

## بکارگیری الگوریتم ویتربی در پیش‌بینی وقوع بارندگی و شبیه‌سازی تداوم‌های خشک و تر - مقایسه با روش‌های متداول

مهدی قمقما<sup>۱</sup> - جواد بذرافشان<sup>۲\*</sup>

تاریخ دریافت: ۱۳۹۲/۶/۱۸

تاریخ پذیرش: ۱۳۹۳/۷/۷

### چکیده

امروزه مدل‌های آماری بسیاری برای شبیه‌سازی سری زمانی گسسته وقوع و عدم وقوع بارش براساس داده‌های تاریخی وجود دارد که تاکید آن‌ها بیشتر بر ساخت آماره‌های اقلیمی است. با این وجود، دقت مدل‌های شبیه‌ساز مذکور باید از نقطه‌نظر لحظه‌ای یا کوتاه‌مدت نیز بهبود یابد. در تحقیق حاضر فرض شده است که ساختار وقوع بارش از مدل مارکف پنهان با یک لایه پنهان (سری گسسته وقوع و عدم وقوع) و یک لایه قابل‌رویت تبعیت می‌کند، که به طور موردی در ایستگاه سینوپتیک خرم‌آباد (دوره آماری ۲۰۰۵-۱۹۶۱) مورد بررسی قرار گرفت. از الگوریتم ویتربی برای حل مساله رمزگشایی سری زمانی توالی حالات تر و خشک استفاده شد. کارکرد پنج متغیر هواشناسی (فشار هوا در ایستگاه، فشار بخار آب، دامنه شبانه‌روزی دمای هوا، رطوبت نسبی و دمای نقطه شبنم) به عنوان توالی قابل‌رویت، براساس معیارهای دقت پیش‌بینی با هدف انتخاب بهترین متغیر در فرآیند رمزگشایی مورد ارزیابی قرار گرفته است. نتایج نشان داد که متغیر دامنه شبانه‌روزی دمای هوا به عنوان مناسب‌ترین توالی قابل‌رویت برای رمزگشایی سری گسسته خشک و تر ارزیابی می‌باشد که می‌تواند به دلیل وجود ارتباط فیزیکی قوی بین آن‌ها باشد. همچنین خروجی الگوریتم ویتربی از دو نقطه‌نظر (آماره‌های اقلیمی و دقت پیش‌بینی) با خروجی مولدهای هواشناسی ClimGen و LARS-WG مقایسه شده است که دقت پیش‌بینی الگوریتم مارکف پنهان مبتنی بر تمامی معیارها بسیار بیشتر از دو مولد هواشناسی است. بر این اساس، جایگزینی الگوریتم ارائه شده در تحقیق حاضر با دو رهیافت دیگر، برای تولید توالی داده‌های خشک و تر توصیه می‌گردد.

**واژه‌های کلیدی:** مدل مارکف پنهان، رمزگشایی، توالی خشک و تر، پیش‌بینی

### مقدمه

شبیه‌سازی اقلیم به‌طور عمده برعهده مدل‌های بزرگ‌مقیاس گردش عمومی جو است که البته دارای توان تفکیک اندکی هستند و براین اساس از تکنیک‌های ریزمقیاس‌نما استفاده می‌شود. این تکنیک‌ها ویژگی‌های شبیه‌سازی شده توسط مدل‌های بزرگ‌مقیاس را با روش‌های فیزیکی یا آماری به سطح منطقه‌ای می‌رسانند. روش‌های فیزیکی مبتنی بر معادلات پایستاری جو هستند و روش‌های آماری، از مدل‌های آماری برای این هدف استفاده می‌کنند که این مدل‌ها مبتنی بر توابع احتمالاتی می‌باشند. به هر حال هر دو روش دارای معایب و مزایایی در اجرا هستند (۳).

کارایی یک مدل شبیه‌ساز اقلیم باید از دو منظر بررسی شود: ۱. دقت در شبیه‌سازی ویژگی‌های اقلیمی (مثلاً فراوانی تداوم‌های خشک و تر) و ۲. دقت در شبیه‌سازی حالت یا مقدار رخداد در یک زمان معین (خشک یا تر بودن روز معین). انتظار می‌رود که مدل‌های آماری به دلیل عدم درج واقعیت فیزیکی جو در ساختار خود، دارای

شبیه‌سازی سری زمانی اقلیم پیش‌رو، برای مدیریت بهتر منابع طبیعی ضروری است چراکه اقلیم یکی از مهم‌ترین و پیچیده‌ترین فاکتورهای موثر بر زیست‌محیط و اجزای آن است. در این بین، مدل‌سازی وقوع و یا عدم وقوع بارش به عنوان یک پدیده هواشناسی، همواره مورد توجه محققان قرار دارد و در بسیاری از مطالعات محیطی مانند مطالعات زراعی و هیدرولوژیکی، این نوع مدل‌سازی ضروری است. به عنوان مثال، وقوع بارش منبعی رطوبتی برای گیاه هنگام مبارزه با تنش رطوبتی است و مطلع بودن تقریبی از وقوع آن منجر به آگاهی از ویژگی‌های مربوط به رشد و عملکرد گیاه می‌شود. وظیفه

۱ و ۲- دانشجوی دکتری و استادیار گروه مهندسی آب، پردیس کشاورزی و منابع طبیعی، دانشگاه تهران، کرج

(Email: jbazr@ut.ac.ir)

\*- نویسنده مسئول:

چندمکانی<sup>۳</sup> که وابستگیهای فضایی و زمانی را به طور همزمان در شبیه‌سازی وقوع بارش دخالت میدهد در کار تاوانگ‌کیم و همکاران (۲۲) مشاهده می‌شود.<sup>۴</sup> این وابستگیها در صورت معنی‌داری در تولید سریهای مصنوعی موثر خواهند بود. روش دیگر برای شبیه‌سازی وقوع بارش استفاده از مدل‌های طول دوره<sup>۵</sup> یا فرآیند تکرار متناوب<sup>۶</sup> یا رهیافت ترتیبی<sup>۷</sup> است. سمنوف و همکاران (۱۹) علت توسعه این روش‌ها را ناکارآمدی مدل‌های خانواده مارکوفی در شبیه‌سازی فراوانی تداوم‌های خشک و تر به خصوص تداوم‌های طولانی توصیف کردند. در این روش‌ها توزیع احتمالاتی به طور مستقیم بر طول دروه‌های خشک و تر برازش می‌یابد و در روند شبیه‌سازی، اعداد تصادفی به صورت متناوب از این توزیع‌ها تولید می‌شود. سری نهایی تولید شده ممکن است از قاعده زنجیر مارکوف مرتبه متناهی تبعیت نماید. توزیع-های آماری مختلفی برای این هدف توسط محققین بکار گرفته شده است که شامل توزیع‌های ساده‌تر نظیر توزیع دو جمله‌ای منفی ناقص<sup>۸</sup> توسط بوشاند (۲) و رولاند و وولسیر (۱۸)، توزیع دو جمله‌ای منفی<sup>۹</sup> توسط ویلبای و همکاران (۳۳) و ویلکز (۲۴)، توزیع هندسی مرکب (۱۵) و توزیع‌های ترکیبی اشاره کرد. علت اصلی توسعه چنین روش-هایی که تاکنون در بسیاری از مطالعات مشاهده می‌شود، ناتوانی مدل مارکوف در شبیه‌سازی فراوانی‌های تداوم‌های خشک و تر در برخی موارد بوده است. راسکو و همکاران (۱۵) و سمنوف و همکاران (۱۹) با این نگاه مولد هواشناسی LARS-WG را توسعه دادند. وجه مشترک تمام این روش‌ها برازش یک مدل آماری بر داده‌های باینری وقوع و عدم وقوع و تولید تصادفی است. ساختار تصادفی، وجه تمایز بین مدل‌های آماری تولید و پیش‌بینی اطلاعات می‌باشد. به عبارت دیگر در فرآیند تولید داده که با هدف ریزمقیاس‌نمایی انجام می‌شود، مقدار یا حالت رخداد در زمان معین با توجه به ماهیت تصادفی فرآیند ممکن است به خوبی شبیه‌سازی نشود و مدل در کارایی دوم دچار ضعف گردد. بر این اساس استفاده از الگوهایی که هم‌قادر به پیش‌بینی رخداد واقعی در زمان معین و هم‌قادر به بازگرداندن آماره‌های درازمدت باشد باید مورد توجه قرار گیرد.

یکی از روش‌های آماری کاربردی در بحث پیش‌بینی توالی‌ها، مدل مارکوف پنهان<sup>۱۰</sup> است. استفاده از این روش در پیش‌بینی توالی‌های ژنی، تشخیص گفتار و کاربردهای مشابه مرسوم‌تر است. با

دامنه‌ی بیشتری از خطا به‌خصوص در دقت نوع دوم باشند. دلیل دیگر این مهم را می‌توان در ساختار تصادفی این مدل‌ها بررسی کرد. اما از طرف دیگر، سادگی محاسبات و روابط حاکم بر این مدل‌ها و در دسترس بودن آن‌ها می‌تواند باعث جایگزینی آن‌ها با روش‌های فیزیک-پایه گردد. همچنین دخالت برخی واقعیات فیزیکی اتمسفر در مدل‌سازی آماری می‌تواند بر دقت درازمدت و کوتاه مدت آن بیفزاید. با این وجود، بیشتر مدل‌های آماری موجود، به انتخاب بهترین مدل آماری و تعمیم آن به آینده صرف‌نظر از ویژگی‌های فیزیکی می‌پردازند.

قدمت مدل‌سازی فرآیند وقوع بارش از نقطه نظر احتمالاتی به سال ۱۸۵۲ برمی‌گردد که سری روزهای خشک و تر برای دوره ۱۸۳۳ تا ۱۸۵۰ با فرض وجود خودهمبستگی شبیه‌سازی شدند (۱۱). همچنین ویلیامز (۲۵) و لانگلی (۱۲) از سری‌های هندسی<sup>۱</sup> برای مدل‌سازی تداوم روزهای خشک و تر استفاده کردند. استفاده از خانواده مارکوف در شبیه‌سازی سری‌های باینری وقوع و عدم وقوع بارش سابقه طولانی دارد. برای این هدف، کارکرد مدل زنجیره مارکوف مرتبه نخست در ابتدا توسط گابریل و نیمن (۴) با مدل برنولی (مارکوف مرتبه صفر) براساس آزمون کای اسکور مقایسه شد. در حقیقت این محققین به انتخاب مرتبه مناسب مدل با آزمون مذکور پرداختند. چرا که در یک مدل‌سازی مارکوفی، مرتبه مدل دارای یک دامنه متناهی است و باید با آزمون انتخاب شود. هاپکینز و رابیلارد (۶) احتمال وقوع بارش روزانه را در یکی از استان‌های کانادا براساس آمار ۴۵ ساله بارش روزانه با مدل مارکوف برآورد نمودند. گیتز و تانگ (۵) برتری مارکوف مرتبه دوم را براساس معیار AIC و کتیز (۱۱) برتری مارکوف مرتبه نخست را براساس معیار BIC برای شبیه‌سازی سری باینری وقوع بارش روزانه گزارش کردند. با این وجود، بررسی‌های محققین مختلف نشان داده که مدل مرتبه نخست برای بسیاری از مناطق دنیا مناسب است (۲۰). بر این اساس ریچاردسون (۱۶) و ریچاردسون و رایت (۱۷) از این مدل برای شبیه‌سازی وقوع بارش در مولد هواشناسی WGEN استفاده کردند. پس از آن نیز، بسیاری از مولدهای هواشناسی نظیر CLIGEN, ClimGen, WGENAL برای شبیه‌سازی وقوع بارش مدل مارکوف مرتبه نخست را در ساختار خود جای دادند. ویلکز (۲۴) یک مدل مارکوف هیبرید<sup>۲</sup> با سه احتمال انتقال شرطی را ارائه داد که در آن، مرتبه مدل بسته به نوع حالت روز معین (خشک/تر) متفاوت است به طوری که در صورت تر بودن روز قبل از وابستگی مرتبه نخست و در صورت خشک بودن آن از وابستگی مرتبه دوم استفاده می‌کند. استفاده از یک مدل مارکوف

- 3-Multi-site markov model
- 4-Space-time rainfall occurrence model
- 5-Spell-length model
- 6-Alternating Renewal Process, ARP
- 7-Serial approach
- 8-Truncated negative binomial distribution
- 9-Negative binomial distribution
- 10-Hidden Markov Model, HMM

- 1-Geometric series
- 2-Hybrid-order markov model

متغیرهای مختلف هواشناسی اعم از بارش، تابش و غلظت لایه ازن بکارگرفته شده است. بر این اساس، یک کاربرد اساسی آن در علم هواشناسی می‌تواند پیش‌بینی توالی رخدادهای گسسته جوئی نظیر وقوع و یا عدم وقوع بارش و شبیه‌سازی تداوم‌های خشک و تر به عنوان یک مدل ریزمقیاس نما باشد.

در تحقیق حاضر به ارزیابی کارکرد الگوریتم ویتربی در رمزگشایی حالات پنهان (وقوع و عدم وقوع بارش در یک روز معین) با فرض دسترسی به توالی قابل‌رویت برخی از متغیرهای هواشناسی شامل فشار هوا ( $Press$ )، دمای نقطه شبنم ( $Tdew$ )، اختلاف دمای بیشینه و کمینه ( $DT$ )، رطوبت نسبی ( $RH$ ) و فشار بخار آب ( $Pv$ ) پرداخته شده است. در واقع در تحقیق حاضر این مهم دنبال می‌شود، آیا با دسترسی به یک سری یا توالی زمانی از یک متغیر هواشناسی می‌توان با حل مساله رمزگشایی مدل مارکف پنهان، حالات وقوع و عدم وقوع بارش را پیش‌بینی نمود؟ و دقت پیش‌بینی چه میزان است و اینکه کدامیک از متغیرهای مذکور دقت بیشتری را دارا هستند. با توجه به دسترسی به سری زمانی مشاهده‌شده از همه متغیرهای مورد بحث، مشکلی در برآورد پارامترهای مدل وجود ندارد و در هر بار اجرای مدل بخشی از اطلاعات به واسنجی مدل و بخش دیگر به صحت‌سنجی مدل تعلق می‌یابد. با رمزگشایی توالی پنهان وقوع و عدم وقوع بارش، دستیابی به سری زمانی روزهای خشک و تر میسر است و بر این اساس می‌توان کارکرد رهیافت در شبیه‌سازی فراوانی تداوم‌های خشک و تر را نیز بررسی نمود. برای این منظور نتایج این الگوریتم در هر دو کارایی با رهیافت مارکف مرتبه نخست ( $ClimGen$ ) و رهیافت ترتیبی ( $LARS-WG$ ) مقایسه شده است.

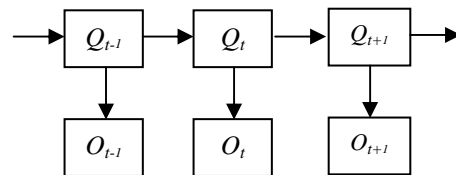
### مواد و روش‌ها

از داده‌های روزانه ایستگاه سینوپتیک خرم‌آباد متعلق به دوره آماری ۱۹۶۱ تا ۲۰۰۵ میلادی برای ارزیابی و مقایسه مدل‌های مورد بحث استفاده شده است. متغیرهای استفاده شده به عنوان توالی قابل‌رویت عبارتند از فشار هوا در ایستگاه، دامنه شبانه‌روزی دمای هوا، رطوبت نسبی، فشار بخار آب و دمای نقطه شبنم که برای دوره ۴۵ ساله مذکور به صورت یک سری زمانی از فایل اطلاعات هواشناسی دریافت شده از سازمان هواشناسی کشور استخراج شدند. بررسی کیفیت دوره آماری و بازسازی داده‌های مفقود با روش‌های آماری انجام و اطلاعات برای ورود به الگوریتم ویتربی مهیا شدند.

### مساله رمزگشایی و الگوریتم ویتربی

یکی از مسائل بنیادین در مدل‌سازی با مارکف پنهان مساله رمزگشایی است که بهترین توالی حالت پنهان متناظر با توالی معین مشاهده  $O$  را به ازای پارامترهای معین مدل جستجو می‌کند. گام

این وجود به دلیل ماهیت گسسته بودن وقوع و عدم وقوع بارش می‌تواند در این مورد نیز کاربردی باشد. در ساده‌ترین حالت مدل مارکف پنهان، دو سری زمانی گسسته (توالی حالات) وجود دارد: ۱- سری زمانی حالات پنهان ( $Q$ ) و ۲- سری زمانی حالات قابل‌رویت ( $O$ ). وجود وابستگی پیاپی مرتبه نخست در توالی حالات پنهان و وابستگی متقاطع همزمان بین حالات پنهان و قابل‌رویت، حداقل فرضیات برای ورود به این مدل است. بر این مبنا دو نوع ماتریس احتمال انتقال و ماتریس احتمال انتشار تعریف می‌شود. شکل ۱، نمای کلی از یک مدل مارکف پنهان با دو لایه است:



شکل ۱- الگوی کلی مدل مارکف پنهان،  $Q$  حالات پنهان و  $O$  حالات قابل رویت،  $t$  زمان رخداد ( $0 \leq t \leq T$ ) که  $T$  طول توالی است.

حل مسائل مربوط به مدل مارکف پنهان شامل مساله ارزیابی (محاسبه احتمال وقوع توالی مشاهده‌ای معین)، مساله رمزگشایی (تعیین محتمل‌ترین توالی حالات پنهان به ازای یک توالی معین مشاهده‌شده) و مساله آموزش مدل (برآورد پارامترهای مدل)<sup>۱</sup>، در ساده‌ترین شکل خود، به روش‌های تحلیلی مشکل و مستلزم صرف وقت و هزینه زیاد است. از این روی، الگوریتم‌های متعددی برای حل این مسائل پیشنهاد شده‌اند. الگوریتم‌های پیشرو<sup>۲</sup> و پسرو<sup>۳</sup> برای حل مساله ارزیابی مدل، الگوریتم ویتربی<sup>۴</sup> برای حل مساله رمزگشایی و رمزگشایی و الگوریتم‌های بام و ولج، امید ریاضی-بیشینه‌سازی و مونت‌کارلو برای حل مساله آموزش ارائه شده‌اند. هر کدام از این الگوریتم‌ها در علم هواشناسی کاربرد خاصی دارند. از الگوریتم‌های حل مساله ارزیابی می‌توان برای محاسبه احتمال وقوع توالی معینی از پدیده‌های هواشناسی استفاده نمود. مساله رمزگشایی نیز چنانچه ذکر شد، به تعیین محتمل‌ترین توالی حالات پنهان به ازای یک توالی مشاهداتی معین می‌پردازد. الگوریتم‌های متعدد رهیافت مارکف پنهان توسط هاگن و گاتورپ (۷) و هاگن و همکاران (۸)، شریکانتن و همکاران (۲۱)، مهرترا و همکاران (۱۳)، ایلویوت و همکاران (۱)، هوکاگلو (۱۰)، هوجانگ و همکاران (۹) و غیره برای مدل‌سازی

- 1-Evaluation, Decoding and Learning problems respectively
- 2-forward
- 3-backward
- 4-viterbi

متغیرهای هواشناسی به صورت خطی بوده و عرض همه طبقات یکسان می‌باشد.

### معیارهای ارزیابی

چنان‌چه در بخش مقدمه به آن اشاره شد، بررسی کارکرد الگوریتم ویتربی به عنوان یک مدل شبیه‌ساز یا پیش‌بینی کننده و مقایسه آن با سایر روش‌ها از دو نقطه نظر می‌باشد: شبیه‌سازی ویژگی‌های اقلیمی سری‌های تر و خشک و قابلیت پیش‌بینی یک رخداد در یک زمان معین. برای ارزیابی ویژگی‌های اقلیمی، از مقایسه فراوانی تداوم‌های تر و خشک و میزان خطای هر مدل و برای ارزیابی توانمندی مدل در پیش‌بینی از معیارهای طبقه‌ای گسسته<sup>۱</sup> استفاده شده است. هر کدام از این معیارها دارای حدود معین و مقداری بهینه هستند که هرگونه انحراف از این مقدار نشان‌دهنده پیش‌بینی اشتباه و یا بیش‌برآوردی و کم‌برآوردی در پیش‌بینی<sup>۲</sup> می‌باشد. تعریف این معیارها، حدود دامنه و فرمول هر معیار در جدول ۱ آورده شده است.

ضرایب موجود در روابط ارائه شده در جدول ۱ توسط ماتریس ۴ آرایه‌ای احتمال وقوع<sup>۳</sup> تعریف می‌شود. در این ماتریس که در شکل ۲ آورده شده است،  $a$  و  $d$  به ترتیب بیانگر تعداد دفعات موفقیت<sup>۴</sup> و رد صحیح<sup>۵</sup>، و  $b$  و  $c$  بیانگر تعداد دفعات شکست مدل هستند. لازم به ذکر است که حالت بهینه تنها در معیارهای ۵ و ۶ به معنای پیش‌بینی کامل است. در حقیقت هرچه توانایی مدل در پیش‌بینی بالاتر باشد انتظار می‌رود که معیار موردنظر به مقدار بهینه نزدیکتر باشد و به این صورت می‌توان قابلیت دو یا چند مدل مختلف را در پیش‌بینی مقایسه نمود. معیار  $B$  یا معیار اریبی نسبت بین رخداد‌های پیش‌بینی شده و رخداد‌های مشاهده شده است و اگر موفقیت‌های پیش‌بینی شده بیشتر از مشاهده شده باشد حالت بیش‌برآوردی و در شرایط عکس حالت کم‌برآوردی اتفاق می‌افتد و تنها به مقادیر حاشیه‌ای ماتریس بستگی دارد. معیار  $POD$  نسبت موفقیت سیستم به رخداد‌های مشاهده شده است که هر چه به یک نزدیکتر باشد به معنای کم‌تر بودن خطای سیستم در بازسازی موفقیت (وقوع و یا عدم‌وقوع) می‌باشد. معیار  $FAR$  کسری از رخداد‌های پیش‌بینی شده است که به‌طور اشتباه به موفقیت تعلق یافته است و هر چه کمتر باشد (به صفر نزدیک‌تر باشد) به معنی کاهش خطا است. معیار  $POFD$  کسری از همه ناموفقیت‌ها در سری مشاهده شده است که سیستم به‌صورت موفقیت پیش‌بینی کرده است و هر چه به صفر نزدیک‌تر باشد خطا کمتر است. معیار

نخست، تعریف توالی حالت بهینه است زیرا چندین معیار بهینگی ممکن وجود دارد. یک تعریف ممکن، توالی حالتی است که از نظر تولید توالی معین مشاهده دارای بالاترین احتمال باشد (۱۴). پس محتمل‌ترین توالی حالت  $Q$  مقدار  $P[Q | O, \lambda]$  را بیشینه می‌کند. برای یک توالی قابل رویت شامل  $T$  رخداد و توالی پنهان با  $N$  حالت، تعداد  $N^T$  توالی ممکن برای  $Q$  وجود خواهد داشت. به عنوان مثال با  $N=3$  و  $T=100$ ، تعداد  $3^{100}$  توالی ممکن وجود خواهد داشت. با این تفسیر حتی برای توالی‌های کوتاه و با تعداد حالات کم نیز، تعداد توالی‌های قابل بررسی زیاد و کار مشکل می‌شود. هم‌چنین در حل این مساله باید از انتخاب توالی نامحتمل پرهیز کرد. یک روش کارآمد برای اجتناب از چنین توالی‌های نامحتمل استفاده از الگوریتم ویتربی است که بر مبنای برنامه‌نویسی پویا استوار است (۱۴). در این الگوریتم دو متغیر  $\delta_t(i)$  و  $\psi_t(j)$  تعریف می‌شوند که به ترتیب عبارتند از، بزرگترین احتمال در مسیر منفردی که برای  $t$  مشاهده اول در نظر گرفته و به حالت پنهان  $t$  ختم می‌شود، و حالت سیستم ( $i$ ) در زمان  $t-1$  که منجر به این مسیر محتمل می‌گردد. این دو کمیت در طی زمان با روابط زیر محاسبه می‌شوند (۱۴):

$$\begin{aligned} \delta_1(i) &= \pi_i b_i(o_1) & 1 \leq i \leq N \\ \psi_1(i) &= 0 \end{aligned} \quad (1)$$

$$\begin{aligned} \delta_t(j) &= \max_{1 \leq i \leq N} \{ \delta_{t-1}(i) p_{ij} \} b_j(o_t) & 1 \leq j \leq N, 2 \leq t \leq T \\ \psi_t(j) &= \arg \max_{1 \leq i \leq N} \{ \delta_{t-1}(i) p_{ij} \} \end{aligned}$$

در این روابط،  $\pi_i, p_{ij}, b_j(o_t)$  به ترتیب عبارتند از: احتمال گسیل مشاهده  $O_t$  از حالت  $i$ ، احتمال انتقال حالت  $i$  به  $j$  و احتمال اولیه حالت  $i$ . در نهایت با پیمایش معکوس، محتمل‌ترین توالی براساس روابط زیر به دست می‌آید:

$$q_t^* = \psi_{t+1}(q_{t+1}^*) \quad t = T-1, T-2, \dots, 1 \quad (2)$$

$$q_T^* = \arg \max_{1 \leq i \leq N} \{ \delta_T(i) \}$$

پیمایش معکوس اجازه می‌دهد تا محتمل‌ترین توالی حالات پنهان تخمین زده شود. برای اجرای الگوریتم ویتربی، سری زمانی به توالی‌های ۵ روزه تقسیم شد و سپس در هر بار اجرا یک توالی برای صحت سنجی مدل و سایر رخداد‌های سری زمانی برای محاسبه پارامترهای مدل یا واسنجی اختصاص یافت. پارامترهای مدل عبارت است از ماتریس‌های احتمال انتقال و احتمال انتشار که از روی داده‌های مربوط به واسنجی مدل متعلق به دوره تاریخی برای هر بار اجرا محاسبه شدند. به دلیل عدم آگاهی از تعداد حالات قابل رویت متغیرهای هواشناسی، کلیه مراحل کار برای ۲ تا ۱۰ تعداد طبقات، انجام گرفت و نتایج برای متغیرها و تعداد حالات قابل‌رویت مختلف مورد ارزیابی قرار گرفت. لازم به ذکر است که طبقه‌بندی توالی

- 1-Discrete categorial criteria
- 2-Overforecast or underforecast
- 3-Contingency table
- 4-Hits
- 5-Correct rejection

همچنین رطوبت نسبی عموماً دارای یک همبستگی مثبت با وقوع بارش است. سایر متغیرها نیز چنین ارتباط فیزیکی را دارا هستند. هرکدام از این متغیرها به طور مجزا به عنوان توالی قابل رویت در نظر گرفته شده و سری گسسته وقوع و عدم وقوع از روی آنها رمزگشایی شده است. با توجه به عدم آگاهی از تعداد حالات لایه قابل رویت (متغیرهای هواشناسی)، این فرآیند برای هر متغیر براساس ۹ تعداد طبقه (از ۲ طبقه تا ۱۰ طبقه) تکرار و هر ۶ معیار برای تمام متغیرها و تمام تعداد طبقات محاسبه شده است. نتیجه محاسبات برای هر متغیر جدولی شده است که در شکل ۳ مشاهده می‌شود.

برای یک ارزیابی کلی از کارکرد توالی قابل رویت متغیرهای هواشناسی، طبق شکل ۳ اختلاف میانگین از مقدار بهینه هر معیار (سطر آخر) محاسبه شد و سپس متوسط کلی آن‌ها به دست آمد. در اینجا می‌توان با یک ارزیابی کلی متغیرهای پرت و نامناسب را شناسایی و آن‌ها را از فرآیند تحلیل کنار گذاشت. در شکل ۴ متوسط این اختلافات (میانگین سطر آخر) برای هر دو نوع موفقیت (وقوع و عدم وقوع) آورده شده است.

ACC دقت مدل در بازسازی رخداد‌های سری مشاهده شده است و مقدار معادل یک به معنی بازسازی کامل سری مشاهده شده طی فرآیند پیش‌بینی است. معیار  $CSI$  نسبت بین موفقیت‌ها و همه رخداد‌های مشاهده شده و یا پیش‌بینی شده است. هرچه این نسبت به یک نزدیک‌تر باشد به معنی کاهش اثر شکست سیستم  $(b,c)$  در فرآیند پیش‌بینی است.

## نتایج

علت انتخاب متغیرهای هواشناسی فشار هوا، رطوبت نسبی، فشار بخار آب، دامنه شبانه‌روزی دمای هوا و دمای نقطه شبنم به عنوان توالی قابل رویت، ارتباط فیزیکی معنی‌دار آن‌ها با وقوع بارش است. تغییرات زمانی فشار هوا در یک نقطه معرف پایداری و ناپایداری هواست، آن چنان‌که تحت شرایط کم‌فشار صعود بسته هوا و وقوع بارندگی و تحت شرایط پرفشار نشست هوا و پایداری جو انتظار می‌رود. اختلاف دمای بیشینه و کمینه به عنوان دامنه اختلاف شبانه‌روزی دمای هوا بسیار متأثر از ابرناکی و یا صاف بودن آسمان است.

جدول ۱- معیارهای ارزیابی سری زمانی پیش‌بینی شده توسط هر مدل‌های مختلف

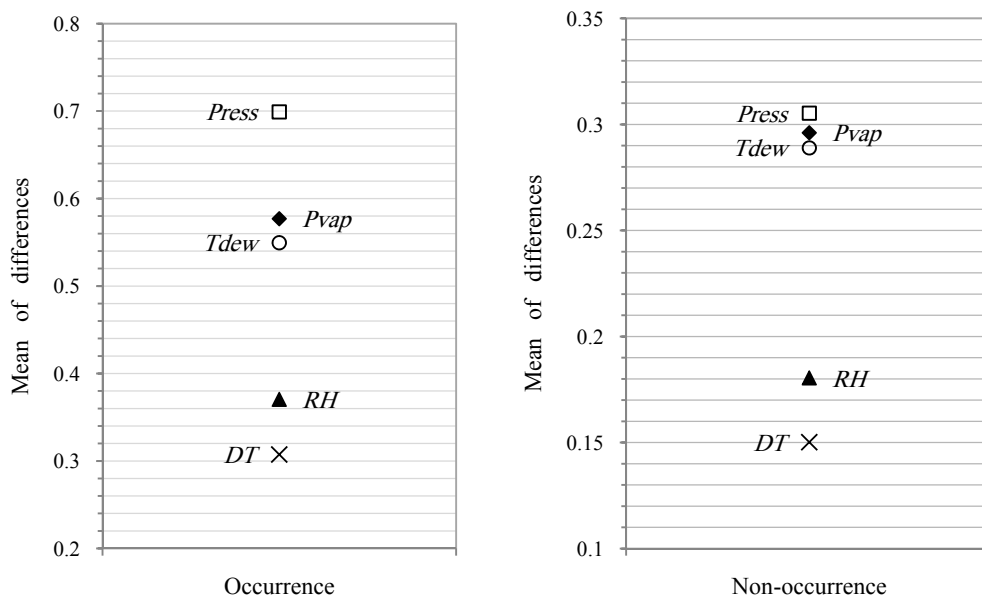
شماره	نام معیار	فرمول	دامنه تغییرات معیار	مقدار بهینه
۱	Bias score	$B = \frac{a+b}{a+c}$	$B < 1$ , underforecast $B > 1$ , overforecast	1
۲	Probability of detection	$POD = \frac{a}{a+c}$	$0 \leq POD \leq 1$	1
۳	False alarm ratio	$FAR = \frac{b}{a+b}$	$0 \leq FAR \leq 1$	0
۴	False alarm rate	$POFD = \frac{b}{b+d}$	$0 \leq POFD \leq 1$	0
۵	Accuracy	$ACC = \frac{a+d}{N}$	$0 \leq ACC \leq 1$	1
۶	Critical success index	$CSI = \frac{a}{a+c+d}$	$0 \leq CSI \leq 1$	1

		Observation		
		Yes	No	
forecast	Yes	$a$	$b$	$a+b$
	No	$c$	$d$	$c+d$
		$a+c$	$b+d$	$N$

شکل ۲- ماتریس احتمال وقوع یک سری باینری

معیارها	<i>B</i>	<i>POD</i>	<i>FAR</i>	...	<i>CSI</i>
تعداد طبقات					
۲					
⋮					
۱۰					
میانگین معیار					
مقدار بهینه	۱	۱	۰	...	۱
اختلاف میانگین از مقدار بهینه					

شکل ۳- جدول به‌دست آمده برای هر توالی قابل رویت



شکل ۴- ارزیابی کلی از کارکرد توالی قابل رویت متغیرهای هواشناسی در شبیه‌سازی سری‌های تر و خشک در ساختار مدل مارکف پنهان

شبم (*Tdew*) دارای یک دقت پایین و متغیرهای رطوبت نسبی (*RH*) و دامنه شبانه روزی دمای هوا (*DT*) برای هر دو حالت وقوع و عدم‌وقوع دارای دقت بالاتری هستند که متغیر دامنه شبانه‌روزی دمای هوا در این میان دارای بالاترین دقت کلی است. همچنین مشاهده می‌شود که دقت کلی مدل در پیش‌بینی حالت عدم وقوع (نمودار سمت راست) بیشتر است چراکه تغییرات آن در دامنه‌های نزدیک‌تر به صفر اتفاق می‌افتد. انتظار می‌رود که چنین موفقیتی خود را در

در نمودارهای شکل ۴ دقت کلی هر کدام از متغیرهای هواشناسی به عنوان لایه قابل رویت در رمزگشایی سری‌های تر و خشک به عنوان لایه پنهان معین شده است. در حقیقت هرچه محور عمودی این نمودارها (متوسط اختلاف میانگین از مقدار بهینه) به صفر نزدیک‌تر باشد نشان‌دهنده کارکرد مناسب‌تر لایه قابل رویت در رمزگشایی لایه پنهان است. چنان‌چه مشاهده می‌شود، متغیرهای هواشناسی فشار هوا (*Press*)، فشاربخار آب (*Pvap*) و دمای نقطه

است.

در شکل ۵ توالی رمزگشایی شده توسط توالی قابل رویت منتخب برای سال ۲۰۰۵ ایستگاه خرم‌آباد به تفکیک ماهانه آورده شده است. در این شکل هر ۱۲ ماه میلادی دارای دو ستون عمودی هستند که ستون اول متعلق به داده‌های تاریخی و ستون دوم متعلق به داده‌های پیش‌بینی شده براساس توالی قابل رویت  $DT$  با سه طبقه گسسته است. سلول‌های خاکستری نشان‌دهنده یک روز تر (وقوع بارش) و سلول‌های سفید نشان‌دهنده یک روز خشک (عدم وقوع بارش) هستند. طبق این شکل، ساختار مارکف پنهان در پیش‌بینی دوره‌های با تداوم بلندتر (چه خشک، چه تر) موفق‌تر است و قادر به پیش‌بینی رخداد‌های نقطه‌ای نمی‌باشد، بطوریکه چنانچه مشاهده می‌شود در ماه می چهار روز تر با تداوم یک روز وجود دارد و مدل نتوانسته آن‌ها را بازگرداند. با این وجود اکثر دوره‌های با تداوم بیش از یک روز با اندکی خطا، به خوبی رمزگشایی شده‌اند. چهار ماه ژوئن تا سپتامبر نیز به صورت کاملاً خشک (بدون دوره تر) در سری ساختگی پیش‌بینی شده‌اند.

شبیه‌سازی فراوانی تداوم‌های خشک و تر نیز نشان دهد. بنابراین متغیر دامنه شبانه‌روزی دمای هوا دارای یک دقت کلی بیشتری نسبت به سایر متغیرها است ( $0/3074$  برای حالت وقوع و  $0/1501$  برای حالت عدم وقوع) و بنابراین نتایج بیشتر این متغیر در ادامه آورده شده است. در جداول ۲ و ۳ مقادیر به دست آمده برای هر معیار و تعداد طبقات مختلف متغیر دامنه شبانه‌روزی دمای هوا به عنوان متغیر قابل رویت برای دو حالت موفقیت وقوع (جدول ۲) و عدم وقوع (جدول ۳) مشاهده می‌شود. در یک نگاه کلی، کمترین انحرافات از مقدار بهینه برای هر دو حالت موفقیت متعلق به تعداد ۳ طبقه است که برای حالت موفقیت وقوع بارندگی (جدول ۲) همه معیارها به استثنای معیار  $POD$  و برای حالت موفقیت عدم وقوع (جدول ۳) همه معیارها به استثنای دو معیار  $FAR$  و  $POFD$  این نکته را نشان می‌دهد. با این وجود، معیار  $POFD$  متعلق به تعداد ۳ طبقه نیز فاصله کمی تا مقدار بهینه معیار مذکور دارد. بنابراین نتیجه‌گیری می‌شود که متغیر دامنه شبانه‌روزی دمای هوا با سه طبقه گسسته بهترین متغیر به عنوان توالی قابل رویت برای رمزگشایی توالی پنهان روزهای تر و خشک

جدول ۲- مقادیر معیارهای طبقه‌ای براساس کارکرد مدل با لایه قابل رویت دامنه شبانه‌روزی دمای هوا (وقوع)

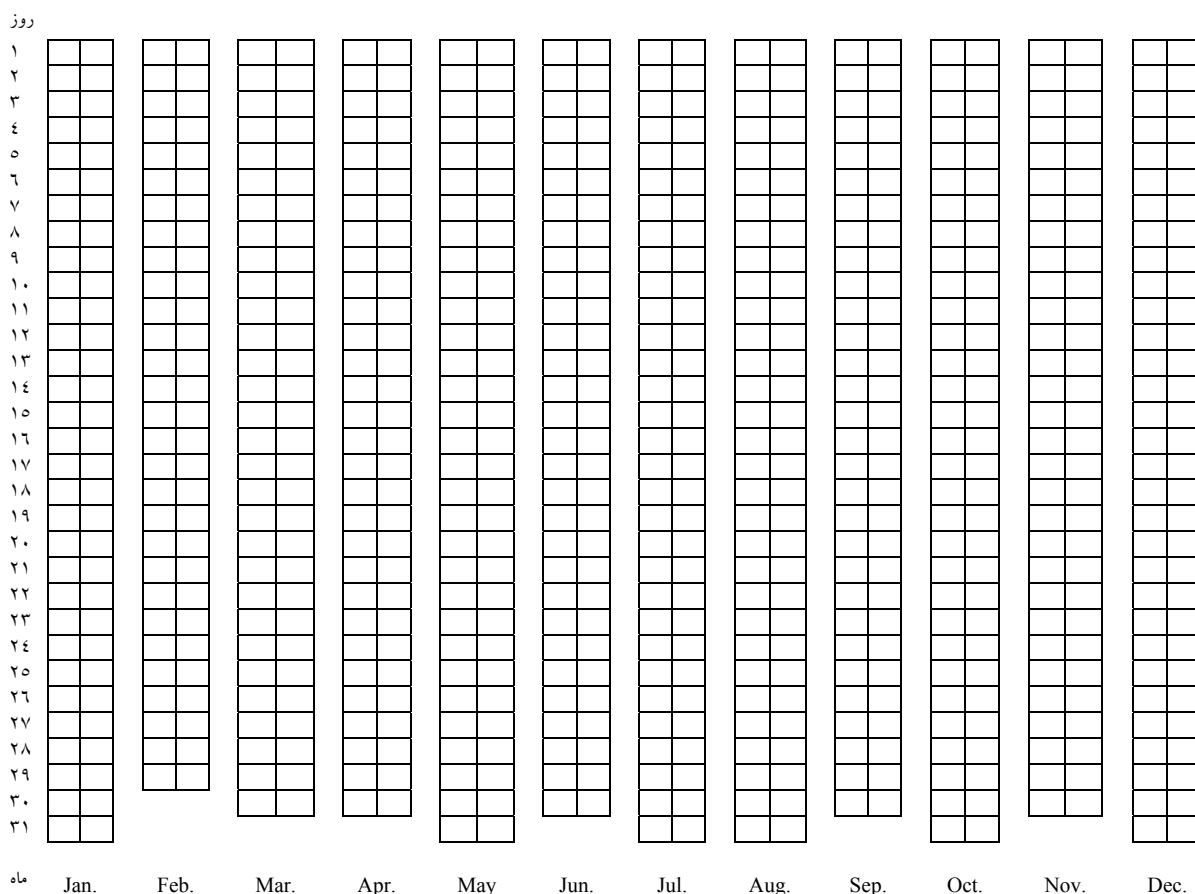
تعداد طبقات	معیارها					
	$B$	$POD$	$FAR$	$POFD$	$ACC$	$CSI$
۲	۱/۶۷۱	۰/۷۸۲	۰/۵۳۲	۰/۲۱۷	۰/۷۸۳	۰/۴۱۴
۳	۰/۹۵۹	۰/۶۵۶	۰/۳۱۶	۰/۰۷۴	۰/۸۷۳	۰/۵۰۳
۴	۱/۶۲۵	۰/۷۸۴	۰/۵۱۷	۰/۲۰۵	۰/۷۹۳	۰/۴۲۶
۵	۱/۲۰۴	۰/۷۲۷	۰/۳۹۶	۰/۱۱۶	۰/۸۵۳	۰/۴۹۲
۶	۱/۱۸۴	۰/۶۹۹	۰/۴۰۹	۰/۱۱۸	۰/۸۴۶	۰/۴۷۱
۷	۱/۳۴۳	۰/۷۶۸	۰/۴۲۸	۰/۱۱۴	۰/۸۴۲	۰/۴۸۸
۸	۱/۲۶۹	۰/۷۴۹	۰/۴۱	۰/۱۲۷	۰/۸۴۹	۰/۴۹۲
۹	۱/۳۶۱	۰/۷۷۷	۰/۴۳	۰/۱۴۳	۰/۸۴۲	۰/۴۹
۱۰	۱/۳۳۹	۰/۷۶۱	۰/۴۲۸	۰/۱۳۸	۰/۸۴۲	۰/۴۸۵

جدول ۳- مقادیر معیارهای طبقه‌ای براساس کارکرد مدل با لایه قابل رویت دامنه شبانه‌روزی دمای هوا (عدم وقوع)

تعداد طبقات	معیارها					
	$B$	$POD$	$FAR$	$POFD$	$ACC$	$CSI$
۲	۰/۸۳۷	۰/۷۸۳	۰/۰۶۴	۰/۲۱۸	۰/۷۸۳	۰/۷۴۴
۳	۱/۰۱	۰/۹۲۶	۰/۰۸۳	۰/۳۴۴	۰/۸۷۳	۰/۸۵۵
۴	۰/۸۴۸	۰/۷۹۵	۰/۰۶۲	۰/۲۱۶	۰/۷۹۳	۰/۷۵۵
۵	۰/۹۵	۰/۸۸۴	۰/۰۷	۰/۲۷۳	۰/۸۵۳	۰/۸۲۹
۶	۰/۹۵۵	۰/۸۸۲	۰/۰۷۷	۰/۳۰۱	۰/۸۴۶	۰/۸۲۲
۷	۰/۹۱۶	۰/۸۶	۰/۰۶۲	۰/۲۳۲	۰/۸۴۲	۰/۸۱۴
۸	۰/۹۳۵	۰/۸۷۳	۰/۰۶۶	۰/۲۵۱	۰/۸۴۹	۰/۸۲۳
۹	۰/۹۱۲	۰/۸۵۷	۰/۰۶	۰/۲۲۳	۰/۸۴۲	۰/۸۱۳
۱۰	۰/۹۲	۰/۸۶۲	۰/۰۶۳	۰/۲۳۹	۰/۸۴۲	۰/۸۱۴

برای استفاده از نتایج رهیافت ترتیبی در شبیه‌سازی تداوم‌های خشک و تر از مولد LARS-WG براساس داده‌های تاریخی، ۴۵ سال خروجی گرفته شد. در شکل ۶ نمودار فراوانی تداوم‌های تر و خشک برای هر سه مدل شبیه‌ساز اقلیم و سری زمانی تاریخی یا مشاهده-شده نشان داده شده است. حداکثر طول تداوم‌های مورد مقایسه ۱۵ روز است، چراکه با افزایش طول دوره، فراوانی نسبی روندی نزولی دارد که در تداوم‌های بزرگتر نزدیک به صفر می‌رسد.

برای آگاهی از میزان دقت مدل در برخورد با سایر روش‌های مرسوم مانند رهیافت مارکف ساده (ریچاردسون، ۱۹۸۱) و رهیافت ترتیبی (سمنوف، ۱۹۹۸) از نظر هر دو کارایی مطروح، نتایج این روش‌ها نیز در مقابل مدل مارکف پنهان با توالی پنهان دو طبقه‌ای (تر و خشک) و توالی قابل رویت سه طبقه‌ای (دامنه شبانه روزی دمای هوا) در این قسمت آورده شده است. برای استفاده از نتایج مدل مارکف ساده یا مارکف مرتبه نخست از مولد هواشناسی ClimGen و



شکل ۵- توالی رمزگشایی شده خشک و تر در کنار توالی تاریخی متعلق به سال ۲۰۰۵ میلادی، ایستگاه خرم‌آباد

های خشک مثلاً تداوم‌های ۲، ۳، ۶ و ۱۰ روزه مدل مارکف پنهان دارای بیشترین دقت است. در مورد دوره‌های یک روزه تر، مدل مارکف ساده دارای دقت بالاتر، و دو مدل دیگر چهار کم‌برآوردی هستند که این کم‌برآوردی برای مارکف پنهان نسبت به مدل LARS-WG کمتر است. بیشترین دقت در مورد تداوم‌های ۲ روزه تر مربوط به الگوریتم ویتربی است. اما این الگوریتم در شبیه‌سازی تداوم‌های بزرگتر بیش‌برآورد بوده و کمترین دقت را دارد. با این وجود، روند کلی تغییرات را بخوبی شبیه‌سازی نموده است. بنابراین در یک

این مطلب به خصوص برای تداوم‌های تر مشهودتر می‌باشد. در شبیه‌سازی دوره‌های یک‌روزه خشک، هر سه مدل دارای کم‌برآوردی هستند که این موضوع برای الگوریتم مربوط به مارکف پنهان در شکل ۵، نیز مشاهده گردید و علت آن را می‌توان در ساختار الگوریتم ویتربی جست‌وجو کرد. در شبیه‌سازی دوره‌های خشک بیش از یک روز، دو رهیافت ترتیبی و مارکف پنهان دارای دقت بیشتری نسبت به مدل مارکف ساده هستند که این موضوع دارای شدت و ضعف برای هر کدام از این دو مدل در تداوم‌های مختلف است. در برخی از تداوم-



می‌گردد (وضعیت دوم). علت نزدیکی مقادیر  $b$  و  $c$  را می‌توان در ساختار تصادفی این دو روش جست‌وجو کرد که در درازمدت تعداد اشتباهات نوع اول و دوم با هم برابر می‌گردد. مقادیر  $a$ ،  $b$  و  $c$  ساخته شده توسط مدل مارکف پنهان به ترتیب معادل ۱۲۲۳۲، ۱۱۰۷ و ۹۷۵ است که مشاهده می‌شود در حالیکه مقدار معیار  $B$  برای این مدل اندکی کمتر از دومولد است، ولی تعداد موفقیت‌ها ( $a$ ) بیشتر و تعداد اشتباهات ( $b$  و  $c$ ) در مدل مورد بحث کمتر است. بنابراین می‌توان نتیجه گرفت که مدل مارکف پنهان در پیش‌بینی رخداد در زمان معین از نظر همه معیارها، نسبت به دو مولد هواشناسی کاربردی، دارای توانمندی قابل توجهی است و این امر می‌تواند سبب برتری الگوریتم رمزگشایی نسبت به دو رهیافت دیگر برای ریزمقیاس نمایی سری زمانی وقوع و عدم‌وقوع گردد.

### نتیجه‌گیری و بحث

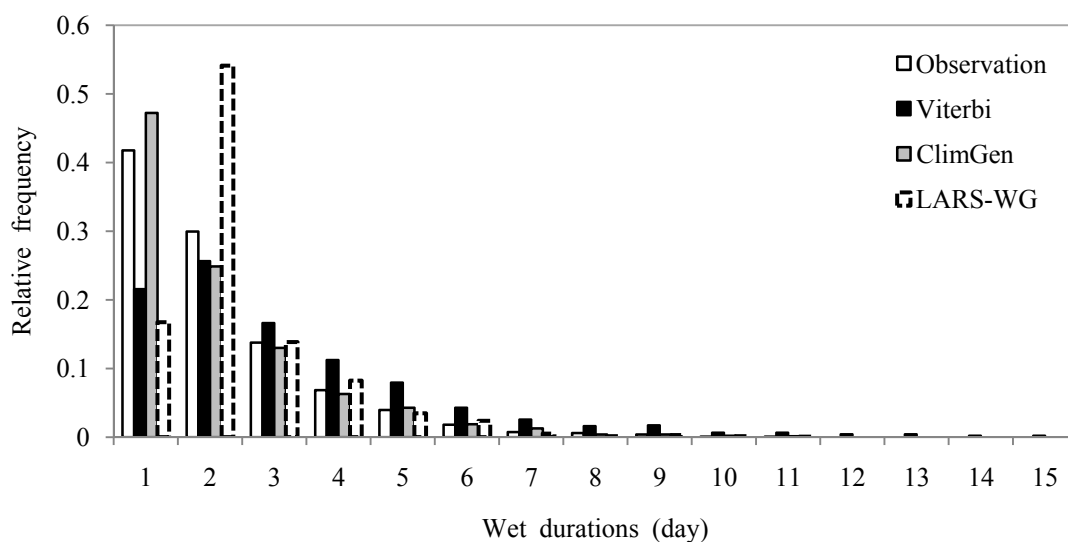
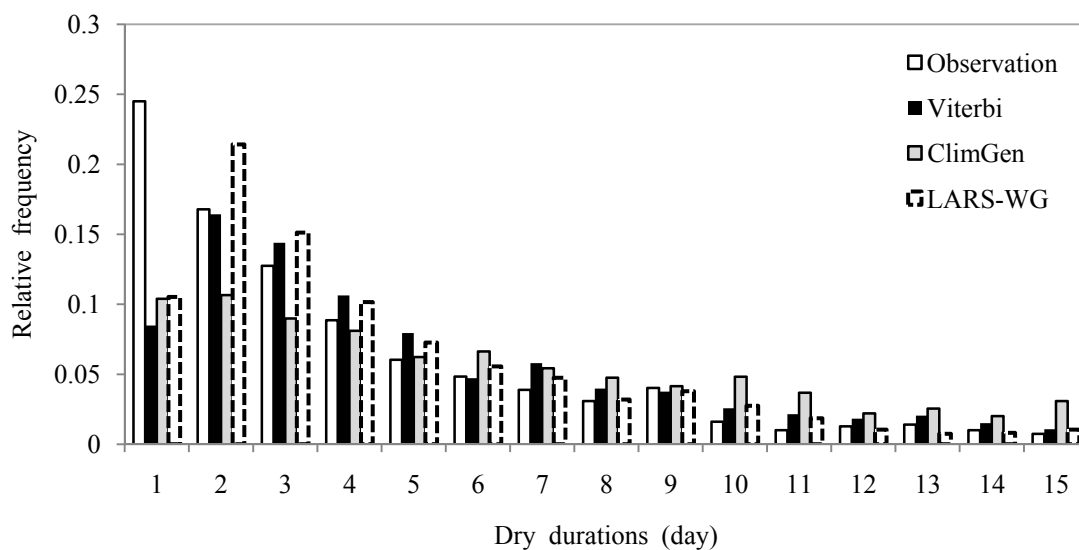
در این مطالعه دو هدف مهم دنبال شده است: (۱) رمزگشایی توالی‌های تر و خشک با استفاده از الگوریتم ویتربی و انتخاب متغیر هواشناسی مناسب به عنوان توالی قابل رویت و (۲) بررسی کارکرد رهیافت منتخب از مرحله یک و مقایسه آن با رهیافت‌های مرسوم در شبیه‌سازی اقلیم و پیش‌بینی رخداد در زمان معین. در مورد هدف اول براساس معیارهای مختلف طبقه‌ای، متغیر دامنه شبانه‌روزی دمای هوا با سه طبقه به عنوان توالی قابل رویت انتخاب شد. انتخاب این متغیر می‌تواند به دلیل ارتباط فیزیکی قوی بین آن و ابرناکی که ممکن است منجر به بارندگی شود، باشد.

نگاه کلی می‌توان کارکرد سه مدل را مناسب دانست و برای هر کدام معایب و مزایایی تعیین نمود.

برای ارزیابی کارکرد سه مدل در ساختن رخداد در زمان معین، در جدول ۴ نتایج اجرای معیارهای طبقه‌ای بر دو مولد هواشناسی و الگوریتم مارکف پنهان آورده شده است. مشاهده می‌شود که الگوریتم ویتربی مربوط به مارکف پنهان از نظر پنج معیار  $ACC$ ،  $CSI$ ،  $POFD$ ،  $FAR$  و  $POD$  دارای بالاترین دقت در پیش‌بینی حالت وقوع یا عدم‌وقوع و به عبارتی دارای کمترین انحراف از مقدار بهینه است که تفاوت‌ها با دو مدل دیگر در برخی معیارها بسیار فاحش می‌باشد. در مورد معیار  $B$  هر سه مدل دارای مقداری بسیار نزدیک به مقدار بهینه هستند که اختلاف این معیار با مقدار بهینه برای مولدهای هواشناسی کمتر است. در اینجا ذکر این نکته الزامی است که در مورد معیار  $B$ ، مقدار بهینه یک، همیشه به معنی پیش‌بینی کامل یا دقت کامل مدل نمی‌باشد. با توجه به رابطه معیار  $B$  در جدول ۱، در دو وضعیت مقدار این معیار معادل یک می‌شود: (۱) اگر مقادیر  $b$  و  $c$  صفر گردد، که به معنی پیش‌بینی کاملاً درست توسط مدل است، و (۲) اگر مقادیر  $b$  و  $c$  با هم برابر باشد، که در اینجا حتی مقدار  $a$  می‌تواند صفر هم باشد. بنابراین نکته، در مورد معیار  $B$  باید بررسی نمود که آیا نزدیکی به مقدار بهینه در اثر وضعیت اول است یا وضعیت دوم. توجه به مقادیر  $a$ ،  $b$  و  $c$  می‌تواند این مطلب را روشن نماید. مقادیر  $a$ ،  $b$  و  $c$  ساخته شده توسط مولد ClimGen به ترتیب معادل ۱۰۶۵۷، ۲۵۸۹ و ۲۵۵۰ و برای مولد LARS-WG به ترتیب معادل ۱۰۹۲۰، ۲۲۵۵ و ۲۲۸۷ می‌باشد. مشاهده می‌شود که مقادیر  $b$  و  $c$  بسیار نزدیک به هم هستند و این موضوع باعث نزدیک شدن معیار به مقدار بهینه خود

جدول ۴- مقایسه معیارهای مدل منتخب مارکف پنهان و رهیافت‌های تریبی و ریچاردسون در بازگرداندن رخداد در زمان معین

نوع موفقیت معیار مدل	عدم وقوع بارش						وقوع بارش					
	$B$	$POD$	$FAR$	$POFD$	$ACC$	$CSI$	$B$	$POD$	$FAR$	$POFD$	$ACC$	$CSI$
Viterbi	۱/۰۱	۰/۹۲۶	۰/۰۸۳	۰/۳۴۴	۰/۸۷۳	۰/۸۵۵	۰/۹۵۹	۰/۶۵۶	۰/۳۱۶	۰/۰۷۴	۰/۸۷۳	۰/۵۰۳
ClimGen	۱/۰۰۳	۰/۸۰۷	۰/۱۹۵	۰/۸۰۵	۰/۶۸۷	۰/۶۷۵	۰/۹۸۸	۰/۱۹۵	۰/۸۰۲	۰/۱۹۳	۰/۶۸۷	۰/۱۰۹
LARS-WG	۰/۹۹۸	۰/۸۲۷	۰/۱۷۱	۰/۷۰۱	۰/۷۲۳	۰/۷۰۶	۱/۰۱	۰/۲۹۹	۰/۷۰۴	۰/۱۷۳	۰/۷۲۳	۰/۱۷۵



شکل ۶- فراوانی تداوم‌های خشک و تر برای داده‌های مشاهده‌شده و سه رهیافت

نوع متغیر مناسب بررسی گردد. بر این اساس توصیه می‌شود که کارکرد الگوریتم ویتربی و نوع متغیر قابل رویت در تعداد ایستگاه‌های بیشتر و با اقلیم متفاوت بررسی شود و این موضوع روشن گردد که آیا می‌توان در همه ایستگاه‌ها بر روی یک متغیر هواشناسی به عنوان توالی قابل رویت به نتیجه رسید یا خیر، و اینکه آیا در همه اقلیم اعم از بیابانی تا بسیار مرطوب مدل مارکف پنهان دارای جواب بهینه است یا خیر. همچنین عرض طبقات در توالی‌های قابل رویت به صورت خطی فرض شده است، به عبارتی دیگر فرض می‌شود که فراوانی طبقات از توزیع یکنواخت تبعیت می‌کند و این فرض در برخی موارد ممکن است صحیح نباشد. به عنوان مثال اگر فراوانی طبقات از توزیع نرمال چوله به راست تبعیت نماید بدین معنی است که مقادیر کمتر

چنانچه جذب تابش مادون قرمز زمین توسط ابرها باعث حبس انرژی ساطع از زمین در جو می‌گردد و این امر به کاهش اختلاف دمای شب و روز کمک می‌کند. در هوای صاف، تابش دریافتی توسط زمین به سرعت از جو زمین خارج شده و این اختلاف تشدید می‌شود. بنابراین یک مبنای فیزیکی در ساختار مدل مارکف پنهان مورد بحث نهفته است که آنرا از سایر روش‌های آماری متمایز می‌سازد. با این مبنا می‌توان برای چشم‌انداز آتی دوره‌های تر و خشک را پیش‌بینی نمود که این امر در ریزمقیاس‌نمایی متغیرهای هواشناسی کاربرد وسیعی دارد. می‌توان گفت که انتخاب متغیر هواشناسی برای توالی قابل رویت یک موضوع مکان-ویژه<sup>۱</sup> است و باید برای هر ایستگاهی

1-site-specific

تداوم‌های خشک و تر دارای قوت و ضعف هستند و مدل مارکف پنهان در برخی موارد دارای برتری و در برخی موارد دارای خطای بیشتر است. با این تفاوت که مدل مارکف پنهان در پیش‌بینی رخداد در زمان معین دارای توانمندی قابل توجهی است و سایر روش‌ها در بیشتر معیارها دارای مقادیر پرت می‌باشند که کارکرد آن‌ها را در این مورد زیر سوال می‌برد. بنابراین استفاده از این مدل به عنوان جایگزینی برای سایر رهیافت‌ها در مدل‌های ریزمقیاس‌نمایی پیشنهاد می‌گردد. در شبیه‌سازی اقلیم برای چشم‌انداز آتی به دلیل عدم دسترسی به سری داده‌های تر و خشک برآورد ماتریس‌های احتمال انتقال و احتمال انتشار به راحتی اقلیم گذشته نمی‌باشد. به این دلیل یا باید به پارامترهای برآوردشده از سری تاریخی بسنده نمود و یا با روش‌های برآورد پارامتر نظیر روش بیشینه‌سازی-امیدریاضی به تخمین پارامترها پرداخت که در مطالعات بعدی توصیه می‌شود.

متغیر دارای فراوانی بزرگتر و مقادیر بیشتر دارای فراوانی کوچکتر در سری تاریخی هستند که بر این اساس عرض طبقات در مقادیر کمتر باید کوچکتر در نظر گرفته شود. از آنجایی که این مساله می‌تواند به دقت کارکرد مدل بیفزاید، توصیه می‌شود که در طبقه‌بندی توالی قابل رویت به توزیع فراوانی آن توجه شود.

مقایسه رهیافت مارکف پنهان ارائه شده در این تحقیق با دو رهیافت مارکف ساده و رهیافت ترتیبی در بخش دوم دنبال شد که این مقایسه در دو نوع کارایی بررسی شد. کارایی اول در شبیه‌سازی ویژگی‌های اقلیمی و کارایی دوم در پیش‌بینی رخداد در زمان معین. در واقع پاسخ به این سوال دنبال شد که آیا می‌توان به مدل شبیه‌سازی دست یافت که علاوه بر پیش‌بینی رخداد بتواند ویژگی‌های درازمدت را مناسب شبیه‌سازی کند؟ مشاهده شد که هر سه رهیافت مارکف ساده، مارکف پنهان و ترتیبی در شبیه‌سازی فراوانی نسبی

## منابع

- 1- Ailliot P., Thompson C. and Thompson P. 2008. Space time modelling of precipitation using a hidden markov model and censored Gaussian distributions. *Journal of Hydrology*, 420:27-51.
- 2- Buishand T.A. 1978. Some remarks on the use of daily rainfall models. *Journal of Hydrology*, 36: 295-308.
- 3- Busuioc A., Tomozeiu R. and Cacciamani C. 2007. Statistical downscaling model based on canonical correlation analysis for winter extreme precipitation events in the Emilia-Romagna region. *International journal of climatology* (in press).
- 4- Gabriel K.R. and Neumann J. 1962. A markov chain model for daily rainfall occurrence at Tel Aviv. *Quarterly journal of the royal meteorological society*, 88(375): 90-95.
- 5- Gates P. and Tong H. 1976. On markov chain modeling to some weather data. *Journal of applied meteorology*, 15: 1145-1151.
- 6- Hopkins J.W. and Robillard P. 1964. Some statistics of daily rainfall occurrence for the Canadian Prairie Provinces. *Journal of applied meteorology*, 3: 104-118.
- 7- Hughes J.P. and Guttorp P. 1994. A class of stochastic models for relating synoptic atmospheric patterns to local hydrologic phenomenon. *Water Resources Research*, 30:1535-1546.
- 8- Hughes J.P., Guttorp P. and Charles S.P. 1999. A non-homogeneous hidden Markov model for precipitation occurrence. *Applied Statistics*, 48(1):15-30.
- 9- Hao Zhang W., Zhang A. and Palazoglu W.S. 2012. Prediction of ozone levels using a hidden markov model (HMM) with gamma distribution. *Atmospheric Environment*, 62: 64-73.
- 10- Hocaoglu F.O. 2011. Stochastic approach for daily solar radiation modeling. *Journal of Solar Energy*, 85: 278-287.
- 11- Katz R.W. 1981. On some criteria for estimating the order of a markov chain. *Technometrics*, 23: 243-249.
- 12- Longley R.W. 1953. The length of dry and wet periods. *Quarterly journal of the royal meteorological society*, 79: 520-527.
- 13- Mehrotra R., Srikanthan R. and Sharma A. 2006. A comparison of three stochastic multi-site precipitation occurrence generators. *Journal of Hydrology*, 331: 280-292.
- 14- Oliver C. Ibe 2009. *Markov process for stochastic modeling*. John Wiley & Sons, New York.
- 15- Racsko P., Szeidl L. and Semenov M. 1991. A serial approach to local stochastic weather models. *Ecological Modeling*, 57: 27-41.
- 16- Richardson C.W. 1981. Stochastic simulation of daily precipitation, temperature, and solar radiation. *Water Resources Research*, 17: 182-190.
- 17- Richardson C.W. and Wright D.A. 1984. WGEN: A model for generating daily weather variables. *USDA/ARS, ARS-8*, 83 pp.
- 18- Rolaand J. and Woolhiser D.A. 1982. Stochastic daily precipitation models: 1) a comparison of occurrence processes. *Water Resources Research*, 18: 1451-1459.
- 19- Semenov M.A., Brooks R.J., Barrow E.M. and Richardson C.W. 1998. Comparison of the WGEN and LARS-WG stochastic weather generators for diverse climates. *Climate Research*, 10: 95-107.
- 20- Srikanthan R., and McMahon T.A. 2001. Stochastic generation of annual, monthly and daily climate data: A review. *Hydrology and Earth System Sciences*, 5(4): 653-670.

- 21- Srikanthan R., Thyer M.A., Kuczera G.A. and McMahon T.A. 2002. Application of hidden state markov model to Australian annual rainfall data. CRCCH Working Document (in press).
- 22- Tae-woongK., Hosung A. and Gunhui C.Ch.Y. 2008. Stochastic multi-site generation of daily rainfall occurrence in south Florida. Stochastic environment research risk assessment, 22:705-717.
- 23- Wilby R.L., Wigley T.M.L., Conway D., Jones P.D., Hewitson B.C., Main J. and Wilks D.S. 1998. Statistical downscaling of general circulation model output: A comparison of methods. Water Resources Research, 34: 2995-3008.
- 24- Wilks D.S. 1999. Interannual variability and extreme-value characteristics of several stochastic daily precipitation models. Journal of Agricultural Meteorology, 93: 153-169.
- 25- Williams C.B. 1952. Sequences of wet and of dry days considered in relation to the logarithmic series. Quarterly Journal of the Royal Meteorological Society, 78: 511-516.

## Application of Viterbi's Algorithm for Predicting Rainfall Occurrence and Simulating Wet\Dry Spells – Comparison with Common Methods

M. Ghamghami<sup>1</sup>- J. Bazrafshan<sup>2\*</sup>

Received:09-08-2013

Accepted:29-09-2014

### Abstract

Today, there are various statistical models for the discrete simulation of the rainfall occurrence/non-occurrence with more emphasizing on long-term climatic statistics. Nevertheless, the accuracy of such models or predictions should be improved in short timescale. In the present paper, it is assumed that the rainfall occurrence/non-occurrence sequences follow a two-layer Hidden Markov Model (HMM) consist of a hidden layer (discrete time series of rainfall occurrence and non-occurrence) and an observable layer (weather variables), which is considered as a case study in Khoramabad station during the period of 1961-2005. The decoding algorithm of Viterbi has been used for simulation of wet/dry sequences. Performance of five weather variables, as the observable variables, including air pressure, vapor pressure, diurnal air temperature, relative humidity and dew point temperature for choosing the best observed variables were evaluated using some measures of error evaluation. Results showed that the variable of diurnal air temperature is the best observable variable for decoding process of wet/dry sequences, which detects the strong physical relationship between those variables. Also the Viterbi output was compared with ClimGen and LARS-WG weather generators, in terms of two accuracy measures including similarity of climatic statistics and forecasting skills. Finally, it is concluded that HMM has more skills rather than the other two weather generators in simulation of wet and dry spells. Therefore, we recommend the use of HMM instead of two other approaches for generation of wet and dry sequences.

**Keywords:** HMM, Decoding, Wet/Dry Spells, Prediction

---

1,2-PhD Student and Assistant Professor of Agrometeorology, Department of Water Engineering, College of Agricultural and Natural Resources, University of Tehran  
(\*- Corresponding Author Email: jbazr@ut.ac.ir)