

ارائه الگوریتم جدید برای پیش‌بینی سرعت باد مبتنی بر مدل پنهان مارکوف

نوید چینی‌فروش^۱، غلامرضا لطیف‌شبهگاهی^{۲*} و مجید آزادی^۳

^۱ دانشجوی دکتری دانشگاه شهید بهشتی، تهران، ایران

^۲ استادیار دانشگاه شهید بهشتی، تهران، ایران

^۳ دانشیار پژوهشکده هواشناسی، تهران، ایران

(تاریخ دریافت: ۱۳۹۷/۱۲/۱۱، تاریخ پذیرش: ۱۳۹۸/۰۹/۱۱)

چکیده

در این مقاله ضمن ارائه مبانی نظری مدل پنهان مارکوف، ساختار مناسب آن برای مدل‌سازی سری زمانی باد پیشنهاد و اجرا شده است. مدل پیشنهادی در شناسایی رژیم‌های حاکم در سری‌های زمانی باد سطح زمین در فرودگاه امام خمینی آزمایش و برای اجرای آن از داده جمع‌آوری شده طی چهار سال متوالی استفاده شده است. ضمن ارائه آزمون ایستایی زمانی برای مدل مارکوف مرتبه اول، این آزمون برای مدل پنهان مارکوف توسعه داده شده است و نتایج آزمون ایستایی دو روش مقایسه شده‌اند. نتایج نشان می‌دهد که آزمون ایستایی زمانی روی داده سرعت باد در مدل پیشنهادی نسبت به مدل مارکوف مرتبه اول در ۷۰ تا ۸۵ درصد موارد بهبود یافته است که این افزایش ایستایی زمانی به معنی به‌دست آوردن دقت بیشتر در پیش‌بینی سرعت باد با استفاده از مدل پنهان مارکوف است. اثر تغییر تعداد رژیم‌ها از دو به سه و چهار، در ماه‌های مختلف سال بررسی و نتایج آن با نتایج اجرای مدل مارکوف مرتبه اول مقایسه شده است. نتایج نشان از این دارد که با تشخیص و تفکیک رژیم با مدل پیشنهادی، در پیش‌بینی ارائه‌شده پراکندگی احتمالات کمتر می‌شود. درنهایت، با به‌دست آوردن پیش‌بینی سرعت باد با روش پیشنهادی و همچنین روش مارکوف مرتبه اول و مقایسه با مقادیر واقعی ثبت‌شده و محاسبه ریشه مجموع مربعات خطا برای هر دو روش، نشان داده شده است که روش پیشنهادی نتایج بهتری تولید می‌کند.

واژه‌های کلیدی: مدل مارکوف پنهان، ایستایی زمانی، سرعت باد، تفکیک رژیم، پیش‌بینی

۱ مقدمه

باد بر بسیاری از شئون زندگی بشری اثرگذار است و منبع مهم انرژی محسوب می‌شود؛ بنابراین، از مهم‌ترین پدیده‌های جوی به‌شمار می‌رود. بسیاری از موارد مانند تصمیم‌گیری‌های مرتبط با کشاورزی، مدیریت آلودگی هوای شهری، حمل و نقل دریایی و هوایی، گردشگری، ورزش و نظایر آن وابستگی مستقیم به سرعت و جهت باد دارند. اثری که وزش باد بر صنایع هوایی دارد، بسیار حیاتی است؛ یکی از مهم‌ترین زمینه‌های متأثر از وزش باد، پرواز هواپیماهاست. در بین این همه، آنچه در سال‌های اخیر بیشتر موجب توجه به باد، اندازه‌گیری و پیش‌بینی آن شده است، مسئله تولید انرژی از باد و بهره‌برداری از نیروگاه‌های بادی است. روش متداول در پیش‌بینی هوا، استفاده از مدل‌های پیش‌بینی عددی وضع هوا است که بر مبنای روابط فیزیکی حاکم بر جو نزدیک زمین بنا نهاده شده‌اند و پیش‌بینی را با استفاده از داده‌های نمونه‌برداری شده جوی و با در نظر گرفتن ناهمواری‌های سطح زمین انجام می‌دهند. هواشناسان مدل‌های پیش‌بینی عددی هوا را که در پیش‌بینی‌های بلندمدت دقت خوبی دارند، برای پیش‌بینی‌های بزرگ‌مقیاس وضع هوا ساخته‌اند (لی و همکاران، ۲۰۰۹). در پیش‌بینی‌های کوتاه‌مدت و خیلی کوتاه‌مدت، مدل‌های پیش‌بینی عددی وضع هوا، به لحاظ حجم زیاد محاسبات کارایی مناسبی ندارند و مدل‌های مبتنی بر داده‌های آماری ترجیح داده می‌شوند (ونگ و همکاران، ۲۰۱۱).

از آنجا که پیش‌بینی‌های خیلی کوتاه‌مدت از نیازهای ضروری نشست و برخاست هواپیماها و همچنین نیروگاه‌های بادی است، تلاش چشمگیری از سوی پژوهشگران به این بخش معطوف شده است. در این روش، داشتن داده‌های آماری و توانایی اندازه‌گیری برخط نقش مهمی دارد. این تحقیق نیز در راستای همین نیاز و با هدف ارائه روشی نوین و اجرای آن روی

داده‌های محلی شکل گرفته است.

به‌طور کلی، اقدامات درخور توجهی که تاکنون برای ایجاد مدل‌های کوتاه‌مدت سرعت باد انجام شده است را می‌توان به دو دسته به‌کارگیری مدل توزیع احتمال و به‌کارگیری مدل تحلیل سری زمانی تقسیم کرد (کارتا و همکاران، ۲۰۰۹). در مدل‌های توزیع احتمال، فرض می‌شود سرعت باد از توزیع احتمال خاص مانند ویبال، نرمال یا رابلی پیروی می‌کند. در این نوع روش‌ها، پارامترهای تابع توزیع احتمال، معمولاً براساس داده تاریخیچه‌ای باد تخمین زده می‌شوند. از آنجا که این داده‌های تاریخیچه‌ای همواره به یک نوع توزیع منجر نمی‌شوند، در نظر گرفتن یک توزیع خاص ممکن است در این شرایط خطای زیادی ایجاد کند.

دو روش مهم در دسته مدل‌های تحلیل سری‌های زمانی، مدل (Auto Regressive Moving Average) و مدل مارکوف هستند. در مدل ARMA، باد به‌عنوان مجموع وزن‌دار فرایند نوفه سفید در نظر گرفته می‌شود که می‌تواند ویژگی‌های خودهمبستگی (autocorrelation) متوالی را حفظ کند (بیلیتون و همکاران، ۱۹۹۶). با توجه به ایستا نبودن سرعت باد، قبل از مدل‌سازی ARMA، داده‌ها با استفاده از روش نرمال‌سازی میانگین-واریانس، پیش‌پردازش و با یک توزیع نرمال تخمین زده می‌شوند (کالوی، ۲۰۱۰). در مواردی که داده‌ها توزیع نرمال ندارند، ممکن است خطای زیادی روی توزیع احتمال اتفاق بیفتد که به محدود شدن کاربرد این روش منجر می‌شود.

زنجیره‌های مارکوف (Markov Chains, MCs) مدل‌های آماری هستند که در بسیاری از کاربردها برای مدل‌سازی سرعت باد استفاده می‌شوند. ویژگی اصلی این مدل‌ها، توانایی آنها در به‌کارگیری هر دو مشخصه آماری و زمانی (temporal) سرعت باد است (تاگلیافری و همکاران، ۲۰۱۶). MCها با وجود سادگی و با توجه به

اینکه بر مبنای وابستگی توزیع احتمال باد در گام بعدی به وضعیت فعلی بنا نهاده شده‌اند، قادرند مشخصات وابستگی زمانی باد را مدل کنند درحالی که مدل‌های دیگر مانند ARMA قادر به دریافت این وابستگی توزیع احتمال نیستند. کاربرد MCها در مطالعات اخیر نشان داده است که سادگی این مدل‌ها، آنها را به ابزاری ارزشمند در مدل‌سازی تبدیل کرده است. با این حال، MCها قادر به دریافت مشخصات باد در بسامدهای بالا نیستند (تاگلیافری و همکاران، ۲۰۱۶). تحقیقات اخیر درباره مدل‌های مارکوف باد، روی اثر مرتبه مدل مارکوف (شمشاد و همکاران، ۲۰۰۵) و تعداد حالت‌ها (هکاگلو و همکاران، ۲۰۰۸) و دقت مدل تمرکز کرده‌اند. شمشاد و همکاران (۲۰۰۵) از MCهای مرتبه دوم و سوم استفاده کردند. در این مدل‌ها توزیع احتمال برای حالت بعدی تنها به حالت فعلی بستگی ندارد، بلکه به حالات گذشته (به ترتیب یک و دو گام زمانی قبل) نیز وابسته است. از آنجاکه MC با مرتبه بالاتر به محاسبات بیشتری احتیاج دارد، مزیت اصلی استفاده از MCها از دست می‌رود.

تاگلیافری و همکاران (۲۰۱۶) برای بهبود دقت MC و مدل‌سازی بهتر همبستگی زمانی در گام‌های زمانی کوچک، بدون افزایش بیش از حد زمان محاسبات، استفاده از زنجیره مارکوف تودرتو (Nested Markov Chain, NMC) را برای مدل‌سازی باد پیشنهاد کردند. NMC را می‌توان یک MC توسعه‌یافته در نظر گرفت که هر حالت در آن، خود یک فرایند MC تصادفی مستقل است و سری زمانی با استفاده از یک MC داخلی (کمکی) تولید می‌شود. البته در این فرایند می‌توان از مدل‌های غیرمارکوف نیز برای تولید سری زمانی لایه داخلی استفاده کرد؛ برای مثال، مدل MC لایه داخلی می‌تواند با حالت AR (Auto Regressive) جایگزین شود که در این صورت، مدل، MSAR (Markov Switching Auto Regressive) خواهد بود (آلیت و

منبت، ۲۰۱۲ و پینسن و مدسن، ۲۰۱۲). در برخی از پژوهش‌ها برای بهبود دقت و خودهمبستگی MCهای استاندارد، از مدل شبه‌مارکوف استفاده شده است. ویژگی این مدل آن است که گام زمانی ثابت نیست بلکه یک متغیر تصادفی است که می‌تواند هرگونه توزیعی داشته باشد. مدت زمانی که در یک حالت می‌گذرد، روی توزیع احتمال گذر تأثیر می‌گذارد (دآمیگو و همکاران، ۲۰۱۴). دآمیگو و همکاران (۲۰۱۲) نشان دادند که چگونه فرایندهای شبه مارکوف با حافظه، خودهمبستگی بهتری نسبت به MCهای متداول ارائه می‌کنند. این امر به دلیل توانایی این مدل برای حفظ حافظه گذرهای گذشته با یک فرایند تصادفی کمکی است.

بدیهی است که باد متأثر از دیگر میدان‌های هواشناسی مانند دمای هوا، رطوبت نسبی و وضع هوا است. این پارامترها در بازه زمانی در طول یک روز و نیز به‌طور فصلی تغییر می‌کنند؛ در نتیجه، تغییرات باد، نظم دوره‌ای روزانه (سومالین و همکاران، ۲۰۱۲) و شولز و همکاران، (۲۰۱۴) و فصلی (کاراتپ و کرسکادن، ۲۰۱۳) دارد. شولز و همکاران (۲۰۱۴) الگوی روزانه باد را با استفاده از یک مدل متغیر با زمان گردشی براساس فضای حالت سه‌بعدی مدل‌سازی کردند. کای و همکاران (۲۰۱۷) یک مدل سرعت باد بر مبنای زنجیره مارکوف غیرهمگن (Non-Homogeneous Markov Chain, NHMC) ارائه کردند که هدف آن، توسعه و ارائه روشی دقیق‌تر برای مدل‌سازی سری‌های زمانی سرعت باد با در نظر گرفتن تغییرات فصلی و روزانه آن است.

مدل پنهان مارکوف (Hidden Markov Model, HMM) در مدل‌سازی پدیده‌هایی که رفتار مجموعه، ناشی از براین رفتار بخش‌هایی با رفتار ناهمگون باشد، کاربرد فراوان دارد. از جمله زمینه‌های مناسب برای به‌کارگیری این مدل، سری زمانی باد است. آلیت و منبت

باد مبتنی بر مدل پنهان مارکوف ارائه و روش اعتبارسنجی بر مبنای آزمون ایستایی زمانی تدوین و تشریح شده است. پس از آن، با معرفی الگوریتمی مناسب، روش با استفاده از داده‌های واقعی اجرا و نتایج با نتایج یک مدل مارکوف مرتبه اول مقایسه شده است. سپس سازگاری نتایج با نتایج چشمداشتی براساس گل‌باد محل و تجارب کارشناسان هواشناسی بررسی شده است. در پایان، مختصری از مدل مارکوف مرتبه اول و نحوه ارزیابی ایستایی آن، مبانی نظری مدل‌های مخلوط و بر مبنای آن مدل پنهان مارکوف در پیوست پ تشریح شده است.

۲ روش تحقیق

۲-۱ الگوریتم پیشنهادی برای اجرای مدل HMM و ارزیابی ایستایی آن

مدل چندجمله‌ای HMM- دسته‌ای از مدل‌های مبتنی بر مارکوف است که در مدل‌سازی سری شمارش‌های محدود استفاده می‌شود و مبانی آن در پیوست پ تشریح شده است. با فرض حالت خاصی از این مدل که در آن تعداد آزمایش‌ها برابر یک باشد ($n_t = 1$) و با قدری ساده‌سازی، بردار مقادیر مشاهده شده X ، احتمالات وابسته به حالت‌ها $p_i(x)$ و درست‌نمایی مطابق زیر به دست خواهد آمد. از آنجا که مجموع احتمالات مشاهده هر یک از مقادیر ممکن برابر یک است ($\sum_{k=1}^q x_{tk} = 1$) و آزمایش یک‌بار انجام می‌شود؛ بردار X با ابعاد q یک درایه برابر ۱ دارد و بقیه درایه‌ها صفر خواهند بود.

با فرض $X = (\underbrace{\cdot, \dots, \cdot}_{j-1}, \underbrace{\cdot, \dots, \cdot}_{q-j})$ در این حالت $p_i(x) = \pi_{ij}$ است:

$$P(X) = \text{diag}(\pi_{1j}, \dots, \pi_{mj}) \quad (1)$$

به عبارت دیگر، $P(X)$ ، با $\pi(j)$ برابر است و با توجه به رابطه (پ-۸) درست‌نمایی گروه‌های مشاهده شده

z_1, z_2, \dots, z_T در زمان‌های $T, \dots, 2, 1$ با رابطه (۲) ارائه

(۲۰۱۲)، از مدل‌های MS-AR برای توصیف سری زمانی باد استفاده کردند. در این روش، از چندین مدل خودبرگشتی برای توصیف ارزیابی زمانی سرعت باد استفاده شده است و انتخاب بین این مدل‌های مختلف با HMM کنترل می‌شود که ارائه‌کننده نوع شرایط جوی است. آپلیت و همکاران (۲۰۱۵) مدل‌های مختلف MS-AR را برای سری زمانی دومتغیره ارائه کردند. ویژگی این مدل‌ها، همگن نبودن HMM و وابسته بودن ارزیابی آنها به شرایط باد گذشته است. روش‌های پیچیده‌تر و دقیق‌تری هم توسعه یافته است که ممکن است برای مثال از مقادیر پارامترهای دیگر مانند فشار و دما استفاده کنند (بیتر و همکاران، ۲۰۱۴)، اما این روش‌ها به محاسبات بیشتری نیاز دارند و هنگامی که مقدار محدودی از داده‌ها در دسترس است، کاربرد چندانی ندارند.

در زنجیره مارکوف که روشی شناخته شده در پیش‌بینی باد است، محدوده‌های مشخصی از سرعت باد به عنوان حالت‌های سامانه در نظر گرفته می‌شوند و با استفاده از داده‌های آموزشی، احتمال گذر از هر حالت به حالت‌های دیگر محاسبه می‌شود. این مقادیر محاسبه شده برای پیش‌بینی سرعت باد در زمان‌های بعدی، یعنی پیش بینی احتمال گذر از هر حالت به حالت دیگر استفاده می‌شوند. در این روش فرض بر این است که احتمالات گذر در نمونه آموزشی با نمونه‌های عملیاتی یکسان هستند و به عبارتی، ماتریس احتمال گذر با زمان تغییر نمی‌کند. یک روش ساده برای تحقیق این مسئله این است که ماتریس احتمال گذر برای نمونه‌های عملیاتی به طور مستقل محاسبه و با ماتریس احتمال گذر به دست آمده از نمونه‌های آموزشی مقایسه شود. برای مقایسه ماتریس‌های احتمال گذر، استفاده از آزمون ایستایی زمانی، روشی مؤثر و کارا است که می‌تواند برای بررسی میزان اثربخشی روش‌ها استفاده شود.

در این مقاله، ابتدا روشی جدید برای پیش‌بینی سرعت

می‌شود:

$$L_T = \delta\pi(j_1) \Gamma\pi(j_2) \Gamma \dots \pi(j_T) \quad (2)$$

اگر زنجیره مارکوف ایستا باشد، احتمال مشاهده گروه دیدبانی l در زمان $t + 1$ ، در صورتی که گروه k در زمان t مشاهده شده باشد، برابر است با:

$$\frac{\delta\pi(k)\Gamma\pi(l)}{\delta\pi(k)} \quad (3)$$

به‌طور مشابه احتمال مشاهده l در زمان $t + 1$ در زمان t و در زمان $t - 1$ برابر است با:

$$\frac{\delta\pi(j)\Gamma\pi(k)\Gamma\pi(l)}{\delta\pi(j)\Gamma\pi(k)} \quad (4)$$

دو رابطه (۳) و (۴) می‌توانند برای محاسبه توزیع‌های پیش‌بینی استفاده شوند.

این مدل ساده‌شده، مبنای روش پیشنهادی پیش‌بینی سرعت باد است. ابتدا مدل چندجمله‌ای-HMM براساس داده سری زمانی جهت باد اجرا و پارامترهای آن تخمین زده می‌شود. با در نظر گرفتن شانزده جهت اصلی و فرعی سرعت باد، یعنی N، NNE، NE، ...، NNW به‌عنوان حالت‌ها یا همان نتایج ممکن برای آزمایش چندجمله‌ای، مدلی تشکیل می‌شود که پارامترهای آن عبارت‌اند از: الف) ماتریس گذر حالت $\Gamma_{m \times m}$ که $m(m-1)$ پارامتر آزاد دارد و هریک از m حالت آن (که پس از این یک رژیم نامیده می‌شود) با یک توزیع چندجمله‌ای بیان می‌شود؛ ب) هریک از چندجمله‌ای‌ها که تعداد $q = 16$ پارامتر، میزان احتمال (نامعلوم) است $(\pi_{i1}, \pi_{i2}, \dots, \pi_{iq}, i = 1, 2, \dots, m)$.

با حل مسئله پیشینه درست‌نمایی برای مدل طرح‌شده مطابق رابطه (۲) براساس الگوریتم ویتربی (Viterbi) (لو، ۱۹۹۵)، علاوه بر احتمال گذر از هریک از رژیم‌ها به رژیم دیگر (درایه‌های ماتریس $\Gamma_{m \times m}$)، احتمال وقوع هریک از حالت‌ها (هریک از شانزده جهت اصلی و فرعی سرعت باد) در هر رژیم به‌دست می‌آید؛ برای نمونه، هنگامی که

$m = 2$ باشد، یعنی دو رژیم وجود داشته باشد، برای هریک از جهات شانزده‌گانه یک مقدار احتمال متناظر با رژیم اول و یک مقدار احتمال متناظر با رژیم دوم به‌دست می‌آید. جهت‌های شانزده‌گانه را می‌توان به دو دسته تقسیم کرد: جهت‌هایی که در رژیم اول محتمل‌ترند و جهت‌هایی که در رژیم دوم محتمل‌ترند. همین‌منا در تعداد رژیم‌های بیشتر برای تفکیک کل سری زمانی باد در نظر گرفته می‌شود. اجرای مدل مارکوف مرتبه اول روی مقدار سرعت هریک از سری‌های زمانی تفکیک‌شده، امکان پیش‌بینی سرعت باد در هر رژیم را به‌صورت جداگانه فراهم می‌کند. نمودار گردش‌شی شکل ۱ به‌ارائه و بیان بهتر روش پیشنهادی کمک خواهد کرد. به‌طور خلاصه، روش پیشنهادی برای پیش‌بینی سرعت باد در مسئله‌ای با تفکیک دو رژیم شامل مراحل زیر است:

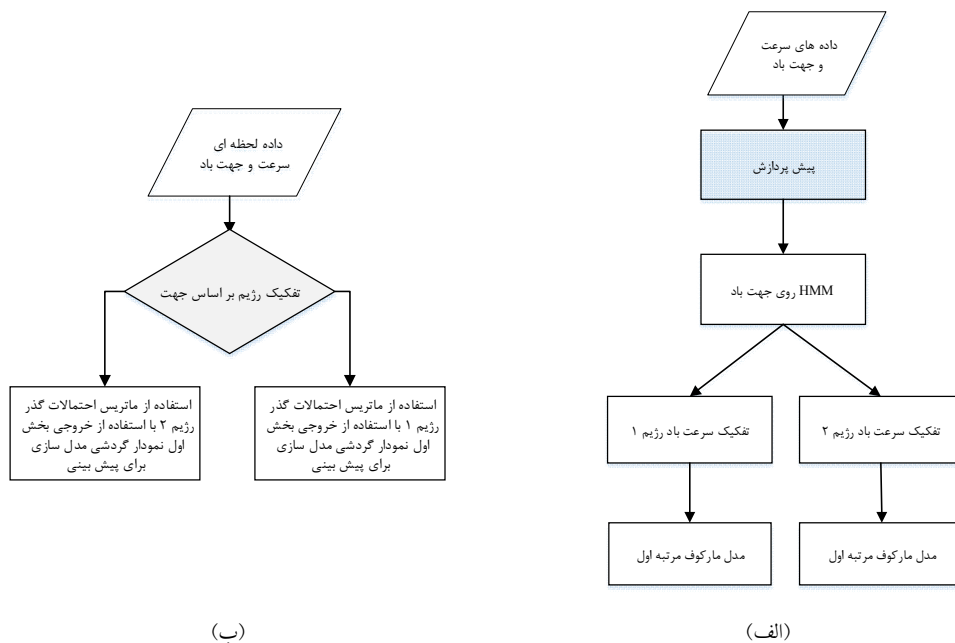
- انتخاب سری زمانی سرعت باد و جهت باد در یک بازه معین در نقطه آزمون به‌عنوان داده‌های آموزشی برای ساختن مدل و تخمین پارامترهای آن؛

- حل مسئله پیشینه درست‌نمایی برای مدل چندجمله‌ای-HMM براساس سری زمانی جهت باد (یعنی تعیین پارامترهای مدل به‌گونه‌ای که رابطه (۲) پیشینه شود. این پارامترها ماتریس احتمال گذر بین دو رژیم و همچنین احتمال متناظر با هریک از جهت‌های شانزده‌گانه باد در هر رژیم است)؛

- تفکیک داده سری زمانی باد به دو دسته با مقایسه کردن احتمال متناظر با هریک از جهت‌های شانزده‌گانه (هر جهتی که احتمال وقوعش در رژیم اول بیشتر باشد، عضو دسته اول و هر جهتی که احتمال وقوعش در رژیم دوم بیشتر باشد، عضو دسته دوم می‌شود)؛

- محاسبه ماتریس احتمال گذر طبق رابطه (پ-۱) و (پ-۲) برای هر دسته متناظر با هر رژیم؛

- اندازه‌گیری برخط سرعت و جهت باد و ملاحظه اینکه جهت باد در کدام یک از رژیم‌های تفکیک‌شده



شکل ۱. نمودار گردش روش پیشنهادی برای اجرای مدل HMM (الف) بخش اول (مدل‌سازی) (ب) بخش دوم (پیش‌بینی).

فوق محتمل‌تر است و انتخاب آن رژیم به‌عنوان رژیم حاکم؛

• پیش‌بینی سرعت باد براساس سرعت حال حاضر و ماتریس احتمال گذر متناظر با رژیم حاکم مطابق رابطه (پ-۴).

مراحل شش‌گانه ذکرشده، برای تعداد رژیم‌های بیشتر نیز اجراشدنی است. در روش‌های مبتنی بر مارکوف تشابه بیشتر بین ماتریس احتمالات گذر به‌دست‌آمده از داده آموزشی و داده آزمون، به معنی عملکرد بهتر مدل است. یکی از کاربردی‌ترین ابزارهای بررسی مشابهت دو ماتریس احتمالات گذر، آزمون ایستایی زمانی است که مفهوم آن در پیوست پ توضیح داده شده است.

با توجه به پیکربندی ارائه‌شده در این روش، می‌توان مقدار ایستایی زمانی را برای کل مدل محاسبه کرد. اگر دو بازه انتخابی وجود داشته باشد $(T = 2)$ ، رابطه (پ-۵) به‌صورت رابطه (۵) ساده‌سازی می‌شود و درجه آزادی

$$k(k-1) \text{ می‌شود:} \quad (۵)$$

$$\beta = \sum_{i,j}^k (n_{i,j}^{(1)} \ln \left(\frac{\gamma_{i,j}^{(1)}}{\gamma_{i,j}} \right) + n_{i,j}^{(2)} \ln \left(\frac{\gamma_{i,j}^{(2)}}{\gamma_{i,j}} \right))$$

با محاسبه مقدار β_1 برای داده‌های رژیم اول و β_2 برای داده‌های رژیم دوم مطابق رابطه (۵)، مقدار $\beta_{Extended}$ از رابطه توسعه‌یافته (۶) محاسبه می‌شود:

$$\beta_{Extended} = \frac{n^{(1)}\beta_1 + n^{(2)}\beta_2}{n^{(1)} + n^{(2)}} \quad (۶)$$

که $n^{(r)}$ تعداد عناصر موجود در گروه تفکیک‌شده براساس رژیم r است و $r = 1, 2$.

برای مقایسه روش پیشنهادی با روش مارکوف مرتبه اول، لازم است مقدار β برای مدل مارکوف مرتبه اول طبق رابطه (۵) با تمام داده سری زمانی باد محاسبه و با مقدار $\beta_{Extended}$ مقایسه شود. مقدار کمتر، به مفهوم ارائه پیش‌بینی مناسب‌تر است.

شده‌اند؛ برای مثال، نمونه کد ۱ به معنی جهت N است (محدوده بین ۳۴۸/۷۵ درجه تا ۱۱/۲۵+ درجه از سمتی که شامل زاویه صفر می‌شود).

پارامترهای HMM دو حالتی شامل ماتریس احتمال گذر بین رژیم‌ها و شانزده احتمال در هریک از رژیم‌ها، برای داده دو ماه مارس و ژوئیه طی چهار سال از ۲۰۱۳ تا ۲۰۱۶ محاسبه شده است و نتایج در جدول‌های ۱ و ۲ ارائه شده است. همان‌طور که در جدول ۱ ملاحظه می‌شود، مقادیر ماتریس‌های گذر حالت برای دو ماه همنام در سال‌های متوالی مشابه خوبی با هم دارند. همین وضعیت برای مقادیر احتمال به‌دست‌آمده برای هریک از جهت‌های شانزده‌گانه نیز وجود دارد که نشان‌دهنده تشابه تقریبی رفتار باد و توانایی پذیرفتنی مدل در شناسایی آن است.

برای مقایسه آسان‌تر مقادیر احتمال متناظر با هر جهت در رژیم‌های تفکیک‌شده در جدول ۲ برای دو سال ۲۰۱۴ و ۲۰۱۵، نتایج با نمودار میله‌ای در شکل ۲ ارائه شده است. البته همان‌طور که انتظار می‌رود، تشابه مقادیر احتمال در دو ماه متناظر به‌صورت کامل برقرار نیست، اما نسبت به دو ماه غیرمتناظر به‌طور چشمگیری بیشتر است.

بر اساس جدول ۲ و نتایج ماه مارس ۲۰۱۳، تعلق هریک از جهت‌های شانزده‌گانه به هریک از رژیم‌های باد به روش گفته‌شده تعیین و نتایج مطابق رابطه (۷) ارائه شد. در جدول ۲ در هر ستون، احتمال داشتن زاویه مربوطه، به‌صورت کسری از ۱۰۰۰ بیان شده است؛ برای مثال، در سطر اول این جدول، عدد صفر و ۴۰ در زیر ستون ماه مارس ۲۰۱۳ نشان می‌دهد که وزش باد در رژیم ۱ به احتمال صفر می‌تواند در جهت شمال باشد (به‌عبارتی، در رژیم ۱ باد شمالی وجود ندارد)، اما در رژیم ۲ وزش باد به احتمال ۴ درصد شمالی است.

$$R_1 = \{5, 6, 7, 8, 9, 10, 11, 12, 13, 14\}, \quad (7)$$

$$R_2 = \{1, 2, 3, 4, 15, 16\},$$

۲-۲ اجرای الگوریتم پیشنهادی مبتنی بر مدل HMM و نتایج آن

داده استفاده‌شده در این مقاله، مقادیر اندازه‌گیری‌شده سمت و سرعت باد با بادسنج مستقر در مجاورت باند ۲۹ فرودگاه امام خمینی تهران (IKIA) است. جهت و سرعت باد در این ایستگاه با بادسنج فراصوت اندازه‌گیری شده است و داده‌ها با گام زمانی ۳۰ ثانیه ذخیره شده‌اند. داده آماری برای چهار سال (۲۰۱۳ تا ۲۰۱۶) جمع‌آوری شده و موجود است. سنجنده‌های فراصوت به لحاظ آستانه اندازه‌گیری بسیار پایین و پایداری و دقت خوب و همچنین توانایی عملکرد در شرایط آب و هوایی سخت و سرما، داده‌های معتبر و مطمئنی ارائه می‌دهند (چینی فروش و شبانگهی، ۲۰۱۸) که باعث شده است خلاء آماری بسیار کمی در داده‌های جمع‌آوری‌شده وجود داشته باشد. گام زمانی ۳۰ ثانیه مبنای تشکیل مدل بوده است. در اجرای مدل به‌جز در مواردی که گام زمانی ۱۵ دقیقه ذکر شده، در بقیه موارد، مدل برای پیش‌بینی ۳۰ ثانیه‌ای طراحی شده است.

۲-۲-۱ تشریح مراحل اجرای مدل

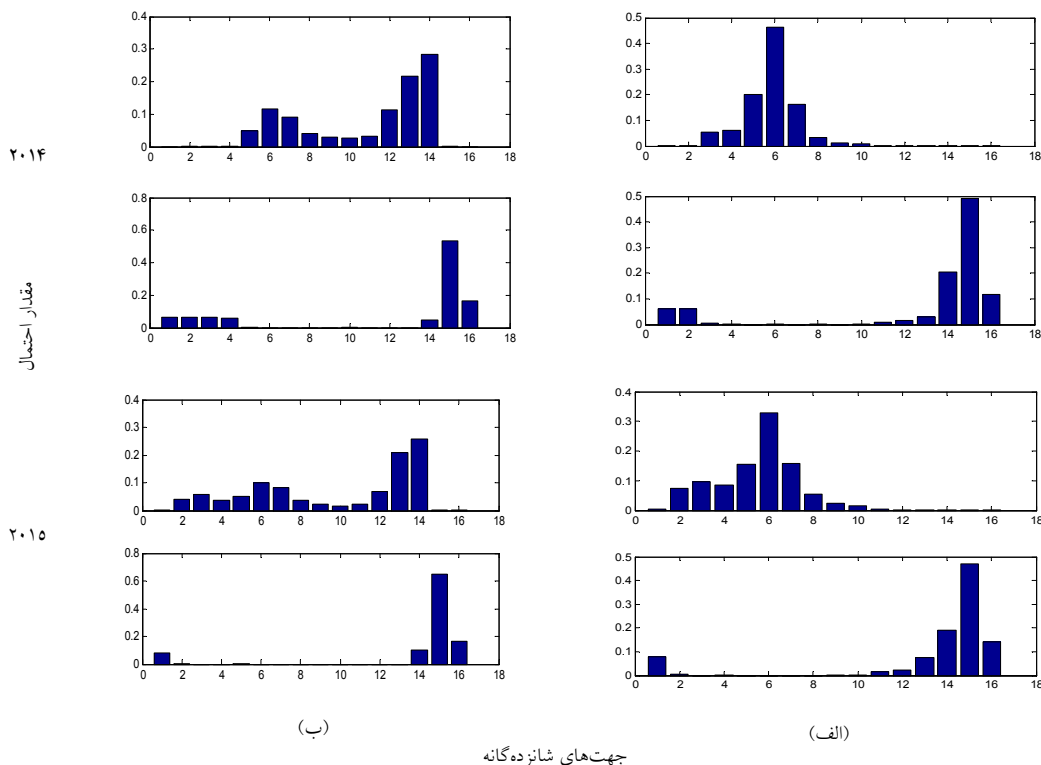
همانند سایر روش‌های مبتنی بر داده، در اینجا نیز یکی از اولین گام‌ها، بررسی داده جمع‌آوری‌شده از لحاظ کامل بودن و توالی مناسب است. سری زمانی داده موجود، بررسی و در صورت جاافتادگی برخی مقادیر، بازیابی آنها با درون‌یابی انجام شده است. در کل بازه جمع‌آوری داده، نسبت تعداد داده‌های جاافتاده به کل داده‌ها کمتر از یک‌دهم درصد است. علاوه‌براین، در سرعت باد صفر، داده جهت باد نامعتبر تلقی و از مجموعه حذف شده است. این مرحله از کار با عنوان پیش‌پردازش در شکل ۱ مشخص شده است. در مرحله پس از پیش‌پردازش، زوایای صفر تا ۳۶۰ درجه به روش گفته‌شده به شانزده جهت اصلی و فرعی تقسیم و از ۱ تا ۱۶ کدگذاری

جدول ۱. مقادیر ماتریس احتمال گذر برای دو ماه متناظر از چهار سال متوالی.

	مارس	ژوئیه
سال ۲۰۱۳	$\Gamma = \begin{bmatrix} 0/9883 & 0/0117 \\ 0/0281 & 0/9719 \end{bmatrix}$	$\Gamma = \begin{bmatrix} 0/9935 & 0/0065 \\ 0/0085 & 0/9915 \end{bmatrix}$
سال ۲۰۱۴	$\Gamma = \begin{bmatrix} 0/9844 & 0/0156 \\ 0/0291 & 0/9709 \end{bmatrix}$	$\Gamma = \begin{bmatrix} 0/9950 & 0/0044 \\ 0/0055 & 0/9945 \end{bmatrix}$
سال ۲۰۱۵	$\Gamma = \begin{bmatrix} 0/9896 & 0/0104 \\ 0/0217 & 0/9783 \end{bmatrix}$	$\Gamma = \begin{bmatrix} 0/9930 & 0/0070 \\ 0/0095 & 0/9905 \end{bmatrix}$
سال ۲۰۱۶	$\Gamma = \begin{bmatrix} 0/9861 & 0/0139 \\ 0/0202 & 0/9798 \end{bmatrix}$	$\Gamma = \begin{bmatrix} 0/9941 & 0/0059 \\ 0/0066 & 0/9934 \end{bmatrix}$

جدول ۲. مقادیر احتمال تخمین زده شده برای HMM دو حالت در دو گروه از ماه‌های همنام در چهار سال متوالی.

	مارس ۲۰۱۳		مارس ۲۰۱۴		مارس ۲۰۱۵		مارس ۲۰۱۶	
N	۰	۴۰	۰	۶۴	۰	۷۹	۰	۵۹
NNE	۰	۳۶	۰	۶۶	۳۹	۶	۵۰	۱۶
NE	۰	۱۴	۰	۶۴	۵۸	۰	۶۰	۱۰
ENE	۲۱	۸۸	۱	۵۸	۳۶	۰	۳۴	۰
E	۴۸	۳	۴۸	۳	۵۱	۰	۴۰	۰
ESE	۱۲۷	۰	۱۱۷	۰	۱۰۰	۰	۱۰۹	۰
SE	۸۱	۰	۹۱	۰	۸۲	۰	۷۳	۰
SSE	۳۰	۰	۴۰	۰	۳۵	۰	۳۷	۰
S	۲۹	۰	۲۹	۰	۲۲	۰	۲۰	۰
SSW	۳۶	۰	۲۶	۰	۱۷	۰	۲۱	۰
SW	۳۰	۰	۳۳	۰	۲۱	۰	۱۷	۰
WSW	۹۳	۰	۱۱۳	۰	۶۸	۰	۷۸	۰
W	۲۱۹	۰	۲۱۶	۰	۲۰۷	۰	۱۹۵	۰
WNW	۲۸۴	۶۸	۲۸۰	۴۸	۲۵۷	۱۰۱	۲۴۷	۱۲۱
NW	۲	۵۹۴	۶	۵۴۴	۱	۶۴۷	۳	۶۷۷
NNW	۰	۱۶۶	۰	۱۶۲	۰	۱۶۵	۰	۱۵۵
	ژوئیه ۲۰۱۳		ژوئیه ۲۰۱۴		ژوئیه ۲۰۱۵		ژوئیه ۲۰۱۶	
N	۱	۹۳	۰	۶۲	۳	۷۷	۵	۷۳
NNE	۳	۳۱	۲	۶۲	۷۴	۴	۲	۲۸
NE	۵۰	۲	۵۳	۵	۹۸	۰	۴۸	۰
ENE	۶۱	۰	۶۲	۰	۸۵	۰	۵۵	۰
E	۱۷۷	۰	۲۰۰	۰	۱۵۵	۰	۱۶۵	۰
ESE	۳۹۵	۰	۴۶۴	۰	۳۲۹	۰	۳۸۰	۰
SE	۱۷۰	۰	۱۶۱	۰	۱۵۳	۰	۱۶۸	۰
SSE	۶۵	۰	۳۳	۰	۵۱	۰	۸۱	۰
S	۵۳	۰	۱۳	۰	۳۲	۰	۷۶	۰
SSW	۱۲	۲	۱۰	۲	۱۶	۲	۱۱	۲
SW	۱۳	۱۹	۱	۱۰	۴	۱۴	۹	۲۴
WSW	۰	۱۹	۰	۱۵	۰	۲۴	۰	۱۴
W	۰	۳۸	۰	۲۸	۰	۷۴	۰	۵۴
WNW	۰	۱۸۶	۰	۲۰۶	۰	۱۹۱	۰	۲۰۱
NW	۰	۴۸۰	۰	۴۹۲	۰	۴۷۰	۰	۴۸۰
NNW	۰	۱۲۹	۰	۱۱۷	۰	۱۴۳	۰	۱۲۳



شکل ۲. مقادیر احتمال تخمین زده شده برای HMM دو حالت در دو گروه از ماه‌های همنام در دو سال متوالی (الف) از بالا به پایین، نمودار اول و دوم به ترتیب مقادیر احتمال رژیم ۱ و ۲ در ماه ژوئیه ۲۰۱۴ و نمودار سوم و چهارم همین مقادیر برای ماه ژوئیه ۲۰۱۵ است. (ب) از بالا به پایین، نمودار اول و دوم به ترتیب مقادیر احتمال رژیم ۱ و ۲ در ماه مارس ۲۰۱۴ و نمودار سوم و چهارم همین مقادیر برای ماه مارس ۲۰۱۵ است.

۳ نتایج اجرای مدل

برای بررسی کارایی روش پیشنهادی، داده باد چهار سال متوالی ایستگاه بادسنجی فرودگاه امام خمینی پردازش اولیه شد. داده جمع‌آوری شده با روش پیشنهادی پردازش شد و میزان ایستایی زمانی بین هر دو ماه همنام در دو سال متوالی و غیرمتوالی (۷۲ حالت) به دست آمد. با به کارگیری روش پیشنهادی با سه مقدار مختلف برای تعداد رژیم‌ها، یعنی $m = 2, 3, 4$ (دو، سه و چهار رژیم) شبیه‌سازی اجرا و نتایج ثبت شده است. میزان ایستایی زمانی برای داده‌های توصیف شده با روش مارکوف مرتبه اول نیز محاسبه و برای مقایسه ارائه شده است. جدول ۳ مقدار ایستایی زمانی را برای دو ماه همنام از

که R_1 مجموعه حالات رژیم ۱ و R_2 مجموعه حالات رژیم ۲ است.

در مثال فوق با تفکیک داده‌های سرعت و جهت باد ماه مارس ۲۰۱۳ بر مبنای رژیم‌های شناسایی شده، دو سری زمانی داده باد ایجاد می‌شود که با اجرای مدل مارکوف مرتبه اول مطابق رابطه (پ-۱) برای هریک از سری‌های زمانی، ماتریس احتمالات گذر به دست می‌آید. از این دو ماتریس احتمالات گذر می‌توان برای پیش‌بینی سرعت باد در ماه مارس سال‌های بعدی استفاده کرد؛ یعنی با مشاهده جهت وزش باد، نوع رژیم حاکم بر باد تشخیص داده می‌شود و با انتخاب ماتریس احتمالات گذر مربوطه طبق رابطه (پ-۴)، سرعت باد در لحظه بعد پیش‌بینی می‌شود.

جدول ۳. مقادیر ایستایی زمانی مدل مارکوف مرتبه اول (بدون تفکیک رژیم) برای دو ماه همنام از دو سال متفاوت (دوازده ماه همنام از چهار سال).

ماه/سال	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲
۲۰۱۳-۲۰۱۴	۱۴۴	۱۵۲	۸۳	۴۸	۲۴۳	۱۰۷	۱۰۷	۲۴۵	۴۰	۱۶۸	۱۲۶	۵
۲۰۱۳-۲۰۱۵	۱۸۷	۶۱	۸۷	۹۱	۲۵۹	۱۴۳	۳۸۱	۲۳۱	۲۴۵	۱۴۳	۱۸۴	۱۲۴
۲۰۱۳-۲۰۱۶	۲۱۵	۹۵	۱۴۷	۹۹	۹۹	۱۸۷	۶۵	۸۸۳	۱۵۷	۸۸	۲۱۵	۱۵۵
۲۰۱۴-۲۰۱۵	۱۹۹	۲۰۹	۵۷	۹۹	۱۲۶	۸۰	۳۲۳	۱۹۶	۲۱۲	۳۴	۶۵	۱۳۵
۲۰۱۴-۲۰۱۶	۳۶۳	۱۴۹	۱۷۰	۱۴۷	۱۶۵	۶۱	۲۲۳	۷۲۸	۱۳۱	۲۰۳	۲۴۶	۱۷۱
۲۰۱۵-۲۰۱۶	۷۸	۷۵	۱۱۵	۱۲۶	۱۹۴	۹۰	۳۷۸	۴۶۷	۲۴۱	۱۲۵	۱۷۷	۴۲

ژانویه سال ۲۰۱۳ و ژانویه سال ۲۰۱۴ عدد ۱۴۴ است درحالی که ایستایی زمانی برای همین دو ماه در حالتی که از روش پیشنهادی با تفکیک دو رژیم استفاده شده، مطابق جدول ۴ عدد ۹۷/۱ است. همین مقدار برای تفکیک سه رژیم مطابق جدول ۵ معادل ۸۱/۷۷ و برای تفکیک چهار رژیم مطابق جدول ۶ برابر ۷۴/۸۱ است. ملاحظه می‌شود که در این مورد با افزایش تعداد رژیم‌ها، ایستایی زمانی بهتر شده است.

در جدول‌های ۴، ۵ و ۶ در مواردی که با رنگ قرمز

دو سال متفاوت نشان می‌دهد. در این جدول، مقادیر بدون به کار بستن روش پیشنهادی و بر مبنای مارکوف مرتبه اول محاسبه شده‌اند.

در جدول ۴، مقادیر ایستایی زمانی برای اجرای مدل پیشنهادی با تفکیک دو رژیم و در جداول ۵ و ۶، مقادیر ایستایی زمانی برای اجرای مدل پیشنهادی به ترتیب با سه و چهار رژیم اعمال شده است؛ برای نمونه، اولین عدد سمت چپ جدول ۳ یعنی ۱۴۴ نشان می‌دهد که ایستایی زمانی مدل مارکوف مرتبه اول (بدون تفکیک رژیم) بین ماه

جدول ۴. مقادیر ایستایی زمانی روش پیشنهادی (با تفکیک دو رژیم) برای دو ماه همنام از دو سال متفاوت (دوازده ماه همنام از چهار سال).

ماه/سال	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲
۲۰۱۳-۲۰۱۴	۹۷/۱۷	۱۲۹/۵	۷۷/۷	۶۴/۳۷	۱۷۵/۱	۱۲۴/۴	۵۹/۰۱	۲۰۱/۱	۵۵/۲۵	۱۶۳/۹	۱۱۱/۵	۲/۵
۲۰۱۳-۲۰۱۵	۱۲۴/۸	۷۱/۲۲	۱۱۱/۹	۹۳/۸۵	۲۰۷/۱	۱۴۵/۵	۲۳۲/۵	۱۶۶	۱۸۹/۶	۱۲۰/۲	۱۳۴/۶	۱۳۴
۲۰۱۳-۲۰۱۶	۱۲۹/۶	۹۱/۳۷	۱۱۷/۹	۱۱۴/۶	۱۰۲/۱	۱۷۳/۸	۱۰۹	۴۳۰/۹	۱۳۳/۱	۷۵/۲۷	۱۲۵/۴	۱۳۶/۴
۲۰۱۴-۲۰۱۵	۱۱۳/۴	۱۵۹/۴	۸۴/۲	۷۹/۵۴	۹۹/۸	۸۴/۶	۲۴۴/۳	۲۴۲/۹	۱۷۶/۵	۹۵/۹۶	۶۲/۰۵	۱۴۵/۱
۲۰۱۴-۲۰۱۶	۱۷۹/۴	۱۵۶/۱	۱۴۶/۱	۹۷/۴۴	۱۱۵/۳	۷۹/۹۷	۱۳۳/۵	۴۱۲/۷	۱۴۲/۶	۲۰۵	۱۸۷/۶	۱۳۸/۹
۲۰۱۵-۲۰۱۶	۷۶/۹۶	۷۸/۹۱	۸۱/۹۹	۱۱۰/۶	۱۴۷/۱	۸۷/۲۶	۱۶۳/۳	۲۰۸/۴	۱۱۵/۱	۱۱۹/۴	۱۴۵/۵	۷۹

جدول ۵. مقادیر ایستایی زمانی روش پیشنهادی (با تفکیک سه رژیم) برای دو ماه همنام از دو سال متفاوت (دوازده ماه همنام از چهار سال).

ماه/سال	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲
۲۰۱۳-۲۰۱۴	۸۱/۷۷	۱۰۹/۱	۶۳/۸۳	۶۹/۸۷	۱۳۲/۱	۱۲۰/۵	۶۲/۷۹	۱۸۵/۵	۵۱/۸۲	۱۱۶/۳	۷۵/۹۱	۱/۸۹۸
۲۰۱۳-۲۰۱۵	۱۲۸/۱	۶۲/۷۳	۷۸/۵۹	۹۴/۳	۱۶۶/۸	۱۳۳/۹	۲۰۷/۸	۱۵۰/۵	۲۱۶/۷	۱۰۲/۹	۸۲/۱۹	۱۱۴/۱
۲۰۱۳-۲۰۱۶	۱۲۲/۱	۹۳/۸۴	۱۰۸/۵	۱۰۳/۵	۹۳/۵۷	۱۵۳/۹	۱۰۲/۱	۳۷۸/۷	۱۱۱/۱	۵۹/۹۶	۱۰۱/۴	۹۰/۳۶
۲۰۱۴-۲۰۱۵	۱۰۷/۷	۱۳۲/۱	۸۲/۰۷	۷۰/۷۱	۸۹/۶۵	۹۱/۲۲	۲۱۸/۴	۱۹۶/۴	۱۶۶/۵	۷۱/۷۶	۵۱/۲۷	۱۲۱/۶
۲۰۱۴-۲۰۱۶	۱۷۵/۵	۱۲۹/۴	۱۱۷/۹	۹۹/۰۶	۸۰/۷۸	۸۳/۳۸	۱۲۹/۸	۳۴۴/۹	۱۱۳/۸	۱۳۱/۸	۱۳۵	۹۹/۱۱
۲۰۱۵-۲۰۱۶	۷۵/۰۸	۹۸/۶۷	۸۲/۵۸	۹۱/۲۸	۱۳۴/۶	۸۲/۴۹	۱۵۳/۷	۱۸۱/۵	۱۴۳/۷	۸۸/۴۹	۱۰۲/۸	۶۸/۳۳

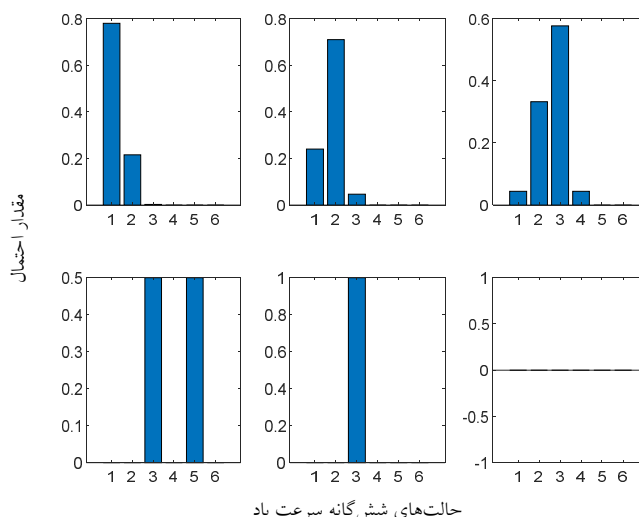
جدول ۶. مقادیر ایستایی زمانی روش پیشنهادی (با تفکیک چهار رژیم) برای دو ماه همنام از دو سال متفاوت (دوازده ماه همنام از چهار سال).

ماه/سال	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲
۲۰۱۳-۲۰۱۴	۷۴/۸۱	۱۰۵/۱	۵۸/۶۲	۵۷/۷۳	۱۳۰/۶	۹۷/۰۳	۶۱/۷	۱۷۸	۶۰/۳۸	۱۱۶/۵	۷۰/۵۲	۱/۸۳۹
۲۰۱۳-۲۰۱۵	۱۱۷/۴	۶۳/۵۹	۷۳/۲۲	۹۱/۲۵	۱۶۸/۱	۹۶/۹۷	۲۱۴/۳	۱۳۶/۹	۱۲۴/۱	۹۹/۳۶	۷۸/۹۵	۱۱۶/۶
۲۰۱۳-۲۰۱۶	۱۱۱/۲	۸۸/۰۴	۱۰۸	۹۸/۲۷	۸۳/۲۱	۱۰۹/۷	۱۰۸	۳۴۲/۵	۴۸/۹۴	۶۰/۷۲	۹۹/۶۹	۸۲/۴۲
۲۰۱۴-۲۰۱۵	۱۰۴/۸	۱۴۶/۹	۸۸/۶۱	۶۹/۵۳	۸۶/۳۵	۸۲/۴۷	۲۰۵/۱	۱۷۸/۲	۱۳۴/۶	۷۷/۲۱	۴۹/۷۷	۱۱۸/۹
۲۰۱۴-۲۰۱۶	۱۶۲/۴	۹۳/۱۳	۱۲۰/۵	۹۵/۸۲	۷۱/۶۳	۷۴/۷۷	۱۲۲/۱	۲۸۶/۷	۱۰۴/۸	۱۴۰	۱۲۷/۳۱	۹۰/۸۷
۲۰۱۵-۲۰۱۶	۶۴/۹۸	۹۱/۳۴	۷۹/۸۹	۸۴/۹۲	۱۲۳/۳	۶۹/۹۴	۱۴۶/۹	۱۷۳/۱	۱۲۹/۵	۸۸/۷۴	۹۷/۰۸	۶۸/۰۶

فرایند پیش‌بینی با روش پیشنهادی (با دو رژیم) و همچنین با روش مارکوف مرتبه اول در ماه سپتامبر ۲۰۱۵ انجام شده و نتایج در شکل‌های ۳، ۴ و ۵ ارائه شده است. در اینجا هر گام زمانی ۱۵ دقیقه است. سرعت باد در محدوده صفر تا ۳۰ متربرثانیه مطابق رابطه (۸) به شش بخش تقسیم (هر بخش معادل ۵ متربرثانیه) و هر بخش، یک حالت مدل مارکوف فرض شده است.

$$\begin{aligned}
 0 < V_i \leq 5 & \quad \text{حالت} = 1 \\
 5 < V_i \leq 10 & \quad \text{حالت} = 2 \\
 10 < V_i \leq 15 & \quad \text{حالت} = 3 \\
 15 < V_i \leq 20 & \quad \text{حالت} = 4 \\
 20 < V_i \leq 25 & \quad \text{حالت} = 5 \\
 V_i > 25 & \quad \text{حالت} = 6
 \end{aligned}
 \tag{۸}$$

مشخص شده‌اند، به کار بستن روش پیشنهادی به بدتر شدن میزان ایستایی نسبت به مدل مارکوف مرتبه اول منجر شده است و مواردی که با رنگ نارنجی مشخص شده‌اند، میزان ایستایی ثابت مانده است. مواردی که رنگ‌آمیزی نشده است (رنگ سفید) به معنای بهبود میزان ایستایی نسبت به مدل مارکوف مرتبه اول است. ملاحظه می‌شود که اجرای مدل پیشنهادی با تفکیک دو رژیم در حدود ۷۰٪ موارد موجب بهبود ایستایی زمانی شده است. این میزان برای تفکیک سه رژیم، برابر ۸۰٪ و برای تفکیک چهار رژیم برابر ۸۵٪ است.



شکل ۳. نتایج پیش‌بینی سرعت باد در گام زمانی بعدی بر مبنای مدل مارکوف مرتبه اول. به ترتیب از چپ به راست و از بالا به پایین متناظر با حالتی است که سرعت فعلی باد در محدوده ۱ تا ۶ باشد.

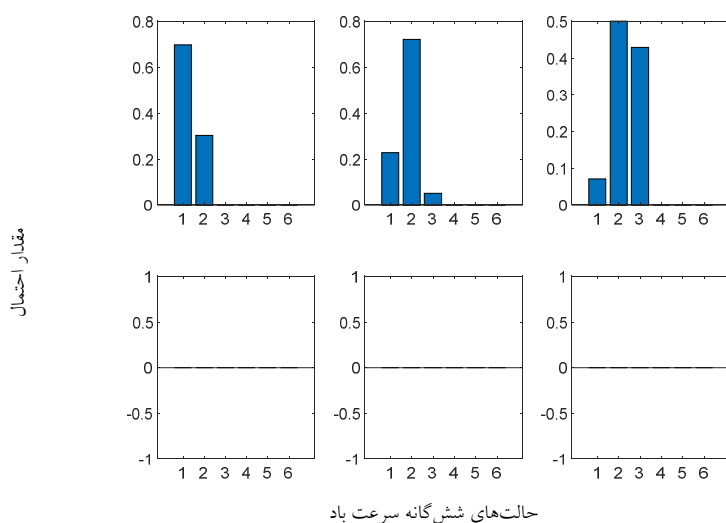
است که سرعت فعلی باد در محدوده سوم باشد. در این شکل، نمودارهای پایین از چپ به راست مبین حالت‌هایی هستند که سرعت فعلی باد در محدوده چهارم تا ششم باشد. در نمودار پایین سمت راست، در همه حالت‌ها احتمال وقوع صفر است و این ناشی از نبود اتفاق متناظر در سری زمانی داده‌های آموزشی است. همان‌طور که در سایر نمودارها نیز مشاهده می‌شود، امکانی برای وقوع سرعت باد در حالت ۶ (بالای ۲۵ متربرثانیه) وجود ندارد که البته با مشاهدات سرعت ثبت‌شده برای باد در ماه سپتامبر مطابقت دارد. شکل‌های ۴ و ۵ مقادیر حاصل از پیش‌بینی با روش پیشنهادی (با تفکیک دو رژیم) روی داده مشابه با شکل ۳ هستند. شکل ۴ پیش‌بینی مربوط به رژیم ۱ و شکل ۵ مربوط به رژیم ۲ تفکیک‌شده به روش پیشنهادی است.

نمودار شکل ۴، مربوط به اولین رژیم تفکیک‌شده، یعنی جهت حدود ۳۰ تا ۲۳۰ درجه، نشان‌دهنده ممکن نبودن وقوع سرعت باد در حالت‌های ۴، ۵ و ۶ است. با ملاحظه مثال فوق دیده می‌شود که با اجرای مدل

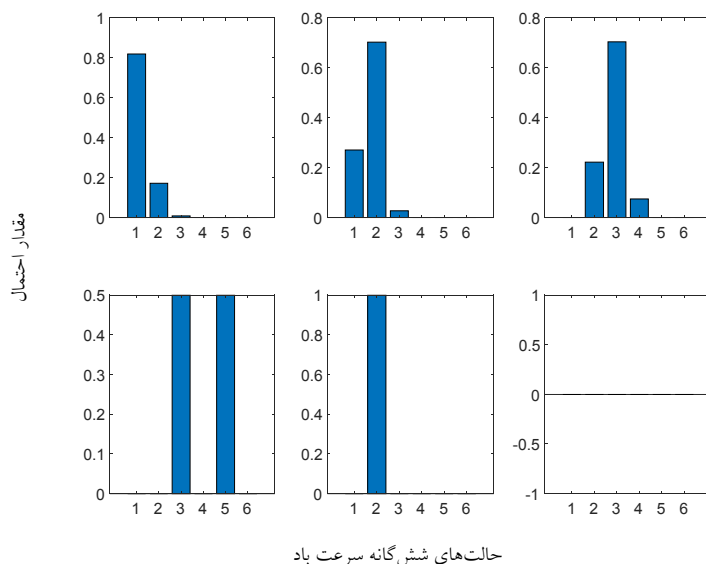
مطابق رابطه (۹)، اجرای مدل HMM با دو رژیم به تفکیک دو مجموعه از جهت‌های باد منجر شده است:

$$R_1 = \{2, 3, 4, 5, 6, 7, 8, 9, 10\}, R_2 = \{1, 11, 12, 13, 14, 15, 16\}$$

در شکل ۳ نتایج بر مبنای مدل مارکوف مرتبه اول محاسبه شده است. در محور عمودی پیش‌بینی احتمالاتی برای سرعت باد در گام زمانی بعدی با فرض معلوم بودن سرعت باد در گام زمانی فعلی آورده شده است. اولین نمودار بالا سمت چپ در شکل ۳ پیش‌بینی سرعت باد برای گام زمانی آینده است، هنگامی که در گام زمانی فعلی سرعت باد در حالت ۱ باشد. این نمودار بیان می‌کند که چنانچه در گام زمانی فعلی، سرعت باد در حالت ۱ باشد (سرعت باد کوچک‌تر یا مساوی ۵ متربرثانیه)، در گام زمانی بعدی به احتمال حدود ۰/۸ در حالت ۱ (سرعت باد کوچک‌تر یا مساوی ۵ متربرثانیه) و با احتمال حدود ۰/۲ در حالت ۲ (بین ۵ تا ده متربرثانیه) است. به‌همین ترتیب شکل وسط بالا نشان‌دهنده پیش‌بینی سرعت باد در گام زمانی بعدی است، هنگامی که سرعت فعلی باد در حالت دوم باشد. نمودار بالا-راست متناظر با حالتی



شکل ۴. مانند شکل ۳ ولی برای نتایج پیش‌بینی سرعت باد بر مبنای روش پیشنهادی مربوط به رژیم ۱ تفکیک‌شده به روش HMM.

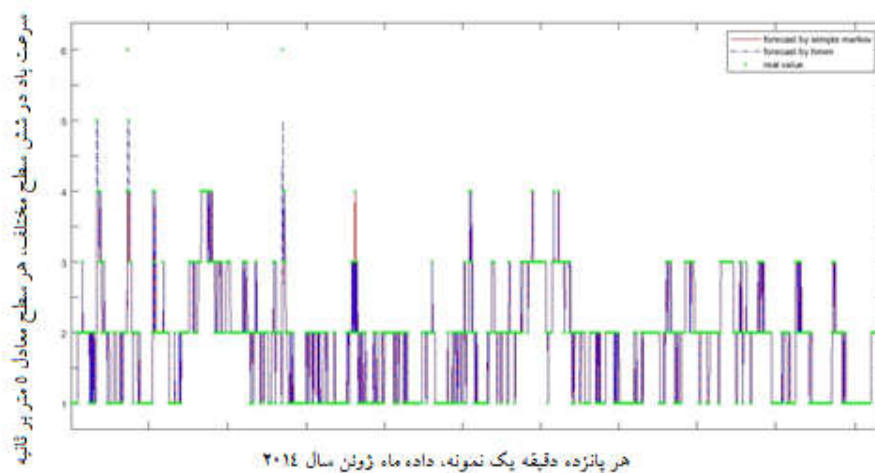


شکل ۵. مانند شکل ۳ ولی برای نتایج پیش‌بینی سرعت باد بر مبنای روش پیشنهادی مربوط به رژیم ۲ تفکیک‌شده به روش HMM.

اول و همچنین با روش پیشنهادی در مقایسه با مقادیر واقعی ثبت‌شده برای داده ۱۵ دقیقه‌ای در ماه ژوئن ۲۰۱۴ در روش محتمل‌ترین مقدار است. ریشه مجموع مربعات خطا در روش مارکوف مرتبه اول عدد $24/7$ و در روش پیشنهادی عدد $24/5$ است که نشان از بهبود ناچیز در میزان خطا دارد. با در نظر گرفتن اهمیت پیش‌بینی در سرعت‌های زیاد باد و حذف مقادیر کمتر از سطح ۴ از محاسبات، مقدار ریشه مجموع مربعات خطا در روش مارکوف مرتبه اول $7/3$ و در روش پیشنهادی $6/3$ است که نشان از بهبود خطای پیش‌بینی با روش پیشنهادی به میزان بیش از ۱۰٪ در سرعت‌های بیش از ۲۰ متر بر ثانیه دارد. این مقایسه برای دست‌کم پانزده نمونه مشابه اجرا شده است که در هر مورد، ریشه مجموع مربعات خطا محاسبه و مشاهده شده است که روش پیشنهادی موجب کاهش این مقدار شده است.

پیشنهادی، پیش‌بینی‌های شکل ۳ که بر اساس مدل مارکوف مرتبه اول ایجاد شده است، به دو نمودار پیش‌بینی مطابق شکل‌های ۴ و ۵ تفکیک شده است که بر اساس تشخیص رژیم حاکم به‌عنوان پیش‌بینی قابل استفاده هستند؛ بنابراین، دامنه احتمالات در پیش‌بینی‌ها محدود شده است که این موضوع باعث کم شدن عدم قطعیت و افزایش مطلوبیت پیش‌بینی می‌شود. این کار به دفعات برای سایر ماه‌ها و سایر سال‌ها در مجموعه داده‌های موجود اجرا شده است که نتایج مشابهی داشته‌اند.

ارزیابی روش پیشنهادی با محاسبه مقدار خطا نسبت به مقادیر واقعی اتفاق افتاده نیز ممکن است. برای این کار لازم است پیش‌بینی به صورت عددی مطلق صادر شود. پیش‌بینی مطلق با روش‌های مختلفی از روی پیش‌بینی احتمالاتی ایجاد می‌شود که اتخاذ محتمل‌ترین مقدار و متوسط مقادیر (mean value) از روش‌های ممکن است. شکل ۶ نشان‌دهنده نتایج پیش‌بینی با روش مارکوف مرتبه



شکل ۶. مقدار پیش‌بینی با روش مارکوف مرتبه اول و روش پیشنهادی در مقایسه با مقادیر واقعی ثبت شده. خط قرمز نشان‌دهنده پیش‌بینی براساس روش مارکوف مرتبه اول، خط چین نشان‌دهنده پیش‌بینی براساس روش پیشنهادی و نقاط سبزرنگ نمایانگر مقادیر واقعی ثبت شده هستند.

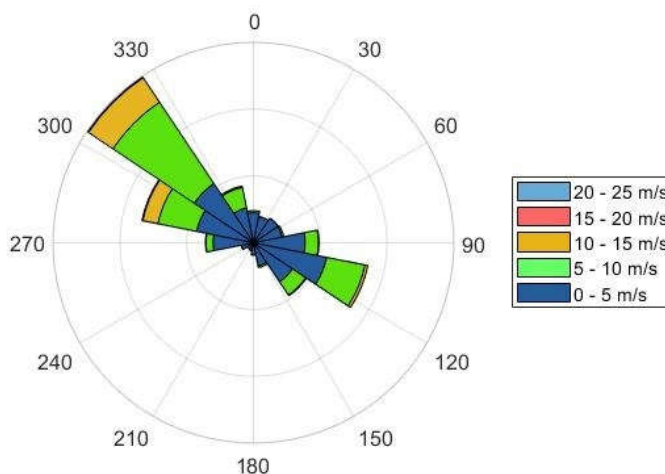
۴ نتیجه‌گیری

در این مقاله، ضمن تشریح مبانی نظری موضوع، روش پیشنهادی برای پیش‌بینی سرعت باد تشریح و با اجرای مدل روی داده موجود، نتایج با مدل مارکوف مرتبه اول مقایسه شد. ایستایی زمانی به‌عنوان اصلی‌ترین روش برای ارزیابی مدل‌های مبتنی بر مارکوف بین تمام بازه آماری چهارساله بررسی و ملاحظه شد که روش پیشنهادی دست‌کم در ۷۰ درصد موارد به بهبود ایستایی زمانی منجر می‌شود.

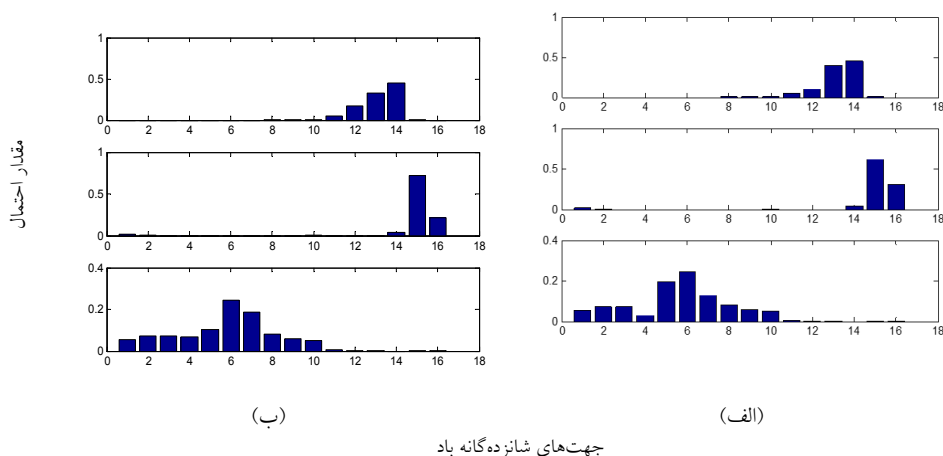
در آزمونی دیگر، وضعیت توابع احتمالی پیش‌بینی در روش پیشنهادی با روش مارکوف مرتبه اول مقایسه و ملاحظه شد که در روش پیشنهادی، دامنه احتمالات ممکن کاهش پیدا می‌کند و به‌عبارتی، پیش‌بینی احتمالاتی سرعت باد در گام زمانی بعدی دقیق‌تر می‌شود. در بررسی دیگر، با استنتاج پیش‌بینی قطعی از پیش‌بینی احتمالاتی موجود، میزان خطای روش پیشنهادی با خطای روش مارکوف مرتبه اول مقایسه و ملاحظه شد که در این معیار نیز بهبود اتفاق افتاده است. این بهبود به‌ویژه در سرعت‌های زیاد بیشتر دیده می‌شود.

بررسی نتایج نشان می‌دهد تفکیک رژیم‌های دوگانه‌ای که در ماه‌های مختلف صورت پذیرفته است، با مشاهدات بلندمدت منطقه سازگاری خوبی دارد. براساس مشاهدات بلندمدت دیدبانان در ایستگاه هواشناسی فرودگاه امام خمینی، دو جهت غالب برای وزش باد وجود دارد که یکی حاصل وزش باد کوه به دشت در ساعات شب است که از سمت ارتفاعات شمال تهران به سمت کویر می‌وزد و جهت آن در محدوده ۲۷۰ تا ۳۶۰ درجه است و دیگری ناشی از وزش باد کویر به سمت کوه است که در جهت مقابل از سمت حدود ۹۰ تا ۱۸۰ درجه است. این پدیده‌ها در نتایج تفکیک رژیم‌های دوگانه که برای تمام ماه‌ها در طی چهار سال دوره آماری انجام شده است، به‌خوبی دیده می‌شود. گل‌باد سه ماه تابستان در سال ۲۰۱۶ نیز این موضوع را تأیید می‌کند.

در ماه‌های زمستان که پدیده‌های بزرگ‌مقیاس فعالیت عمده‌ای دارند، علاوه بر دو جهت پیش‌گفته که ناشی از شرایط محلی هستند، جهت باد همراه بیشتر پدیده‌های بارشی که اغلب از سمت جنوب‌غربی است نیز به جهت‌های تأثیرگذار اضافه می‌شود. در نمودار میله‌ای



شکل ۷. گل‌باد سه ماه تابستان سال ۲۰۱۶ در ایستگاه هواشناسی فرودگاه امام خمینی. زوایا برحسب درجه هستند.



شکل ۸. مقادیر احتمال تخمین زده شده برای HMM سه‌حالتی در (الف) ماه مارس ۲۰۱۴ (ب) ماه فوریه ۲۰۱۴. نمودارها از بالا به پایین مقادیر متناظر برای رژیم ۱ تا ۳ هستند. در تمامی نمودارها، محور افقی، جهت‌های شانزده‌گانه باد (N, NNE, NE, ..., NNW) و محور عمودی مقدار احتمال است.

نتایج نشان می‌دهد که مدل پیشنهادی، بهبود چشمگیری نسبت به مدل مارکوف مرتبه اول در پیش‌بینی خیلی-کوتاه‌مدت باد ایجاد می‌کند. در هواشناسی، پیش‌بینی‌های خیلی کوتاه‌مدت سمت و سرعت باد (بیشتر، یک‌ساعته) برای برنامه‌ریزی‌های روزانه کاربرد فراوانی دارند. زمان-های کوتاه‌تر معمولاً کاربردهایی در سایر زمینه‌ها از جمله مدیریت توربین‌های بادی دارند. در این کاربردها،

شکل ۸ تفکیک سه رژیم در جهت باد برای ماه مارس ۲۰۱۴ ملاحظه می‌شود که مؤلفه یکی از رژیم‌ها در جهت گفته شده است درحالی‌که دو رژیم دیگر هنوز متأثر از عوامل محلی همانند قبل هستند.

مدل پیشنهادی برای داده با گام زمانی ۳۰ ثانیه و همچنین ۱۵ دقیقه ارزیابی و عملکرد آن طی دوره‌های یک‌ماهه با مدل مارکوف مرتبه اول مقایسه شده است.

منابع

- چینی‌فروش، ن. و لطیف‌شبانگاهی، غ.، ۲۰۱۸، تجهیزات اندازه‌گیری سرعت و جهت باد: روش‌ها، چالش‌ها و روند فناوری: پژوهش‌های اقلیم شناسی، ۳۳، ۴۳-۶۲.
- Ailliot, P., Bessac, J., Monbet, V., and Pene, F., 2015, Non-homogeneous hidden Markov-switching models for wind time series: Journal of Statistical Planning Inference, **160**, 75-88.
- Ailliot, P., and Monbet, V., 2012, Markov-switching autoregressive models for wind time series: Environmental Modelling Software, **30**, 92-101.
- Billinton, R., Chen, H., and Ghajar, R., 1996, Time-series models for reliability evaluation of power systems including wind energy: Microelectronics Reliability, **36**(9), 1253-1261.
- Bitner-Gregersen, E. M., Bhattacharya, S. K., Chatjigeorgiou, I. K., Eames, I., Ellermann, K., Ewans, K., Hermanski, G., Johnson, M. C., Ma, N., and Maisondieu, Ch., 2014, Recent developments of ocean environmental description with focus on uncertainties: Ocean Engineering, **86**, 26-46.
- Callaway, D. S., 2010, Sequential reliability forecasting for wind energy: Temperature dependence and probability distributions: IEEE Transactions on Energy Conversion, **25**(2), 577-585.
- Carta, J. A., Ramirez, P., and Velazquez, S., 2009, A review of wind speed probability distributions used in wind energy analysis: Case studies in the Canary Islands: Renewable Sustainable Energy Reviews, **13**(5), 933-955.
- D'Amico, G., Petroni, F., and Prattico, F., 2012, Wind speed modeled as a semi-Markov process: Environmetrics, **24**(6), 367-376.
- D'Amico, G., Petroni, F., and Prattico, F., 2014, Wind speed and energy forecasting at different time scales: A nonparametric approach: Statistical Mechanics and its Applications, **406**, 59-66.
- Hocaoglu, F. O., Gerek, O. N., and Kurban, M., 2008, The effect of Markov chain state size for synthetic wind speed generation: paper presented at the Proceedings of the 10th International Conference on Probabilistic Methods Applied to Power Systems, IEEE, 1-4.
- Jiang, Y., Song, Z., and Kusiak, A., 2013, Very short-term wind speed forecasting with
- پیش‌بینی‌هایی در حد چندین ثانیه نیز می‌تواند به امکان اعمال الگوی مناسب کنترلی برای بیشینه کردن دریافتی و جلوگیری از وارد شدن خسارت به تجهیزات منجر شود (جیانگ و همکاران، ۲۰۱۳) و اسکلیف و همکاران، (۲۰۱۳). جدای از مصارف متناظر با تولید انرژی الکتریکی از باد، با این روش می‌توان برای نشست و برخاست هواپیماها نیز اطلاعات مفیدی تولید کرد. مدل پیشنهادی همانند سایر مدل‌های آماری در پیش‌بینی‌های بلندمدت توانایی خوبی ندارد.
- در این مقاله تفکیک رژیم در روش پیشنهادی با انتخاب رژیم متناظر با احتمال بیشتر صورت پذیرفته است که خود خالی از خطا نیست و بهتر است اطلاعات کمکی این فرایند را بهبود بخشند. از جمله اطلاعات کمکی، می‌تواند ساعت مورد نظر طی بیست و چهار ساعت شبانه‌روز، میزان پوشش ابر و میزان بارندگی باشد که از داده‌های رادار یا داده‌های شبکه فشرده باران‌سنجی می‌توان استخراج کرد. علاوه بر این، فرض نرخ ثابت گذر در ماتریس احتمال گذر، فرضی تقریبی است و با توجه به اثرپذیری این نرخ از سایر پارامترها، می‌توان به جای نرخ ثابت گذر، از نرخ متغیر با زمان (برای مثال، یک متغیر سینوسی با دوره تناوب بیست و چهار ساعت) برای مدل‌سازی تغییرات احتمال گذر استفاده کرد. در این حالت، مسئله مارکوف پنهان به مسئله شبه مارکوف پنهان (Hidden-Semi Markov) تبدیل می‌شود که حل آن احتیاج به اجرای الگوریتم‌های پیچیده‌تر دارد.

سپاسگزاری

از آقای مهندس پرویز رضازاده که در تدوین این مقاله از دانش هواشناسی ایشان بهره بردیم و همچنین از خانم مهندس مرضیه تفرشی که در ویرایش کمک بسیار کردند، سپاسگزاری می‌کنیم.

- Bayesian structural break model: Renewable Energy, **50**, 637-647.
- Karatepe, S., and Corscadden, K., 2013, Wind speed estimation: incorporating seasonal data using Markov chain models: ISRN Renewable Energy.
- Lei, M., Shiyen, L., Chuanwen, J., Hongling, L., and Yan, Z., 2009, A review on the forecasting of wind speed and generated power: Renewable and Sustainable Energy Reviews, **13**(4), 915-920.
- Leroux, B., 1989, Maximum likelihood estimation for mixture distributions and hidden Markov models: University of British Columbia.
- Lou, H. L., 1995, Implementing the Viterbi algorithm: IEEE Signal processing magazine, **12**(5), 42-52.
- Picard, F., 2007, An introduction to mixture models, **7**.
- Pinson, P., and Madsen, H., 2012, Adaptive modelling and forecasting of offshore wind power fluctuations with Markov switching autoregressive models: Journal of forecasting, **31**(4), 281-313.
- Schlipf, D., Schlipf, D. J., and Kühn, M., 2013, Nonlinear model predictive control of wind turbines using LIDAR: Wind Energy, **16**(7), 1107-1129.
- Scholz, T., Lopes, V. V., and Estanqueiro, A., 2014, A cyclic time-dependent Markov process to model daily patterns in wind turbine power production: Energy, **67**, 557-568.
- Shamshad, A., Bawadi, M. A., Hussin, W. W., Majid, T. A., and Sanusi, S., 2005, First and second order Markov chain models for synthetic generation of wind speed time series: Energy, **30**(5), 693-708.
- Suomalainen, K., Silva, C. A., Ferrão, P., and Connors, S., 2012, Synthetic wind speed scenarios including diurnal effects: Implications for wind power dimensioning: Energy, **37**(1), 41-50.
- Tagliaferri, F., Hayes, B. P., Viola, I. M., Djokić, S. Z., 2016, Wind modelling with nested Markov chains: Journal of Wind Engineering Industrial Aerodynamics, **157**, 118-124.
- Wang, X., Guo, P., and Huang, X., 2011, A review of wind power forecasting models: Energy Procedia, **12**, 770-778.
- Xie, K., Liao, Q., Tai, H.M., and Hu, B., 2017, Non-Homogeneous Markov wind speed time series model considering daily and seasonal variation characteristics: IEEE Transactions on Sustainable Energy, **8**(3), 1281-1290.

پیوست پ

مدل مارکوف مرتبه اول و ارزیابی ایستایی آن

روش‌های مختلفی مبتنی بر احتمال در زمینه پیش‌بینی باد استفاده شده است که یکی از مهم‌ترین و رایج‌ترین این روش‌ها، مدل مارکوف است (شمشاد و همکاران، ۲۰۰۵). مارکوف مرتبه اول مدلی است که در آن احتمال وقوع هر حالت تنها به حالت قبلی بستگی دارد. ماتریس احتمالات گذر Γ یک ماتریس مربعی از احتمالات است که برای k حالت، یک ماتریس $k \times k$ است (شمشاد و همکاران، ۲۰۰۵):

$$\Gamma = \begin{bmatrix} \gamma_{1,1} & \gamma_{1,2} & \dots & \gamma_{1,k} \\ \gamma_{2,1} & \gamma_{2,2} & \dots & \gamma_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{k,1} & \gamma_{k,2} & \dots & \gamma_{k,k} \end{bmatrix} \quad (پ-۱)$$

در ماتریس گذر مرتبه اول، هر درایه $\gamma_{i,j}$ احتمال گذر از حالت i به حالت j را نشان می‌دهد. اگر $n_{i,j}$ تعداد گذرها

از حالت i به حالت j در سری داده‌ها باشد، احتمال گذر

از حالت i به حالت j برابر است با:

$$\gamma_{i,j} = \frac{n_{i,j}}{\sum_j n_{i,j}} \quad (پ-۲)$$

مقدار احتمال گذر، عددی است بین صفر و یک و مجموع احتمالات گذر در هر سطر در ماتریس گذر Γ برابر با یک است:

$$\sum_{j=1}^k \gamma_{ij} = 1 \quad (پ-۳)$$

با استفاده از این روش و انتساب هر محدوده سرعت باد به یک حالت، می‌توان با دانستن سرعت فعلی (حالت i)، بیان کرد که سرعت در گام زمانی آینده با چه احتمالی در حالت j خواهد بود؛ به بیان دیگر، با دانستن سرعت باد در زمان حال می‌توان پیش‌بینی کرد که در گام زمانی بعدی، سرعت باد با چه احتمالی در چه بازه‌ای است. پیش‌بینی

که T تعداد بازه‌های زمانی، $n_{i,j}(t)$ تعداد دفعات گذر و $\gamma_{i,j}(t)$ احتمال گذر از i به j برای سری داده در بازه زمانی t ام است $t = 1, 2, \dots, T$.

$\gamma_{i,j}$ احتمال گذر از حالت i به j برای کل داده‌های سری زمانی است. برای اینکه این زنجیره مارکوف ایستا باشد، پارامتر β باید یک توزیع χ^2 با درجه آزادی $(T-1)(k-1)$ داشته باشد. این فرایند در سطح معنی‌داری ۵٪ ایستاست اگر (درجه آزادی ۵٪) $\beta < \chi^2$ باشد.

مدل مارکوف پنهان

مدل مارکوف پنهان بر پایه مدل‌های مخلوط بنا نهاده شده است. در مدل مخلوط فرض بر این است که عوامل محیطی و فیزیکی مختلف در پدیده‌ها اثر می‌گذارند به گونه‌ای که روال‌ها و رژیم‌های مختلفی بر این پدیده‌ها حاکم می‌شوند. این عوامل اثرگذار یا نحوه تعامل آنها اغلب ناشناخته هستند یا تعداد آنها به قدری زیاد است که اثرگذاری آنها به‌طور واضح بیان‌شدنی نیست. دسته مهمی از این نوع پدیده‌ها از نوع شمارش نامتناهی هستند؛ مانند تعداد زلزله‌های اتفاق افتاده یا تعداد طوفان‌های شدید در یک سال. یک مدل استاندارد برای شمارش‌های نامتناهی، توزیع پواسون با تابع احتمال $P(x) = e^{-\lambda} \lambda^x / x!$ است. ویژگی این مدل این است که در آن مقدار واریانس با مقدار میانگین برابر است. در مواردی که اختلاف واریانس با میانگین زیاد است، نشان‌دهنده پراکندگی بیش از حد این داده‌ها نسبت به توزیع پواسون و ناسازگاری توزیع این داده‌ها در این مدل است؛ برای نمونه، با بررسی داده‌های باد ایستگاه شاهد در یک بازه یک‌ماهه در سپتامبر ۲۰۱۴ (که با گام زمانی ۳۰ ثانیه جمع‌آوری شده‌اند) و رسم بافت‌نگار (histogram) مربوطه (شکل پ-۱) ملاحظه می‌شود که یک تابع منفرد پواسون،

احتمالاتی سرعت باد در گام زمانی آینده به‌سادگی با رابطه (پ-۴) بیان می‌شود:

$$V_1 = V \cdot T, \quad (\text{پ-۴})$$

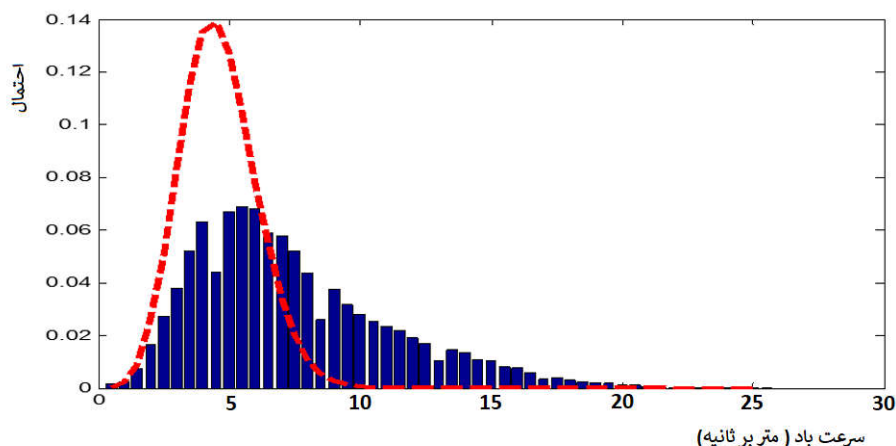
که در آن، V بردار شرایط اولیه است و نشان‌دهنده تعلق سرعت باد به یکی از بازه‌های تفکیک شده است؛ برای نمونه، اگر سرعت باد به سه ناحیه (زیر ۱۰ متربرثانیه، بین ۱۰ و ۲۰ متربرثانیه و بیشتر از ۲۰ متربرثانیه) تقسیم شده باشد، یعنی سه حالت برای سرعت باد مفروض باشد، $V_1 = [0 \ 1 \ 0]$ نشان‌دهنده اینست که در شرایط اولیه سرعت باد بین ۱۰ و ۲۰ متربرثانیه است. در این صورت طبق رابطه (پ-۴)، مقادیر احتمال زیر به‌دست می‌آید:

$$V_1 = [\gamma_{2,1} \ \gamma_{2,2} \ \gamma_{2,3}],$$

یعنی به احتمال $\gamma_{2,1}$ در گام زمانی بعدی سرعت باد در ناحیه اول، با احتمال $\gamma_{2,2}$ در ناحیه دوم و با احتمال $\gamma_{2,3}$ در ناحیه سوم است که البته همیشه در تعریف مسئله و بنا به نیاز می‌توان به تعداد دلخواه ناحیه تعریف کرد.

استفاده از این روش در صورتی معتبر است که فرایند مارکوف به زمان وابسته نباشد، یا به عبارتی، ایستا باشد؛ یعنی با گذشت زمان، ماتریس احتمالات گذر فرایند مارکوف تغییر نکند. یک روش معمول برای بررسی ایستایی زمانی یک فرایند مارکوف به‌دست آمده از یک سری زمانی این است که سری زمانی به چند بازه تقسیم شود. سپس با تشکیل فرایند مارکوف برای هر بازه زمانی به‌صورت جداگانه، ماتریس احتمالات گذر محاسبه شود. در این حالت، تحقق شرط ایستایی زمانی مشروط بر این است که ماتریس‌های احتمالات گذر به‌دست آمده برای هر بازه تقریباً با یکدیگر برابر باشند. در آمار، برای بررسی تحقق ایستایی زمانی، پارامتر β تعریف می‌شود (شمشاد و همکاران، ۲۰۰۵):

$$\beta = \sum_{i,j} \sum_{t=1}^T n_{i,j}(t) \ln \left(\frac{\gamma_{i,j}(t)}{\gamma_{i,j}} \right), \quad (\text{پ-۵})$$



شکل پ-۱. بافت‌نگار داده‌های سرعت باد در ایستگاه هواشناسی فرودگاه امام برای سپتامبر ۲۰۱۴ (نمودار ستونی آبی‌رنگ) و نمودار پواسون برازش شده بر آن (خط چین قرمز)

m برابر ۲، توزیع مخلوط به دو تابع چگالی احتمال مطابق زیر بستگی دارد:

۲	۱	جزء انتخابی
	$P_1(x)$	تابع توزیع احتمال

برای تعیین هر جزء، به یک متغیر تصادفی گسسته C نیاز است که عمل ترکیب یا مخلوط را انجام می‌دهد:

$$C = \begin{cases} 1 & \text{with Probability } \delta_1 \\ 2 & \text{with Probability } \delta_2 = 1 - \delta_1 \end{cases}$$

یک روش قدرتمند برای تخمین پارامترهای یک توزیع مخلوط، روش بیشینه درست‌نمایی (Maximum Likelihood, ML) است. درست‌نمایی یک مدل مخلوط با m جزء برای هر دو حالت گسسته و پیوسته با رابطه (پ-۶) بیان می‌شود (لروکس، ۱۹۸۹):

$$L(\theta_1, \dots, \theta_m, \delta_1, \dots, \delta_m | x_1, \dots, x_n) = \prod_{j=1}^n \sum_{i=1}^m \delta_i P_i(x_j, \theta_i), \quad (\text{پ-۶})$$

که در این رابطه، $\theta_1, \dots, \theta_m$ یک بردار شامل پارامترهای توزیع هر توزیع جزئی، بردار x_1, \dots, x_n مجموعه n داده

تقریب خوبی برای داده‌های سرعت باد نیست. نگاهی به مقدار متوسط و مقدار واریانس داده‌ها نیز همین موضوع را نشان می‌دهد. میانگین داده‌های باد ۴/۱۱ و واریانس آن ۸/۷۵ است که نشان‌دهنده نامناسب بودن مدل پواسون منفرد برای شبیه‌سازی است. یک روش برای کار با داده‌های با پراکندگی زیاد یا توزیع دو یا چند مدل، استفاده از مدل مخلوط است. مدل مخلوط برای تطابق با ناهمگونی مشاهده‌نشده در نمونه‌ها طراحی شده است. به عبارتی، کارایی آن برای مواردی است که نمونه‌ها شامل گروه‌های مشاهده‌نشده باشند و هر یک از گروه‌های مشاهده‌نشده توزیعی متفاوت داشته باشند؛ برای نمونه، با این روش هنگامی که توزیع یک داده از پراکندگی زیادی برخوردار باشد و توزیع آن یک توزیع منفرد (مثلاً پواسون) نباشد، آن داده‌ها را می‌توان به صورت چند گروه در نظر گرفت که هر یک از آنها توزیع منفردی دارند.

در حالت کلی، یک توزیع آماری مخلوط، شامل تعداد متناهی (مانند m) توزیع مستقل و یک توزیع ترکیبی می‌شود که این توزیع ترکیبی، هر یک از اجزاء (هر یک از m جزء) را انتخاب می‌کند (پیکارد، ۲۰۰۷). توزیع هر یک از اجزاء می‌تواند پیوسته یا گسسته باشد. با فرض

می‌شوند که به‌ویژه برای سری‌های گسسته‌مقدار، شامل سری‌های گروهی یا سری‌های شمارشی کاربرد دارند.

مدل پنهان مارکوف $\{X_t : t \in \mathbb{N}\}$ نوع خاصی از مدل‌های مخلوط وابسته است. اگر $C^{(t)}$ و $X^{(t)}$ تاریخچه تغییرات از ۱ تا t باشند، این مدل با رابطه (پ-۷) بیان می‌شود:

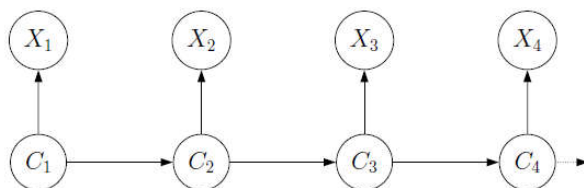
$$\begin{aligned} P_r(C_t | C^{(t-1)}) \\ = P_r(C_t | C_{t-1}), \quad t = 2, 3, \dots \\ P_r(X_t | X^{(t-1)}, C^{(t)}) \\ = P_r(X_t | C_t), \quad t \in \mathbb{N}. \end{aligned} \quad (\text{پ-۷})$$

مدل از دو بخش تشکیل شده است؛ یکی فرایند پارامتر (parameter process) $\{C_t : t = 1, 2, \dots\}$ که خواص مارکوف را برآورده می‌کند و دیگری فرایند وابسته به حالت $\{X_t : t = 1, 2, \dots\}$ ، به‌گونه‌ای که وقتی C_t معلوم باشد، توزیع X_t فقط به حالت فعلی C_t بستگی دارد و به مقادیر قبلی وابسته نیست. به‌عبارتی، با توجه به مجموعه فرایندهای پارامتر ارائه‌شده، مدل در هر گام زمانی می‌تواند در یکی از حالت‌های $\{C_t : t = 1, 2, \dots\}$ قرار گیرد که گذر بین این حالت‌ها براساس فرایند مارکوف انجام می‌شود. هرگاه مدل در یکی از این حالت‌ها قرار گرفت، توزیع مدل برابر X_t متناظر آن، یعنی $\{X_t : t = 1, 2, \dots\}$ خواهد بود. این ساختار با گراف جهت‌دار شکل پ-۲ بیان می‌شود: درست‌نمایی یک HMM (L_T) برای دنباله داده‌های

دیدبانی و $\delta_1, \dots, \delta_m$ پارامترهای مخلوط (انتخاب) هستند که مجموع آنها یک است. هدف از حل مسئله بیشینه درست‌نمایی، تعیین پارامترهای مختلف از جمله θ و δ ها به‌گونه‌ای است که با فرض وجود داده‌های x_1, \dots, x_n عبارت سمت راست در رابطه (پ-۶) بیشینه شود.

مدل مخلوط مستقل اجازه وابستگی مقادیر مختلف متوالی را به هم نمی‌دهد، درحالی‌که سری زمانی باد یک سری به هم وابسته است. همان‌طور که در شکل ملاحظه می‌شود، وقتی به اندازه یک نمونه (۳۰ ثانیه) در سرعت باد تأخیر ایجاد می‌شود، مقدار تابع خودهمبستگی، عددی حدود ۹۰٪ مقدار اولیه به‌دست می‌آید که نشان‌دهنده این است که سری زمانی مربوطه یک سری به‌هم‌وابسته است. البته نمودار با افزایش تأخیر کاهش یافته است؛ یعنی هرچه مشاهدات با تأخیر بیشتری باشند، همبستگی کمتری پیدا می‌کنند. صرف‌نظر کردن از شرط مستقل بودن در مدل مخلوط و در نظر گرفتن مدل مارکوف برای رفتن از یک توزیع به توزیع دیگر، راهی مناسب برای مدل کردن سری زمانی باد به‌صورت مدل مخلوط است. این روش مبنای معرفی مدل مارکوف پنهان است.

مدل‌های مارکوف پنهان Hidden Markov Model (HMM) مدل‌هایی هستند که مقادیر ناشی از دیدبانی یا اندازه‌گیری در آنها، توزیعی وابسته به حالت‌هایی دارند که این حالت‌ها، وابسته به حالت‌های یک فرایند مارکوف دیده‌نشده هستند. این مدل‌ها به‌عنوان مدل‌هایی با کاربرد عمومی برای سری‌های زمانی یک یا چند متغیره شناخته



شکل پ-۲. گراف جهت‌دار ساختار مدل HMM. در این مدل فرایند پارامتر یکی از C_1, C_2, C_3, C_4 را انتخاب می‌کند و توزیع متناظر آن یعنی X_1, X_2, X_3, X_4 براساس آن مشخص می‌شود.

π_i دارد که i نشان دهنده هر یک از حالت‌ها است؛ برای مثال، اگر شیر آمدن در پرتاب سکه موفقیت لحاظ شود و در آزمایش اول، پنج بار سکه بیندازیم که دو بار A شیر \equiv بیاید، $n_1 = 5$ و $x_1 = 2$ خواهد بود.

یک مدل چندجمله‌ای-HMM، مفهوم توسعه یافته تعریف فوق است. در این حالت فرض می‌شود که به جای دو حالت (موفقیت و شکست) در هر بار آزمایش q حالت وجود داشته باشد که $q \geq 2$ است. در این صورت، تعداد نتایج مشاهده شده، q برابر حالت قبل است و عبارتست از $\{x_{tj}; t = 1, \dots, T; j = 1, \dots, q\}$ البته که $x_{t1} + x_{t2} + \dots + x_{tq} = n_t$ تعداد انجام آزمایش در زمان t است؛ بنابراین x_{13} نشان دهنده تعداد نتیجه‌های برابر حالت ۳ در آزمایش دوم است؛ برای نمونه، در پرتاب تاس شش‌حالتی، اگر در آزمایش اول هفت بار تاس بیندازیم که دو بار عدد ۳ بیاید، $n_1 = 7$ و $x_{13} = 2$ خواهد بود.

می‌توان بردار X_t که شامل تمام مقادیر مشاهده شده در زمان t است را به صورت زیر نوشت:

$$X_t = (x_{t1}, x_{t2}, \dots, x_{tq}). \quad (\text{پ-۱۰})$$

x_1, x_2, \dots, x_T که با m حالت، با توزیع اولیه δ و ماتریس احتمال گذر Γ محاسبه شده‌اند، با رابطه (پ-۸) بیان می‌شود که در آن $P(x_T)$ احتمال وقوع x_T است:

$$L \quad (\text{پ-۸}) \\ =$$

هنگامی که خروجی یک آزمایش می‌تواند دو حالت داشته باشد (پیروزی یا شکست)، مدل دوجمله‌ای برای محاسبه احتمال موفقیت در تعداد مشخصی از آزمایش به کار می‌رود. احتمالات دوجمله‌ای عبارت است از:

$$p_i(x_t) \quad (\text{پ-۹}) \\ = \binom{n_t}{x_t} \pi_i^{x_t} (1 - \pi_i)^{n_t - x_t},$$

که x_t تعداد موفقیت‌ها، n_t تعداد آزمایش‌ها در زمان t و π_i احتمال موفقیت است.

مدل دوجمله‌ای-HMM بر مبنای همین روش برای مدل‌سازی سری شمارش‌های محدود به کار می‌رود. در واقع، مدل دوجمله‌ای-HMM نوعی زنجیره مارکوف است که در آن مقادیر مشاهده شده $\{x_t; t = 1, \dots, T\}$ تعداد موفقیت‌ها در n_1, n_2, \dots, n_T آزمایش مستقل برنولی هستند. مدلی با m حالت، m مقدار برای احتمال موفقیت

A Novel Algorithm for Wind Forecasting Based on Hidden Markov Model

Navid Chiniforush¹, Gholamreza Latif Shabgahi^{2*} and Majid Azadi³

¹ *PHD candidate, Shahid Beheshti University, Tehran, Iran*

² *Associate Professor, Shahid Beheshti University, Tehran, Iran*

³ *Assistant professor, Atmospheric Science and Meteorological Research Center, Tehran, Iran*

(Received: 02 March 2019, Accepted: 01 December 2019)

Summary

Meteorological time series are used as important input for risk forecasting and related warning systems. Wind is one of the most important atmospheric parameters because of its extensive effects in many industries and fields of human life. Many researches have been carried out to improve forecasting of the wind with the aim of improving output of wind farms, issuing warning for public, detection of wind shear and turbulence in the airports and so on. Generally, there are two main groups of meteorological forecasting methods, one is based on physical relation of atmospheric parameters, and the other is based on historical data. For a long time, time series of wind have been used for forecasting the wind speed. ARMA (Auto-Regressive Moving Average) and Markov model are two important groups of time series analyzing methods. In this paper, the capability of HMM (Hidden Markov Model) is described and used for identification and classification of wind time series. Based on theoretical concept of HMM, a proper method is proposed, and utilized for simulation with real data. The proposed method is based on constructing a multinomial-HMM on wind direction time series. The whole range of possible wind direction (360 degrees) is divided into 16 groups and then categorized to different regimes. Wind forecasting is then carried out based on these separated categories. Temporal stationary test which is well known for Markov chain, is extended for the proposed method and used for its efficiency evaluation. Efficiency of the proposed model is investigated by using real data of IKIA (Imam Khomeini International Airport). A part of the collected data including wind speed and direction is used for constructing of the proposed model and another part is used for its evaluation. The achieved results show that there is improvement in temporal stationary for HMM vs simple Markov model, in 70 to 80 percent of cases. History of the observations in IKIA shows that there are two major wind directions in the area which are related to the local condition: from mountain to the desert in the day times from north-west and from the opposite direction at nights. These are the only important directions in the area in summer when there are no important meteorological phenomena, while in winter one major direction would be added from south-west because of the large scale meteorological systems. Increasing the number of regimes has also significant improvement in temporal stationary in winter times, while there is no important improvement in summer times. This has a good harmony with long term recorded data.

Keywords: Hidden Markov Model, temporal stationary, wind speed, regime separation, forecasting

*Corresponding author:

gh_latif@sbu.ac.ir