

توسعه مدل پیش‌بینی غلظت ازن در هوا با استفاده از شبکه عصبی مصنوعی

مهرداد رفیع پور گتابی^{۱*}، علی اصغر آل‌شیخ^۲، عباس علیمحمدی^۳، ابوالقاسم صادقی نیارکی^۴

۱- کارشناسی ارشد GIS، دانشکده مهندسی نقشه‌برداری، دانشگاه صنعتی خواجه نصیرالدین طوسی

۲- استاد، دانشکده مهندسی نقشه‌برداری، دانشگاه صنعتی خواجه نصیرالدین طوسی

۳- دانشیار، دانشکده مهندسی نقشه‌برداری، دانشگاه صنعتی خواجه نصیرالدین طوسی

۴- استادیار، دانشکده مهندسی نقشه‌برداری، دانشگاه صنعتی خواجه نصیرالدین طوسی

(تاریخ دریافت: ۱۳۹۲/۲/۱۹ - تاریخ تصویب: ۱۳۹۴/۲/۲۷)

چکیده

با توجه به مضرات ازن بر سلامت انسان و محیط‌زیست و افزایش آن در دهه‌های گذشته، بررسی و پیش‌بینی میزان آن در هوا از اهمیت بالایی برخوردار است. پیش‌بینی غلظت ازن در هوا می‌تواند برای پیشگیری و کنترل توسط مسئولان استفاده شود. در این مقاله پارامترهای مهم و تأثیرگذار بر غلظت ازن با استفاده از داده‌های پایش کیفیت هوا ایستگاه‌های آزادی و امام خمینی طی سال‌های ۲۰۰۹ تا ۲۰۱۰، بررسی شده است. در این راستا متغیرهای هواشناسی شامل رطوبت نسبی، دما، فشار، سرعت و جهت باد با غلظت ازن به کمک همبستگی خطی و تحلیل مؤلفه‌های اصلی تحلیل شد. مطالعه همبستگی بین متغیرهای مختلف هواشناسی با غلظت ازن نشان داد که غلظت ازن تحت تأثیر پارامترهای رطوبت نسبی، دما و سرعت باد است. تأثیر پارامترهای رطوبت نسبی و دما را می‌توان با توجه به عملکرد فتوشیمیایی و نقش این فرآیند در تولید ازن توصیف کرد. در حالی که همبستگی ازن با پارامتر سرعت باد نشان‌دهنده انتقال ازن از مناطق دیگر است. در ادامه سعی شد از یک شبکه عصبی پرسپترون چندلایه جهت پیش‌بینی غلظت ازن استفاده شود. ورودی‌های این شبکه عبارتند از رطوبت نسبی، دما و سرعت باد که قبلاً تأثیر آنها بر غلظت ازن اثبات شده است. نتایج اجرای شبکه عصبی توسعه داده شده در ایستگاه‌های میدان آزادی و امام خمینی در تهران نشان داد که مدل طراحی شده میزان غلظت ازن را با دقت ۶۷ تا ۹۷ درصد در ۲۴ ساعت آینده تخمین می‌زند. نتایج این تحقیق و مدل توسعه داده شده می‌تواند برای مدیریت بهتر آلودگی هوا و کنترل غلظت ازن توسط مدیران شهری مورد استفاده قرار گیرد.

واژگان کلیدی: ازن، مدل‌سازی، تحلیل مؤلفه‌های اصلی، شبکه عصبی

برای ایجاد مدل قطعی جهت پیش‌بینی غلظت ازن، فرآیند پیچیده‌ای است؛ که نیاز به استفاده از تعداد زیادی متغیرهای شیمیایی و فیزیکی، و نیاز به داده‌های ورودی بسیار با دقت دارد. به همین دلیل توسعه این نوع از مدل‌ها غالباً دشوار است (Coman *et al.*, 2008). از جمله این مدل‌ها می‌توان به تحقیقی که Finardi و همکاران (۲۰۰۸) انجام دادند، اشاره کرد که سیستمی جهت مدلسازی آلودگی هوا با استفاده از پارامترهای هواشناسی و مدل انتقال شیمیایی ایجاد کرده‌اند.

بعضی تحقیقات به مدلسازی آلودگی هوا با استفاده از مدل‌های خطی پرداخته‌اند در حالی که مدل‌های آماری متداول قادر به شناسایی و مدلسازی روابط غیر خطی، موثر بر آلودگی هوا نیستند. برای مثال Duenas و همکاران (۲۰۰۲) به تحلیل یک ساله غلظت ازن در منطقه مالاگا پرداخته‌اند. آنها در این تحقیق از رگرسیون چند متغیره برای پیش‌بینی غلظت ازن به همراه پارامترهای هواشناسی استفاده کرده‌اند. نتیجه تحقیق نشان داد که مدل می‌تواند در شرایطی که غلظت ازن نامعلوم است با استفاده از پارامترهای هواشناسی آن را پیش‌بینی کند. غلظت ازن پیش‌بینی شده توسط مدل با غلظت مشاهده شده ازن، دارای ضریب همبستگی بین ۵۵ تا ۸۵ درصد می‌باشد.

در سال‌های اخیر استفاده از شبکه‌های عصبی مصنوعی که نیاز به مهارت زیادی در استفاده از آنها وجود ندارد (Abbaspour *et al.*, 2005)؛ همچنین قادر به شناسایی روابط غیر خطی بین مقادیر ورودی و خروجی هستند، برای مدلسازی آلودگی هوا مورد استفاده قرار گرفته‌اند (Singh *et al.*, 2012) و همکاران (۲۰۰۹) اقدام به مدلسازی آلودگی هوا با استفاده از شبکه عصبی و پارامترهای همبسته با غلظت آلاینده‌ها نمودند. این پارامترها شامل حجم ترافیک، ساعت روز، روز هفته، غلظت آلاینده در ۱ تا ۳ ساعت گذشته، سرعت و جهت باد، دما، میزان تابش، بارش باران، رطوبت، فاصله تا محور خیابان و جهت خیابان بودند. نتایج

آلودگی هوا یکی از مشکلات محیط زیستی در بسیاری از شهرهای جهان است. میزان آلودگی هوا معمولاً با غلظت آلاینده‌های هوا مانند اکسیدهای نیتروژن، دی اکسید سولفور، مونوکسیدکربن، ذرات معلق هوا و ازن سنجیده می‌شود (Chan & Jian, 2013). ازن یکی از آلاینده‌های مهم هوا است که در تحقیقات زیادی مورد توجه قرار گرفته است (Alalawi *et al.*, 2008). در بین آلاینده‌های هوا، در دهه‌های اخیر میزان غلظت جهانی ازن افزایش یافته است (Sousa *et al.*, 2007). ازن از آلاینده‌های ثانویه است؛ که برخلاف آلاینده‌های اولیه، به صورت غیر مستقیم و بر اثر واکنش جوی آلاینده‌های اولیه تشکیل می‌شود. غلظت ازن بستگی به واکنش‌های فوتوشیمیایی بین آلاینده‌های دیگر مانند NO₂ و پارامترهای هواشناسی مانند شدت تابش خورشیدی دارد (Ibarra-Berastegi *et al.*, 2008). به همین دلیل پیش‌بینی غلظت این آلاینده در هوا و تعیین پارامترهای تأثیر گذار بر غلظت آن بسیار مشکل است. یکی از راه‌های غلبه بر این مشکل استفاده از تحلیل مؤلفه‌های اصلی است که روشی قابل قبول برای شناسایی الگو و تعیین عوامل موثر است (Sousa *et al.*, 2007). به عنوان مثال Olajire و همکاران (۲۰۱۱) برای تشخیص درصد مشارکت هر یک از پارامترهای تأثیرگذار بر آلودگی هوا از تحلیل مؤلفه‌های اصلی استفاده کردند. اثرات مضر ازن بر سلامتی انسان و گیاهان باعث افزایش نگرانی در ارتباط با این آلاینده شده است (Olajire *et al.*, 2011)؛ سیستم‌های مختلفی با ابزارها و مدل‌های متفاوتی (Siwek & Osowski, 2012) برای پیش‌بینی ۲۴-ساعته و آگاه‌سازی عمومی راجع به این آلاینده توسعه داده شده است (Schlink *et al.*, 2003). در ادامه به چند مورد از آنها اشاره می‌شود.

برای مدلسازی و پیش‌بینی ازن، دو نوع مدل می‌تواند مورد استفاده قرار گیرد: مدل‌های قطعی و آماری. از جمله مدل‌های قطعی معادلات دیفرانسیل جزئی است. استفاده از معادلات دیفرانسیل جزئی

در مورد جهت باد تبدیلی به شرح زیر انجام شد تا این پارامتر از حالت غیرخطی و دوره‌ای به خطی تبدیل شود (Voukantsis et al., 2011):

$$\sin WD = \sin\left(\frac{2\pi(v - \min(v))}{\max(v) - \min(v)}\right)$$

$$\cos WD = \cos\left(\frac{2\pi(v - \min(v))}{\max(v) - \min(v)}\right)$$

که در آن v پارامتر زمانی جهت باد است در حالی که $\sin WD$ و $\cos WD$ مقادیر خطی شده جهت باد هستند.

۲-۲ روش تحقیق

در این تحقیق از تحلیل مؤلفه‌های اصلی برای شناسایی پارامترهای مؤثر بر غلظت ازن استفاده شده است. سپس از شبکه عصبی مصنوعی برای پیش‌بینی غلظت ازن در ۲۴ ساعت بعد با استفاده از پارامترهای تأثیر گذار استفاده شده است. لازم به ذکر است که تمامی تحلیل‌های صورت گرفته در این تحقیق با استفاده از نرم افزار MATLAB 2013 صورت گرفته است. در ادامه هریک از روش‌های به کار گرفته شده در این تحقیق توضیح داده شده است.

تحلیل مؤلفه‌های اصلی

برای مدلسازی و پیش‌بینی آلودگی هوا نیاز به حجم عظیمی از داده‌ها می‌باشد. در مورد داده‌های مورد استفاده در این مقاله، ۸۷۶۰ ساعت در تعداد ۸ پارامتر (غلظت ازن، سرعت باد، سینیوس و کسینوس جهت باد، دما، فشار، رطوبت نسبی و ساعت برداشت داده)، تعداد ۷۰۸۰ داده را شامل می‌شود. یکی از سوالاتی که باید پاسخ داده شود یقیناً پارامترهای مؤثر بر غلظت ازن می‌باشد. تحلیل مؤلفه‌های اصلی، روش مناسبی جهت استخراج پارامترهای مهم است؛ این روش داده‌های وابسته را حذف می‌کند. این افزونگی با انتقال به مجموعه‌ای جدید از متغیرها (مؤلفه‌های اصلی) که نسبت به هم همبستگی ندارند،

نشان داد، مدل استفاده شده قادر به پیش‌بینی دقیق غلظت آلاینده‌ها در ۱۰ ساعت آینده است. شبکه عصبی به کار گرفته شده برای پیش‌بینی غلظت ازن دارای میانگین مجذور مربعات خطا برابر 10 ppb^1 تا 16 ppb و ضریب همبستگی بین $0/85$ تا $0/95$ بین مقادیر واقعی و مقادیر پیش‌بینی شده، می‌باشد.

هدف اصلی مقاله حاضر شناسایی پارامترهای هواشناسی تأثیرگذار بر غلظت ازن و سپس مدلسازی و پیش‌بینی ۲۴ ساعت آینده غلظت ازن با استفاده از پارامترهای تأثیرگذار است. پارامترهای تأثیرگذار با استفاده از ضریب همبستگی و تحلیل مؤلفه‌های اصلی مشخص می‌شوند؛ در حالی که مدلسازی و پیش‌بینی با استفاده از پارامترهای هواشناسی و به کمک شبکه عصبی انجام می‌گیرند.

۲- مواد و روش‌ها

۱-۲ معرفی منطقه مورد مطالعه

شهر تهران بزرگ‌ترین و پرجمعیت‌ترین شهر ایران است. شهری که به دلیل جمعیت زیاد و ترافیک شدید تحت تأثیر آلودگی شدید هوا قرار دارد. در این مطالعه از داده‌های ایستگاه‌های پایش کیفیت هوای آزادی و امام خمینی متعلق به سازمان محیط زیست، استفاده شده است. شکل ۱ موقعیت ایستگاه‌های سنجش کیفیت هوا را نشان می‌دهد.

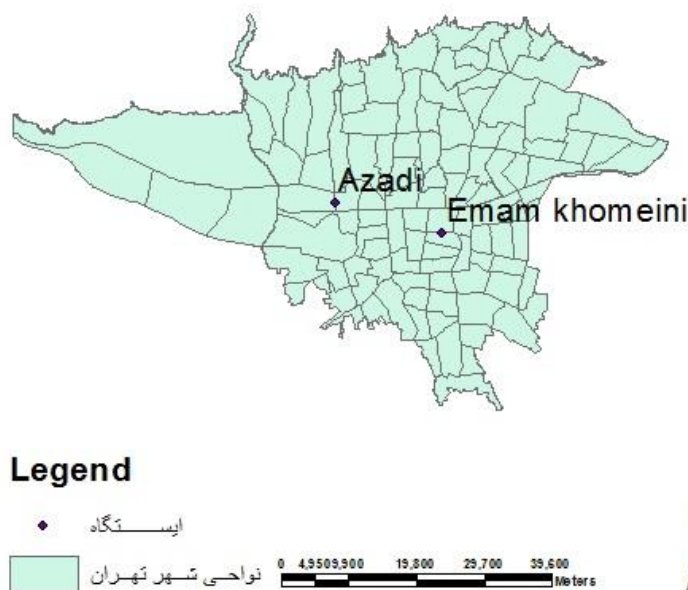
در این تحقیق از داده‌های غلظت ازن و پارامترهای هواشناسی سال‌های ۲۰۰۹-۲۰۱۰ استفاده شده است. پارامترهای هواشناسی مورد استفاده در این تحقیق شامل دما، رطوبت نسبی، سرعت و جهت باد می‌باشد. لازم به ذکر است که داده‌های غلظت ازن و پارامترهای هواشناسی مستقیماً توسط دستگاه‌های پایش کیفیت هوا به صورت ساعتی اندازه‌گیری می‌شود.

¹ Part Per Billion

سپس مقادیر ویژه و مقادیر ویژه نرمالیزه شده محاسبه می‌شود؛ و ورودی‌ها به فضای جدید برای کاهش ابعاد داده‌ها، انتقال می‌یابد (Jolliffe, 2002).

انجام می‌شود (Lu *et al.*, 2004). روش تحلیل مؤلفه‌های اصلی به این صورت است که ابتدا اطلاعات را برای نشان دادن اینکه فضای داده‌ها در جهت کدام ورودی‌ها بیشترین تغییرات را دارد، بررسی شود.

مکان ایستگاه‌های مورد استفاده در تحقیق



شکل ۱: ایستگاه‌های پایش کیفیت هوا

۲۰۱۰) دو ایستگاه استفاده شده است. اگر داده‌ها به صورت خام مورد استفاده قرار گیرند، تحلیل مؤلفه‌های اصلی به واحد متغیرها بستگی خواهد داشت (Karatzas & Kaltsatos, 2007). پس در گام اول برای ورود داده‌ها به پردازش‌های بعدی باید داده‌ها نرمالیزه شوند. داده‌ها در این تحقیق طبق فرمول ۲ نرمال‌سازی شدند:

$$\text{normalized data} = \frac{\text{raw data} - \text{mean}(\text{data})}{\text{std}(\text{data})}$$

که در آن برای بدست آوردن داده‌های نرمال

با انجام تحلیل مؤلفه‌های اصلی می‌توان پارامترهای موثر بر غلظت ازن را برای تحلیلها و بررسی‌های بعدی تعیین کرد. قبل از استفاده از تحلیل مؤلفه‌های اصلی باید میزان تناسب داده‌ها را برای استفاده از این تحلیل سنجید. در این پژوهش به کمک شاخص ^۱ KMO تناسب داده‌ها برای استفاده از تابع تحلیل مؤلفه اصلی سنجیده شده است. در صورتی که مقدار شاخص KMO کمتر از ۰,۵ باشد، استفاده از تحلیل مؤلفه‌های اصلی برای داده‌های مورد نظر جایز نیست (Williams *et al.*, 2012).

در این تحقیق از داده‌های یک ساله (۲۰۰۹-

¹ Kaiser-Meyer-Olkin

همانطور که در این شکل دیده می‌شود در لایه اول (Input layer) به تعداد پارامترهای ورودی نرون (۴) وجود دارد؛ در حالی که لایه دوم (Hidden layer) شامل ۱۰ نرون است. مقادیر ورودی در ماتریس وزن (w) ضرب می‌شوند و سپس با مقادیر بایاس^۸ (b) جمع می‌شوند این مقادیر ورودی^۹ لایه پنهان هستند. مقادیر ورودی در لایه پنهان مشابه مرحله قبل در مقادیر ماتریس وزن ضرب شده و با مقادیر بایاس جمع می‌شوند تا ورودی برای لایه آخر شبکه عصبی تامین شود. همچنین نرون‌های لایه آخر (Output layer) برابر تعداد خروجی‌ها (غلظت ازن) است. در لایه میانی تابع انتقال سیگموئید^{۱۰} و لایه خروجی تابع انتقال خطی دارد. دلیل استفاده از این توابع انتقال T این است که به‌کارگیری تابع انتقال سیگموئید در لایه پنهان و تابع خطی در لایه خروجی می‌تواند برای تخمین هر تابع غیرخطی به کار گرفته شود (Menhaj, 2011).

دقت مدل‌سازی می‌تواند با استفاده از پارامترهای میانگین مطلق خطا (MAE)، میانگین مربعات خطا (RMSE)، شاخص توافق (IA) و ضریب همبستگی (R) طبق فرمول‌های زیر، تعیین شود (Willmott et al., 1985).

$$MAE = \frac{\sum |O_i - P_i|}{n} \quad (3)$$

$$RMSE = \sqrt{\frac{\sum (O_i - P_i)^2}{n}} \quad (4)$$

$$IA = 1 - \frac{\sum (P_i - O_i)^2}{\sum (|P_i - \bar{O}| + |O_i - \bar{O}|)^2} \quad (5)$$

$$R = \frac{cov(O, P)}{\sigma_O \sigma_P} \quad (6)$$

در فرمول‌های بالا O_i مقدار واقعی، P_i مقدار پیش‌بینی شده، n تعداد نمونه‌ها، \bar{O} میانگین مقادیر

شده^۱، میانگین داده‌ها^۲ را از داده‌های اولیه^۳ کسرنموده و سپس مقادیر به دست آمده بر واریانس داده‌ها^۴ تقسیم شده است (Stephenson, 2003).

شبکه عصبی

شبکه عصبی موضوعی نوین در زمینه تحلیل داده‌ها است که می‌تواند به عنوان روشی شبه پارامتریک^۵ دسته‌بندی شود. شبکه عصبی روشی داده‌مبنا است که توانایی تشخیص ساختار غیر خطی داده‌ها را دارد (Balestrassi et al., 2009). شبکه عصبی دارای خصوصیات ویژه‌ای همچون قابلیت یادگیری، تطبیق، تعمیم، خوشه‌بندی و یا مرتب‌سازی داده‌ها با استفاده از محاسبات موازی^۶ می‌باشد (Haykin, 1998) ایده شبکه‌های عصبی اولین بار توسط مک کالو و پیت مطرح گردید (McCulloch & Pitts, 1943). ساده‌ترین نوع از شبکه‌های عصبی، شبکه عصبی پرسپترون^۷ است؛ این نوع شبکه فقط می‌تواند برای مسائل خطی مورد استفاده قرار گیرد اما با اضافه کردن یک لایه پنهان شبکه عصبی می‌تواند برای مسائل غیر خطی نیز مورد استفاده قرار گیرد (Kröse & van der Smagt, 1996). شبکه عصبی چندلایه پرکاربردترین شبکه عصبی مصنوعی است، چون در صورتی این شبکه عصبی به درستی آموزش داده شود، با یک لایه میانی قادر به حل بسیاری از توابع است. هر چند که استفاده از چند لایه انعطاف‌پذیری شبکه را بیشتر می‌کند (Ibarra-Berastegi et al., 2008). در این مقاله از شبکه عصبی پرسپترون چندلایه برای مدل‌سازی غلظت ازن با استفاده از پارامترهای موثر و سپس پیش‌بینی ۲۴ ساعته آن استفاده شده است. ساختار شبکه عصبی مورد استفاده در این تحقیق در شکل ۲ دیده می‌شود.

¹ Normalized data

² Mean of data

³ Raw data

⁴ Variance of data

⁵ Semi parametric

⁶ Parallel processing

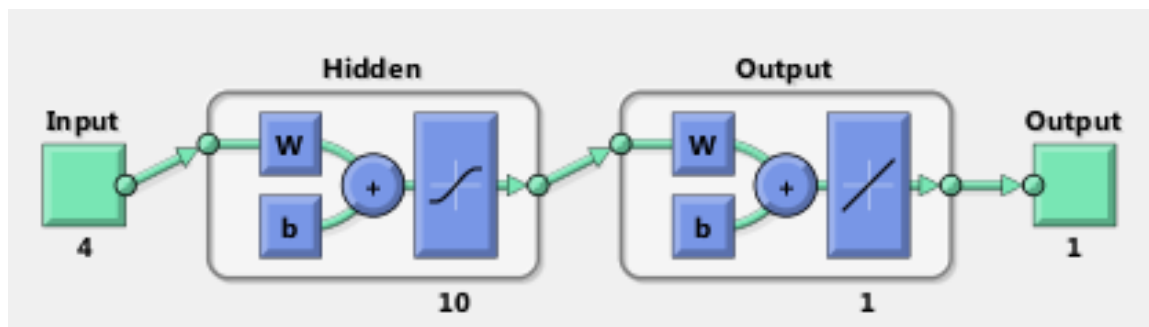
⁷ Perceptron

⁸ Bias

⁹ Input values

¹⁰ Sigmoid transfer function

واقعی، $COV(O, P)$ کوواریانس بین مقادیر واقعی و پیش‌بینی شده، σ_O واریانس مقادیر واقعی و σ_P واریانس مقادیر پیش‌بینی شده است.



شکل ۲: ساختار شبکه عصبی مورد استفاده در این تحقیق

مشاهده می‌شود، غلظت ازن در ایستگاه امام خمینی بسیار بیشتر از ایستگاه آزادی است.

همانطور که در ماتریس همبستگی (جدول ۲) قابل مشاهده است؛ این ماتریس یک ماتریس متقارن است که میزان همبستگی مقادیر هر ردیف را با مقادیر هر ستون نشان می‌دهد. غلظت ازن (در ستون اول ماتریس) با دما (سطر چهارم ماتریس) همبستگی مثبت و با رطوبت نسبی (سطر دوم ماتریس) همبستگی منفی دارد. این مساله با توجه به عملکرد فرآیند فتوشیمیایی، منطقی به نظر می‌رسد. زیرا افزایش دما، نتیجه افزایش میزان تابش خورشیدی است و کاهش رطوبت نسبی، نشان‌دهنده کاهش پوشش ابرها می‌باشد. هر دوی این اثرات باعث تولید ازن ناشی از فرآیند فتوشیمیایی هستند (EPA, 2006). علاوه بر این غلظت ازن با سرعت باد همبستگی مثبت دارد که بیانگر انتقال ازن از مناطق دیگر به منطقه آزادی است. بنابراین، در این مطالعه ازن فقط به صورت محلی و توسط عملکرد فتوشیمیایی تولید نشده است. این در حالی است که جهت باد تأثیر بسیار کمی در غلظت ازن دارد؛ که با توجه به اثر فتوشیمیایی که مبدا نقطه‌ای ندارد و در هر مکانی ممکن است به وجود بیاید، می‌تواند توجیه

۳- نتایج

معمولا داده‌های یک سری زمانی دارای نقصان و یا داده پرت ۱ آنهم به علت خرابی و یا کالیبره نبودن دستگاه اندازه‌گیری هستند (Kitagawa, 2010). درصد نقصان داده در ایستگاه آزادی در سال ۲۰۰۹-۲۰۱۰ برابر با ۳ درصد و در مورد ایستگاه امام خمینی ۵ درصد می‌باشد. این نقصان داده در شکل ۳ و ۴ بین ماه‌های ۶ و ۷ سال به خوبی قابل مشاهده است. در این تحقیق برای پیدا کردن داده‌های پرت از میانگین و انحراف معیار طبق فرمول ۷ استفاده شده است (Matlab, 2007).

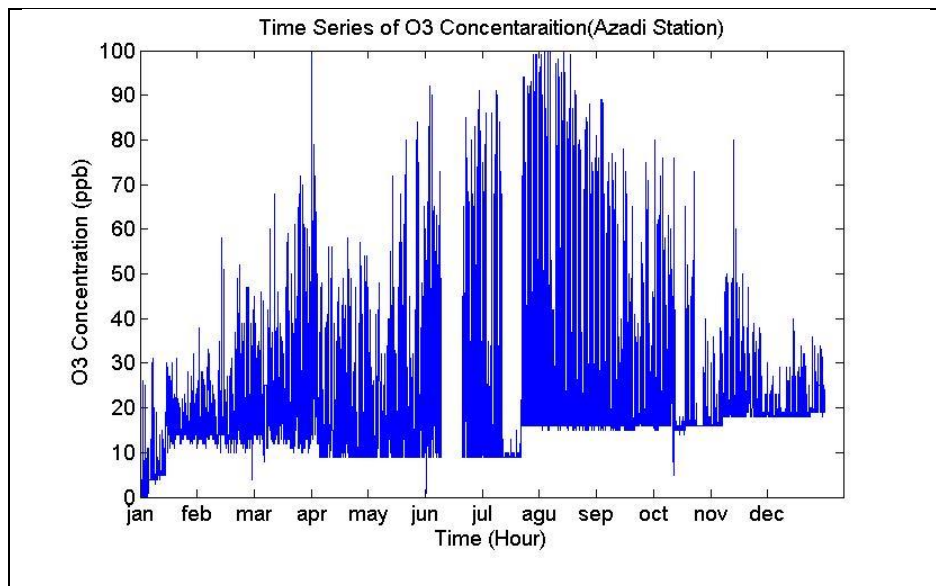
$$out = |x - \text{mean}| > 3 \times \sigma \quad (7)$$

که در آن mean میانگین داده‌ها، σ انحراف معیار داده‌ها و x هر کدام از مقادیر اولیه داده‌ها می‌باشد. البته در مقالات مشابه از میانه به جای میانگین استفاده شده است (Matlab, 2007).

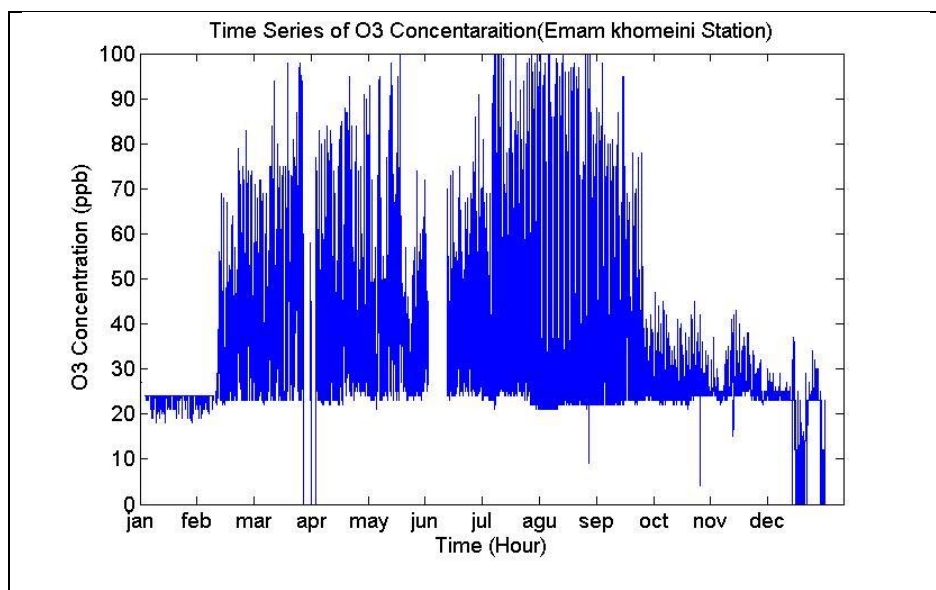
مشخصات آماری داده‌های مورد استفاده در این تحقیق در جدول ۱ قابل مشاهده است. همانطور که

¹Outlier

شود.



شکل ۳: نمودار غلظت ازن در ایستگاه آزادی در ماه های مختلف سال ۲۰۰۹-۲۰۱۰



شکل ۴: نمودار غلظت ازن در ایستگاه امام خمینی در ماه های مختلف سال ۲۰۰۹-۲۰۱۰

جدول ۱: پارامترهای آماری داده‌های مربوط به اندازه‌گیری‌های ازن و پارامترهای هواشناسی در سال ۲۰۰۹-۲۰۱۰

		میانگین	میانه	انحراف معیار	کمینه	بیشینه
ایستگاه آزادی	ازن (ppb)	۲۲/۵۲	۱۸	۱۴/۷	۱	۷۲
	رطوبت نسبی (%)	۳۶/۵۶	۳۳	۱۹/۶۵	۲/۱۹	۹۸/۹۳
	فشار (میلیمتر جیوه)	۹۹۶	۱۰۰۷	۴۵/۹۳	۴۶۳	۱۱۲۲
	دما (سانتی‌گراد)	۲۰/۰۳	۱۹/۵۱	۹/۳۶	-۰/۲	۴۲/۴۹
	سرعت باد (متر بر ثانیه)	۱/۶۱	۱/۳۶	۱/۱۳	۰/۰۰۱	۷/۳۶
	سینوس جهت باد	۰/۰۲۲	۰/۰۴	۰/۷۰	-۱	۱
	کسینوس جهت باد	-۰/۰۰۸	-۰/۰۳۳	۰/۷۱	-۱	۱
ایستگاه امام خمینی	ازن (ppb)	۳۵/۲۸	۲۵	۱۸/۱۷	۴	۹۱
	رطوبت نسبی (%)	۳۳/۹۴	۲۹/۴۱	۱۹/۴۷	۳/۰۵۱	۹۸/۵۴
	فشار (میلیمتر جیوه)	۸۸۱/۱	۸۸۱/۵	۱۲	۴۳۶/۱	۸۹۹/۸
	دما (سانتی‌گراد)	۱۸/۳۷	۱۷/۹۴	۹/۶۴	-۰/۵۳	۴۰/۹۴
	سرعت باد (متر بر ثانیه)	۱/۰۷۲	۱/۰۲۶	۰/۵۸	۰/۰۰۲	۵/۰۴۳
	سینوس جهت باد	-۰/۰۰۷	-۰/۰۱۲	۰/۷۰	-۱	۱
	کسینوس جهت باد	۰/۰۰۳	۰/۰۱۶	۰/۷۱	-۱	۱

جدول ۲: ماتریس همبستگی پارامترهای هواشناسی و غلظت ازن در ایستگاه آزادی

	ازن	رطوبت نسبی	فشار	دما	سرعت باد	سینوس جهت باد	کسینوس جهت باد
ازن	۱						
رطوبت نسبی	-۰/۳۱	۱					
فشار	-۰/۲۹	۰/۱۳	۱				
دما	۰/۳۴	-۰/۷۴	-۰/۰۶	۱			
سرعت باد	۰/۳۳	-۰/۱۸	-۰/۲۹	۰/۰۷	۱		
سینوس جهت باد	۰/۰۱	۰/۰۱	-۰/۰۱	۰	۰	۱	
کسینوس جهت باد	۰/۰۲	۰	-۰/۰۱	۰/۰۱	۰/۰۳	۰	۱

مؤلفه‌های اصلی میزان واریانس نشان‌دهنده درصدی از داده‌ها است که رفتاری به خصوص دارند. برای تعیین تعداد مؤلفه‌های مهم در تحقیق، روش‌های مختلفی وجود دارد. استفاده از حد آستانه می‌تواند در انتخاب تعداد مؤلفه‌های موثر مناسب باشد. معمولاً استفاده از حد آستانه به طوری که ۷۰ تا ۹۰ درصد داده‌ها را شامل شود، می‌تواند موثر باشد (Jolliffe, 2002).

شاخص KMO در مورد داده‌های استفاده شده در ایستگاه آزادی ۰/۵۹ است که بیشتر از ۰/۵۰ است. پس می‌توان تحلیل مؤلفه‌های اصلی را در مورد داده‌های ایستگاه آزادی به کار گرفت. نتایج تحلیل مؤلفه‌های اصلی در جدول ۳ نشان داده شده است. دو ردیف آخر این جدول به ترتیب واریانس و واریانس تجمعی را نشان می‌دهد. می‌توان این‌گونه برداشت کرد که در جدول تحلیل

جدول ۳: نتیجه تحلیل مؤلفه‌های اصلی در ایستگاه آزادی. هر سطر از هر ستون وزن هر پارامتر را در هر PC نشان می‌دهد.

	مؤلفه اول	مؤلفه دوم	مؤلفه سوم	مؤلفه چهارم	مؤلفه پنجم	مؤلفه ششم	مؤلفه هفتم
ازن	-۰/۴۷	-۰/۲۴	-۰/۰۴	۰	۰/۰۶۴	۰/۸۳	-۰/۱۴
رطوبت نسبی	۰/۵۴	-۰/۳۸	-۰/۰۱	۰/۰۰۱	۰/۰۰۸	۰/۳۲	۰/۶۷
فشار	۰/۲۹	۰/۵۵	۰/۰۶۹	۰/۰۴۷	۰/۷۲	۰/۲۶	-۰/۰۶
دما	-۰/۵۳	۰/۴۶	۰/۰۱۱	۰/۰۲۹	-۰/۰۷	-۰/۰۴	۰/۷۰
سرعت باد	-۰/۳۳	-۰/۵۲	-۰/۰۳	-۰/۰۱	۰/۶۸	-۰/۳۷	۰/۱۴
سینوس جهت باد	-۰/۰۳	-۰/۰۲	۰/۷۳	-۰/۶۸	۰	۰/۰۱۱	۰/۰۱۲
کسینوس جهت باد	-۰/۰۳	-۰/۰۸	۰/۶۷	۰/۷۳	-۰/۰۴	-۰/۰۱	-۰/۰۱
درصد انحراف معیار	۳۰/۴۹	۱۸/۲۲	۱۴/۳۸	۱۴/۱۵	۱۰/۱۴	۹/۰۳۷	۳/۵۹
درصد انحراف معیار تجمعی	۳۰/۴۹	۴۸/۷۱	۶۳/۰۹	۷۷/۲۴	۸۷/۳۸	۹۶/۴۱	۱۰۰

قابل مشاهده است، مانند ایستگاه آزادی، ازن با رطوبت نسبی نسبت عکس و با دما نسبت مستقیم دارد. همچنین همبستگی بالا و مثبت ازن با این سرعت باد نشان‌دهنده این موضوع است که این پارامتر عاملی جهت انتقال ازن از مناطق دیگر به منطقه امام خمینی است. همچنین غلظت ازن با فشار همبستگی منفی دارد. این در حالی است که غلظت ازن با پارامترهای سینوس و کسینوس جهت باد همبستگی کمی نشان می‌دهد؛ این مساله می‌تواند ناشی از این موضوع باشد که در نزدیکی منطقه امام خمینی، عوارضی که از انتقال ازن توسط باد جلوگیری کند وجود ندارد؛ همچنین فرآیند فتوشیمیایی و تولید ازن به منبع خاصی وابسته نبوده و به عبارتی دارای منبع نقطه‌ای نمی‌باشد، که این موضوع می‌تواند دلیل همبستگی ضعیف ازن با جهت باد باشد.

همانطور که در جدول ۳ دیده می‌شود، واریانس تجمعی در مؤلفه‌ی چهارم برابر با ۷۷ درصد است؛ پس می‌توان رفتار داده‌ها چهار مؤلفه‌ی اول را در نظر گرفت. بیشترین مقدار مثبت در مؤلفه اول، مربوط به رطوبت نسبی است؛ در حالی که در این مؤلفه، تأثیر غلظت ازن و سرعت باد و دما به صورت مقادیر منفی می‌باشد. این به آن معنی است که مقادیر سرعت باد، غلظت ازن و دما در یک جهت عمل می‌کنند. در مؤلفه دوم بیشترین تأثیر مربوط به فشار و جهت باد است. در مؤلفه سوم و چهارم نیز سینوس و کسینوس جهت باد تأثیر گذارترین پارامترها هستند، در حالی که پارامترهای دیگر در این دو مؤلفه تأثیر کمی دارند. این موضوع با مستقل بودن پارامترهای جهت باد از پارامترهای دیگر هواشناسی قابل توجیه است.

ماتریس همبستگی پارامترهای مختلف اندازه‌گیری شده در ایستگاه امام خمینی، در جدول ۴ نشان داده شده است. همانگونه که در این ماتریس

جدول ۴: ماتریس همبستگی پارامترهای هواشناسی و غلظت ازن در ایستگاه امام خمینی

	ازن	رطوبت نسبی	فشار	دما	سرعت باد	سینوس جهت باد	کسینوس جهت باد
ازن	۱						
رطوبت نسبی	-۰/۴۳	۱					
فشار	-۰/۱۸	۰/۲۱	۱				
دما	۰/۴۳	-۰/۷۲	-۰/۲۶	۱			
سرعت باد	۰/۵۳	-۰/۲۸	-۰/۰۸	۰/۲۳	۱		
سینوس جهت باد	-۰/۰۳	۰/۰۱۷	۰/۰۰۶	۰	-۰/۰۲	۱	
کسینوس جهت باد	۰/۰۰۳	۰/۰۱۲	-۰/۰۱	۰/۰۰۵	۰	۰	۱

پارامترهای ازن، رطوبت نسبی، فشار و سرعت باد در مؤلفه‌های اصلی اول و دوم به خوبی قابل مشاهده است؛ در حالی که تأثیر سینوس و کسینوس جهت باد در این دو مؤلفه بسیار کم است. در مؤلفه‌های اصلی ۳ و ۴ عکس حالت قبل اتفاق افتاده است و فقط پارامترهای جهت باد نقش اصلی را در این مؤلفه‌ها دارند؛ و می‌توان این پارامترها را مستقل از پارامترهای دیگر در نظر گرفت.

مقدار شاخص KMO برای داده‌های ایستگاه امام خمینی ۰/۶۶ است، که نشان دهنده این موضوع است که تحلیل مؤلفه‌های اصلی می‌تواند برای داده‌های ایستگاه امام خمینی انجام پذیرد. همانطور که در جدول ۵ دیده می‌شود، واریانس مؤلفه‌های ۲ تا ۴ تقریباً با هم برابر است؛ برابری این مؤلفه‌ها می‌تواند نشان دهنده مستقل بودن پارامترها از هم باشد. با در نظر گرفتن حد آستانه ۷۰ درصد، ۴ مؤلفه اصلی را باید در نظر گرفته شود. تأثیر

جدول ۵: نتیجه تحلیل مؤلفه‌های اصلی در ایستگاه امام خمینی. هر سطر از هر ستون وزن هر پارامتر را در هر PC نشان می‌دهد.

	مؤلفه اول	مؤلفه دوم	مؤلفه سوم	مؤلفه چهارم	مؤلفه پنجم	مؤلفه ششم	مؤلفه هفتم
ازن	-۰/۴۹	۰/۳۵	۰/۰۴۶	۰/۰۷۷	-۰/۱۵	۰/۷۷	-۰/۰۹
رطوبت نسبی	۰/۵۲	۰/۲۲	۰/۰۳۸	۰/۰۷۱	-۰/۳۹	۰/۲۲	۰/۶۸
فشار	۰/۲۵	۰/۵۵	۰/۰۳۷	۰/۰۶۲	۰/۷۸	۰/۰۶۴	۰/۰۴۹
دما	-۰/۵۲	-۰/۳	-۰/۰۳	-۰/۰۷	۰/۳۴	-۰/۰۴	۰/۷۱
سرعت باد	-۰/۳۹	۰/۶۲	۰/۰۵۳	۰/۱۳	-۰/۲۹	-۰/۵۹	۰/۱۰
سینوس جهت باد	-۰	۰/۰۶۵	۰/۷۳	-۰/۶۸	-۰/۰۳	-۰/۰۱	۰
کسینوس جهت باد	۰/۰۰۲	-۰/۲	۰/۶۷	۰/۷۱	۰/۰۵۳	-۰/۰۱	-۰/۰۲
درصد انحراف معیار	۳۴/۸	۱۴/۵۵	۱۴/۳	۱۴/۲۷	۱۲/۰۶	۶/۰۹۶	۳/۹۲
درصد انحراف معیار تجمعی	۳۴/۸	۴۹/۳۵	۶۳/۶۵	۷۷/۹۲	۸۹/۹۸	۹۶/۰۸	۱۰۰

پارامترهای موثر بر غلظت ازن و غلظت ازن در لحظه t و خروجی آن غلظت ازن از t+24 می‌باشد.

$$O3_{t+24} = F(O3_t, \text{Effective Meteorological Parameters}_t)$$

بعد از مشخص شدن پارامترهای تأثیرگذار بر غلظت ازن، می‌توان اقدام به مدلسازی کرد. در این مقاله برای مدلسازی غلظت ازن از شبکه عصبی چندلایه استفاده شده است. این مدل برای پیش‌بینی ۲۴ ساعته ازن استفاده می‌شود؛ و مقادیر ورودی آن

مدل‌های ایجاد شده برای پیش‌بینی مقادیر روز بعد استفاده شدند.

برای مدل‌سازی از داده‌های سه ماه استفاده شد؛ و مدل‌سازی برای هر ۴ فصل انجام گشت، و سپس

جدول ۶: نتایج پیش‌بینی ۲۴ ساعته غلظت ازن برای یک هفته از هر فصل

جدر میانگین مربعات خطا (ppb)	ضریب همبستگی	شاخص توافق	میانگین مطلق خطا (ppb)		
۱۶/۵۸	۰/۶۷	۰/۶۳	۴/۶۱	زمستان	ایستگاه امام خمینی
۹/۷۷	۰/۹۴	۰/۹۱	۳/۵۹	بهار	
۳/۱۹	۰/۸۹	۰/۸۲	۱/۳۸	تابستان	
۵/۳۷	۰/۷۲	۰/۶۸	۲/۳۰	پاییز	
۱۳/۶۲	۰/۷۱	۰/۶۹	۴/۴۳	زمستان	ایستگاه آزادی
۱۱/۱۶	۰/۸۸	۰/۸۱	۱/۹۴	بهار	
۱۱/۵۳	۰/۸۱	۰/۷۳	۲/۶۱	تابستان	
۳/۹۷	۰/۶۸	۰/۷۱	۱/۷۷	پاییز	

ازن و پارامترهای هواشناسی مربوط به ایستگاه‌های سنجش کیفیت هوا آزادی و امام خمینی شهر تهران در سال ۲۰۰۹ و تحلیل مؤلفه‌های اصلی رابطه میان پارامترهای هواشناسی و غلظت ازن مورد بررسی قرار گرفته است. این تحلیل نشان داد که پارامترهای رطوبت نسبی، دما و سرعت باد بیشترین تأثیر را در غلظت ازن دارند؛ پارامترهای رطوبت نسبی و دما می‌توانند در تولید ازن توسط عملکرد فتوشیمیایی مؤثر باشند. همبستگی غلظت ازن با جهت باد می‌تواند نشان‌دهنده این موضوع است که قسمتی از غلظت ازن توسط باد از مناطق دیگر به این منطقه انتقال پیدا می‌کند. نتایج این تحقیق نشان می‌دهد استفاده از روش تحلیل مؤلفه‌های اصلی در شناسایی پارامترهای مؤثر بر غلظت ازن، کارایی مناسبی دارد. بعد از شناسایی پارامترهای تأثیرگذار، غلظت ازن با استفاده از شبکه عصبی مدل‌سازی و پیش‌بینی شده است. ورودی شبکه عصبی مورد نظر پارامترهای

نتایج پیش‌بینی برای فصل‌های مختلف سال ۲۰۰۹ در ایستگاه‌های آزادی و امام خمینی در جدول ۶ ارائه شده است.

همانطور که در جدول ۶ دیده می‌شود؛ مقدار شاخص توافق و ضریب همبستگی در فصل‌های زمستان و پاییز کمتر از فصل‌های دیگر است. این موضوع می‌تواند نشان‌دهنده این مساله باشد که مقدار ازن در پاییز و زمستان تحت تأثیر عوامل دیگری نیز هست. همچنین پارامترهای تأثیرگذار بر غلظت ازن از جمله میزان تابش خورشیدی در این فصل‌ها کمتر می‌شوند. این مساله می‌تواند در تحقیقات آتی مورد بحث قرار گیرد. با توجه به پارامترهای آماری محاسبه شده در این تحقیق به وضوح قابل مشاهده است که مدل در هر چهار فصل به خوبی عمل می‌کند.

۴- بحث و نتیجه‌گیری

در این پژوهش با استفاده از داده‌های غلظت

تاثیر غلظت پارامترهای هواشناسی بر غلظت آلاینده‌ها به صورت عددی بیان نشده است (Safavi & Alijani, 2007). اما در مقاله حاضر تاثیر پارامترهای هواشناسی بر غلظت ازن با استفاده از تحلیل مولفه‌های اصلی و ضریب همبستگی بررسی شده است. همچنین در تحقیق دیگری که توسط Abbaspour و همکاران (۲۰۰۵) انجام گردید غلظت ساعتی مونوکسیدکربن برای ۲۴ ساعت آینده با استفاده از شبکه عصبی مصنوعی پیش‌بینی شد؛ مدل آن‌ها دارای ضریب همبستگی ۶۷ درصد بین مقادیر واقعی آلاینده و مقادیر پیش‌بینی شده می‌باشد؛ که از دقت مدل پیشنهادی در این تحقیق کمتر است.

نتایج نشان می‌دهد که مدل استفاده شده در این تحقیق کارایی خوبی در مدلسازی و پیش‌بینی ۲۴ ساعته غلظت ازن دارد. پیش‌بینی انجام شده برای ۲۴ ساعت آینده می‌تواند برای کنترل و مدیریت آلودگی، و همچنین آگاهی عمومی مورد استفاده قرار گیرد.

هواشناسی و غلظت ازن در لحظه t و خروجی آن غلظت ازن در $t+24$ می‌باشد. شبکه عصبی برای مدلسازی غلظت ازن در فصول مختلف سال استفاده شد. پیش‌بینی غلظت ازن در فصل بهار بهترین نتایج را داشت؛ در فصل بهار همبستگی مقایر پیش‌بینی شده و غلظت واقعی ازن ۸۸ و ۹۴ درصد بود و شاخص توافق ۸۱ و ۹۱ درصد به ترتیب برای ایستگاه‌های امام خمینی و آزادی بود. در حالی که در فصل زمستان مدل در پیش‌بینی غلظت ازن دقت کمتری داشت. در فصل زمستان شاخص توافق ۶۳ و ۶۹ درصد و ضریب همبستگی ۶۷ و ۷۱ درصد به ترتیب برای ایستگاه‌های امام خمینی و آزادی بود. ممکن است علت دقت بهتر مدل در فصل بهار و دقت پایین مدل در فصل پاییز زمستان، کاهش پارامترهای تاثیرگذار بر غلظت ازن از جمله کاهش تابش خورشیدی در فصل زمستان و پاییز باشد. تحقیقات دیگر در مورد تاثیر پارامترهای هواشناسی بر آلودگی هوا در تهران انجام شده است از جمله تحقیقی که صفوی و علیجانی انجام دادند. اما در تحقیق آن‌ها

REFERENCES

Abbaspour, M., Rahmani, A., Teshnehlab, M., 2005. Carbon monoxide prediction using novel intelligent network. *International Journal of Environmental Science and Technology* 1(4), 257-264.

Alalawi, S., Abdulwahab, S., Bakheit, C., 2008. Combining principal component regression and artificial neural networks for more accurate predictions of ground-level ozone. *Environmental Modelling & Software* 23(4), 396-403.

Balestrassi, P. P., Popova, E., Paiva, A. P., Marangon Lima, J. W., 2009. Design of experiments on neural network's training for nonlinear time series forecasting. *Neurocomputing* 72(4-6), 1160-1178.

Cai, M., Yin, Y., & Xie, M. (2009). Prediction of hourly air pollutant concentrations near urban arterials using artificial neural network approach. *Transportation Research Part D: Transport and Environment*, 14(1), 32-41. doi: 10.1016/j.trd.2008.10.004

Chan, K.Y., Jian, L., 2013. Identification of significant factors for air pollution levels using a neural network based knowledge discovery system. *Neurocomputing* 99, 564-569.

Cheng, S., Li, L., Chen, D., Li, J., 2012. A neural network based ensemble approach for improving the accuracy of meteorological fields used for regional air quality modeling. *Journal of Environmental Management* 112, 404-414.

Coman, A., Ionescu, A., Candau, Y., 2008). Hourly ozone prediction for a 24-h horizon using neural networks. *Environmental Modelling & Software* 23(12), 1407-1421.

EPA, U., 2006. Air Quality Criteria for Ozone and Related Photochemical Oxidants (2006 Final) Public Review Draft. U.S. Environmental Protection Agency, Washington, DC.

Finardi, S., De Maria, R., D'Allura, A., Cascone, C., Calori, G., & Lollobrigida, F. (2008). A deterministic air quality forecasting system for Torino urban area, Italy. *Environmental Modelling & Software*, 23(3), 344-355.

Haykin, S., 1998. *Neural Networks: A Comprehensive Foundation* (2nd Edition). Prentice Hall, 842 p

Ibarra-Berastegi, G., Elias, A., Barona, A., Saenz, J., Ezcurra, A., Diaz de Argandoña, J.,

2008. From diagnosis to prognosis for forecasting air pollution using neural networks: Air pollution monitoring in Bilbao. *Environmental Modelling & Software*, 23(5), 622-637.
- Jolliffe, I. T., 2002. *Principal Component Analysis*. Springer. New York. 488 p
- Karatzas, K. D., Kaltsatos, S., 2007. Air pollution modelling with the aid of computational intelligence methods in Thessaloniki, Greece. *Simulation Modelling Practice and Theory*, 15(10), 1310-1319.
- Kitagawa, G., 2010. *Introduction to Time Series Modeling*. Taylor & Francis, 5 p.
- Kröse, B., van der Smagt, P., 1996. *An Introduction to Neural Networks*. University of Amsterdam. 135 p
- Lu, W. Z., Wang, W. J., Wang, X. K., Yan, S. H., Lam, J. C., 2004. Potential assessment of a neural network model with PCA/RBF approach for forecasting pollutant trends in Mong Kok urban air, Hong Kong. *Environmental Research*, 96(1), 79-87.
- Matlab. 2007. *Matlab toolbox for Data analysis*. Natick, Massachusetts: The MathWorks Inc.
- McCulloch, W., Pitts, W., 1943. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4), 115-133.
- Menhaj, M. B., 2011. *Fundamentals of neural networks*. Tehran: Amirkabir University of Technology. (in Persian).
- Safavi, Y., Alijani, B., 2007. Analysing geographical parameters on air pollution in Tehran. *Geographical Research*, 58, 99-112.
- Schlink, U., Dorling, S., Pelikan, E., Nunnari, G., Cawley, G., Junninen, H., Doyle, M., 2003. A rigorous inter-comparison of ground-level ozone predictions. *Atmospheric Environment*, 37(23), 3237-3253.
- Singh, K. P., Gupta, S., Kumar, A., Shukla, S. P., 2012. Linear and nonlinear modeling approaches for urban air quality prediction. *Science of the Total Environment*, 426, 244-255.
- Siwek, K., Osowski, S., 2012. Improving the accuracy of prediction of PM10 pollution by the wavelet transformation and an ensemble of neural predictors. *Engineering Applications of Artificial Intelligence*, 25(6), 1246-1258.
- Sousa, S., Martins, F., Alvimferraz, M., & Pereira, M. (2007). Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations. *Environmental Modelling & Software*, 22(1), 97-103.
- Stephenson, D. B., 2003. *Statistical concepts in environmental science*. University of Reading, Reading. 98 p
- Voukantsis, D., Karatzas, K., Kukkonen, J., Rasanen, T., Karppinen, A., Kolehmainen, M., 2011. Intercomparison of air quality data using principal component analysis, and forecasting of PM(10) and PM(2.5) concentrations using artificial neural networks, in Thessaloniki and Helsinki. *Science of the Total Environment*, 409(7), 1266-1276.
- Williams, B., Brown, T., Onsmann, A., 2012. Exploratory factor analysis: A five-step guide for novices. *Australasian Journal of Paramedicine*, 8(3), 1.13.
- Willmott, C. J., Ackleson, S. G., Davis, R. E., Feddema, J. J., Klink, K. M., Legates, D. R., Rowe, C. M., 1985. Statistics for the evaluation and comparison of models. *Journal of Geophysical Research: Oceans*, 90(C5), 8995-9005.

Developing a model for ozon concentration prediction using artificial neural network

Mehrdad Rafiepour^{1*}, Ali Asghar Alesheikh², Abbas Alimohammad³, Abolghasem Sadeghi Niaraki⁴

1- Master Student of GIS, K.N.Toosi University of Technology

2- Professor, K.N.Toosi University of Technology

3- Associate Professor, K.N.Toosi University of Technology

4- Assistant Professor, K.N.Toosi University of Technology

Received: 10-Mar.-2014

Accepted: 17-May-2015

Abstract

Ozone concentration in metropolitan areas frequently exceed regulatory standards now. Since it is harmful on human health and environment, its modeling and visualization are of vital important. Modeling and prediction of ozone is required by urban managers to control and prevent its effects. In this paper, important parameters influencing the hourly ozone concentration are estimated by using data collected in Azadi and Imam Khomeini air quality stations in 2009-2010. In this context, the correlations between ozone concentration and meteorological parameters such as relative humidity, temperature, pressure, wind speed, and its directions are determined by linear regression and principle component analysis. Results showed that ozone concentration is mostly affected by relative humidity, temperature and wind speed. The effects of relative humidity and temperature on ozone concentration are attributed to photochemical processes. While the correlation between the ozone and wind speed is due to the ozone transfer from nearby regions. After determining the important parameters, a neural network was used for forecasting of ozone concentration of the next 24 hours for a week in four different seasons. The input to the neural network model was relative humidity, temperature and wind speed. Results of the implementation showed that the model can predict the ozone concentration in Azadi and Imam Khomeini air quality stations by an accuracy of 67 to 97 percent for the next 24 hours. Urban managers for more efficient management and control of the ozone concentration can use results of this research work.

Keywords: Ozone, modeling, principal component analysis, neural network

* Corresponding Author: E-mail: mehردادrafiepour@gmail.com Phone: +98-9127752751