



## پیش‌بینی داده‌های جا افتاده ایستگاه‌های پایش آلودگی هوا به کمک فیلتر کالمن

ندا کفاش چرن‌دابی<sup>۱\*</sup>

<sup>۱\*</sup> استادیار، گروه نقشه برداری، دانشکده فنی و مهندسی مرنند، دانشگاه تبریز، تبریز، ایران

(n\_kaffash@tabrizu.ac.ir)

(تاریخ دریافت مقاله: ۹۸/۰۶/۱۴، تاریخ پذیرش مقاله: ۹۸/۰۸/۲۵)

### چکیده

در سالهای اخیر با توجه به افزایش چشمگیر آلاینده‌های هوا، تهیه نقشه‌های دقیق آلودگی هوا و توسعه سامانه‌های آبی پایش آلودگی هوا؛ توجه محققان زیادی را به خود جلب کرده است. با توجه به اینکه داده‌های به دست آمده از ایستگاه‌های پایش آلودگی هوا، عاری از خطا نبوده و برخی از مشاهدات آنها جا افتاده است؛ لذا بهره‌گیری از مدلی توانمند جهت پالایش داده‌ها و پیش‌بینی مشاهدات جا افتاده ضروری بنظر می‌رسد. در تحقیق حاضر، مقادیر آلاینده‌های هوا برای ایستگاه‌های موجود در منطقه ۳ تهران به کمک فیلتر کالمن تصحیح و پیش‌بینی شد. نتایج حاکی از دقت متوسط ۹۲ درصدی برای فیلتر کالمن بوده است.

### کلمات کلیدی

آلودگی هوا، فیلتر کالمن، پیش‌بینی مشاهدات، GIS.



# Forecasting Air Pollution based on Monitoring Station with using Kalman Filter

Neda Kaffash Charandabi <sup>1\*</sup>

<sup>1\*</sup>Assistant Professor, Department of Survey, Marand Faculty of Engineering, Tabriz University, Tabriz, Iran (n\_kaffash@tabrizu.ac.ir)

(Date of received: 05/09/2019, Date of accepted: 16/11/2019)

## ABSTRACT

*In recent years, due to the dramatic increase of air pollutants preparation of right air pollution maps and the development of instantaneous air pollution monitoring systems attention of researchers has attracted. According to the data obtained from air pollution monitoring stations, they were not isolated from the error and some of their observations are missing; therefore, it is necessary to use a robust model to refine the data and predict the observations. In the present study, the data's of three stations in Tehran of air pollutants corrected and predicted by Kalman filter. Results showed that an average accuracy of 92% for the Kalman filter.*

## Keywords:

*Air pollution, Kalman filter, Predict of observation, GIS.*



## ۱- مقدمه

در سالهای اخیر، آلاینده‌های موجود در محیط زیست از مهمترین مشکلات شهرهای بزرگ دنیا بوده‌اند که مسئله آلودگی هوا یکی از بحرانی‌ترین این موارد می‌باشد. افزایش میزان آلاینده‌های موجود در هوا علل گوناگونی دارد که از آن جمله می‌توان به افزایش جمعیت، پیشرفت صنعت و فناوری، توسعه شهرها و افزایش وسایط نقلیه اشاره کرد [۱]. تهیه داده‌های آلودگی هوا از ایستگاه‌های پایش آلاینده‌ها، جهت مطالعه و بررسی راهکارهای مناسب برای کاهش و کنترل آن، از اهمیت ویژه‌ای برخوردار است. ایستگاه‌های متعددی برای سنجش آلاینده‌ها در مناطق مختلف کلانشهرها واقع شده‌اند و به جمع‌آوری داده می‌پردازند. اخذ دقیق، آنی و پیوسته این داده‌ها برای مطالعات و کاربردهای گوناگون نظیر طراحی سیستم‌های پایش آنی آلودگی هوا، پایش بیماران مبتلا به آسم و تنگی نفس، در بستر سامانه‌های اطلاعات مکانی، ضروری می‌باشد. لیکن متأسفانه در موارد زیادی جداول به دست آمده از این ایستگاه‌ها، فاقد مشاهدات آلاینده‌های مختلف در ساعات گوناگون می‌باشد. عدم ثبت مقادیر مشاهدات علل گوناگونی دارد که از آن جمله می‌توان به خراب یا غیرفعال بودن حسگر و قطع ارتباط حسگر با مرکز پایش اشاره کرد [۱]. پیش‌بینی آلودگی هوا و مقادیر آلاینده‌ها در تحقیقات متعددی مورد بررسی قرار گرفته است. Brunelli و همکارانش در مطالعه‌ای با استفاده از تکنیک شبکه‌های عصبی و طراحی ۲ شبکه، پرسپترون چندلایه و المان به پیش‌بینی حداکثر غلظت روزانه  $PM_{10}$ ،  $CO$ ،  $SO_2$ ،  $NO_2$ ،  $O_3$  در مناطق شهری ایتالیا اشاره کردند [۲]. در تحقیق دیگری یک مدل شبکه عصبی برای پیش‌بینی غلظت  $PM_{10}$  هوای شهر نتیاگو در شیلی طراحی شد و با استفاده از این مدل آلودگی  $PM_{10}$  هوای این شهر را برای سال ۲۰۰۳، پیش‌بینی کردند [۳]. بوداقپور و چرخستانی به پیش‌بینی میزان غلظت آلاینده‌های هوای تهران با استفاده از شبکه عصبی مصنوعی پرداختند [۴]. نتایج مطالعه آن‌ها حاکی از آن است که دقت پیش‌بینی با استفاده از شبکه عصبی بالاتر از رگرسیون خطی می‌باشد. علیاری و همکاران در مطالعه‌ای به پیش‌بینی کوتاه مدت هوا با کمک شبکه‌های عصبی پرسپترون چندلایه و هوش جمعی پرداختند و در نهایت این روش‌های هوشمند را با هم مقایسه کردند. روش پیشنهادی این تحقیق که ترکیبی بر اساس PSO و فیلتر کالمن برای آموزش شبکه عصبی بود، توانایی مناسبی در بهبود عملکرد پیش‌بینی را نشان داد [۵]. در اکثر مطالعات انجام شده، تخمین دقیق یک آلاینده و یا میزان حداکثر آلاینده‌ها مورد بررسی قرار گرفته است. حال آنکه بررسی مقادیر جاقفاده تمامی آلاینده‌ها در جداول به دست آمده از ایستگاه‌های پایش آلودگی هوا و بهره‌گیری از مدلی مناسب برای پیش‌بینی مقادیر آنها با دقت مطلوب کمتر مورد توجه قرار گرفته است. در تحقیق حاضر، داده‌های به دست آمده برای آلاینده‌های مختلف در ایستگاه‌های پایش آلودگی هوای منطقه سه شهر تهران؛ مورد تحلیل قرار گرفتند. داده‌های به دست آمده از حسگرهای مختلف عاری از خطا نبوده و در برخی از ساعت‌ها، داده‌هایی ثبت نشده است (داده‌های جاقفاده). لذا بایستی مقادیر جاقفاده پیش‌بینی، داده‌های پرت حذف و داده‌های زمانی حسگرها تصحیح گردند. در این تحقیق برای نیل به این اهداف از فیلتر کالمن استفاده شد. دلیل این انتخاب کارایی اثبات شده فیلتر کالمن در پالایش داده‌ی حسگرها در تحقیقات پیشین بوده است. همچنین فیلتر کالمن برای پیش‌بینی و تصحیح، یک فیلتر بازگشتی تک مرحله‌ای است، یعنی به داده‌های خیلی قبل نیاز ندارد. این مزیت فیلتر کالمن برای سیستم‌های آنی بسیار مؤثر است؛ چراکه جهت پاسخگویی در هر زمان و هر حالت تنها به کمک داده‌های حالت قبل، با کیفیت مطلوب داده حسگرها تصحیح و پیش‌بینی می‌شوند [۶] [۷].



## ۲- فیلتر کالمن

مشاهدات سری زمانی غالباً دارای مقادیر پرت، مشاهدات همراه با خطا و مقادیر نامشخص می‌باشند. فیلتر کالمن برای نخستین بار توسط Rudolph Kalman، برای برآورد مقادیر نامشخص از مشاهدات انجام یافته در گذر زمان معرفی گردید. الگوریتم روش کالمن یک فیلتر بازگشتی دو مرحله‌ای (مرحله پیش‌بینی و مرحله تصحیح) می‌باشد. در مرحله پیش‌بینی مقادیر مشاهدات همراه با عدم قطعیت برآورد شده و در مرحله تصحیح برآوردها بر مبنای میانگین وزنی (وزن بزرگ‌تر متعلق به مقادیر با درجه اطمینان بالاتر) به‌روز می‌گردند [۶] [۸]. فیلتر کالمن از مقادیر حال اندازه‌گیری‌ها، برآوردهای قبلی و ماتریس عدم قطعیت استفاده می‌کند. در موارد خاصی که توزیع تمام خطاها گوسین<sup>۱</sup> باشد، تخمین فیلتر برای تمام برآوردها دقیق خواهد بود. فیلتر کالمن مشابه مدل مخفی مارکوف عمل می‌کند، با این تفاوت که فضای متغیرهای پنهان پیوسته است. مهم‌ترین مزیت فیلتر کالمن پیش‌بینی و تصحیح مقادیر اندازه‌گیری شده توسط حسگرها می‌باشد؛ به ویژه زمانی که داده‌ها پیوسته و دارای حجم زیادی می‌باشند [۶] [۸]. معادله وضعیت برای هر سیستم مطابق رابطه (۱) فرض می‌شود (Kalman, 1960; Wolpert and Ghahramani, 2000).

$$\hat{X}_k = K_k . Z_k + (1 - K_k) \hat{X}_{k-1} \quad (1)$$

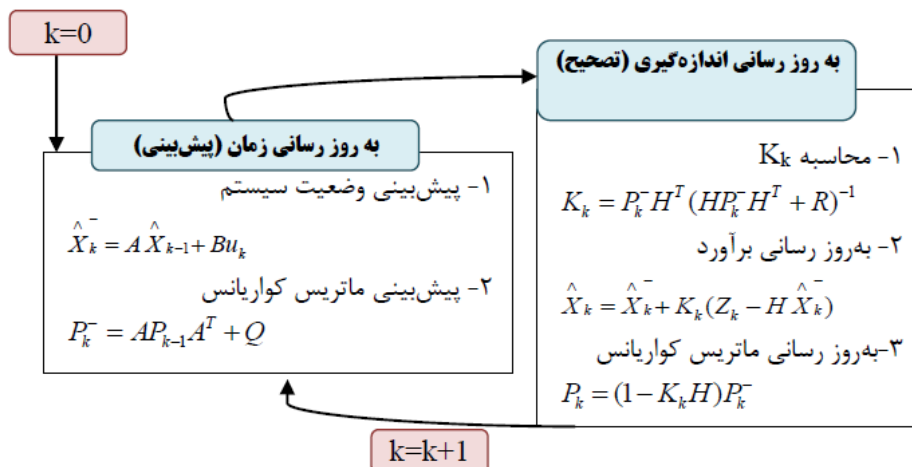
که در آن  $\hat{X}_k$  برآورد فعلی،  $\hat{X}_{k-1}$  برآورد پیشین،  $Z_k$  مقدار اندازه‌گیری شده،  $k$  گام زمانی و  $K_k$  ضریب حاصله کالمن یا بهره کالمن<sup>۲</sup> می‌باشد. تنها مقدار مجهول رابطه فوق  $K_k$  می‌باشد که بایستی با توجه به فضای مسئله مشخص گردد. با فرض  $K_k$  برابر ۰/۵ معادله فوق به یک میانگین‌گیری ساده تبدیل می‌شود. در حالی که فیلتر کالمن بهترین مقادیر  $K_k$  را در هر وضعیت یافته و میانگین‌گیری بهینه انجام می‌دهد. حال با در نظر گرفتن دو معادله کلی مطابق رابطه (۲) و (۳) فیلتر کالمن به حل مسأله می‌پردازد [۶] [۹].

$$X_k = A.X_{k-1} + Bu_k + w_{k-1} \quad (2)$$

$$Z_k = H.X_k + v_k \quad (3)$$

در معادله دوم  $u_k$  مقادیر کنترلی (در صورت وجود) و  $w_{k-1}$  و  $v_k$  میانگین و انحراف معیار توابع خطا و  $A$ ،  $B$  و  $H$  ضرایب ثابت برگرفته از ماتریس عدم قطعیت می‌باشند. هر چقدر تخمین پارامترها دقیق‌تر باشد، تخمین مقادیر مسئله بهتر خواهد شد. برای شروع، دو دسته معادلات مجزا وجود دارد؛ معادلات در فاز پیش‌بینی و معادلات در فاز تصحیح؛ که هر دو دسته معادلات در هر گام ( $K^{th}$ ) اجرا می‌شوند. شکل (۱) نشان دهنده روند بازگشتی فیلتر کالمن و روابط محاسبه شده در مرحله پیش‌بینی و تصحیح داده‌ها می‌باشد [۶] [۹].

<sup>۱</sup> . Gaussian<sup>۲</sup> . Kalman Gain



شکل ۱: نحوه محاسبات روابط در مرحله پیش‌بینی و تصحیح فیلتر کالمن.

$R$  ماتریس عدم قطعیت اندازه‌گیری و  $Q$  ماتریس عدم قطعیت جریان است.  $\hat{X}_k^-$  برآورد قبلی (قبل از اعمال تصحیح در هر مرحله) و  $P_k^-$  کواریانس خطای پیشین می‌باشد.  $\hat{X}_k^-$  مقدار برآورد شده از  $X$  در زمان  $k$  بر مبنای مشاهدات  $(Z_k)$  و  $P_k^-$  مقدار کواریانس در زمان  $k$  است. همچنین با انتخاب مقادیر بهینه برای  $K_k$  می‌توان تخمین‌های مناسب برای اندازه‌گیری‌ها انجام داد. با مقادیر اولیه  $X_0$ ،  $P_0$  الگوریتم و مقادیر مشاهدات، فیلتر شروع بکار کرده و معادلات دو مرحله تا رسیدن به گام  $k$  ام مشاهدات تکرار می‌گردد [۶] [۸]. فیلتر کالمن یک حل بهینه برای داده‌های خطی بدون همبستگی با توزیع خطای گاوس و میانگین صفر می‌باشد. بنابراین فرضهای فیلتر کالمن خطی بودن برآوردها و توزیع گاوسی خطا می‌باشد. خطی بودن مشاهدات به این معنی است که وضعیت هر مشاهده در زمان  $t$  به طور خطی بر وضعیت قبلی آن و تابع کنترل  $u_k$  (در صورت وجود) وابسته است. این فرض تا حد زیادی صحیح است، لیکن توزیع گاوسی خطا در برخی از مسائل صادق نیست [۶] [۹]. در روند بازگشتی این الگوریتم ضمن پیش‌بینی و تصحیح مشاهدات جا افتاده و دارای خطا، کلیه اندازه‌گیری‌ها پالایش شده و مشاهدات نرم‌تری ۳ حاصل می‌گردد. برگشتی بودن این فیلتر بسیار کارآمد است؛ به این معنی که برای برآورد زمان حال مشاهدات، نیازی به حفظ تمام مشاهدات قبلی ندارد [۶] [۸] [۹].

### ۳- پیاده سازی

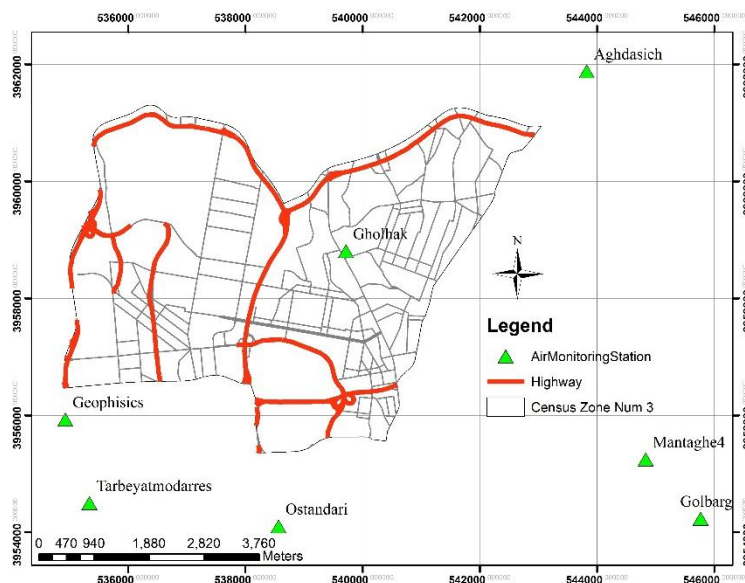
کلان شهر تهران به عنوان یکی از آلوده‌ترین شهرهای جهان شناخته شده است [۱۰] [۱۱]. این امتیاز منفی، دلایل عمده‌ای دارد که یکی از آنها موقعیت جغرافیایی و اقلیمی آن است. در فصل پاییز و زمستان به سبب بروز وارونگی دما<sup>۴</sup>، آلودگی تشدید شده و جان تمام مردم را در معرض خطر قرار می‌دهد. واحدهای صنعتی مستقر در تهران و به ویژه آلودگی ناشی از وسایل نقلیه از دیگر علل مهم و اساسی در آلودگی هوای شهر تهران است. ضمن اینکه عوامل اجتماعی و فرهنگی و توسعه شهری نیز در آلودگی سهم

<sup>3</sup> . Smooth

<sup>4</sup> Inversion



قابل توجهی دارند [۱۰] [۱۱]. آلودگی هوای تهران به یک معضل زیست محیطی در این شهر تبدیل شده و سبب افزایش شمار بیماران تنفسی، قلبی و عروقی با سرعت بسیار زیادی شده است. بنابراین منطقه ۳ شهر تهران برای پیاده سازی انتخاب گردید که در شکل (۲) نشان داده شده است. نقشه شکل ۲، در مقیاس ۱:۲۰۰۰، سیستم تصویر UTM<sup>۵</sup> با بیضوی مرجع WGS 1984 تهیه شده است. در منطقه سه از بیماران در حوالی خیابان های دولت، میرداماد، شریعتی، ولیعصر و میدان ونک نمونه برداری انجام و داده ها ذخیره شد.



شکل ۲: منطقه مطالعاتی به همراه بزرگراه ها و ایستگاه های پایش آلودگی هوا.

پس از انتخاب منطقه مطالعاتی، داده های آلاینده های هوا شامل؛  $SO_2$ ،  $O_3$ ،  $NO_2$  و  $PM_{10}$  از ایستگاه های پایش آلودگی هوا جمع آوری شد (۷ ایستگاه پایش هوای واقع در پیرامون منطقه مطالعاتی در شکل ۲ نشان داده شده اند). داده های مورد نیاز از سازمان حفاظت محیط زیست و شرکت کنترل کیفیت هوا جمع آوری گردید. در این مرحله از تحقیق، بایستی داده های ورودی مورد بررسی قرار بگیرند، داده های پرت حذف، مشاهدات جا افتاده پیش بینی و داده ها تصحیح گردند. چرا که برخی از حسگرهای مورد استفاده در این تحقیق (حسگر ثابت ایستگاه های پایش آلودگی هوا) در برخی از ساعات فاقد مشاهده ثبت شده بودند (به عنوان مثال، شکل (۳)).

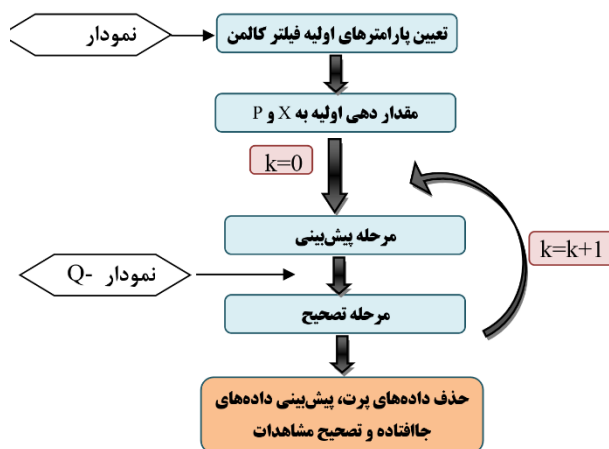
<sup>5</sup> Universal Transverse Mercator



| CO   | ETHYLBENZENE | M-P XYLENE | NO2   | NO    | NOX    | O3     | O-XYLENE | PM10      | PM2.5     |
|------|--------------|------------|-------|-------|--------|--------|----------|-----------|-----------|
| ppm  | ppb          | ppb        | ppb   | ppb   | ppb    | ppb    | ppb      | microg/m3 | microg/m3 |
| 3.96 | 4.85         |            | 50.16 | 70.10 | 120.73 | 17.03  | 0.75     | 126.67    | 43.96     |
| 3.44 | 4.00         |            | 46.51 | 51.70 | 98.66  | 22.19  | 1.00     | 114.49    | 38.57     |
| 3.42 | 3.35         |            | 48.10 | 77.81 | 126.45 | 22.74  | 0.75     | 90.14     | 32.02     |
| 3.79 | 3.91         |            | 48.22 | 55.29 | 104.14 | 12.42  | 1.00     | 78.82     | 26.40     |
| 2.71 | 2.94         |            | 38.58 | 15.00 | 54.17  | 24.41  | 0.75     | 77.16     | 28.06     |
| 2.83 | 2.00         |            | 41.28 | 21.17 | 62.93  | 18.31  |          | NaN       | 31.92     |
| 2.92 | 2.00         |            | 41.48 | 17.98 | 60.00  | 17.32  |          | NaN       | 35.54     |
| 2.69 | 2.00         |            | 39.50 | 21.46 | 61.34  | 18.98  |          | 73.19     | 25.51     |
| 2.46 | 2.00         |            | 37.52 | 15.53 | 53.50  | 25.09  |          | 82.62     | 23.74     |
| 2.14 | 1.85         |            | 27.85 |       | 35.03  | 70.51  |          | 70.62     | 35.45     |
| 1.87 | 1.00         |            | 20.51 |       | 28.68  | 93.21  |          | 46.38     | 25.34     |
| 1.83 | 1.00         |            | 21.82 |       | 30.70  | 97.12  |          | 31.71     | 25.15     |
| 1.72 | 1.00         |            | 20.30 |       | 29.46  | 101.81 |          | 24.79     | 19.86     |
| 1.54 | 1.00         |            | 18.13 |       | 26.92  | 111.66 |          | 32.82     | 12.73     |
| 1.53 | 1.00         |            | 19.52 |       | 28.75  | 115.39 |          | 35.56     | 11.12     |

شکل ۳: نمونه‌ای از داده‌های جا افتاده آلاینده PM<sub>10</sub> در داده‌های ایستگاه اقدسیه.

برای تصحیح و پیش‌بینی مقادیر آلاینده‌ها در این تحقیق از فیلتر کالمن به همراه نمودارهای تابع خود همبستگی<sup>۶</sup> و چندک<sup>۷</sup> استفاده شد. فیلتر کالمن برای داده تمام آلاینده‌ها در ایستگاه‌های موجود در منطقه‌ی مطالعاتی و حریم آن اعمال گردید. این فیلتر در محیط Matlab<sup>®</sup> 7.12.0 برنامه‌نویسی شد. هدف از تولید این نمودارها بررسی نتایج کالمن و تنظیم پارامترهای آن به صورت دقیق‌تر بوده است (به صورت تجربی) [۷]. نمودار ACF جهت بررسی مانایی داده‌ها ترسیم گردید. در این نمودار میزان وابستگی هر مشاهده به مشاهده قبل مشخص می‌شود. در شروع مشاهدات وابستگی داده با خودش یک و در ادامه کمتر می‌شود. با توجه به نرخ همبستگی (مثبت و منفی) ضریب Q و R فیلتر تعیین می‌گردد. اگر همبستگی یک باشد مقدار R برابر صفر، برای همبستگی کمتر از یک و مثبت یا منفی مقادیر R بیش‌تر از صفر یا کمتر از صفر و برای Q منفی یا مثبت به ترتیب انتخاب می‌شود [۷]. نمودار Q-Q نیز جهت بررسی توزیع داده‌ها (مقادیر مشاهده شده در قیاس با مقادیر پیش‌بینی شده) بسیار موثر است. یکی از فرضیات اصلی فیلتر کالمن خطی بودن  $X_i$  هاست لیکن این توزیع خطی لزوماً  $y=x$  نیست. وجود داده‌های پرت سبب می‌شود کارایی فیلتر کالمن کاهش یابد. لذا به کمک نمودار Q-Q داده‌های پرت تشخیص داده شده و حذف می‌گردند. پس از پالایش این مقادیر، داده‌های نرم‌تری پیش‌بینی می‌گردد [۷]. روند پالایش مقادیر پرت، استنتاج داده‌های جا افتاده و تصحیح مشاهدات حسگرها در این تحقیق مطابق روند شکل (۴) بوده است.



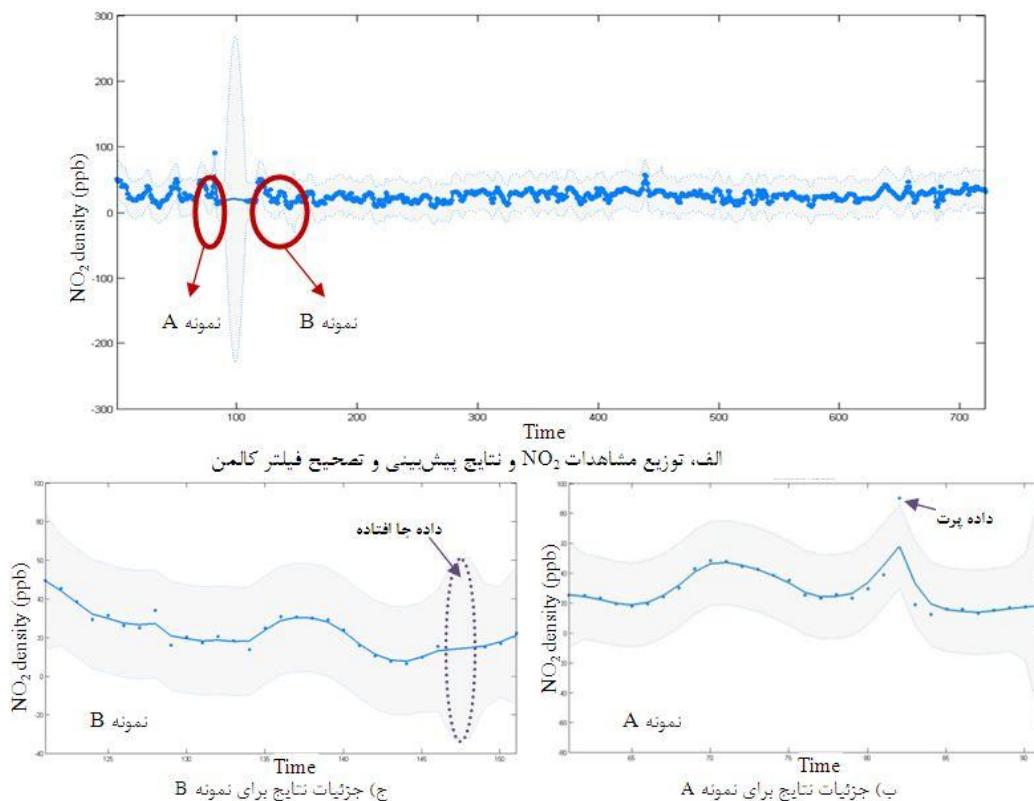
شکل ۴: فرآیند تصحیح و پیش‌بینی داده‌ی حسگرها.

<sup>۶</sup> . Auto Correlation Function (ACF)

<sup>۷</sup> . Empirical Quantiles-Theoretical Quantiles (Q – Q)



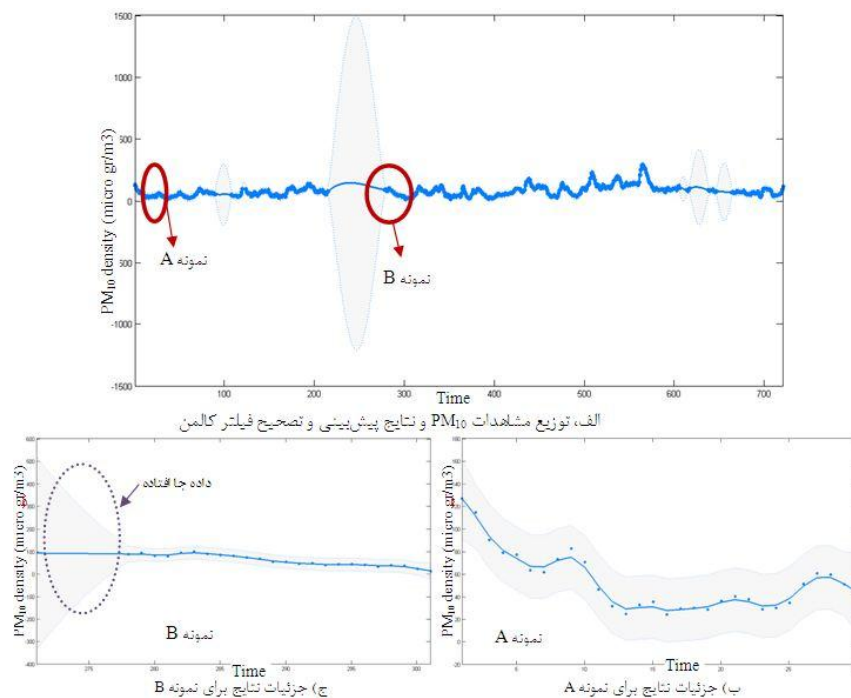
با توجه به توزیع داده‌ها در هر آلاینده، نمودار ACF و Q-Q ضرایب ثابت کالمن به صورت تجربی تعیین گردید. به عنوان نمونه، برای آلاینده NO<sub>2</sub> ایستگاه اقدسیه، فیلتر با ۷۲۰ گام زمانی ( $k=720$ ) و مقدار اولیه  $X_0=10$ ،  $H=1$ ،  $v_0 = 0.1$  و  $P_0=1$  اجرا گشته و نتایج در شکل (۵) ارائه شده است.



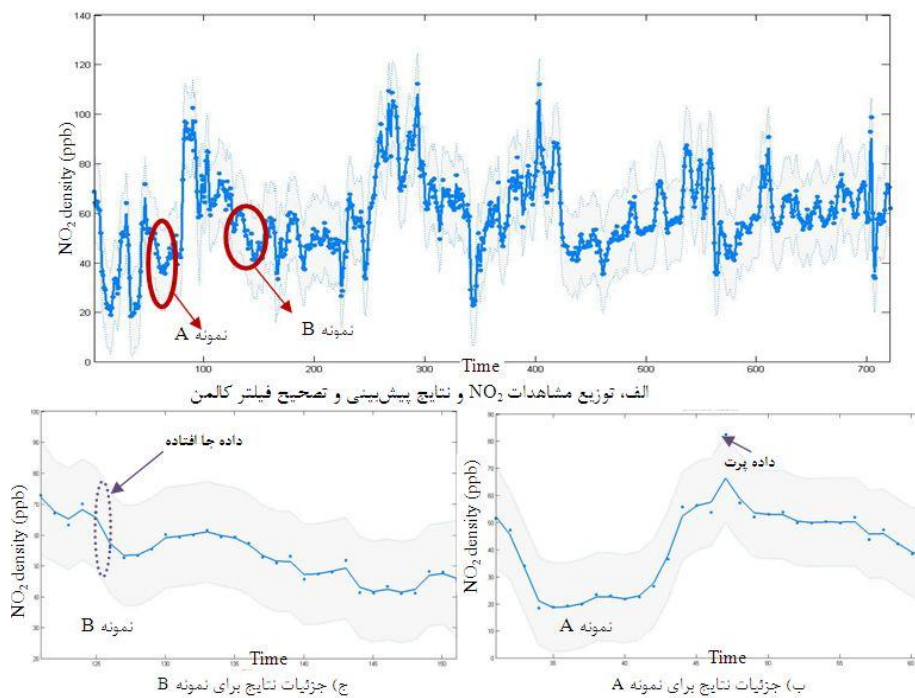
شکل ۵: نتایج فیلتر کالمن در پایش و تصحیح داده‌های آلاینده NO<sub>2</sub> در ایستگاه اقدسیه.

در این تحقیق هر گام زمانی یک ساعت است. همچنین، انتخاب مقدار اولیه  $X$  با توجه به توزیع مقادیر NO<sub>2</sub> و کواریانس اولیه  $P$  به دلیل همبستگی کامل هر مشاهده با خودش در شروع کار بوده است. مقدار  $H$  برابر یک به دلیل اطمینان از وجود نویز در مشاهدات و مقدار اولیه  $v_0$  به دلیل انحراف اندک مشاهدات در شروع کار از میانگین بوده است. در شکل (۵-الف)؛ نقاط روی نمودار، مقادیر آلاینده در هر گام زمانی و خط پیوسته، نتایج حاصل از پیش‌بینی و پالایش فیلتر کالمن است. محدوده‌ی طوسی رنگ نشان دهنده محدوده‌ی اطمینان ۰٫۹۵، لازم به ذکر است که هر گام زمانی با توجه به داده حسگر ایستگاه‌های ثابت برابر یک ساعت بوده است. اشکال قسمت ب و ج نشان‌دهنده گام زمانی ۹۰-۲۵ و ۱۵۰-۱۰۰ با جزئیات دقیق‌تر می‌باشند. (نمونه A و نمونه B). در نمونه A به طور نمونه یکی از داده‌های پرت نشان داده شده است که در روند پیش‌بینی و تخمین فیلتر حذف شده است. همچنین در نمونه B محل داده جا افتاده مشخص است که توسط فیلتر پیش‌بینی شده است. تصحیح و پیش‌بینی انجام شده برای آلاینده PM<sub>10</sub> در این ایستگاه و NO<sub>2</sub> و PM<sub>10</sub> در ایستگاه قلهک انجام و در اشکال (۶)، (۷) و (۸) نشان داده شد. در روندی مشابه داده‌های مربوط به تمام آلاینده‌ها در تمام ایستگاه پالایش شد.

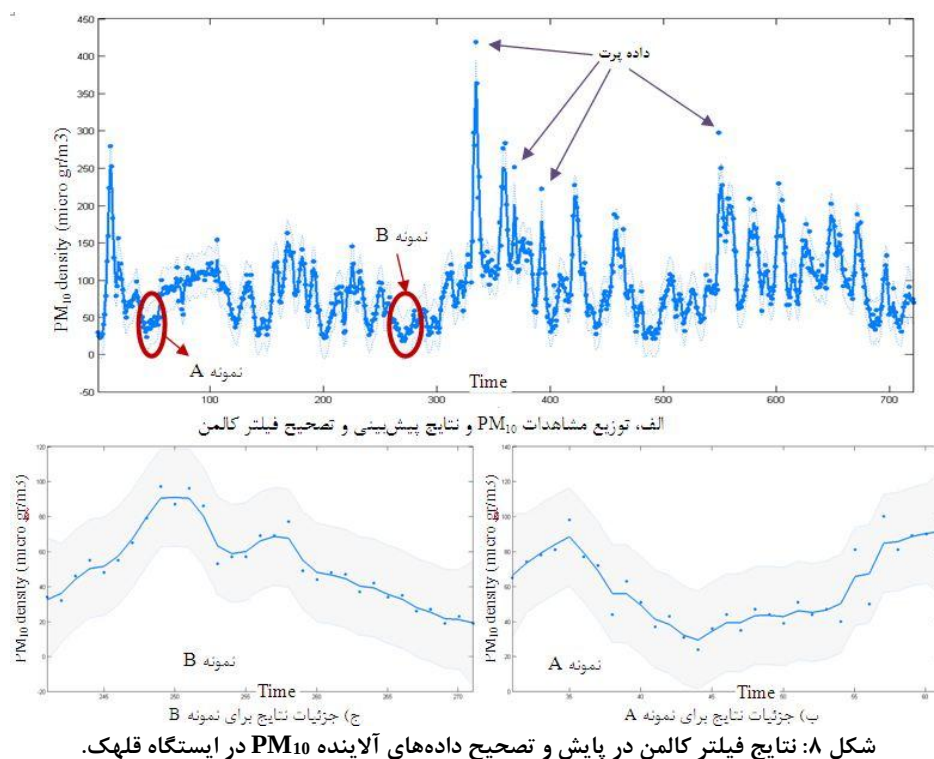




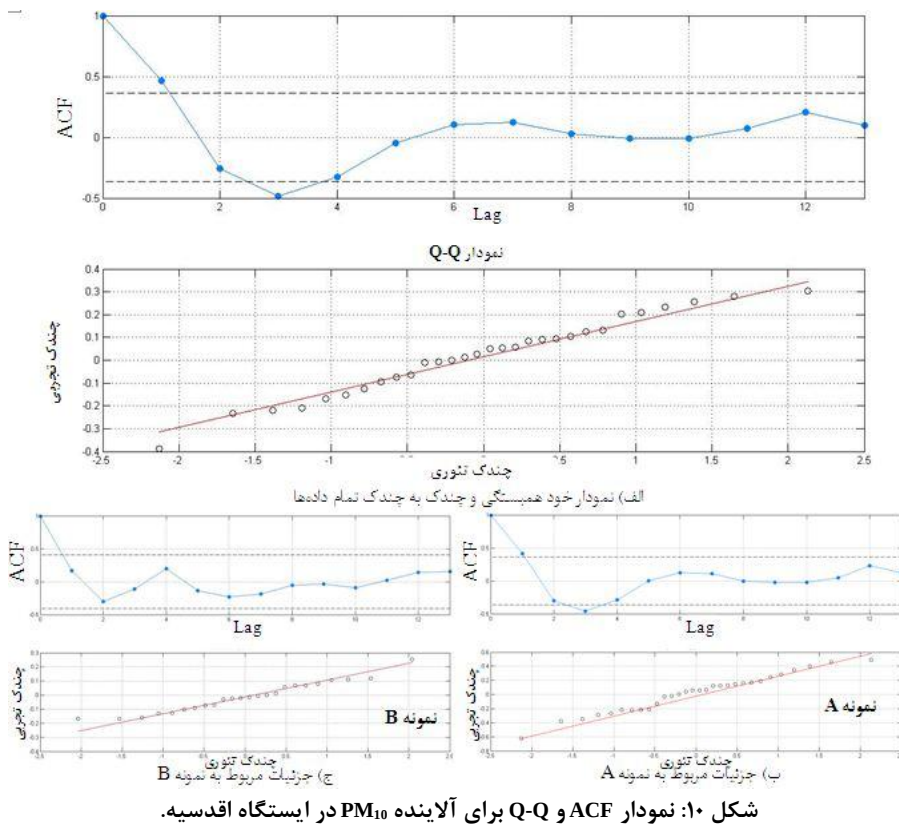
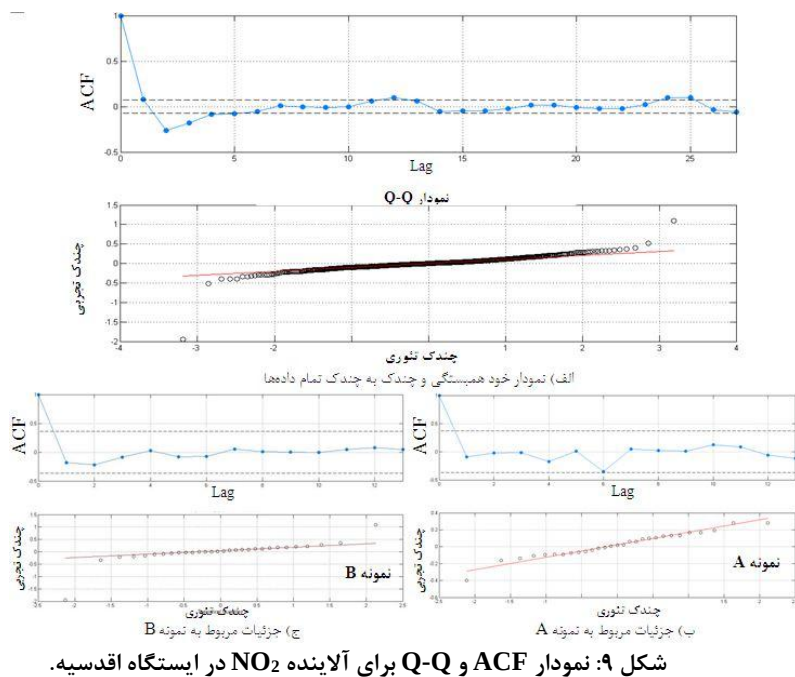
شکل ۶: نتایج فیلتر کالمن در پایش و تصحیح داده‌های آلاینده  $PM_{10}$  در ایستگاه اقدسیه.



شکل ۷: نتایج فیلتر کالمن در پایش و تصحیح داده‌های آلاینده  $NO_2$  در ایستگاه قلهک.

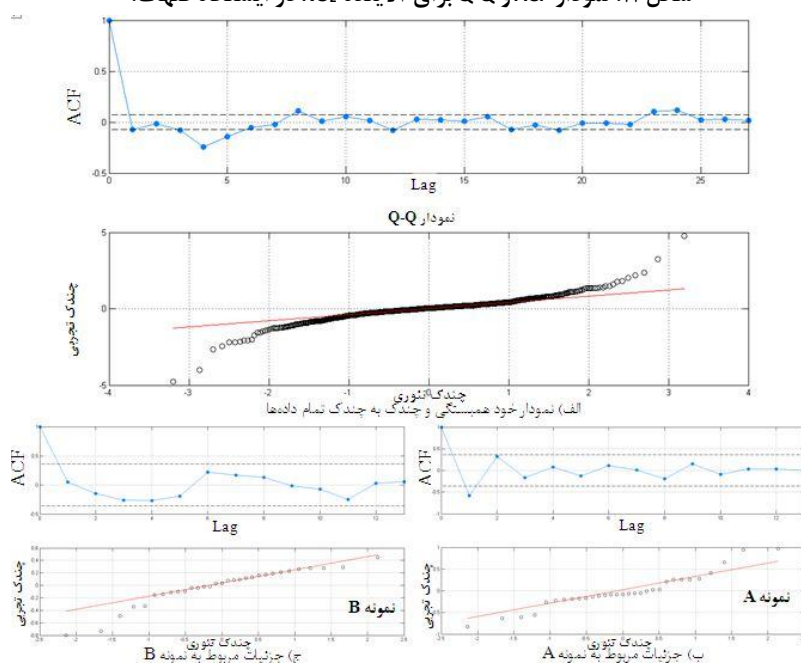
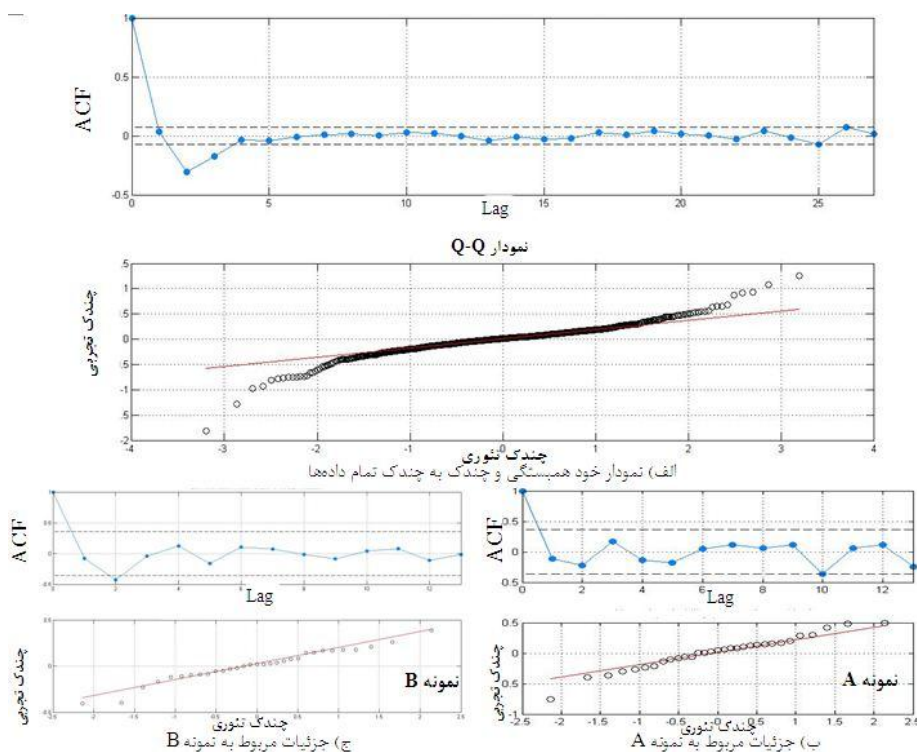
شکل ۸: نتایج فیلتر کالمن در پایش و تصحیح داده‌های آلاینده  $PM_{10}$  در ایستگاه قلهک.

همزمان با اعمال فیلتر کالمن نمودارهای ACF و Q-Q نیز ترسیم گردید که برای مشاهدات  $NO_2$  ایستگاه اقدسیه در شکل ۹ نشان داده شده است. نمودار ACF همبستگی بین مشاهدات در یک بازه زمانی سی روزه را نشان داده است. از این نمودار جهت استخراج الگوی احتمالی مشاهدات و تنظیم پارامترهای فیلتر کالمن استفاده می‌شود. در شروع ( $lag=0$ ) وابستگی داده با خودش یک و در ادامه کمتر می‌شود. با توجه به نرخ تغییرات همبستگی برای هر روز ماتریس R و Q تعیین می‌شود. نمودار Q-Q نیز جهت بررسی توزیع داده‌ها بسیار موثر است. نمودار Q-Q (شکل ۹) مؤید توزیع خطی داده‌هاست که یکی از فرضیات اصلی فیلتر کالمن است. ضمن اینکه داده‌های پرت قابل تشخیص می‌باشند؛ با حذف داده‌های پرت کارایی روش کالمن بهبود می‌گردد. اشکال ب و ج نیز نمودار ACF و Q-Q را برای نمونه‌های A و B با جزئیات بیشتری نشان می‌دهند. نمودار ACF و Q-Q برای آلاینده  $PM_{10}$  ایستگاه اقدسیه در شکل (۱۰) و برای آلاینده‌های  $NO_2$  و  $PM_{10}$  در ایستگاه قلهک به ترتیب در اشکال (۱۱) و (۱۲) ارائه شده است. در قسمت ب و ج هر یک از این اشکال، جزئیات نمونه‌های A و B آورده شده است تا با جزئیات دقیق‌تری رفتار همبستگی مشاهدات و توزیع آنها نشان داده شود.





محور افقی و عمودی در نمودار ACF به ترتیب نشانگر بازه زمانی روزانه و مقادیر همبستگی می‌باشد. همچنین در نمودار Q-محور افقی و عمودی به ترتیب نشان‌دهنده مقادیر محاسبه شده و مشاهده شده آلاینده‌ها می‌باشد که در بازه (۱۰ و ۱۰-) نرمال شده‌اند. این نرمال‌سازی برای یکسان نمودن بازه محور افقی و عمودی نمودار Q-Q برای آلاینده‌های مختلف است.





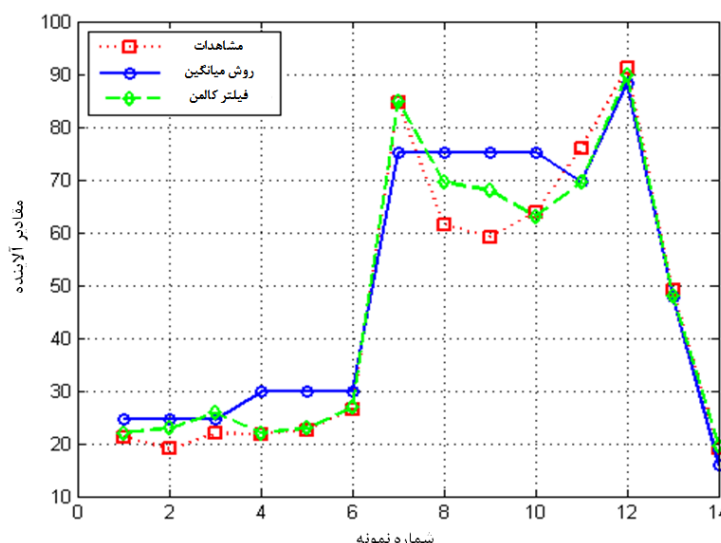
بنابراین، در این گام از تحقیق داده‌ی ورودی حسگرها پالایش شده و جهت بهره‌گیری در سایر فرآیندهای استنتاج و پیش‌بینی آماده شده است.

#### ۴- ارزیابی نتایج

برای ارزیابی فیلتر کالمن در پیش‌بینی مقادیر نامشخص، داده‌های معلوم آلاینده‌ها بطور تصادفی حذف شد و نتایج حاصل از پیش‌بینی فیلتر کالمن با آنها مقایسه گردید. بطور نمونه، ۱۴ داده‌ی معلوم از سری داده NO<sub>2</sub> به طور تصادفی حذف شد. سپس نتایج پیش‌بینی با داده‌ی واقعی مقایسه گردید. همچنین برای بررسی دقیق‌تر، نتایج حاصل با نتایج روش میانگین مقایسه شد؛ چراکه در بیشتر موارد برای تخمین داده‌های جاافتاده از روش میانگین استفاده می‌شود. نتایج پیش‌بینی و تصحیح فیلتر کالمن با نتایج روش میانگین و مقادیر واقعی آن داده‌ها به کمک شاخص متوسط خطای مطلق (MAPE)<sup>۸</sup> مطابق رابطه‌ی (۴) مقایسه گردید [۱۲].

$$MAPE = \left(1 - \frac{\text{Predicted Value}}{\text{Actual value}}\right) \times 100 \quad (۴)$$

این شاخص از تقسیم مقادیر پیش‌بینی شده به مقادیر مشاهده شده به دست می‌آید. نتایج حاصل از مقایسه مقادیر به دست آمده از فیلتر کالمن، روش میانگین و داده‌های مشاهده شده در شکل (۱۳) آمده است.



شکل ۱۳: نتایج ارزیابی فیلتر کالمن در پیش‌بینی مقادیر مشاهده شده.

بررسی مقادیر MAPE نشان دهنده‌ی دقت بالای روش کالمن در بهترین حالت برابر ۹۹/۵۳ درصد و در بدترین حالت برابر ۷۸/۹۴٪ بوده است. در حالی که روش میانگین به ترتیب در بهترین و بدترین حالت دارای دقت ۹۶/۹۵ درصد و ۷۳/۵۳ درصد بوده است. شاخص MAPE به طور متوسط برای فیلتر کالمن برابر ۶/۴۵ درصد و برای روش میانگین ۱۷/۵۷ درصد بوده است. با تکرار

<sup>۸</sup> . Mean Absolute Percentage Error (MAPE)





این عمل برای سایر آلاینده‌ها، دقت متوسط ۹۲ درصد برای فیلتر کالمن بدست آمد که نشان‌دهنده دقت بالای این فیلتر و نزدیکی به مقادیر مشاهده‌شده (صحت) بوده است.

#### ۵- جمع‌بندی و نتیجه‌گیری

در تحقیق حاضر مشاهدات مربوط به آلاینده‌های هفت ایستگاه پایش آلودگی هوا در منطقه سه تهران که در یک بازه تقریباً دو ماهه (آذر و دی ماه سال ۱۳۹۳) جمع‌آوری شده بود، به کمک فیلتر کالمن و نمودارهای خودهمبستگی تصحیح و داده‌های جا افتاده پیش‌بینی گردید. نتایج حاصل از ارزیابی این روش، بیانگر دقت متوسط ۹۲ درصدی برای فیلتر کالمن بوده است. با نگاهی به داده‌های آلودگی سایر ایستگاه‌ها، نظیر ایستگاه‌های پایش آلودگی هوای شهر تبریز، وجود خطا در مشاهدات و عدم وجود داده‌ها در ساعات مختلف مشاهده می‌گردد. این امر بر اهمیت تصحیح و پیش‌بینی این داده‌ها، قبل از بهره‌گیری در تحلیل‌های مکانی تاکید می‌نماید. پیشنهاد می‌شود با بهره‌گیری از مدل‌های ترکیبی نظیر فیلتر کالمن و روش‌های هوش جمعی بر سرعت و دقت پیش‌بینی داده‌های آلودگی هوا افزوده گردد.

#### ۶- مراجع

- ۱- صادقی، ح.، خاکسار آستانه، س.، ۱۳۹۳، پیش‌بینی کوتاه‌مدت آلودگی ذرات معلق شهر اهواز با کمک شبکه‌های عصبی، پژوهش‌های محیط زیست، سال ۵، شماره ۹، ۱۷۷-۱۸۶.
- 2-Brunelli, U., Piazza, V., Pignato, L., Sorbello, F. and Vitabile, S., (2007), **Two-days ahead prediction of daily maximum concentrations of So<sub>2</sub>, O<sub>3</sub>, Pm<sub>10</sub>, No<sub>2</sub>, Co in the urban area of Palermo, Italy**, Atmospheric Environment, Vol.41, pp.2967-2995.
- 3- Perez, P. and Reyes, J., 2006, **An integrated neural network model for PM<sub>10</sub> forecasting**, Atmospheric Environment, Vol.40, pp.2845-2851.
- ۴- بوداقپور، س. و چرخستانی، ا.، ۱۳۹۰، پیش‌بینی میزان غلظت آلاینده‌های هوای تهران با استفاده از شبکه عصبی مصنوعی، علوم و تکنولوژی محیط‌زیست، دوره سیزدهم، شماره یک.
- ۵- علیاری شوره‌دلی، م.؛ تشنه لب، م. و خاکی، ع.، ۱۳۸۷، پیش‌بینی کوتاه مدت آلودگی هوا با کمک شبکه‌های عصبی پرسپترون چندلایه، خط حافظه‌دار تاخیر، گاما و ANFIS با روش‌های ترکیبی آموزشی مبتنی بر PSO، مجله کنترل، جلد ۲، شماره ۱، ۱-۱۹.
- 6- Kalman, R. E., 1960, **A new approach to linear filtering and prediction problems**, Journal of basic Engineering, Vol.82, No.1, pp.35-45.
- 7- Laine, M., Latva-Pukkila, N. and Kyrölä, E., 2014, **Analysing time-varying trends in stratospheric ozone time series using the state space approach**, Atmospheric Chemistry and Physics, Vol.14, No.18, pp.9707-9725.
- 8- Ishihara, J. Y., Terra, M. H. and Campos, J. C., 2006, **Robust Kalman filter for descriptor systems**, IEEE Transactions on Automatic Control, Vol.51, No.8, pp.1354-1358.
- 9- Wolpert, D. M. and Ghahramani, Z., 2000, **Computational principles of movement neuroscience**, nature neuroscience, Vol.3, pp.1212-1217.

۱۰- انتظاری، ع.، ۱۳۸۴، مطالعه آماری و سینوپتیکی آلودگی هوای تهران. (رساله دکتری)، دانشگاه تربیت مدرس.



۱۱- زبردست، ا.، ریاضی، ح.، ۱۳۹۴، شاخص‌های محیط انسان ساخت و تاثیرات آن بر آلودگی هوا (مطالعه موردی: محدوده پیرامونی چهارده ایستگاه سنجش کیفیت هوا در شهر تهران)، نشریه هنرهای زیبا- معماری و شهرسازی، دوره ۲۰، شماره ۱، ۵۵-۶۶.

12- Tofallis, C., 2015, **A better measure of relative prediction accuracy for model selection and model estimation**, Journal of the Operational Research Society, Vol.66, No.8, pp.1352-1362.