

پیش بینی دمای سطح آب خلیج فارس با استفاده از رگرسیون چندگانه و تحلیل مؤلفه‌های اصلی

سید محمد جعفر ناظم‌السادات^۱ و امین شیروانی^۲

چکیده

با توجه به آن که نوسانات دمای سطح آب خلیج فارس (PGSST) (Persian Gulf Sea Surface Temperature) تأثیر قابل ملاحظه‌ای بر بارش زمستانه، منابع آب و تولیدات کشاورزی نواحی جنوب غربی کشور دارد، امکان پیش بینی دمای زمستانه دمای سطح آب این گستره آبی با استفاده از مدل رگرسیون چندگانه مورد ارزیابی قرار گرفت. سری‌های زمانی PGSST برای کلیه فصول در دوره ۱۹۹۲-۱۹۴۷ به عنوان پیشگوکننده و سری زمانی این متغیر برای دوره ۱۹۹۳-۱۹۴۸ به عنوان پیشگو شونده در نظر گرفته شدند. تحلیل مؤلفه‌های اصلی (PCA) (Principal Component Analysis) به منظور استخراج مؤلفه‌های اصلی و کاهش حجم داده‌ها به کار برده شد. چهار سری زمانی جدید (PC1، PC2، PC3 و PC4) که ۷۳/۵٪ از واریانس کل را شرح دادند به عنوان مؤلفه‌های اصلی انتخاب شده و بقیه به عنوان اختلال در نظر گرفته شد. دوران واریامکس به منظور یافتن وابستگی هر مؤلفه اصلی با سری‌های زمانی اولیه PGSST استفاده شد. این دوران نشان داد که PC1، PC2، PC3 و PC4 به ترتیب معرف تغییرات دما در فصل زمستان، پاییز، بهار و تابستان می‌باشند. در مدل رگرسیونی، مؤلفه‌های اول، دوم و چهارم در سطح ۵٪ معنی‌دار شدند و مؤلفه سوم معنی‌دار نگردید. نتایج نشان داد که متغیرهای معنی‌دار، ۳۳/۵٪ از واریانس PGSST زمستانه را شرح می‌دهند. روشن شد که جهت پیش بینی دمای زمستانه سطح آب خلیج فارس، دمای سطح آب این گستره آبی در زمستان سال قبل از اهمیت ویژه‌ای برخوردار است. در مرحله بعدی دمای پاییزه و تابستانه برای پیش‌بینی دمای زمستانه نقش دارند.

واژه‌های کلیدی: دمای سطح آب خلیج فارس، رگرسیون چندگانه، تحلیل مؤلفه‌های اصلی

مقدمه

بسیار زیادی بر شرایط آب و هوایی بسیاری از نقاط ایران مخصوصاً مناطق جنوبی دارد. تقریباً در تمامی مواقعی که استان فارس و مناطق جنوبی کشور همراه با بارش فراوانی بوده است ابرهای ورودی از مرزهای جنوب غربی و خلیج فارس وارد

آب‌های خلیج فارس علاوه بر آن که از دیدگاه‌های تجاری، نظامی، شیلات و آبیان، صنایع و تأسیسات دریایی نقش بسیار مهمی در حیات اقتصادی-اجتماعی ایران ایفا می‌نمایند، تأثیر

۱. دانشیار مهندسی آب، دانشکده کشاورزی، دانشگاه شیراز

۲. کارشناس ارشد مرکز مطالعات اقلیمی، دانشکده کشاورزی، دانشگاه شیراز

بارش این فصل اهمیت بالایی دارد. نتایج مربوط به مجموعه مطالعات انجام شده مؤید این نکته می‌باشند که درک مکانیزم حاکم بر نوسانات SST در پهنه آب‌های جنوبی کشور می‌تواند برای پیش‌بینی‌های دراز مدت بارش، ارزیابی بارش و شناخت وضعیت منابع آب و فعالیت‌های زیست محیطی در نواحی جنوب غربی کشور مورد استفاده قرار گیرد.

گرچه تحقیقات گذشته بر اهمیت دمای سطح آب زمستانه خلیج فارس بر بارش و شرایط زیست محیطی مناطق جنوبی و جنوب غربی کشور تأکید دارند، پیش‌بینی مقدار این دما (PGSST) کمتر مورد توجه بوده است. این پیش‌بینی برای ارزیابی و نیز برنامه ریزی وضعیت منابع آب در مناطق جنوب غرب کشور از اهمیت فوق‌العاده‌ای برخوردار است. به منظور اطلاع از شرایط بارش زمستانه مناطق جنوبی کشور لازم است که دمای سطح آب خلیج فارس از چند ماه قبل پیش‌بینی شود تا بر اساس این پیش‌بینی بتوان برآوردی از مقدار بارش زمستانه به دست آورد. بنابراین، هدف اصلی این پژوهش آن است که با در نظر گرفتن وضعیت SST در یک سال گذشته نسبت به پیشگویی آن در فصل زمستان آینده اقدام نمود. این نوع بررسی‌ها اخیراً مورد توجه مراکز علمی پیش‌جهان در آمریکا و اروپا قرار گرفته است (۸). با وجود جستجوی فراوان، نگارندگان موفق نشدند که مقاله‌ای علمی که حاکی از انجام این گونه تحقیقات در ایران باشد پیدا نمایند.

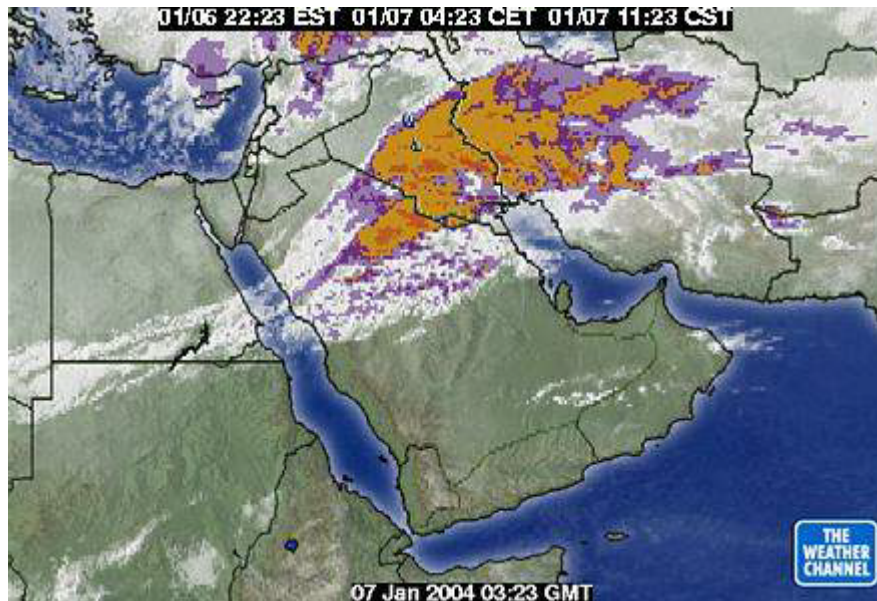
مواد و روش‌ها

داده‌ها

داده‌های ماهانه دمای سطح آب در پهنه خلیج فارس برای سال‌های ۱۹۹۳-۱۹۴۷ میلادی از بانک‌های اطلاعاتی سازمان NOAA (National Oceanic and Atmospheric Administration) استخراج گردید. مطابق استاندارد این بانک، خلیج فارس حاوی ۸ گره به ابعاد ۲ درجه طولی در ۲ درجه عرضی تقسیم شده است که پس از بررسی وضعیت داده‌های تمام گره‌ها، پنج گره (از هشت گره) جهت محاسبات بعدی انتخاب گردیدند

کشور شده‌اند (شکل ۱). مطالعه تصاویر ماهواره‌ای از شرایط آب و هوایی کشور نشانگر آن است که بسیاری از بارش‌های مناطق جنوبی کشور (در ماه‌های سرد سال) ناشی از ورود توده هوای گرم اقیانوسی است که پس از عبور از خلیج فارس وارد ایران می‌شوند (شکل ۱). این توده‌های هوا که به آنها کم فشارهای سودانی گفته می‌شود پس از برخورد با سلسله جبال زاگرس و یا برخورد جبهه‌های سردتر دریایی و یا قاره‌ای که از نواحی شمالی و غربی وارد کشور شده‌اند موجب ریزش‌های جوی فراوان می‌شوند. بدین ترتیب علاوه بر توپوگرافی زمین و فیزیک توده‌های هوای ورودی، شرایط اقلیم سطحی خلیج فارس (مانند جهت و سرعت باد، سیستم فشار و نیز دمای سطح آب) نیز نقش مؤثری در جهت حرکت و نیز توانایی تولید بارش کم فشارهای سودانی دارد. تأثیر دمای سطح آب دریاها بر بارش مناطق مختلف، محدود به ایران نبوده و تحقیقات انجام شده بیانگر آن است که نوسان‌های این دما سهم قابل توجهی از واریانس بارش در مناطق مختلف جهان را توجیه می‌نماید (۹، ۸ و ۱۳). گرچه خلیج فارس یک ناحیه کاملاً آزاد دریایی نمی‌باشد، ولی نتایج جدید نشان دهنده تأثیر شرایط اقلیمی این گستره دریایی بر آب و هوای ایران دارد.

ناظم السادات و همکاران (۱۰) و ناظم السادات (۱۱) نشان داده‌اند که هم‌بستگی منفی و معنی‌داری بین دمای زمستانه سطح آب خلیج فارس و بارش زمستانه در استان‌های جنوبی و جنوب غرب کشور (فارس، بوشهر و خوزستان و هرمزگان) وجود دارد. بر اساس این بررسی‌ها، هنگامی که دمای سطح آب (Sea Surface Temperature, SST) زمستانه سطح خلیج فارس کمتر از میانگین می‌شود، بارش در استان‌های مذکور عموماً بیشتر از حد معمول خواهد شد. هم‌چنین در تحقیقات اخیر ناظم السادات و شیروانی (۱)، تکامل زمانی - مکانی دمای سطح آب خلیج فارس (Persian Gulf SST, PGSST) برای پیش‌بینی بارش زمستانه مناطق نام‌برده مورد بررسی قرار گرفت. نتایج این تحقیقات نیز نشان می‌دهد که در مقایسه با سایر فصول، نوسانات SST زمستانه در سطح خلیج فارس در برآورد مقدار



شکل ۱. شمای مسیر عمومی توده‌های هوای اقیانوسی که پس از عبور از شمال افریقا، دریای سرخ از طریق خلیج فارس وارد کشور می‌گردند.

زمانی پیشگو شونده در مدل پیش بینی قرارگرفت.

روش‌های محاسباتی

الف) رگرسیون چندگانه

تحلیل رگرسیونی یک روش آماری است که مشخص می‌نماید دو یا چند متغیر تا چه اندازه به یکدیگر وابسته هستند. روش‌ها و معادلات بسیاری جهت بیان درجه وابستگی علت و معلولی بین متغیرهای مستقل و وابسته بیان شده که هر یک از آنها در شرایط خاصی مورد استفاده قرار می‌گیرند (۳، ۴، ۶ و ۱۴).

چنانچه در نظر باشد که میزان وابستگی یک متغیر مستقل (مانند SST زمستانه) به چند متغیر وابسته دیگر (مانند SST در چهار فصل قبل) مورد سنجش قرار گیرد می‌توان از رگرسیون چندگانه استفاده نمود. به منظور استفاده از متغیرهای وابسته باید دقت کرد که این متغیرها مستقل از یکدیگر باشند. در صورت عدم استقلال متغیرهای وابسته، تحلیل رگرسیون چندگانه ضمن انتخاب چند متغیر مهم‌تر بقیه را از گردونه محاسبات حذف خواهد کرد. مثلاً از بیست متغیر پیشگوکننده ممکن است دو یا سه متغیر وارد مدل رگرسیون چندگانه شده و بقیه به علت

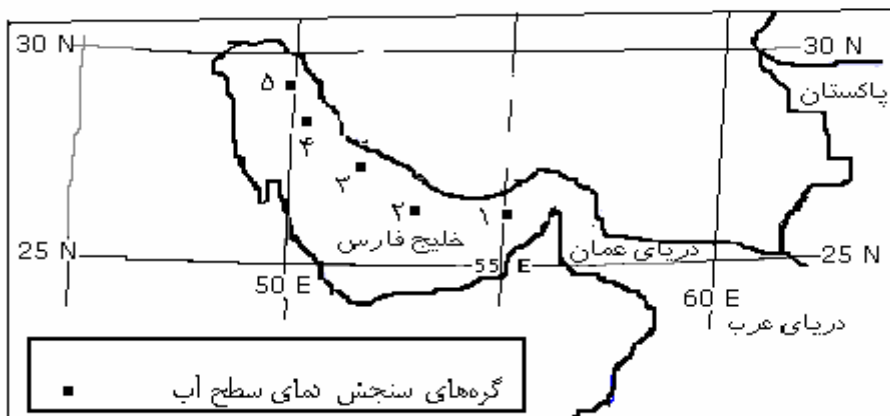
(شکل ۲). ارقام گم شده در گره‌های انتخابی از روش هم‌بستگی خطی با ایستگاه مجاور تکمیل گردیدند. برای محاسبه مقادیر فصلی PGSST میانگین سری زمانی ماه‌های ژانویه، فوریه، مارس به عنوان فصل زمستان، ماه‌های آوریل، می، ژوئن به عنوان فصل بهار، ماه‌های جولای، اگوست و سپتامبر به عنوان فصل تابستان و ماه‌های اکتبر و نوامبر و دسامبر به عنوان فصل پاییز در نظر گرفته شدند. برای هر یک از ۵ گره انتخابی ۴ سری زمانی و برای مجموعه آنها ۲۰ سری زمانی فصلی تولید گردید. عناصر تشکیل دهنده ماتریس پیشگو کننده با بیست ستون به شرح زیر تشکیل گردید:

$$SST = [SST_{11}, SST_{12}, \dots, SST_{ij}, \dots, SST_{45}] \quad [1]$$

به طوری که ستون SST_{ij} بیانگر سری زمانی (۱۹۹۲-۱۹۴۷) دمای سطح آب در فصل i ام و گره j ام می‌باشد ($i=1, \dots, 4$ و $j=1, \dots, 5$). میانگین PGSST زمستانه در ۵ گره ذکر شده (شکل ۲) که به صورت زیر قابل محاسبه است:

$$PGSST = \frac{SST_{21} + SST_{22} + SST_{23} + SST_{24} + SST_{25}}{5} \quad [2]$$

برای دوره زمانی ۱۹۴۸-۱۹۹۳ تعیین شده و به عنوان سری



شکل ۲. گره‌های انتخابی در خلیج فارس

متخصصین علوم در رشته‌های مختلف به منظور کاهش حجم داده‌ها استفاده می‌گردد (۲، ۵، ۷ و ۱۲). در این بررسی با توجه به تعداد زیاد متغیرهای پیشگوکننده (۲۰ متغیر پیشگوکننده) و وابستگی خطی که بین آنها موجود بود، کاربرد مستقیم آنها در مدل رگرسیونی مورد توجه قرار نگرفت. همان طور که در بخش‌های بعدی اشاره خواهد شد، وقتی که تمام بیست متغیر اولیه به طور مستقیم وارد مدل گردید، به علت وابستگی بین آنها تنها تعداد دو سری از کل بیست سری توسط مدل پذیرفته شده و بقیه حذف شدند. این امر به مفهوم آن است که تنها نوسان‌های تعداد بسیار معدودی از ۲۰ متغیر پیشگوکننده به منظور پیشگویی مورد استفاده قرار می‌گیرد. به همین دلیل تحلیل مؤلفه‌های اصلی به شرح زیر مورد توجه قرار گرفت.

فرض کنید P تعداد کل متغیرهای پیشگوکننده (در این مطالعه بیست متغیر) و T_i متغیر دمای سطح آب (در این جا PGSST) باشند. بردار T با p مؤلفه به صورت $T = (T_1, T_2, \dots, T_p)'$ و ماتریس واریانس-کواریانس Σ به صورت زیر در نظر گرفته می‌شود:

$$\Sigma = \begin{pmatrix} \sigma_{11} & \text{COV}_{12} & \dots & \text{COV}_{1p} \\ \text{COV}_{21} & \sigma_{22} & \dots & \text{COV}_{2p} \\ \dots & \dots & \dots & \dots \\ \text{COV}_{p1} & \text{COV}_{p2} & \dots & \sigma_{pp} \end{pmatrix} \quad [3]$$

همبستگی زیاد به یکدیگر در محاسبات منظور نگردند. به همین دلیل، در صورتی که بین داده‌های خام متغیرهایی که به عنوان متغیرهای مستقل در نظر گرفته می‌شوند یک رابطه هم خطی وجود داشته باشد، لازم است که این داده‌ها به روش‌های مناسب فیلتر شوند. اثر فیلتر کردن آن است که رابطه هم خطی بین آنها را از بین خواهد برد. استفاده از تحلیل مؤلفه اصلی که در زیر توضیح داده خواهد شد بدین منظور (فیلتر کردن داده‌ها و استخراج سری‌های جدید مستقل از هم) مورد استفاده قرار گرفت.

ب) تحلیل مؤلفه‌های اصلی

با توجه به آن که داده‌های اقلیمی، هواشناسی و اقیانوسی که از ایستگاه‌های مجاور تهیه شده‌اند، عموماً به یکدیگر وابسته بوده و هم خطی شدیدی بین آنها موجود می‌باشد، کاربرد این گونه سری‌های زمانی در مدل‌های رگرسیونی محل اشکال بوده و موجب بروز خطای زیادی در نتایج به دست آمده می‌گردد. برای استفاده بهینه از داده‌های موجود و نیز کاستن از منابع خطا لازم است که از تعداد زیاد سری‌های زمانی وابسته کاسته شده و با به کارگیری روش‌های مناسب ریاضی تعداد معدودی سری زمانی جدید تولید شود که بتواند درصد بالایی از واریانس موجود در سری زمانی اولیه را توجیه نماید. تحلیل مؤلفه‌های اصلی از جمله اقداماتی است که به طور مکرر توسط

واریانس (λ_1) و y_p دارای واریانس بعدی (λ_p) و آخرین مؤلفه (λ_p) دارای کمترین مقدار واریانس می‌باشد. معمولاً y_1 را اولین مؤلفه اصلی و y_p را دومین مؤلفه اصلی گویند. به y_i امین بردار ویژه گفته می‌شود. طبق رابطه زیر:

$$\lambda_1 + \lambda_p + \dots + \lambda_p = \text{trace} \Sigma = \sigma_{11} + \sigma_{22} + \dots + \sigma_{pp} \quad [11]$$

مجموع واریانس‌های p مؤلفه اصلی برابر مجموع واریانس‌های متغیرهای اولیه T_1, \dots, T_p است. یادآوری می‌گردد که Trace یک ماتریس به مفهوم مجموع عناصر روی قطر اصلی می‌باشد. پس مؤلفه‌های با واریانس کوچک‌تر را بدون آن که اثر معنی‌داری روی کل واریانس بگذارد، می‌توان نادیده گرفت. به این ترتیب p متغیر اولیه به $k (\leq p)$ مؤلفه اصلی اولیه، تقلیل پیدا می‌کند و $p - k$ متغیرهای بعدی به عنوان اختلال (Noise) در نظر گرفته شده و وارد محاسبات نمی‌گردد. بنابراین از روش PCA به منظور استخراج تعداد کمتری سری زمانی مستقل و غیر وابسته که قسمت اعظم واریانس مجموعه PGSST را توجیه نمایند استفاده شد. این روش به عنوان پیش پردازش (فیلتر) برای داده‌هایی که در مدل رگرسیونی استفاده می‌گردد مورد لزوم بود. سری زمانی متغیرهای مستقل و متعامد تولید شده به عنوان پیش‌گوکننده‌ها در نظر گرفته شدند.

نتایج

مؤلفه‌های اصلی PGSST

با کاربرد روش تحلیل مؤلفه‌های اصلی، بیست مؤلفه تولید شد که نخستین مؤلفه بیشترین واریانس و آخرین آن کمترین واریانس را نشان می‌دهد. انتخاب چند مؤلفه اول که بیشترین واریانس را دارند و به عنوان مؤلفه اصلی شناخته می‌شوند، از اساسی‌ترین اقدامات در تجزیه و تحلیل مؤلفه‌های اصلی می‌باشد. با انتخاب چند مؤلفه‌های اصلی اول سایر مؤلفه‌ها از محاسبات بعدی حذف می‌شوند و بنابراین باید دقت زیادی در انتخاب آستانه حذف نمود. رسم تغییرات مقادیر ویژه در مقابل

اعضای روی قطر اصلی ماتریس Σ واریانس عناصر بردار T و دیگر اعضای آن کواریانس بین هر دو عنصر از بردار T می‌باشند. اگر مقادیر ویژه (Eigen values) مرتب شده Σ به صورت $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ باشند، آنگاه یک ماتریس متعامد مانند T به صورت زیر

$$\Gamma = \begin{pmatrix} \gamma_{11} & \gamma_{21} & \dots & \gamma_{p1} \\ \gamma_{12} & \gamma_{22} & \dots & \gamma_{p2} \\ \vdots & \vdots & \dots & \vdots \\ \gamma_{1p} & \gamma_{2p} & \dots & \gamma_{pp} \end{pmatrix} \\ = (\gamma_1 \quad \gamma_2 \quad \dots \quad \gamma_p) \quad ; \quad \gamma_i = (\gamma_{i1} \quad \gamma_{i2} \quad \dots \quad \gamma_{ip})' \\ \Gamma \Gamma' = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix} = I_p$$

وجود دارد به طوری که

$$\Gamma' \Sigma \Gamma = D \lambda \quad ; \quad D \lambda = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & \lambda_p \end{pmatrix} \\ = \text{Diag}(\lambda_1, \lambda_2, \dots, \lambda_p) \quad [5]$$

بنابراین، اگر داشته باشیم:

$$y = \Gamma' T = \begin{pmatrix} \gamma_1' \\ \gamma_2' \\ \vdots \\ \gamma_p' \end{pmatrix} T = \begin{pmatrix} \gamma_1' T \\ \gamma_2' T \\ \vdots \\ \gamma_p' T \end{pmatrix} \quad [6]$$

آن گاه

$$\text{cov}(y) = \Gamma' \Sigma \Gamma = D \lambda \quad [7]$$

و مؤلفه‌های

$$y_1 = \gamma_1' T = \gamma_{11} T_1 + \gamma_{12} T_2 + \dots + \gamma_{1p} T_p \quad [8]$$

$$y_2 = \gamma_2' T = \gamma_{21} T_1 + \gamma_{22} T_2 + \dots + \gamma_{2p} T_p \quad [9]$$

و

$$y_p = \gamma_p' T = \gamma_{p1} T_1 + \gamma_{p2} T_2 + \dots + \gamma_{pp} T_p \quad [10]$$

ناهمبسته‌اند (برهم عمودند). به متغیرهای جدید y_1, \dots, y_p ، مؤلفه‌های اصلی گویند و واریانس y_i برابر λ_i است. چون $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ می‌باشد، مؤلفه y_i دارای بزرگ‌ترین

شماره مؤلفه‌ها که اصطلاحاً به آن Sreet plot گفته می‌شود یکی از راه‌های تشخیص آستانه حذف می‌باشد. همان طور که در شکل ۳ نشان داد شده است، اولین مقدار برابر $3/78$ می‌باشد که $34/5$ درصد از کل واریانس موجود در سری داده‌ها را توجیه می‌نماید. شکل ۳ نشان دهنده آن است که دومین و سومین مقدار ویژه به ترتیب برابر $1/9$ و $1/34$ بودند که هر یک از آنها $17/6$ و $12/2$ درصد از کل واریانس را توجیه می‌نمایند گروهی از محققین بیان می‌کنند که آستانه حذف جایی است که نمودار Sreet حالت افقی پیدا می‌نماید و بنابراین با توجه به شکل ۳ تقریباً می‌توان ۵ و ۶ مؤلفه اول را به عنوان مؤلفه اصلی قلمداد نمود. در عین حال، انتخاب آستانه حذف تنها بر اساس قضاوت تخمینی دقیق نبوده است و لازم است که آزمون‌های دیگری نیز به عمل آید. هدف از آزمون‌های بعدی آن است که یک مبنای فیزیکی برای هر یک از مؤلفه‌های اصلی پیدا نمود. مثلاً روشن شود که هر یک از مؤلفه‌ها نماینده چند ایستگاه واقع در یک ناحیه خاص می‌باشند. برای نمونه اگر مؤلفه اول نماینده یک ناحیه خاص گردید نباید سایر مؤلفه‌ها عضوی در این ناحیه داشته باشند.

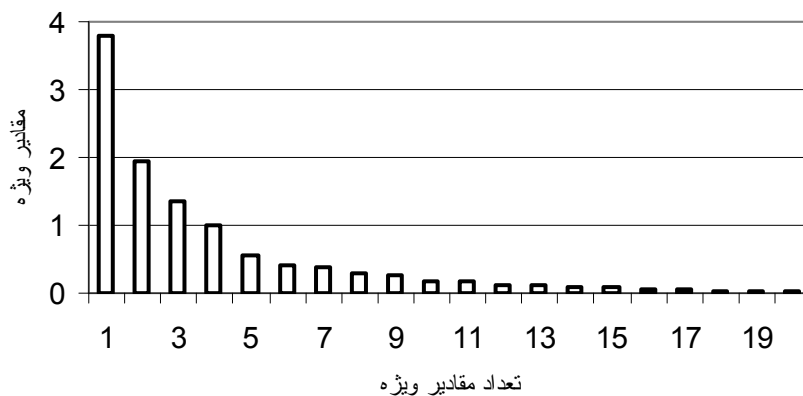
برای پیدا نمودن آستانه حذف، تعداد دو، سه، چهار، پنج و شش مؤلفه اصلی (مؤلفه‌های داوطلب) در نظر گرفته شد و آزمون لازم برای تشخیص تعداد مؤلفه‌های مطلوب برای نگهداری به عمل آمد. برای این منظور عوامل بارگذاری (Loadings) مربوط به هر یک از مؤلفه‌ها به روش واریماکس (Varimax) دوران داده شدند و وابستگی هر یک از مؤلفه‌های دوران داده شده به فصول مختلف مورد ارزیابی قرار گرفت. دوران عوامل بارگذاری باعث می‌شود که وابستگی این عوامل با مؤلفه‌های اصلی واضح‌تر شود و تشخیص آن که تغییرات هر یک از مؤلفه‌های اصلی متأثر از کدام عوامل بارگذاری می‌باشد آسان‌تر می‌شود.

نتایج نشان داد که وقتی تعداد مؤلفه‌ها چهار عدد است مقادیر عوامل بارگذاری از تمام گره‌ها برای یک فصل نزدیک به یک و برای سایر فصول برای تمامی گره‌ها نزدیک صفر است.

در سایر موارد که تعداد مؤلفه‌ها کمتر یا بیشتر از چهار بود چنین وضعیتی مشاهده نشد. با توجه به آن که الگوی نوسان سطح آب در هر فصل تا حد زیادی از فصل دیگر متفاوت است نتایج به دست آمده از نظر فیزیکی نیز معنی‌دار تشخیص داده شد و چهار مؤلفه به عنوان مؤلفه اصلی در نظر گرفته شدند. براین اساس، چهار PC اولیه مربوط به PGSST (ماتریس SST، معادله ۱) برای تجزیه و تحلیل‌های بعدی و ورود به محاسبات رگرسیونی در نظر گرفته شدند. مؤلفه‌های PC1 تا PC4 اولیه قادر به توجیه $73/5$ درصد از کل واریانس عناصر ماتریس SST بودند. نتایج به دست آمده بیانگر آن بود که سری‌های زمانی PC1، PC2، PC3 و PC4 به ترتیب هم‌بستگی بسیار بالایی با نوسانات PGSST در فصول زمستان، پاییز، بهار و تابستان دارند (جدول ۱). به بیان دیگر، سهم عمده‌ای از نوسانات هر پنج گره انتخابی در خلیج فارس توسط سری زمانی PC1 (نوسانات SST در فصل زمستان) توجیه می‌گردد. بنابراین سری زمانی PC1 نمایشگر تغییرات دمای سطح آب در پهنه خلیج فارس در فصل زمستان می‌باشد. به همین ترتیب، PC2، PC3 و PC4 معرف مجموعه ویژگی‌های PGSST در فصول پاییز، بهار و تابستان می‌باشند (جدول ۱). با توجه به نتایج به دست آمده تشخیص داده شد که چهار مؤلفه اولیه که هر یک نماینده یک فصل است بهترین گزینه برای انتخاب مؤلفه‌های اصلی می‌باشد.

رگرسیون چند گانه

در ابتدا بیست متغیر اولیه پیشگو کننده وارد مدل رگرسیونی گردید که تنها دو متغیر از این مجموعه توسط مدل رگرسیونی معنی‌دار بودند وارد محاسبات گردیدند. بدین ترتیب، همان‌طور که پیش بینی می‌شد ۱۸ متغیر بعدی به علت رابطه هم خطی با این دو متغیر اولیه اجازه ورود به محاسبات را پیدا نمودند. این امر به مفهوم آن است که هجده متغیر یاد شده که بیانگر وضعیت دمایی نقاط مختلف خلیج فارس می‌باشند نمی‌توانند در امر پیش‌گویی اقلیمی، نقش مؤثری داشته باشند که چنین



شکل ۳. نمودار Screenshot

جدول ۱. مقدار واریانس تعریف شده توسط هر یک از مولفه‌های اول تا چهارم

نام مؤلفه اصلی	درصد واریانس تعریف شده	درصد تجمعی واریانس	شماره گره و فصل متغیرها مرتبط با PC ها
PC1	۳۴/۵	۳۴/۵	تمامی گره‌های انتخابی در فصل زمستان
PC2	۱۷/۶	۵۲/۱	تمامی گره‌های انتخابی در فصل پاییز
PC3	۱۲/۲	۶۴/۳	تمامی گره‌های انتخابی در فصل بهار
PC4	۹/۲	۷۳/۵	تمامی گره‌های انتخابی در فصل تابستان

$$PGSST = 2/54 - 0/129PC_1 + 0/154PC_2 + 0/048PC_3 + 0/18PC_4$$

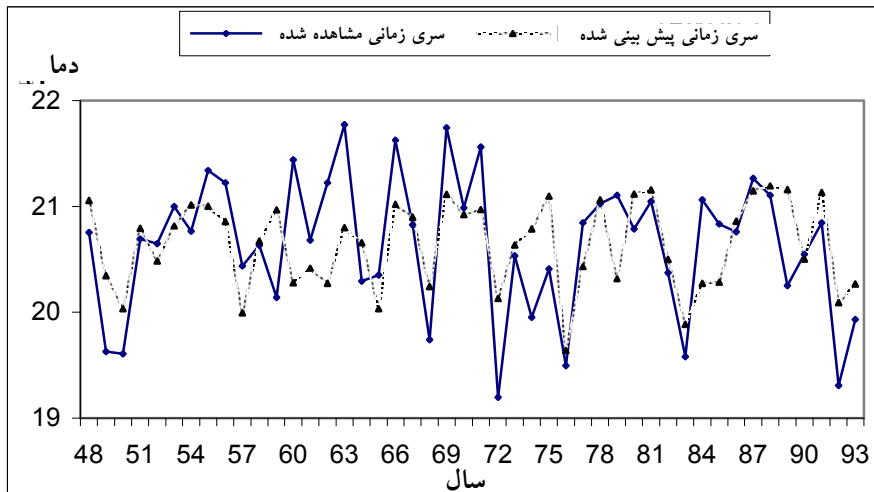
$$R^2 = 34/5\%$$

$$RMSE = 0/5499 \quad [12]$$

در این معادله آماره t برای مقدار ثابت (Intercept) رگرسیون و نیز هر یک از متغیرهای PC_1 ، PC_2 ، PC_3 و PC_4 به ترتیب برابر $0/51$ ، $-3/06$ ، $2/62$ ، $0/69$ و $2/21$ تعیین گردیدند. با استفاده از جدول t ، برای سطح معنی‌داری 5% و درجه آزادی 41 ($n-5$) داریم $t_{(0/975, 41)} = 2/019$. مقایسه بین آماره‌ها و مقدار به دست آمده از جدول نشان می‌دهد که مقدار عرض از مبدأ و PC_3 معنی‌دار نمی‌باشند. یادآوری می‌شود هنگامی که مقدار آماره t بزرگ‌تر از عدد جدول باشد فرض صفر (فرض صفر بیانگر صفر بودن مقدار ثابت یا ضرایب رگرسیونی می‌باشد) رد می‌شود. جذر میانگین توان دوم خطا (Root Mean Square Error, RMSE) نیز برابر $0/5499$ بود. بنابراین با حذف مقدار ثابت رگرسیونی و سری زمانی

موضوعی نمی‌تواند مورد قبول باشد. چون استفاده از این شیوه محاسباتی اعتبار پیش‌بینی را کاهش می‌دهد، مدل رگرسیونی ارائه شده مورد توجه قرار نگرفت. برای آن که بتوان مشکل هم خطی را حل کرد و سهم عمده‌ای از واریانس کل مجموعه متغیرهای پیشگوکننده را در پیش‌بینی‌ها دخالت داد، از نتایج به دست آمده از تجزیه و تحلیل مؤلفه‌های اصلی استفاده شد.

سری‌های زمانی PC_1 ، PC_2 ، PC_3 و PC_4 (PC Score) به عنوان متغیرهای مستقل (پیشگوکننده) وارد مدل رگرسیون چندگانه گردید. یعنی به جای استفاده از بیست متغیر اولیه پیشگوکننده چهار سری زمانی تولید شده (PC_1 ، PC_2 ، PC_3 و PC_4) به عنوان متغیر پیشگوکننده وارد مدل گردیدند. در این مدل سری زمانی معدل زمستانه دمای سطح آب خلیج فارس (رابطه ۲) در گره‌های انتخابی به عنوان متغیر وابسته (پیشگو شونده) در نظر گرفته شد. معادله رگرسیون چندگانه با وارد کردن چهار مؤلفه اصلی اول به صورت زیر به دست آمد:



شکل ۴. سری زمانی مشاهده شده و پیش بینی شده دمای زمستانه سطح آب خلیج فارس

شده PGSST در دوره مورد مطالعه نشان داده شده است. ضریب هم‌بستگی (F) بین این دو سری زمانی برابر 0.579 بود که در سطح 5% از نظر آماری معنی‌دار می‌باشد. توان دوم این ضریب (R^2) حاکی از آن است که در مجموع حدود 33.5% از مقدار واریانس PGSST زمستانه توسط سه مؤلفه اصلی تعریف می‌شود. همان‌طور که در شکل ۴ ملاحظه می‌شود کمترین دقت مدل در پیش‌بینی مربوط به سال‌های ۱۹۶۰ و ۱۹۶۳ بوده است. در مقابل، در بسیاری از سال‌ها از جمله ۱۹۸۷، ۱۹۶۶، ۱۹۵۷ و ۱۹۸۹ پیش‌بینی‌ها از دقت بالایی برخوردار می‌باشند. در شکل ۵ نمودار پراکنندگی، خط یک به یک و خط رگرسیونی بین سری مشاهدات و پیش‌بینی شده رسم شده است.

نتایج به دست آمده بیانگر آن است که با استفاده از مقادیر ۱۲ ماه قبل دمای سطح آب خلیج فارس می‌توان دمای سه ماه آتی را پیش‌بینی نمود. با توجه به آن که این دما نقش مؤثری در پیش‌بینی بارش دارد، استفاده از مدل ارائه شده نقش مهمی در پیش‌بینی منابع آب کشور و به خصوص نواحی جنوبی دارد. ارائه این مدل به مفهوم آن است که بدون در نظر گرفتن هیچ شاخص اقلیمی دیگر، دمای خلیج فارس در طول فصول گذشته می‌تواند به طور معنی‌داری نوسان‌های سه ماه آینده خود را پیش‌بینی نماید. بدیهی است که چنانچه بتوان با نصب

PC3، معادله رگرسیونی براساس سه مؤلفه اصلی که معنی‌دار شدند، به صورت زیر محاسبه گردید:

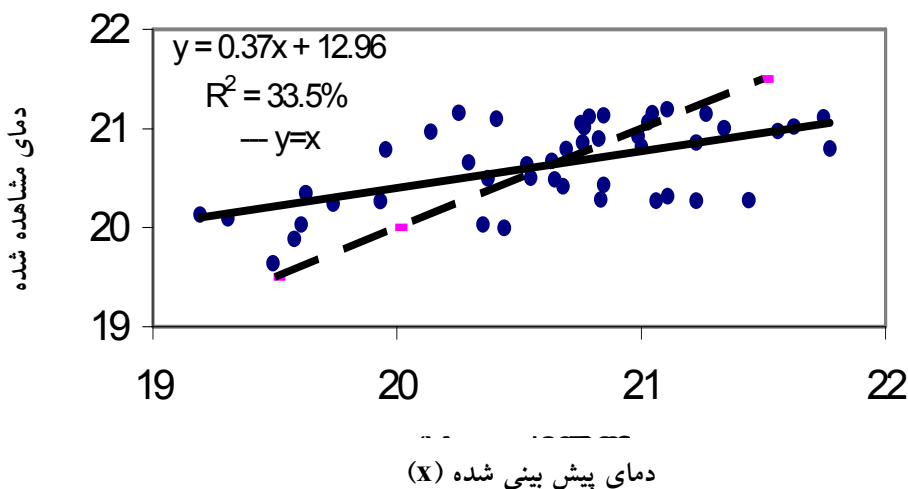
$$PGSST = -0.153PC_1 + 0.165PC_2 + 0.177PC_4 \quad [13]$$

$$R^2 = 33.5\%$$

$$RMSE = 0.5424$$

آماره t برای PC1، PC2 و PC4 به ترتیب برابر -11.75 ، 2.99 و 2.2 بود. جدول مقادیر t بیانگر آن است که برای سطح معنی‌داری 5% و درجه آزادی 43 ($n-3$)، $t_{(0.975, 43)} = 2.016$ می‌شود. بنابراین مقایسه بین آماره‌ها و مقدار به دست آمده از جدول نشان می‌دهد که تمامی ضرایب معنی‌دار می‌باشند. بزرگی قدر مطلق آماره t نیز اهمیت آنها را در مدل نشان می‌دهد. از این دیدگاه، PC1 از اهمیت بیشتری نسبت به دو مؤلفه دیگر دارد و در مرحله بعدی PC2 و PC4 دارای اهمیت می‌باشند. بنابراین سری زمانی اولین مؤلفه که بیانگر نوسانات PGSST در زمستان می‌باشد از اهمیت بیشتری در پیش‌بینی PGSST زمستانه سال بعد برخوردار بود. به بیان دیگر برای تشخیص پیش‌بینی SST زمستانه سال $i+1$ ، مقدار PC2 و PC3 که به ترتیب بیانگر تغییرات پاییزه و بهاره بودند در پیش‌بینی‌ها نقش دوم و سوم را ایفا می‌نمایند.

در شکل ۴ سری زمانی مشاهده شده و پیش‌بینی



شکل ۵. نمودار پراکنندگی

که ۷۳/۵٪ از کل واریانس PGSST در کلیه فصول را توجیه می‌نمودند، از داده‌های اولیه استخراج گردید. نشان داده شد که مؤلفه‌های اول، دوم، سوم و چهارم به ترتیب معرف نوسانات SST در فصول زمستان، پاییز، بهار و تابستان می‌باشند که به ترتیب ۳۴/۵، ۱۷/۶، ۱۲/۲ و ۹/۲ درصد از واریانس کل متغیرهای پیشگوکننده اولیه را شرح می‌دهند. این چهار مؤلفه اصلی به عنوان متغیرهای مستقل وارد مدل رگرسیونی شدند. مشخص گردید که متغیرهای مستقل در مجموع ۳۳/۵٪ از کل واریانس PGSST زمستانه سال بعد را شرح می‌دهند. نتایج به دست آمده حاکی از آن است که دمای سطح آب خلیج فارس در فصول قبل و به خصوص در زمستان سال قبل می‌تواند به عنوان یک شاخص مهم در پیش‌بینی‌های دمای سطح آب در زمستان سال بعد مورد توجه قرار گیرد.

سپاسگزاری

بدین وسیله از مدیریت صندوق بیمه محصولات کشاورزی به خاطر کمک در اجرای طرح تشکر و قدردانی می‌شود.

بویه‌هایی سایر مشخصات اقلیمی سطح آب مانند فشار و باد را اندازه‌گیری کرده و وارد مدل نمود، دقت پیش‌بینی‌ها به طور قابل توجهی افزایش خواهد یافت. با توجه به آن که این گونه پیش‌بینی‌ها نقش مؤثری در دقت پیش‌گویی بارش دارد، بنابراین نصب بویه‌های شناور در پهنه‌های خلیج فارس مورد تأکید این پژوهش است.

نتیجه‌گیری

با توجه به اهمیت نوسانات دمای سطح آب خلیج فارس در فصل زمستان و تأثیر آن بر میانگین وقوع بارش مناطق جنوبی کشور، پیش‌بینی این دما با استفاده از مدل رگرسیون چندگانه مورد بررسی قرار گرفت. میانگین دمای سطح آب خلیج فارس در فصل زمستان به عنوان متغیر پیشگوشونده و دمای سطح آب همین گستره آبی در یک دوره یک ساله قبل (چهار فصل قبل) به عنوان پیشگوکننده انتخاب شد. به علت معنی‌دار نشدن تعداد زیادی از متغیرهای پیشگوکننده اولیه و مشکل هم خطی بین متغیرهای پیشگوکننده، از روش تحلیل مؤلفه‌های اصلی به عنوان پیش پردازش داده‌ها استفاده شد. چهار مؤلفه اصلی اولیه

منابع مورد استفاده

۱. ناظم السادات، س. م. ج. و ا. شیروانی. ۱۳۸۲. کاربرد مدل تحلیل هم‌بستگی متعارف برای مطالعه تأثیر دمای سطح آب خلیج فارس بر بارندگی زمستانه مناطق جنوبی ایران. پذیرفته شده در مجله علمی کشاورزی دانشگاه اهواز.
۲. ناظم السادات، س. م. ج.، ب. بیگی و س. امین. ۱۳۸۲. پهنه بندی بارش زمستانه استان های بوشهر، فارس و کهگیلویه و بویر احمد. علوم کشاورزی و منابع طبیعی ۷ (۱): ۶۱-۷۲.
۳. کانتی، م. و ب. جان. ۱۳۷۶. تحلیل چند متغیره (ترجمه م. م. طباطبائی). انتشارات مرکز نشر دانشگاهی، تهران.
4. Cohen, J., S. G. west, L. Ailen and P. Cohen. 1975. Applied Multiple Regression/correlation Analysis for the Behavioral Sciences. Lawrence Erlbaum Pub., New Jersey.
5. Emery, W. and R. E. Thomson. 1997. Data Analysis Method in Physical Oceanography. Pergamon Pub., Oxford.
6. Kerlinger, F.N. and E. J. Pedhazur. 1973. Multiple Regression in Behavioral Research. Holt, Rinehart and Winston Inc., New York.
7. Landman, W. A. and S. J. Mason. 2001. Forecasts of near-global sea surface temperature using canonical correlation analysis. J. Climate 14: 3819-3833.
8. Landman, W. A. and S. J. Mason. 1999. Operational long-lead prediction of South African rainfall, using canonical correlation analysis. Int. J. Climatol. 19: 1073-1090.
9. Lough, J. M. 1997. Regional indices of climate variation: temperature and rainfall. Int. J. Climatol. 17: 55-66.
10. Nazemosadat, M. J., I. Cordery and S. Slamian. 1995. The impact of Persian Gulf sea surface temperatures on Iranian rainfall. The Proceeding of The First International Conference of Iranian Water Resources, Esfahan, Iran.
11. Nazemosadat, M. J. 1998. The Persian Gulf sea surface temperature as a drought diagnostic for southern parts of Iran. Drouth News Network 10:12-14.
12. Preisendorfer, R. M. 1988. Principal Component Analysis in Meteorology and Oceanography. Elsevier Pub., New York.
13. Smith, I. 1994. Indian ocean sea surface temperature patterns and Australian winter rainfall. Int. J. Climatol. 14(3): 287-305.
14. Vindo, H. D. and A. Ullah. 1981. Recent Advances in Regression Methods. Marsel Dekker Inc., New York.