-1301.

OLIVEIRA, A. P. e SILVA DIAS, P. L. (1982). Aspectos Observacionais da Brisa Marítima em São Paulo; *Anais do 2o Congresso Brasileiro de Meteorologia, Pelotas*, vol.2, p. 129-161, 1982.

PASTOR-BÁRCENAS, O.; SORIA-OLIVAS, E.; MARTÍN-GUERRERO, J.D.; CAMPS-VALLS, G.; CARRASCO-RODRÍGUEZ, J.L.; DEL VALLE-TASCÓN, S. (2005). Unbiased sensitivity analysis and pruning techniques in neural networks for surface ozone modelling. *Ecological Modelling* 182, 149-158

PEREIRA, L.A.A.; LOOMIS, D.O.; CONCEIÇÃO, G.M.S. (1998). Association between air pollution and intrauterin mortality in Sao Paulo, Brazil. *Environ Health Perspect* 1998; 106:325-9.

RUIZ-SUÁREZ, J.C.; MAYORA-IBARRA, O.A.; TORRES-JIMÉNEZ, J.; RUIZ-SUÁREZ, L.G. (1995). Short-term ozone forecasting by artificial neural networks. *Advances in Engineering Software* 23, 143-149

RUMELHART, D.; MCCLELLAND, J. (1986). *Parallel distributed processing explorations in the microstruture of cognition* , v. 1, cap. 8 , MIT , Cambridge , Mass., 1986

SÁNCHEZ-CCOYLLO, O. R.; YNOUE, R.Y.; MARTINS, L.D.; ANDRADE, M.F. (2006). Impacts of ozone precursor limitation and meteorological variables on ozone concentration in São Paulo, Brazil. *Atmospheric Environment* 40, S552-S562.

SANCHEZCCOYLLO, O. R.; YNOUE, R. Y.; MARTINS, L. D.; ASTOLFO, R.; MIRANDA, R. M.; FREITAS, E.; BORGES, A. S.; FORNARO, A.; Freitas, H.; MOREIRA, A.; ANDRADE, M. F. (2008). VEHICULAR PARTICULATE MATTER EMISSIONS IN ROAD TUNNELS IN SAO PAULO, BRAZIL. *Environmental Monitoring and Assessment* . , v.1, 2008.

SEINFELD, J.H. (1986). *Atmospheric Chemistry and Physics of Air Pollution*, Wiley-Interscience, New York.

SCHLINK, U.; DORLING, E.; PELIKAN, E.; NUNNARI, G.; CAWLEY, G.; JUNNISEN, H.; GREIG, A.; FOXALL, R.; EBEN, K.; CHATTERTON, T. (2003). A rigorous inter-comparison of ground-level ozone predictions, *Atmospheric Environment*, Volume 37, Issue 23, July 2003, Pages 3237-3253

UNEP-WHO (1994). Air pollution in the world's megacities - a report from the United-Nations Environment Program and the World Health Organization. *Environment* 36 (2), 2-37.

VIVANCO, M.; ANDRADE, M.F. (2006). Validation of the emission inventory in the Sao Paulo Metropolitan Area of Brazil, based on ambient concentrations ratios of CO, NMOG and NOx and on a photochemical model. *Atmospheric Environment* 40, 1189-1198.

WANG, W.; LU, W.; WANG, X.; LEUNG, A.Y.T (2003). Prediction of maximum daily ozone level using combined neural network and statistical characteristics. *Environment International* 29, 555-562

WHO (1976). Photochemical oxidants: Environmental Health Criteria 7. World Health Organization, Geneva.

YNOUE, R.Y (1999). O impacto do uso de campos meteorológicos diagnósticos e prognósticos na produção e dispersão dos oxidantes fotoquímicos em São Paulo. Dissertação de Mestrado, Instituto Astronômico e Geofísico - Universidade de São Paulo.

ZOLGHADRI, A.; MONSION, M.; HENRY, D.; MARCHIONINI, C.; PETRIQUE, O. (2004). Development of an operational model-based warning system for tropospheric ozone concentrations in Bordeaux, France. *Environmental Modelling & Software* 19, 369-382