

مقایسه روش‌های پس پردازش برونداد مدل WRF برای دمای روزانه در ایستگاه مهرآباد تهران

فرحناز تقوی^{۱*}، مونا کوثری^۲، مجتبی جلالی^۳

۱- استادیار، گروه فیزیک فضا، موسسه ژئوفیزیک دانشگاه تهران، تهران، ایران

۲- دانشجوی دکتری نقشه برداری دانشکده فنی، دانشگاه تهران، تهران، ایران

۳- دانشجوی دکتری هواشناسی، موسسه ژئوفیزیک دانشگاه تهران، تهران، ایران

چکیده

پیش بینی دمای هوا و شرایط جوی با توجه به تاثیر آن بر روی زندگی روزمره انسان همیشه بسیار مهم بوده و یکی از مباحث چالش برانگیز می باشد. در این راستا استفاده از مدل های پیش بینی عددی وضع هوا برای پیش بینی دمای سطح زمین توجه زیادی را به خود جلب کرده است. معمولاً این مدل‌ها دارای خطاهای سامان مند است که عمده آن به خاطر پایین بودن میزان تفکیک توپوگرافی و نقص در پارامترسازی فرایندهای فیزیکی متفاوت در مدل است. در این تحقیق تاثیر داده‌گذاری بر اساس سه روش تبدیل فوریه، شبکه عصبی و پالایه کالمن به منظور پس پردازش دمای روزانه سطح زمین برای ایستگاه تهران و مدل WRF در بهبود پیش بینی دما مورد بررسی قرار گرفته است. بررسی‌های آماری نشان می‌دهد که خطای مدل با توجه به فصل در ایستگاه متفاوت است و پیش‌بینی مدل در همه روزها به صورت کم برآورد یا بیش برآورد می‌باشد به این معنا که در همه روزها خطا مثبت یا در همه روزها منفی است؛ درحالی‌که پس از اعمال پالایه کالمن، این برآورد برای بعضی روزها مثبت و بعضی روزها منفی می‌شود، این مطلب در کاهش قابل ملاحظه خطای میانگین که ارزیابی را اندازه‌گیری می‌کند، مشهود است. مقدار جذر میانگین مربع خطاها که پاشندگی خطا را اندازه‌گیری می‌کند نیز پس از اعمال پالایه کالمن کاهش قابل توجه را نشان می‌دهد. هر چند با صفر فاصله دارد و بیانگر وجود برآورد اضافی و نقصانی است. در میان روش‌ها، پالایه کالمن توانست پیش‌بینی مدل را تا حد قابل قبولی اصلاح کند و مقدار خطا را تا حد فراوانی به اندازه ۹۰٪ کاهش دهد.

کلید واژه‌ها: پس پردازش، دمای روزانه، مدل WRF، شبکه عصبی، تبدیل فوریه، پالایه کالمن.

مقدمه

پیش بینی را کاهش دهد. لازم به ذکر است که دقت دمای سطح زمین ادغام شده با داده‌های ماهواره‌ای با استفاده از روش‌های آنالیز داده نسبت به خطاهای موجود در مدل فیزیکی از جمله پردازش ابرها، رطوبت خاک و بارش حساس هست (نوریس و دوسیلوا، ۲۰۰۷). از طرف دیگر گوارد داده دما به دلیل دریافت حرارتی تابشی توسط دستگاه‌ها و پایین بودن سطح ظرفیت گرمایی مدل امری پیچیده تلقی می‌شود (بنالی و همکاران، ۲۰۱۲). همچنین چرخه روزانه دما، متفاوت بودن خطاهای شب و روز، این مسئله را پیچیده تر می‌کند (دیاز و بردلی، ۱۹۹۷). بدین ترتیب برون‌داد مستقیم مدل را نمی‌توان به صورت خام در اختیار کاربران قرار داد. لذا پس پردازش برون‌داد خام مدل‌های مورد استفاده در پیش‌بینی عددی وضع هوا از گام‌های اصلی و ضروری فرآیند پیش‌بینی‌های عملیاتی است (گالانیس و همکاران، ۲۰۰۶). در ایران نیز مطالعاتی در مورد پس پردازش برون‌داد مدل‌های پیش‌بینی عددی وضع هوا برای کمیت‌هایی مانند دما و بارش، صورت گرفته است. آزادی و همکاران با انجام پژوهشی در سال ۱۳۹۱ نشان دادند که استفاده از روش شبکه‌های عصبی مصنوعی برای پس پردازش برون‌داد خام مدل و سپس واسنجی کردن اعضای سامانه همادی قبل از ترکیب اعضا می‌تواند پیش بینی احتمالاتی دقیق تری را تولید کند. همچنین آزادی و عربلی (۱۳۸۸) به مقایسه پیش‌بینی‌های رسمی و پیش‌بینی‌های مدل WRF بعد از پس پردازش برای دمای بیشینه و کمینه روزانه در ایستگاه‌های همدیدی مراکز استان‌ها پرداختند. نتایج به دست آمده، نشان داد که پیش‌بینی‌های پالاییده بر پیش‌بینی‌های رسمی برتری دارد و نیز لزوم استفاده از پس پردازش برون‌داد خام مدل را نشان داد. با توجه به مطالب بیان شده، برون‌داد خام مدل‌های پیش‌بینی فصلی در مقایسه با مدل‌های پیش‌بینی کوتاه مدت دارای تفکیک پذیری کمتر و خطای بیشتری بوده و نیازمند پس پردازش آماری هستند. هدف از انجام این تحقیق بررسی روش‌های متفاوت به منظور بهبود پیش بینی دمای سطحی زمین در یک فصل می‌باشد. از این رو در این تحقیق به بررسی پالایش داده‌ها با استفاده از سه روش پالایه کالمن (ولچ، ۱۹۹۵)، شبکه عصبی (هانسن و سالامون، ۱۹۹۰) و تبدیل فوریه (برسول، ۱۹۸۶) پرداخته می‌شود. پالایه کالمن که به عنوان تخمین خطی مرتبه دوم نیز

در سال‌های اخیر، پیش بینی دقیق آب و هوا امری ضروری محسوب می‌شود و این موضوع به بهبود روش‌های مختلف پیش بینی‌های هواشناسی در مقیاس‌های جهانی و منطقه‌ای کمک می‌کند. پیش‌بینی هوا از سامانه‌های غیرخطی و پیچیده و فاقد مدل ریاضی است که به دلیل تغییرپذیر بودن با زمان، روش‌های معمول پیش‌بینی امکان پیش‌بینی را غیرممکن و از طرفی با توجه به دارا بودن اهمیت پیش‌بینی در زمینه‌های مختلف از جمله امور اقتصادی، نظامی و کشاورزی، ضرورت پیش‌بینی مطرح می‌شود. مدل‌های پیش‌بینی عددی وضع هوا بر پایستاری جرم، تکانه و انرژی ترمودینامیکی استوارند. ایده اصلی این است که اگر حالت فعلی جو معلوم باشد می‌توان با حل معادله‌های حاکم بر دینامیک و فیزیک جو، حالت آینده جو را پیش‌بینی کرد (رایز و همکاران، ۲۰۰۳). پارامتر دمای سطح زمین به عنوان یکی از مهم‌ترین شرایط مرزی مدل‌های هواشناسی مطرح می‌شود از این رو به عنوان یکی از قدیمی‌ترین و چالش برانگیزترین پارامتر در پیش‌بینی وضع هوا پیش‌بینی محسوب می‌گردد. دمای سطحی جزء ضروری فرایندهای تابشی سطحی هست که جهت بررسی جریان‌های حرارتی چرخشی سطح، لایه مرزی بین سیاره‌ای و همرفت به عنوان یک پارامتر حیاتی شناخته شده است (سیمونز و همکاران، ۲۰۰۴). با وجود اینکه دقت پیش‌بینی وضع هوا توسط مدل‌های عددی در طی سال‌های اخیر بهبود یافته است و استفاده از مدل‌های پیش‌بینی عددی وضع هوا (NWP) می‌تواند به پیش بینی هر چه صحیح‌تر وضعیت هوا کمک کند؛ اما مدل‌ها در پیش‌بینی پارامترهای گوناگون به ویژه پارامترهای نزدیک سطح زمین خطا دارند و نیز قادر به پیش بینی صریح برخی پدیده‌ها مانند احتمال وقوع رعد و برق یا میزان دید نیستند. مهم‌ترین دلیل این امر وجود نقص در پارامتری سازی فیزیکی مورد استفاده در مدل‌های عددی و تفکیک ضعیف توپوگرافی در مدل نسبت به شرایط واقعی و ناتوانی مدل در شبیه‌سازی موفق پدیده‌های زیر شبکه‌ای است. همچنین وجود نادرستی‌هایی که در پیش‌بینی‌های مدل‌های عددی دارد، ممکن است به علت خطاهایی باشد که در تعیین شرایط اولیه و جانبی ایجاد می‌شود. بدین ترتیب عدم قطعیت مرتبط با موارد ذکر شده می‌تواند اعتماد پذیری

بوده و ریزش‌های جوی آن‌که از ماه‌های آبان و آذر آغاز و تا اواسط اردیبهشت ماه ادامه دارد، تابعی از فعالیت‌های سامانه‌های فوق می‌باشد.

روش تحقیق

برای استفاده کاربردی برون‌داد مدل، روش‌های گوناگون پس پردازش داده‌ها مورد استفاده قرار می‌گیرد که از بین آن‌ها می‌توان به روش‌های آماره برون‌داد مدل MOS (گلان و لاری، ۱۹۷۲)، روش پیش‌یابی کامل PPM (کلین و همکاران، ۱۹۵۹)، روش واسنجش برون‌داد مدل MOC (مائو و همکاران، ۱۹۹۹) اشاره کرد. محدودیت اصلی این روش‌ها لزوم دسترسی به داده‌های طولانی مدت است که همیشه در دسترس نیست. روش‌های دیگری در سال‌های اخیر توسعه یافته و مورد استفاده قرار گرفته‌اند که می‌توان برون‌داد مدل را بدون نیاز به داده‌های طولانی مدت تصحیح کرد. از آن جمله می‌توان به روش‌های پالایه کالمن (موناچی، ۲۰۱۱؛ گالانیس و همکاران، ۲۰۰۶؛ هاملید، ۱۹۹۵؛ براون، ۱۹۹۲؛ کالمن و باسی، ۱۹۶۱؛ کالمن، ۱۹۶۰)، شبکه عصبی مصنوعی (مرزبان، ۲۰۰۳)، میانگین متحرک (مکولر و استال، ۲۰۰۸) و روش‌های بیزی (جلمن و همکاران، ۲۰۱۳) اشاره کرد.

برون‌داد مدل مورد استفاده در این تحقیق مربوط به مدل WRF می‌باشد. مدل WRF سامانه‌ای با قابلیت فراوان است که می‌توان وضعیت‌های مختلف جو را با استفاده از آن شبیه‌سازی نمود. مدل مذکور، قابل نصب بر روی سامانه‌های مختلف و رایانه‌ای شامل خوشه‌ی رایانه می‌باشد. محدوده‌ی کاربرد آن بسیار وسیع است و گسترده‌ی بین چندین متر تا هزاران کیلومتر را تحت پوشش قرار می‌دهد. مدل پیشرفته WRF یک مدل غیرهیدرواستاتیک است. شبکه‌بندی مدل شبکه آراکاوای-C می‌باشد. در این مدل از طرحواره‌های انتگرال‌گیری مرتبه دوم و سوم زمانی رنگ-کوتا و طرحواره‌های مرتبه دو تا شش برای فرارفت در دو راستای افقی و قائم استفاده می‌شود. این مدل از گام‌های کوچک زمانی برای مدهای اکوستیک و امواج گرانشی استفاده می‌کند. در این قسمت از تحقیق روش‌های مورد استفاده (تبدیل فوریه، شبکه عصبی و پالایه کالمن) به‌طور مجزا بیان می‌شود.

از آن یاد می‌شود، الگوریتمی است که حالت یک سیستم پویا را با استفاده از مجموعه‌ای از اندازه‌گیری‌های شامل خطا در طول زمان برآورد می‌کند. روش شبکه عصبی با ایده گرفتن از سیستم عصبی بدن انسان و باهدف شبیه‌سازی هر چه بیشتر کامپیوتر به انسان شکل گرفت و تا به حال به خوبی پیشرفت کرده است. این روش قابلیت بسیار بالایی در استنتاج نتایج از داده‌های مبهم و پیچیده جهت استخراج الگوها دارند و در شناسایی روش‌هایی که آگاهی از آن‌ها برای انسان و فن‌های کامپیوتری دیگر پیچیده و دشوار است، بسیار متمرکز هستند. یک شبکه عصبی آموزش دیده می‌تواند به‌عنوان یک متخصص در مقوله‌ای که برای تجزیه و تحلیل به آن داده شده، عمل کند و برآوردهایی از وضعیت‌های دلخواه و پیش‌بینی‌های آینده را ارائه دهد. روش تبدیل فوریه بر این مبنا است که یک تابع با دوره تناوب T را می‌توان به یک تابع ثابت و تعداد نامحدودی از توابع سینوسی تجزیه نمود. در تمامی این روش‌ها دمایی که توسط ایستگاه هواشناسی ثبت می‌شود، با دمای سطحی حاصل از مدل فیزیکی ادغام می‌شود و منجر به بهبود تخمین دما می‌شود (بچتولد و همکاران، ۲۰۱۹). این نوشتار شامل پنج بخش می‌شود. بخش دوم شامل داده‌ها و روش‌های مورد استفاده به‌منظور مدل‌سازی خطای مشاهدات و مدل می‌باشد. در بخش سوم پارامترهای ارزیابی و سری زمانی روش‌های مختلف حاصل از این تحقیق نشان داده شده است. نتایج و بحث در بخش چهارم بیان می‌شود. و در نهایت در بخش پنجم نتیجه‌گیری ارائه شده است.

مواد و روش‌ها

داده‌ها و منطقه مورد مطالعه

داده‌های مورد استفاده در این تحقیق مربوط به پارامتر دمای ایستگاه هواشناسی مهرآباد می‌باشد. این ایستگاه از نوع سینوپتیک فرودگاهی واقع در طول جغرافیایی ۵۱.۱۹ درجه و عرض جغرافیایی ۳۵.۴۱ است. داده‌ها مربوط به ۴ ماه (از ۲۰۱۸/۰۸/۰۱ تا ۲۰۱۸/۱۱/۲۹) برابر یک فصل هواشناسی می‌باشد. ویژگی‌های اقلیمی در مقیاس کلی منطقه، همانند سایر بخش‌های استان تهران در فصول سرد سال متأثر از سامانه‌های شمالی، شمال غربی، غربی و به‌ویژه جنوب غربی

تبدیل فوریه

جنبه متمایز کننده رویدادهای تناوبی، الگوی تکراری در زمان است. دما نیز همانند تمامی عناصر اقلیمی این قبیل رویدادها را در بازه‌های زمانی مختلف تکرار نموده و دو وضعیت یکسان متوالی را در یک فاصله زمانی کمابیش برابر

تجربه می‌کند. به طوری که مثلاً دمای ژانویه هر سال تقریباً مشابه دمای ژانویه سال‌های پیش و آتی است. به هر حال یک سری زمانی تناوبی مثلاً میانگین ماهانه دما به طول n را می‌توان به فرم زیر نوشت:

$$z_i = a_0 + \sum_{i=1}^q (a_i \cos 2\pi f_i t + b_i \sin 2\pi f_i t) \quad i = 1, 2, 3, \dots, q \quad (1)$$

صورت تعیین تعداد معدودی از هارمونیک‌ها که از جمع آن‌ها تقریب خوبی برای یک سری زمانی به دست می‌آید، مدل‌سازی فوریه نامیده می‌شود. به عنوان مثال فرض کنید که هارمونیک k ام یعنی $h_k(t)$ تقریب خوبی برای متوسط‌های دمای ماهانه در یک ایستگاه باشد. در این صورت مطابق با معادله (۵) می‌توان نوشت:

$$z_i = a_0 + a_k \cos\left(2\pi \frac{k}{n} t\right) + b_k \sin\left(2\pi \frac{k}{n} t\right) + e_i \quad (5)$$

که در آن (e_i) مجموع $(q-1)$ هارمونیک دیگر است که در مدل منظور نمی‌شود. ضروری است یادآوری شود که هرچند نمایش فوریه یک دنباله متناهی مقوله‌ای صرفاً ریاضی است ولی تعیین آن تعداد معدود از هم سازها که به لحاظ آماری معنی‌دار بوده است، تقریب خوبی به دست دهد.

شبکه عصبی مصنوعی

در این تحقیق، شبکه عصبی چندلایه، شامل یک لایه ورودی و یک لایه میانی و یک لایه خروجی در پس پردازش خروجی مدل استفاده شد. در هر لایه یک یا چند عنصر پردازشگر (نرون) وجود دارد که به همه نرون‌های لایه بعدی با اتصالات وزن‌دار مربوط می‌شود. پردازش هر نرون در دو مرحله انجام می‌شود. در مرحله نخست به هر یک از ورودی‌های نرون وزنی داده می‌شود که با مقداری بایاس جمع جبری می‌شوند. حاصل جمع در مرحله دوم وارد پالایه‌ای غیرخطی می‌شود که معمولاً تابع انتقال یا تابع محرک نامیده می‌شود. تابع محرک مثل تابع فشرده عمل می‌کند؛ طوری که خروجی نرون در یک شبکه عصبی بین دو مقدار معین (معمولاً بین 0 و 1 یا بین -1 و 1) قرار گیرد. شکل ۱ فرایند ریاضی مراحل فوق را به صورت شمایی نشان

که در آن z_i عنصر اقلیمی مورد بررسی در زمان t ، f_i فراوانی تکرار مشاهدات است. همان‌گونه که در معادله (۱) دیده می‌شود، مؤلفه‌های سینوسی و کسینوسی حول یک میانگین ثابت رفتار سری را تعیین می‌کنند. بدین دلیل است که رابطه فوق در توصیف پدیده‌های نوسانی اهمیت زیادی دارد. در واقع مؤلفه‌های سینوسی و کسینوسی، هارمونیک‌هایی هستند که در شکل‌گیری رفتار سری تناوبی مؤثرند. تعداد این مؤلفه‌های نوسانی حداکثر نصف طول داده‌ها است. چرا که رفتار نوسانی حداقل از دو مؤلفه سینوسی و کسینوسی تشکیل شده است. در واقع هر هارمونیک گویای یک روند رو به بالا و یک روند رو به پایین در یک سری زمانی است؛ بنابراین هر طول موج متوالی در سری زمانی تناوبی با یک همساز نشان داده می‌شود.

چنان‌که آشکار است الگوریتم فوریه را می‌توان به عنوان یک الگوریتم رگرسیون خطی در نظر گرفت که در آن تعداد ضرایب رگرسیون برابر با طول سری می‌باشد. در رابطه (۱)، a_0 ، a_i و b_i ضرایب فوریه نامیده می‌شوند و همانند یک رگرسیون چندمتغیره با استفاده از روش کمترین مربعات خطا به دست می‌آیند که به صورت معادله (۲) تا (۴) می‌باشد (بوینگتون، ۱۹۶۹؛ جانستون، ۱۹۶۰).

$$a_0 = \frac{1}{n} \sum_{t=1}^n z_t \quad (2)$$

$$a_i = \frac{2}{n} \sum_{t=1}^n (z_t \cos 2\pi f_i t) \quad (3)$$

$$b_i = \frac{2}{n} \sum_{t=1}^n (z_t \sin 2\pi f_i t) \quad (4)$$

تحت شرایط کلی معمولاً یک سری زمانی از جمع یک مؤلفه تصادفی (e_i) و یک مؤلفه تناوبی (h_i) تشکیل شده است؛ بنابراین جمع معدودی از (q) هارمونیک ممکن، تقریب خوبی برای توصیف این عناصر تلقی می‌شود. در این

شبکه با داده‌هایی که در آموزش و ارزیابی از آن‌ها استفاده نشده آزمایش و عملکرد آن با استفاده از شاخص‌های آماری بررسی می‌شود.

کارایی اجرای مدل شبکه عصبی به تعداد نرون‌های لایه میانی بستگی دارد. ولی تعداد زیاد آن‌ها سبب برآزش بیش‌ازحد داده‌های آموزش می‌شود و بر کیفیت نتایج مدل تأثیر می‌گذارد. بنابراین انتخاب تعداد نرون‌های میانی بسیار مهم است. با توجه به اینکه روش مشخصی برای تعیین تعداد بهینه نرون‌های میانی وجود ندارد، با استفاده از روش سعی و خطا تعداد آن‌ها تعیین می‌شود. در این تحقیق مدل شبکه عصبی ۱۰ بار با تعداد نرون ۱ تا ۱۰ در لایه میانی آموزش یافت و سپس با استفاده از داده‌های اختصاص‌یافته برای آزمون نتایج مدل‌ها با داده‌های واقعی مقایسه و مناسب‌ترین تعداد نرون انتخاب شد.

از آنجاکه دامنه تغییرات داده‌های ورودی و خروجی اختلاف زیادی با هم دارند، واردکردن آن‌ها به شبکه عصبی باعث کاهش سرعت و دقت مدل می‌شود. نرمال‌سازی داده‌ها، قبل از آموزش شبکه، طوری که همه داده‌ها در بازه‌ای مناسب قرار گیرند، به آموزش شبکه کمک زیادی می‌کند. در این مقاله با استفاده از رابطه ۷ همه داده‌های ورودی و خروجی در بازه ۰ تا ۱ نرمال شدند و سپس بعد از شبیه‌سازی مدل به مقادیر اصلی برگشتند:

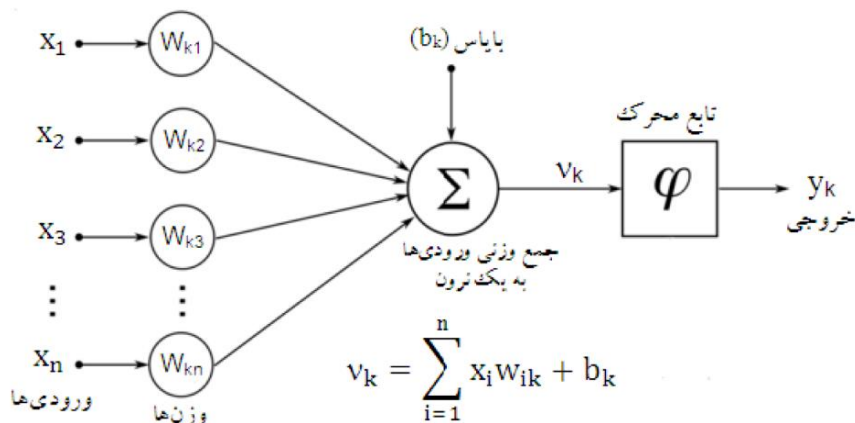
$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

در این رابطه X_{norm} مقدار نرمال شده، X مقدار اصلی و X_{min} و X_{max} به ترتیب بیشترین و کمترین مقادیر اصلی‌اند.

می‌دهد. در این تحقیق، تابع محرک لوگ سیگموئید برای هر دو لایه میانی و خروجی استفاده شد. این تابع از توابع محرک رایج است که پیوسته است و بین دو مقدار 0 و 1 تغییر می‌کند. تابع لوگ سیگموئید با رابطه 6 تعریف می‌شود:

$$y_k = 1 / (1 + \exp(V_k)) \quad (6)$$

V_k جمع وزنی ورودی‌ها به k امین نرون در لایه بعدی و y_k خروجی نرون است. وزن‌ها ارتباط‌دهنده نرون‌های شبکه و بایاس با آموزش تعیین می‌شوند؛ به صورتی که خطا بین خروجی‌های مدل با مقادیر واقعی در مجموعه‌ای که برای آموزش اختصاص‌یافته به حداقل برسد. در این تحقیق، از الگوریتم پس انتشار خطا برای آموزش شبکه عصبی استفاده شد. کیفیت شبکه عصبی به کلیت بخشی آن برای دامنه وسیعی از داده‌های ورودی و خروجی بستگی دارد که با اعمال نظارت بر آموزش به دست می‌آید. نظارت بر آموزش باعث می‌شود خروجی‌های دقیقی برای ورودی‌هایی که در شبکه دیده نشده برآورد گردد. عدم کلیت بخشی به علت آموزش بیش‌ازحد پیش می‌آید که در این شرایط شبکه فقط روی داده‌های آموزش خوب کار می‌کند و برای سایر داده‌ها مناسب نیست. توقف به موقع آموزش قبل از همگرا شدن شبکه از روش‌های شبکه از روش‌های رایج دور زدن آموزش بیش‌ازحد است (پریچلت، ۱۹۹۸). در این تحقیق از تابع لونیبرگ مارکووات به دلیل سرعت بالاتر و نتایج دقیق‌تر نسبت به سایر روش‌های آموزشی برای تصمیم به توقف مرحله آموزش شبکه استفاده شد (هاگن و منهج، ۱۹۹۴). در این تابع از یکسری داده به عنوان داده‌های ارزیابی استفاده شد. پس از هر بار تکرار فرایند یادگیری، شبکه با وزن جدید برای داده‌های ارزیابی اجرا می‌شود. به‌طور معمول در مراحل اولیه آموزش خطای برآورد خروجی مدل برای داده‌های ارزیابی کاهش می‌یابد؛ ولی زمانی که آموزش بیش‌ازحد اتفاق می‌افتد با شروع افزایش خطا آموزش متوقف و وزن‌های شبکه در شرایط حداقل خطا برای داده‌های ارزیابی تعیین می‌شوند. به عبارت دیگر، آموزش شبکه با استفاده از داده‌های آموزش و ارزیابی صورت می‌گیرد. بعد از آموزش،



شکل ۱- روند محاسبات در یک نرون شبکه عصبی

ماتریس‌های M_i (عملگر معادله سیستم) و H_i (عملگر معادله دیدبانی) باید قبل از اعمال پالایه تعیین شوند (گالانیس و همکاران، ۲۰۰۶). در این روش بردار حالت نامعلوم X_i بر مبنای همه مقادیر دیدبانی تا زمان t_i برآورد می‌شود. در گام اول، پیش‌بینی بردار حالت X و ماتریس کوواریانس خطای آن یعنی ماتریس p ، بر اساس مقادیر تحلیلی در گام زمانی t_{i-1} توسط معادلات پیش‌بینی محاسبه می‌شوند:

$$x^f(t_i) = M_{i-1}[x^a(t_i)] \quad (10)$$

$$p^f(t_i) = M_{i-1}p^a(t_{i-1})M_{i-1}^T + Q(t_{i-1}) \quad (11)$$

که در آن‌ها ماتریس Q ، ماتریس کوواریانس خطای فرایند $\eta(t_i)$ می‌باشد. در گام بعدی، وقتی دیدبانی جدید در زمان t_i یعنی y_i^o انجام شود، می‌توان برآورد جدید بردار حالت یعنی $x^a(t_i)$ را با ترکیب اطلاعات قبلی از طریق "معادلات تصحیح کننده" زیر به روز کرد:

$$x^a(t_i) = x^f(t_i) + K_i(y_i^o - H_i[x^f(t_i)]) \quad (12)$$

$$p^a(t_i) = (I - K_i H_i)p^f(t_i)$$

که در آن‌ها ماتریس بهره کالمن، K_i از رابطه زیر محاسبه می‌شود:

$$K_i = p^f(t_i)H_i^T[H_i p^f(t_i)H_i^T + R_i]^{-1} \quad (14)$$

در معادله‌های فوق، ماتریس R_i ، ماتریس کوواریانس خطای اندازه‌گیری (ε_i) و $p^a(t_i)$ کوواریانس خطای $x^a(t_i)$ در زمان t_i می‌باشد. بهره کالمن تنظیم می‌کند که پالایه چگونه خودش را با شرایط جدید و ممکن سازگار کند.

پالایه کالمن

پالایه‌های کالمن بر اساس سامانه‌های خطی پویا گسسته در بازه زمانی هستند و معمولاً تخمین دقیق‌تری را نسبت به تخمین بر مبنای یک اندازه‌گیری واحد را بر مبنای استنباط بیزی (جلمن و همکاران، ۲۰۱۳) و تخمین توزیع احتمال مشترکی از یک متغیر تصادفی در یک مقطع زمانی ارائه می‌کند. این پالایه از نام رودولف ای کالمن، یکی از پایه‌گذاران این تئوری گرفته شده است. یک پالایه بازگشتی است به این معنا که تنها تخمین حالت قبل و مشاهده فعلی برای محاسبه تخمین حالت فعلی لازم است و با کمک مجموعه‌ای از اندازه‌گیری‌های ناقص دارای نوفه و نیز با استفاده از دینامیک حاکم بر مدل، حالت سامانه دینامیکی را برآورد می‌کند. اساس این پالایه جبر خطی و مدل مخفی مارکوف است. سامانه دینامیکی موردنظر به صورت یک زنجیره مارکوف مدل‌سازی می‌شود. این زنجیره با عملگرهای خطی پریشیده شده با نوفه گاوسی ساخته شده است. در این پالایه، حالت سامانه با برداری از اعداد حقیقی نمایش داده می‌شود. هدف اصلی، شبیه‌سازی یک فرایند نامعلوم (بردار حالت) با زمان است که مقدار معلوم آن در زمان (t_i) در این مقاله با $x^t(t_i)$ مشخص شده است. y_i^o نیز به عنوان بردار دیدبانی در همان زمان به کار برده شده است. معادله‌های سیستم و دیدبانی که به ترتیب نشان‌دهنده تحول زمانی بردار حالت و ارتباط بردار حالت با دیدبانی متناظر می‌باشد، به صورت زیر نوشته می‌شوند:

$$x^t(t_{i+1}) = M_i[x^t(t_i)] + \eta(t_i) \quad (8)$$

$$y_i^o = H_i[x^t(t_i)] + \varepsilon_i \quad (9)$$

میانگین خطای مطلق

میانگین خطای مطلق برای یک پیش‌بینی از رابطه زیر محاسبه می‌شود:

$$MAE = \frac{1}{N} \sum_{i=1}^N |F_i - O_i| \quad (16)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |F_i - O_i| \quad (16)$$

که در آن F_i مقدار پیش‌بینی شده متغیر موردنظر و O_i مقدار مشاهداتی متناظر با آن است. این شاخص نسبت به میانگین خطا، برتری دارد زیرا در مورد شاخص میانگین خطا گفته شد که این احتمال وجود دارد که مقادیر پیش‌بینی شده دارای خطای زیادی باشند؛ ولی تفاضل مقدار مشاهداتی از آن‌ها گاهی مثبت و گاهی منفی شود و در نتیجه مقدار میانگین خطای پیش‌بینی صفر یا نزدیک به صفر شود، ولی در اینجا با در نظر گرفتن قدر مطلق تفاضل مقادیر مشاهداتی و پیش‌بینی این عیب برطرف شده و از صفر شدن میانگین خطای مطلق جلوگیری می‌شود.

جذر میانگین مربع خطاها

مقدار عددی جذر میانگین مربع خطاها از رابطه زیر به دست می‌آید:

$$RMSE = \left[\frac{1}{N} \sum_{i=1}^N (F_i - O_i)^2 \right]^{1/2} \quad (17)$$

باید توجه کرد که RMSE در مقایسه با MAE برای خطاهای بزرگ‌تر جریمه بیشتری قائل می‌شود، چون در این شاخص مقادیر خطا به توان ۲ رسانده می‌شود. به عبارت دیگر این کمیت به پیش‌بینی‌های پرت حساس است. RMSE نیز مانند MAE مقدار خطا را برحسب همان یکای کمیت مورد سنجش بیان می‌کند.

امتیاز مهارتی میانگین خطای مطلق

برای محاسبه امتیاز مهارتی میانگین خطای مطلق از رابطه زیر استفاده می‌شود:

$$MAE \text{ skill score} = 1 - \frac{MAE_{pp}}{MAE_{DMO}} \quad (18)$$

معادلات (۸) تا (۱۴) به منظور به‌روزرسانی الگوریتم کالمن از زمان t_{i-1} به t_i استفاده می‌شود. در این روش تعیین $\eta(t_i)$ و ε_i از اهمیت ویژه‌ای برخوردار است و مقدار آن‌ها می‌تواند خروجی را به شدت تحت تأثیر قرار دهد؛ این موضوع که بعد از تغییرات زیاد در داده‌های ورودی، پالایه کالمن با چه سرعتی با وضعیت جدید انطباق پیدا کند، به نسبت $\eta(t_i)/\varepsilon_i$ بستگی دارد. انتخاب درست این مقادیر باعث پایداری نتایج می‌شود و داده‌های اندازه‌گیری شده با خطای زیاد در یک روز، خطای بزرگی در خروجی ایجاد نمی‌کنند.

پارامترهای آماری مورد استفاده جهت راستی آزمایی

راستی آزمایی یک پیش‌بینی، فرایندی است که کیفیت پیش‌بینی را تعیین می‌کند. در این فرایند پیش‌بینی با دیدبانی متناظر که واقعاً اتفاق افتاده است یا خیر، برآوردهای خوب از برونداد مدل مقایسه می‌شود. برای ارزیابی و راستی آزمایی پیش‌بینی‌ها، از شاخص‌ها و کمیت‌های راستی آزمایی گوناگونی استفاده می‌شود. پارامترهایی که در این تحقیق مورد استفاده قرار گرفته است، در ادامه تشریح می‌شود.

میانگین خطا

میانگین خطا نشان‌دهنده بخش سامان مند خطا است و با استفاده از رابطه زیر محاسبه می‌شود:

$$\text{mean error} = \frac{1}{N} \sum_{i=1}^N (F_i - O_i) \quad (15)$$

که در آن F_i مقدار پیش‌بینی متغیر موردنظر و O_i مقدار مشاهده متناظر با آن است. این کمیت نمی‌تواند بزرگی و اندازه خطاها را اندازه‌گیری کند، هم‌چنین نمی‌تواند رابطه و تناظر بین مقادیر پیش‌بینی و مشاهده را مشخص کند. هرچند به نظر می‌رسد که بهترین حالت مقدار میانگین خطای پیش‌بینی برابر صفر باشد، اما واقعاً این‌گونه نیست؛ چه‌بسا مقادیر پیش‌بینی دارای خطای زیادی باشند ولی تفاضل مقدار مشاهدات از آن‌ها، گاهی مثبت و گاهی منفی شود و در نتیجه مقدار میانگین خطای پیش‌بینی صفر یا نزدیک به صفر شود.

می‌شود. خروجی مدل و مشاهده بدون در نظر گرفتن پلایه مطابق با شکل (۲) می‌باشد.

تصحیح اختلاف بین مدل و مشاهده با استفاده تبدیل فوریه

همان‌طور که در قسمت روش تحقیق بیان شد، در تبدیل فوریه بسامد با توان طیفی بالا استخراج می‌شود. شکل ۳ توان طیفی را بر حسب بسامدها نشان می‌دهد. در این روش دو راهکار وجود دارد:

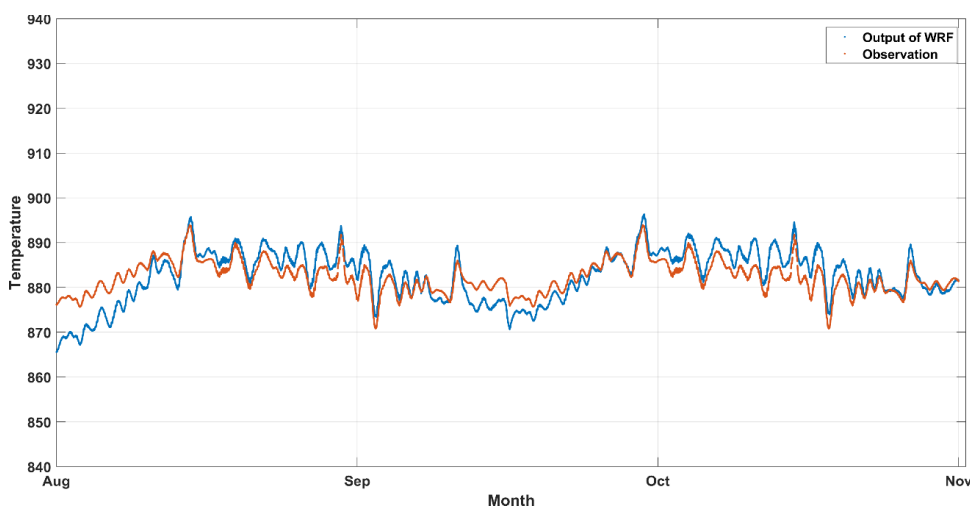
(۱) در این حالت از تمام بسامدها جهت مدل‌سازی استفاده می‌شود.

(۲) در این حالت تنها از ۱۰ بسامد اول استفاده می‌شود.

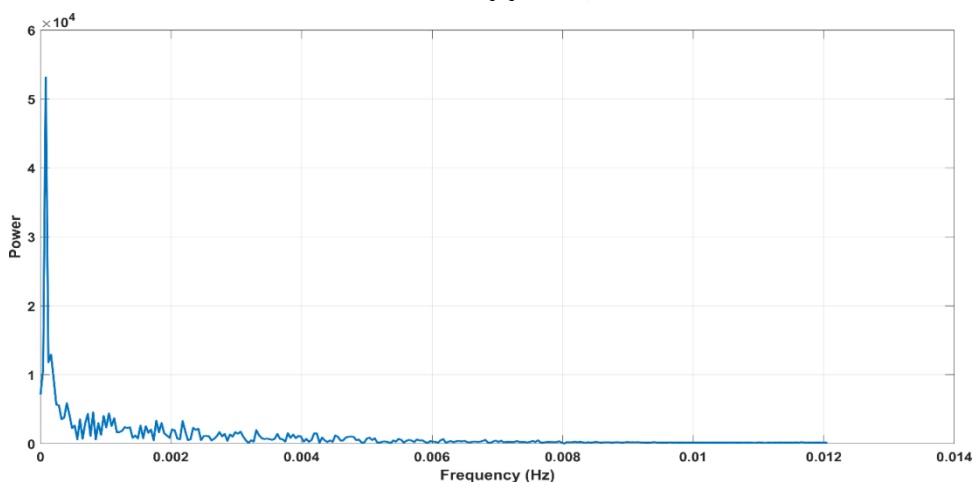
که در آن MAE_{DMO} میانگین خطای مطلق قبل از پردازش و MAE_{pp} میانگین خطای مطلق پس از پردازش می‌باشد. این شاخص اشاره به اطلاعاتی درباره مقدار با ارزش یک پیش‌بینی نسبت به یک پیش‌بینی جایگزین دارد.

نتایج و بحث

همان‌طور که پیش‌ازین ذکر شد، هدف از انجام این تحقیق مقایسه بین روش‌های پس پردازش می‌باشد و تصحیح اختلاف بین مدل و مشاهدات با استفاده از روش‌های مختلف می‌باشد که در ادامه نتایج هر یک از روش‌ها بیان



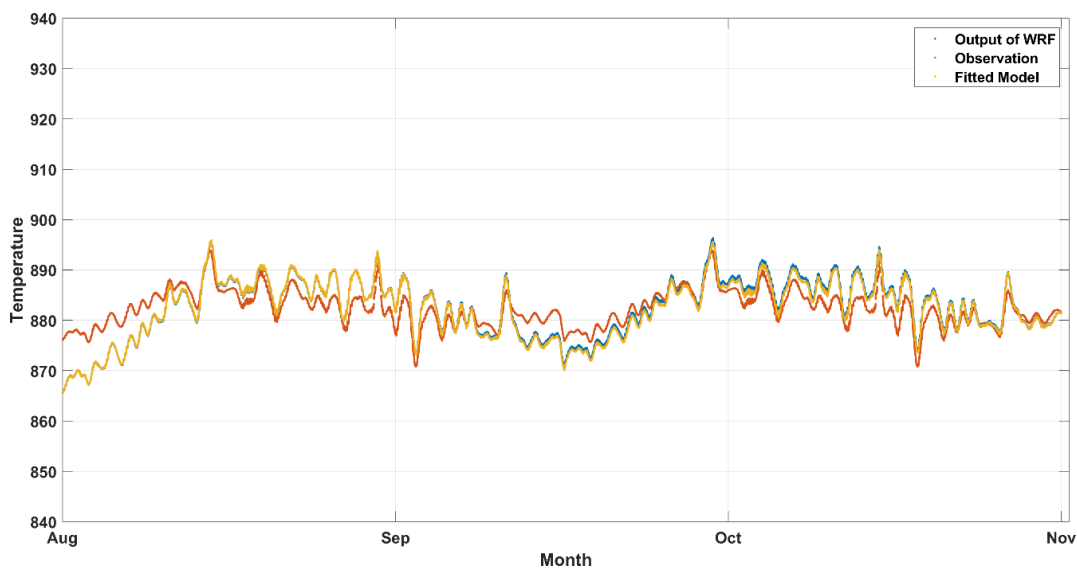
شکل ۲: مقایسه مدل و مشاهدات پارامتر دما برای ایستگاه مهرآباد در ۴ ماه از سال ۲۰۱۸. منحنی خروجی مدل با رنگ آبی و منحنی مشاهدات با رنگ قرمز نشان داده شده است.



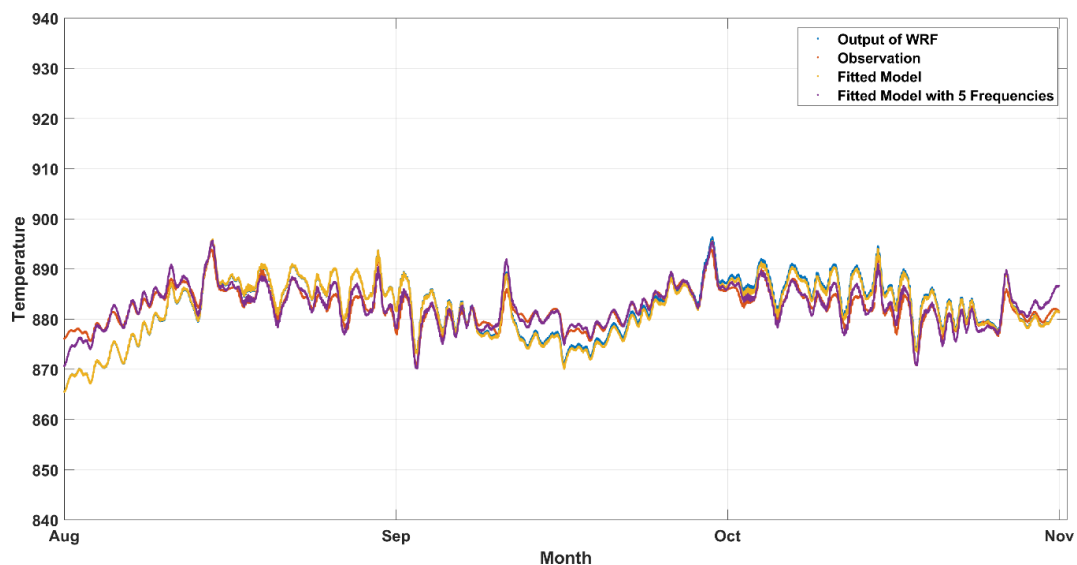
شکل ۳: توان طیفی سیگنال نسبت به بسامدها

تبدیل فوریه با تمام بسامدها با رنگ زرد نشان داده شده است و در شکل ۵ مدل‌سازی دما با استفاده از ده بسامد اول به‌دست‌آمده از تبدیل فوریه نشان داده شده است. همان‌طور که در شکل ۵ مشاهده می‌شود، در حالتی که تنها از ۱۰ بسامد اول جهت مدل‌سازی پارامتر دما استفاده شود، اختلاف بین مشاهده و پارامتر مدل‌سازی کمتر می‌باشد.

پس از استخراج بسامدهای مؤثر، مدل‌سازی را با تمام بسامدها و ۱۰ بسامد اولیه انجام می‌شود و خروجی مدل‌سازی مطابق با شکل ۴ و ۵ می‌باشد. در شکل ۴، مدل‌سازی دما با استفاده از تمامی بسامدهای به‌دست‌آمده از تبدیل فوریه نشان داده شده است. منحنی خروجی مدل WRF با رنگ آبی، منحنی مشاهدات با رنگ قرمز و منحنی مدل



شکل ۴: مدل‌سازی دما با استفاده از تمامی بسامدهای به‌دست‌آمده از تبدیل فوریه. منحنی خروجی مدل WRF با رنگ آبی، منحنی مشاهدات با رنگ قرمز و منحنی مدل تبدیل فوریه با تمام بسامدها با رنگ زرد نشان داده شده است.

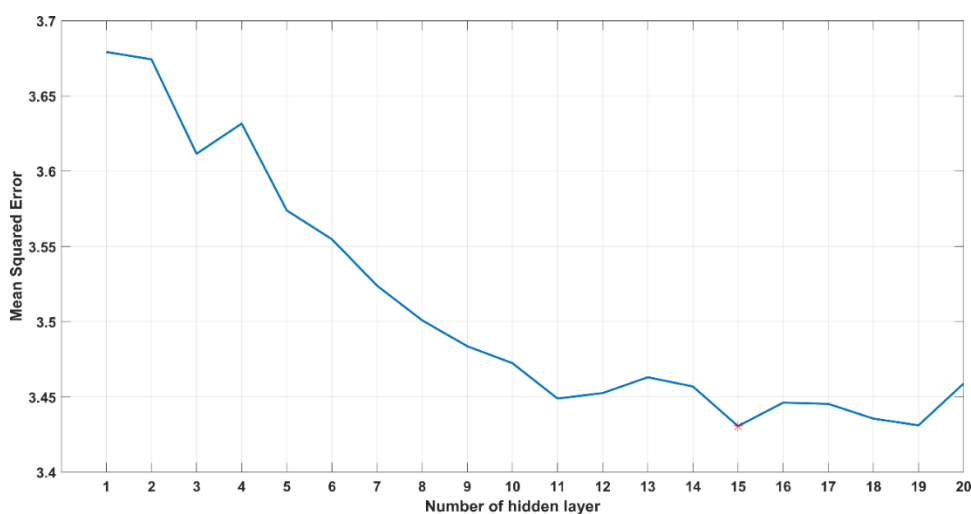


شکل ۵: مدل‌سازی دما با استفاده از ده بسامد اول به‌دست‌آمده از تبدیل فوریه. منحنی خروجی مدل WRF با رنگ آبی، منحنی مشاهدات با رنگ قرمز و منحنی مدل تبدیل فوریه با تمام بسامدها با رنگ زرد و منحنی مدل تبدیل فوریه با ده بسامد اول با رنگ بنفش نشان داده شده است.

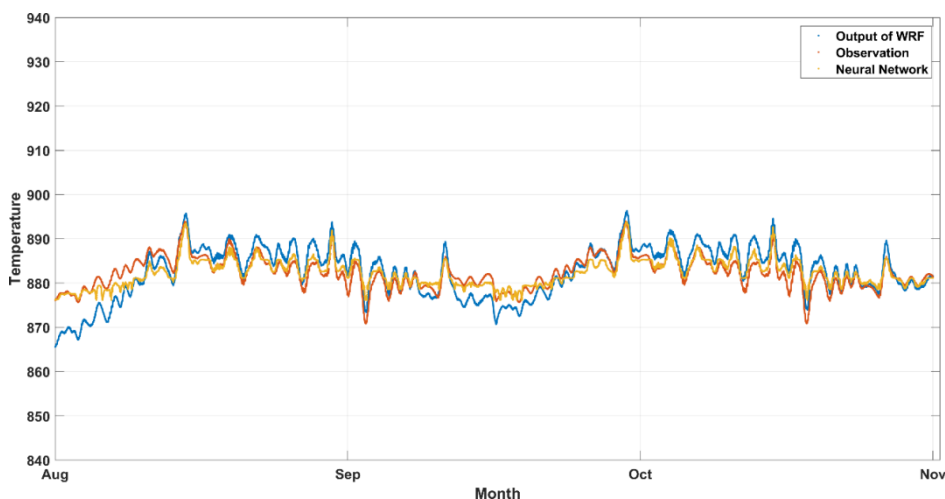
به بهبود مدل کمک کرد. جهت محاسبه تعداد لایه بهینه در این روش، مقدار لایه‌ها را از ۱ تا ۲۰ به صورت تجربی تغییر دادیم و لایه‌ای که مقدار کمینه RMSE دارد، به عنوان تعداد لایه بهینه انتخاب شد. شکل ۶ تغییرات RMSE برای تعداد لایه‌های مختلف را نشان می‌دهد. در نهایت خروجی حاصل از این روش مطابق با شکل ۷ می‌باشد. در این شکل مشاهده، خروجی حاصل از مدل و خروجی حاصل از شبکه عصبی نشان داده شده است.

تصحیح اختلاف بین مدل و مشاهده با استفاده شبکه عصبی

در این روش، اختلاف بین پارامتر دمای حاصل از مدل WRF و مشاهدات ایستگاه مهرآباد با استفاده از شبکه عصبی مصنوعی مدل‌سازی گردید. یکی از پارامترهای مورد استفاده در شبکه عصبی تعداد لایه می‌باشد. این پارامتر در سرعت و کارایی حل مسئله تأثیر به سزایی دارد. تعداد لایه بیش از نیاز موجب کاهش سرعت و افزایش زمان محاسبات می‌شود. از این رو با تعیین تعداد لایه بهینه می‌توان



شکل ۶: تغییرات جذر میانگین مربع خطاها نسبت به تعداد لایه‌ها.



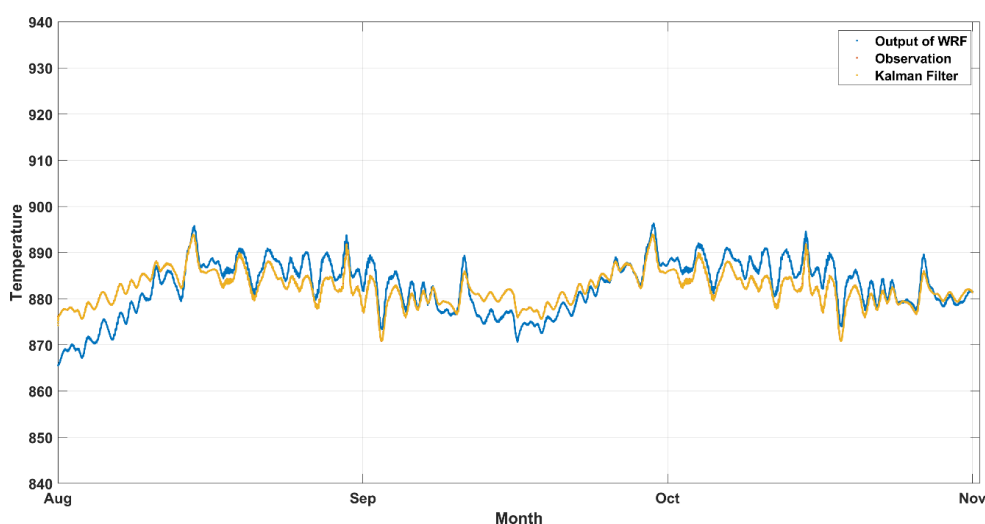
شکل ۷: مدل‌سازی اختلاف بین دمای حاصل از مدل و مشاهده با استفاده از شبکه عصبی. منحنی خروجی مدل WRF با رنگ آبی، منحنی مشاهدات با رنگ قرمز و منحنی مدل شبکه عصبی با رنگ زرد نشان داده شده است.

تصحیح اختلاف بین مدل و مشاهده با استفاده از پالایه کالمن

در این روش از پالایه کالمن جهت تصحیح بین مدل و مشاهده استفاده شد. در این روش در مرحله اول برای مدت ۷ روز پالایه کالمن اجرا گردید و مقدار خطای مشاهده و مدل برآورد گردید. سپس، در مرحله بعد از مقادیر محاسباتی مرحله قبل این پالایه برای تمامی مشاهدات به کار برده شد. خروجی حاصل از این روش مطابق با شکل ۸ می‌باشد.

ارزیابی پارامترهای آماری مختلف جهت انتخاب روش بهینه

در این قسمت، پارامترهای آماری راستی آزمایی که در بخش قبل توصیف شدند، جهت انتخاب روش بهینه مقایسه می‌شوند. در این بخش چهار پارامتر ارزیابی از جمله ME خطای میانگین، MSE خطای مطلق میانگین و جذر میانگین مربع خطاها، $RMSE$ و $Nash$ (امتیاز مهارتی میانگین خطای مطلق) به منظور ارزیابی کمی دمای مدل و دمای پس از پس پردازش مورد استفاده قرار گرفت. نتایج هر یک از روش‌ها و پارامترها ایستگاه مورد بررسی مهرآباد تهران در جدول ۱ نشان داده شده است.



شکل ۸: مدل‌سازی اختلاف بین دمای حاصل از مدل و مشاهده با استفاده از پالایه کالمن. منحنی خروجی مدل wrf با رنگ آبی و منحنی مدل پالایه کالمن با رنگ زرد نشان داده شده است.

جدول ۱- مقایسه چهار پارامتر ارزیابی از جمله ME (خطای میانگین)، MSE (خطای مطلق میانگین) و $RMSE$ (جذر میانگین مربع خطاها) و $Nash$ (امتیاز مهارتی میانگین خطای مطلق).

روش	ME	MSE	RMSE	Nash
مدل	0.2876	2.9442	3.5344	48.108
فوریه با تمام بسامد	0.0001	2.8799	3.4822	76.003
فوریه با ده بسامد	0.0001	2.8799	1.2834	87.4489
شبکه عصبی	0.0039	1.4178	1.8532	73.8312
پالایه کالمن	0.0007	0.0665	0.1019	99.9209

است؛ درحالی‌که پس از اعمال پالایه کالمن، این برآورد برای بعضی روزها اثر متفاوتی دارد و این مطلب در کاهش قابل‌ملاحظه خطای میانگین که ارزیابی را اندازه‌گیری می‌کند، مشهود است. مقایسه مقادیر امتیاز مهارتی در روش‌های مختلف که به اطلاعاتی درباره مقدار با ارزش یک پیش‌بینی نسبت به یک پیش‌بینی جایگزین دارد بیانگر تاثیر مثبت پس پردازش با پالایه کالمن را نشان می‌دهد. مقدار RMSE (جذر میانگین مربع خطاها) از عدد ۳.۵۴ به حداقل در روش کالمن و به عدد ۰.۱۰۱۹ کاهش پیدا کرد و در واقع میزان اختلاف بین دمای مدل و مشاهده از حدود ۲.۹۴ درجه سلسیوس به ۰.۱ درجه سلسیوس رسید.

منابع

1. Azadi, M., Vashani, S., Hajjam, S., 2012, Probabilistic precipitation forecast using post processing of output of ensemble forecasting system, Volume 38, Issue 3 - Pages 203-216, 10.22059/JESPHYS.2012.29126
2. Azadi, M., Arabli, P., 1388, Comparison of official forecasts and WRF model forecasts for maximum and minimum daily temperatures in the synoptic stations of their centers, 8th Numerical Weather Prediction Conference, Tehran, Iran.
3. Bracewell, R. N., (1986). The Fourier transform and its applications (Vol. 31999). New York: McGraw-Hill.
4. Bechtold, M., De Lannoy, G.J., Koster, R.D., Reichle, R.R., Mahanama, S.P. and Roose, D., 2019. Peatland Hydrology in a Global Land Surface Modeling and Data Assimilation Framework. In Geophysical Research Abstracts (Vol. 21).
5. Benali, A., Carvalho, A.C., Nunes, J.P., Carvalhais, N. and Santos, A., 2012. Estimating air surface temperature in Portugal using MODIS LST data. Remote Sensing of Environment, 124, pp.108-121.

نتیجه‌گیری

همان‌طور که در جدول ۱ مشاهده می‌شود، استفاده از پالایه کالمن خطای میانگین، خطای مطلق میانگین و جذر میانگین مربع خطاها را برای دمای تهران به میزان قابل‌توجهی کاهش داده است. میزان اختلاف بین مدل و مشاهده از حدود ۳ درجه سلسیوس به ۰.۱ درجه سلسیوس رسید و هم‌چنین امتیاز مهارتی پس از استفاده از پالایه کالمن به ۹۰ درصد دست یافت. این مقادیر نشان می‌دهد که پالایه کالمن خطای قاعده‌مند مدل را تقریباً حذف کرده است؛ به‌طوری‌که خطاها به‌صورت تصادفی حول صفر توزیع شده‌اند. پس از روش پالایه کالمن، روش فوریه با استفاده از ده بسامد مؤثر مقدار خطا نسبت به مدل WRF حدود ۲ درجه سلسیوس کاهش یافته است و افزایش ۷۰٪ امتیاز مهارتی را نسبت به مدل اصلی نشان می‌دهد.

نتیجه‌گیری

استفاده از مدل‌های پیش‌بینی عددی وضع هوا می‌تواند به پیش‌بینی هر چه صحیح‌تر وضعیت هوا کمک کند؛ اما مدل‌ها در پیش‌بینی پارامترهای گوناگون به‌ویژه پارامترهای نزدیک سطح زمین خطا دارند و به همین دلیل قادر به پیش‌بینی درست برخی پدیده‌ها مانند احتمال وقوع رعد و برق و دید نیستند. عواملی هم چون وجود نقص در پارامتر سازی فرایندهای فیزیکی متفاوت در مدل منجر به ایجاد خطا در برون‌داد مدل می‌شود. از این رو برای استفاده کاربردی برون‌داد مدل، نیاز به پس پردازش خروجی مدل امری ضروری به حساب می‌آید. در این مطالعه تاثیر داده گواری به عنوان ابزار پس پردازش بین خروجی مدل WRF و مشاهدات دمای روزانه سطح زمین مربوط به ایستگاه مهرآباد تهران به مدت ۴ ماه مورد ارزیابی قرار گرفت. فرایند داده گواری بر مبنای روش‌های تبدیل فوریه، شبکه عصبی و پالایه کالمن جهت پس پردازش پارامتر دمای روزانه استفاده گردید. تحلیل آماری نتایج نشان می‌دهد که خطای مدل با توجه به فصل برای ایستگاه متفاوت است و پیش‌بینی مدل در همه روزها اثر یکسان با مقادیر متفاوت دارد؛ یعنی در همه روزها خطا مثبت یا در همه روزها منفی

- using the kalman filter, *Weather Forecast.*, 10, 689-707.
16. Kalman, R. E., 1960, A new approach to linear filtering and prediction problems, *Trans. ASME, Ser. D, J. Basic Eng.*, 82, 34-45.
Kalman, R. E., and Bucy, R. S., 1961, new results in linear filtering and prediction problems, *Trans. ASME, Ser. D.*, 83, 95-108.
 17. Klein, William H., Billy M. Lewis, and Isadore Enger, (1959). "Objective prediction of five-day mean temperatures during winter." *Journal of Meteorology* 16.6, 672-682.
 18. Johnston, J., 1960, *Econometric Methods*. McGraw Hill Book Co . New York PP, 134-135.
 19. Marzban, C., 2003, Neural network for post processing model output, *ARPS, Mon. Wea. Rev.*, 131, 1103-1111.
 20. Mao, Q., McNider R.T., Mueller S. F., and Jaung, H. H., (1999), An optimal model output calibration algorithm suitable for objective temperature forecasting, *Weather Forecast.*, 14, 190-202.
 21. Mccollor, D. and Stull, R., 2008, Hydro meteorological accuracy enhancement via post processing of numerical weather forecasts in complex terrain, *Wea. forecasting*, 23, 131-144.
 22. Norris, P.M. and Da Silva, A.M., 2007. Assimilation of satellite cloud data into the GMAO finite-volume data assimilation system using a parameter estimation method. Part I: Motivation and algorithm description. *Journal of the Atmospheric Sciences*, 64(11), pp.3880-3895.
 23. Prechelt, L. (1998). Automatic early stopping using cross validation: quantifying the criteria. *Neural Networks*, 11, 761-767.
 6. Bevington, P. R. (1969). *Data Reduction and Error Analysis for the Physical Sciences*. McGraw-Hill Book Co. New York PP. 164-176
 7. Brown, R. G., and Hwang, P. Y., 1992, *Introduction to random signals and applied Kalman filtering*, second edition, John Wiley
 8. Diaz, H.F. and Bradley, R.S., 1997, Temperature variations during the last century at high elevation sites. In *Climatic change at high elevation sites* (pp. 21-47). Springer, Dordrecht.
 9. Galanis, G., and Anadrastakis, M., 2002, A one-dimensional Kalman filter for the correction of near surface temperature forecasts, *Meteorol. Appl.* 9, pp.473-441.
 10. Galanis, G., Louka, P., Katsafados, P., Kallos, G., Pytharoulis, I., (2006): Applications of Kalman filter based on non-linear functions to numerical weather predictions. *Ann. Geophys.* 24, 2451-2460.
 11. Glahn, H. R., and Lowry, D. A., 1972. The use of Model Output Statistics (MOS) in objective weather forecast. *J. Appl. Meteorol.*, 11, 1203-1211.
 12. Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A. and Rubin, D.B., 2013, *Bayesian data analysis*. Chapman and Hall/CRC.
 13. Hagan, M. T. and Menhaj, M. (1994). Training feedforward networks with the Marquardt algorithm. *IEEE Transactions on Neural Networks*, 5, 989-993.
 14. Hansen, L.K. and Salamon, P., 1990, Neural network ensembles. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (10), pp.993-1001
 15. Homleid, M., 1995, Diurnal corrections of short-term surface temperature forecasts

26. Welch, G. and Bishop, G., 1995, an introduction to the Kalman filter. University of North Carolina at Chapel Hill Department of Computer Science Chapel Hill, NC 27599-3175.
24. Rayner, N.A.A., Parker, D.E., Horton, E.B., Folland, C.K., Alexander, L.V., Rowell, D.P., Kent, E.C. and
25. Simmons, A.J., Jones, P.D., da Costa Bechtold, V., Beljaars, A.C.M., Kållberg, P.W., Saarinen, S., Uppala, S.M., Viterbo, P. and Wedi, N., 2004, Comparison of trends and low-frequency variability in CRU, ERA-40, and NCEP/NCAR analyses of surface air temperature. *Journal of Geophysical Research: Atmospheres*, 109(D24).