

## Technical Note

## یادداشت فنی

### Regionalization of precipitation in Iran using principal components and cluster analysis

### پهنه بندی بارش در ایران با استفاده از تحلیل مولفه‌های اصلی و تحلیل خوشه‌ای

A. Shirvani<sup>1\*</sup> and S.M.J. Nazemosadat<sup>2</sup>

امین شیروانی<sup>۱\*</sup> و سید محمدجعفر ناظم السادات<sup>۲</sup>

#### Abstract

Principal components and cluster analysis were used for grouping the monthly precipitation over Iran into some homogeneous regions. Such grouping is considered as an essential and primary step for pre-process in downscaling of the general circulation models outputs. The monthly precipitation data of 42 stations for the period of 1967-2003 (37 years) forming a 42×444 matrix was used as input file of the principal components analysis. Using PCA, this matrix was reduced into a 42×33 matrix that accounted for about 96% of total variance in the observational time series. The reduced matrix was used as the input file of the Cluster analysis. The results regionalized the country into six different regions.

**Keywords:** Principal components analysis, Cluster analysis, Precipitation, Iran

Received: July 30, 2007  
Accepted: February 19, 2012

#### چکیده

برای گروه‌بندی بارش ماهانه ایران به نواحی همگن از تحلیل مولفه‌های اصلی و خوشه‌ای استفاده شد. چنین گروه بندی به عنوان مرحله اول و مهم در پیش پردازش فرو مقیاس نمودن خروجی‌های مدل‌های عمومی جو در نظر گرفته می‌شود. داده‌های بارش ماهانه ۴۲ ایستگاه همدیدی برای دوره ۱۹۶۷-۲۰۰۳ (۳۷ ساله) تشکیل یک ماتریس ۴۲ در ۴۴۴ داد که به عنوان فایل ورودی تحلیل مولفه‌های اصلی بکار برده شد. با استفاده از تحلیل مولفه‌های اصلی این ماتریس به یک ماتریس ۴۲ در ۳۳ که ۹۶٪ از کل واریانس در سری‌های زمانی مشاهده شده را توجیه کرد، کاهش یافت. ماتریس کاهش یافته به عنوان فایل ورودی تحلیل خوشه‌ای استفاده شد. نتایج نشان داد که کشور به شش ناحیه مختلف پهنه بندی می‌شود.

**کلمات کلیدی:** تحلیل مولفه‌های اصلی، تحلیل خوشه‌ای، بارش، ایران.

تاریخ دریافت مقاله: ۸ مرداد ۱۳۸۶

تاریخ پذیرش مقاله: ۳۰ بهمن ۱۳۹۰

1- Assistant Professor, Water Engineering Department and the Oceanic and Atmospheric Research Center, College of Agriculture, Shiraz University, Shiraz, Iran, Email: am\_shirvani@hotmail.com.

2- Professor, Water Engineering Department, Director of the Oceanic and Atmospheric Research Center, College of Agriculture, Shiraz University, Shiraz, Iran, Email: jafar@shirazu.ac.ir.

\*- Corresponding Author

۱- استادیار بخش مهندسی آب و مرکز پژوهش‌های جوی اقیانوسی، دانشکده کشاورزی، دانشگاه شیراز، شیراز، ایران

۲- استاد بخش مهندسی آب و رییس مرکز پژوهش‌های جوی اقیانوسی، دانشکده کشاورزی، دانشگاه شیراز، شیراز، ایران

\*- نویسنده مسئول

## ۱- مقدمه

تحقیقات بیشتر برای پهنه بندی اقلیمی کشور بکار برده شده است و از چندین متغیر مختلف اقلیمی به طور همزمان استفاده شده است.

Domroes et al. (1998) پنج ناحیه همگن بارش برای ایران تشخیص دادند. در این تحقیق یک پهنه همگن به دو قسمت شکسته شده است به طوری که پهنه همگن دیگری بین آنها قرار گرفته است و در حقیقت مشکل ناپیوستگی (discontinuity) در پهنه بندی بوجود آمده است و استفاده از این نوع پهنه بندی برای مدل سازی منطقه ای بارش استفاده نمی شود (مکاتبه شخصی با Simon Mason). مسعودیان نیز چندین تحقیق در زمینه پهنه بندی انجام داده است که به یکی از آنها اشاره می شود. مسعودیان (۱۳۸۴) با تهیه نقشه های رقومی بارش ماهانه کشور و کاربرد تحلیل خوشه ای سه رژیم بارش پاییزی، زمستانی، زمستانی-بهاری تعریف نمود.

استفاده از تحقیق های انجام شده قبل از فرو مقیاس نمودن خروجی های GCM معمول نیست. هدفی که در تحقیق حاضر دنبال می شود آنست که سری های زمانی قابل کاربرد در فرو مقیاس نمودن خروجی های GCM بر روی ایران تشکیل شود. از این رو به منظور فراهم نمودن زیر ساخت های لازم تحقیقاتی به منظور کاربرد مدل های عمومی جو در پیشگویی بارش کشور، پژوهش حاضر به انجام رسید. نتایج این تحقیق می تواند به عنوان فایل های ورودی در فرو مقیاس نمودن خروجی های GCM جهت پیش بینی بارش در نظر گرفته شود.

بر این پایه، کشور بر حسب رژیم های مختلف بارش به نواحی تقریباً مستقل از یکدیگر تقسیم بندی خواهد گردید. در تحقیق حاضر، با بهره از مطالعات Mason (1998)، Mimmack et al. (2001)، Wilks (2006) روش هایی برای انتخاب بهینه تعداد مولفه های اصلی و تعداد گروه های همگن بارش بکار برده شد.

## ۲- داده ها و روش های محاسباتی

سری های زمانی دوازده ماهه بارش ۴۲ ایستگاه همدیدی کشور برای دوره زمانی ۲۰۰۳-۱۹۶۷ از سالنامه هوا شناسی انتشارات سازمان هواشناسی کشور استخراج گردید. داده های گمشده در هر ماه به روش میانگین دراز مدت ماهانه جایگذاری گردید. داده های ۳۷ سال بارش ۱۲ ماهه هر ایستگاه بصورت یک ستون با  $۳۷ \times ۱۲$  ردیف پشت سر هم قرار داده شدند. در مواردی که درصد بالایی از ماه های سال بدون بارش باشند، استاندارد کردن معمولی (تفاضل

در سال های اخیر استفاده از روش های چند متغیره آماری نظیر تحلیل مؤلفه های اصلی (PCA) و تحلیل خوشه ای (CA) مورد توجه بسیاری از محققین علوم هوا و اقلیم شناسی قرار گرفته است. با توجه به حجم بسیار زیاد داده های هوا و اقلیم شناسی و نیز اقیانوس شناسی که عموماً در حوضه های وسیع جغرافیایی و در فواصل زمانی کوتاه جمع آوری می گردد، از روش PCA به منظور کاهش حجم داده ها و نیز از روش CA برای پهنه بندی مناطق جغرافیایی استفاده می گردد. تحلیل خوشه ای یکی از روش های کارا در جهت شناخت مناطق یا الگوهای مشابه اقلیمی در یک پهنه وسیع جغرافیایی می باشد. استخراج الگوهای همگن بارش، برآورد بارش منطقه ای را میسر ساخته و بررسی عوامل موثر بر نوسان های آن را آسان می نماید. با توجه به اهمیت ویژه ای که پیش بینی بارش منطقه ای در مدیریت منابع آب کشور دارد، تشخیص نواحی همگن بارش نقش موثری در توسعه مدل های بهره برداری و توزیع منابع آب دارد. تحقیق های متعددی در خارج از ایران برای پهنه بندی بارش انجام شده است که در این مقاله از ذکر آنها خودداری می شود و تنها به یکی از آنها که در فرو مقیاس نمودن خروجی های مدل های عمومی جو<sup>۳</sup> (GCM) استفاده می شود، اشاره می شود. Mason (1998) با استفاده از داده های بارش ماهانه، قاره آفریقا (در نیمکره جنوبی) را به هشت ناحیه همگن تقسیم کرد و نمایه بارش را برای هر یک از ناحیه ها استخراج نمود. سری های زمانی این نمایه ها هم اکنون به عنوان متغیر پیشگو شونده در فرو مقیاس نمودن خروجی های مدل های عمومی جو در این منطقه مورد استفاده مجامع علمی است.

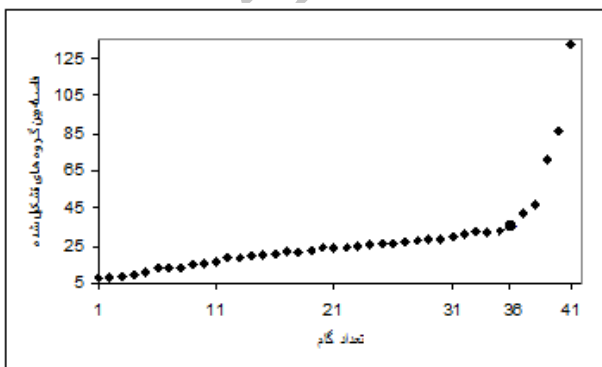
در ایران، با در نظر گرفتن متغیرهایی مانند بارش، دما، فشار هوا و طول و عرض جغرافیایی، پهنه بندی اقلیمی با بکارگیری روش های چند متغیره آماری مورد مطالعه قرار گرفته است که در ادامه به معدودی از آنها اشاره می شود. علیچانی (۱۳۷۴) با در نظر گرفتن تعداد زیادی از ویژگی های مهم بارش از قبیل بارش سالانه، تمرکز فصلی، حداکثر بارش روزانه و با توجه به ویژگی های سینوپتیکی، پنج ناحیه بارشی را تشخیص داده است.

جهان بخش و ذوالفقاری (۱۳۸۱) نسبت به پهنه بندی بارش های روزانه ۲۲ ایستگاه سینوپتیکی در غرب ایران در ۸ ماه مرطوب سال (اکتبر تا می) اقدام کردند و پنج ناحیه بارشی متمایز را برای غرب کشور مشخص نمودند. Dinpashoh et al. (2004) با استفاده از روش تحلیل عاملی و تحلیل خوشه ای، ۱۲ متغیر از ۵۷ متغیر کاندید را انتخاب نمودند و کشور را به ۷ پهنه اقلیمی تقسیم نمودند. این

واریانس مشاهده‌ها را تعریف می‌نمودند به عنوان مناسب‌ترین تعداد مولفه وارد محاسبات تحلیل خوشه‌ای شد. تحلیل خوشه‌ای بر پایه فاصله ماهالونوبیس و روش کمترین واریانس وارد انجام گرفت. فاصله بین گروه‌های تشکیل شده در هر مرحله از روش سلسله مراتبی بر حسب تابعی از هر گام رسم گردید (شکل ۱). رسم این تابع در تشخیص تعداد مناسب گروه‌ها مفید می‌باشد. اولین تغییر معنی‌دار در روند تابع به عنوان نقطه مناسب توقف روش سلسله مراتبی به منظور تشخیص گروه‌ها در نظر گرفته می‌شود (Wilks, 2006). این روش در تشخیص تعداد گروه‌ها مناسب عمل می‌کند.

از این‌رو، اولین تغییر معنی‌دار در فاصله بین خوشه‌ها در هر گام، نشان دهنده تعداد خوشه‌ها یا گروه‌های مناسب می‌باشد. بر این پایه، در مرحله ۳۶ از روش سلسله مراتبی اولین تغییر معنی‌دار در فاصله بین گروه‌ها مشاهده گردید (شکل ۱). این بدان معنی می‌باشد که روش سلسله مراتبی در مرحله ۳۶ باید متوقف شود. نتیجه توقف در این مرحله متناظر با انتخاب شش گروه می‌باشد. بنابراین تعداد شش گروه به عنوان شش ناحیه همگن (شکل ۲) تشخیص داده شد و سپس عضو هر گروه نیز مشخص شد. این نتایج در زمینه فرو مقیاس نمودن خروجی‌های GCM قابل استفاده می‌باشد.

میانگین بارش سالانه ناحیه R1 که در برگزیده ایستگاه‌های اردبیل، ارومیه، خوی تبریز و زنجان است برابر ۳۱۱ میلی متر می‌باشد. بارش‌های بهاره این گستره نزدیک به ۳۵ تا ۴۰ درصد از کل بارش سالانه را در بر می‌گیرد که بیشترین سهم بارش بهاره در همه کشور می‌باشد. بارش‌های زمستانه، پاییزه و تابستانه با نزدیک به ۳۰، ۱۵ و ۵ درصد در جایگاه بعدی قرار دارند. بدین گونه نزدیک به ۷۰٪ از بارش سالانه شمال باختری ایران در فصول زمستان و بهار فراهم می‌شود.



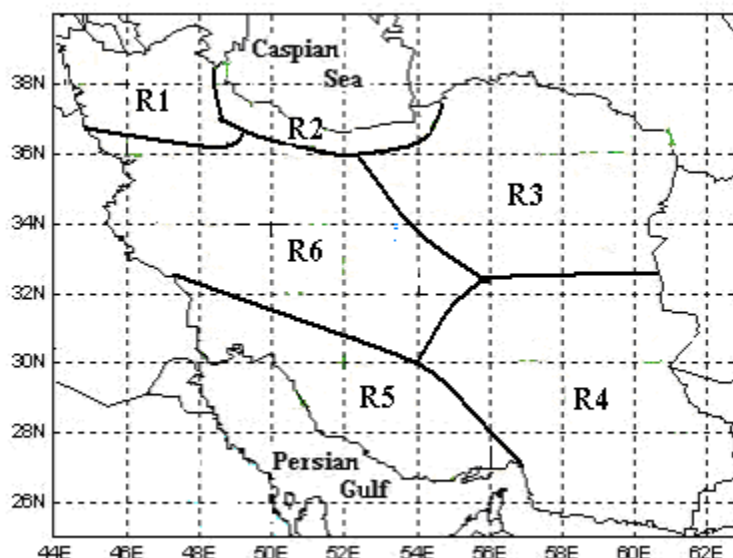
شکل ۱- فاصله بین خوشه‌های تشکیل شده به عنوان تابعی از گام‌ها

داده‌ها به وسیله میانگین و تقسیم بر انحراف معیار) نا مناسب می‌باشد (Wilks, 2006). از این‌رو، از میانگین سالانه برای استاندارد کردن استفاده می‌شود. قبل از ورود ماتریس داده‌های استاندارد شده جهت تحلیل خوشه‌ای، با بکارگیری تحلیل مولفه‌های اصلی تعداد مشاهدات کاهش داده می‌شود. در این تحقیق از فاصله ماهالونوبیس و روش وارد در محاسبات تحلیل خوشه‌ای استفاده می‌شود. در هر گام از گروه بندی که فاصله‌ها محاسبه می‌شوند، روش سلسله مراتبی با استفاده از ملاک کمترین واریانس وارد<sup>۴</sup> به منظور تشخیص گروه بندی بعدی بکار برده خواهد شد. هر گام از روش وارد با هدف کمینه وجه تشابه بین دو خوشه بصورت مجموع مربعات برای همه متغیرها (ماه‌ها) و ایستگاه‌های آن دو خوشه تعریف می‌شود. انتخاب تعداد مناسب خوشه‌ها بر پایه تشخیص تغییر معنی‌دار در فاصله بین گروه‌های تشکیل شده در هر مرحله انجام خواهد گردید.

### ۳- نتیجه گیری

با کاربرد روش تحلیل مولفه‌های اصلی بر پایه ماتریس همبستگی، ۴۴۴ مولفه تولید گردید و از بین آنها تعدادی که درصد بالایی از واریانس بارش ماهانه را توجیه می‌نمایند نگهداری شده و جهت محاسبات بعدی استفاده می‌شوند. نتایج نشان داد که مقدار ویژه برای اولین، دومین و سومین مولفه به ترتیب برابر ۱۱۲/۷، ۵۴ و ۳۱ می‌باشد که ۲۵، ۱۲ و ۷ درصد از کل واریانس را تعریف می‌نمایند. ۴۰ مولفه اول در مجموع ۹۹/۷ درصد از کل واریانس را تعریف نموده و بقیه ۴۰۲ مولفه نقشی در این تعریف ندارند.

انتخاب کافی تعداد مولفه‌های اصلی نقش موثری در نتایج حاصل از پردازش تحلیل خوشه‌ای دارد (Mimmack et al., 2001). برای نیل به یک نتیجه مطلوب و منطقی در مرحله نخست وقتی تنها تعداد ۱۰ مولفه نگهداری شود، تحلیل خوشه‌ای انجام گردید و سپس با استفاده از روش وارد تعداد گروه‌ها (مناطق بارش) و ایستگاه‌های مربوط به هر گروه مشخص شد. این تحلیل برای حالتی که مولفه‌های اصلی ورودی به تحلیل خوشه‌ای برابر ۱۱، ۱۲، ۱۳، ...، ۳۹ و ۴۰ مولفه باشد به طور جداگانه انجام شد و نتایج هر انتخاب بر مناطق بارش و اعضای هر منطقه مشخص، ارزیابی گردید. مقایسه نتایج انتخاب‌ها با یکدیگر نشان داد که وقتی تعداد مولفه‌ها بیشتر از ۳۲ باشد هیچ تاثیری در تعداد نواحی بارش تولید شده ندارد. برای مثال ورود ۳۳ یا ۳۸ مولفه اصلی منجر به پهنه بندی یکسانی می‌گردد. در سایر موارد که تعداد مولفه‌ها کمتر از ۳۳ بود چنین وضعیتی مشاهده نشد و نتایج متفاوتی برای پهنه بندی بارش حاصل گردید. بنابراین ۳۳ مولفه اصلی که مجموعاً ۹۶ درصد از کل



شکل ۲- شنش ناحیه همگن بارش

ناحیه R5 ایستگاه‌های آبادان، اهواز، شیراز، بوشهر، فسا، بندر لنگه و بندر عباس را در خود جای داده و نزدیک به ۸۰٪ از بارش سالانه این گستره در ماه‌های دسامبر هر سال تا مارچ سال رخ می‌دهد.

ناحیه R6 ایستگاه‌های سقز، سنندج، شهرکرد، همدان، خرم‌آباد، کرمانشاه، تهران، قزوین، اراک، کاشان، اصفهان و یزد را دربرمی‌گیرد. این ناحیه نشانگر پهنه‌ای است که بارش موثر آن کم و بیش از ماه اکتبر آغاز شده و تا ماه می سال بعد ادامه دارد. در این دوران بارش سالانه در یک ماه بخصوص چندان متمرکز نگردیده است و این رژیم بارش می‌تواند جایگاه ارزنده‌ای بر موفقیت کارکردهای کشاورزی و رشد گیاهان مرتعی و خودرو داشته باشد.

#### ۴- جمع‌بندی

پهنه‌بندی بارش ماهانه در ایران با استفاده از روش‌های چند متغیره آماری انجام شد. کشور به شش ناحیه همگن بارش تفکیک شد. بارش منطقه‌ای پهنه‌های شناسایی شده می‌تواند به عنوان متغیرهای پیشگو شونده در فرو مقیاس نمودن خروجی‌های GCM جهت پیش‌بینی بارش در نظر گرفته شود.

#### پی‌نوشت‌ها

- 1- Principal Components Analysis, PCA
- 2- Cluster Analysis, CA
- 3- General Circulation Models, GCM
- 4- Ward's Minimum Variance Method

ناحیه R2 دربرگیرنده آستارا، بابلسر، بندر انزلی، نوشهر و گرگان است. این گستره رژیم نمایانگر بارش درپهنه جنوبی دریای خزر می‌باشد که ناهمسانی فراوانی با رژیم بارش در دیگر پهنه‌های کشور دارد. ناحیه R2 سیمای بارش در بخشی از کشور است که در همه ماه‌های سال و بویژه در فصل تابستان، که دیگر نقاط کشور کم و بیش خشک می‌باشند، از بارش خوبی بهره‌مند می‌باشند. در پهنه جنوبی دریای خزر مقدار بارش بهار به میزان ۱۲/۵٪ و بارش پاییزه به میزان ۴۳/۷٪ بترتیب کمترین و بیشترین سهم بارش سالانه را در برمی‌گیرد. (Dinpashoh et al. (2004) نیز در تحقیقات خود اقلیم این گستره از کشور را از دیگر مناطق کشور جدا نموده است.

ناحیه R3 ایستگاه‌های سرخس، مشهد، بیرجند، بار نیشابور، سبزوار، تربت حیدریه، سمنان و شاهرود را در بر می‌گیرد و با ۲۲۰ میلی‌متر بارش پس از R4 از خشک‌ترین پهنه‌های کشور می‌باشد. بیشترین بخش بارش سالانه این گستره در بازه زمانی نوامبر هر سال تا می سال بعد می‌باشد و در دیگر ماه‌های سال خشکی بر منطقه چیره است. کویر ایران منطقه R3 را از دو منطقه R4 و R6 که در نیمه باختری کشور جای دارد، جدا می‌نماید.

ناحیه R4 در برگیرنده ایستگاه‌های زابل، خاش، زاهدان، بم و کرمان است که در ناحیه کم‌پست و پایین کشور جای دارد. در سنجش با دیگر پهنه‌های مورد بررسی، این ناحیه کم‌باران‌ترین ناحیه بوده به گونه‌ای که میانگین بارش سالانه آن ۱۰۰ میلی‌متر است که بیش از ۹۰ درصد آن در بازه زمانی دسامبر تا مارچ ریزش می‌نماید.

## ۵- مراجع

- Dinpashoh Y, Fakheri-Farda A, Moghaddamb M, Jahanbakhsh S, Mirnia M. (2004). Selection of variables for the purpose of regionalization of Iran's precipitation climate using multivariate methods. *Journal of Hydrology*, 297(1), pp. 109-123.
- Mason, S.J. (1998). Seasonal forecasting of South African rainfall using a non-linear discriminant analysis model. *International Journal of climatology*, 18(2), pp. 147-164.
- Mimmack GM, Mason SJ, Galpin JS. (2001). Note and correspondence choice of distance matrices in cluster analysis: Defining regions. *Journal of Climate*, 14(12), pp. 2790-2797.
- Wilks, DS. (2006). *Statistical methods in the atmospheric sciences*. Second Edition, Academic Press: San Diego, 630p.
- جهانبخش اصل، س. و ذوالفقاری، ح. (۱۳۸۱)، "بررسی الگوهای سینوپتیک بارشهای روزانه در غرب ایران". فصلنامه تحقیقات جغرافیایی ۶۴، ۲۳۴-۲۵۸.
- علیجانی، ب. (۱۳۷۴)، "آب و هوای ایران". انتشارات دانشگاه پیام نور، ۲۲۱ صفحه.
- مسعودیان، س. ا. (۱۳۸۴)، "شناسایی رژیمهای بارش ایران به روش تحلیل خوشه‌ای"، مجله علوم انسانی دانشگاه اصفهان، جلد هجدهم، شماره ۱، ۱-۱۳.
- Domroes, M., Kaviani, M., Schaefer, D. (1998), An analysis of regional and intra- annual precipitation variability over Iran using multivariate statistical methods, *Theoretical and Applied Climatology*, 61(3), pp. 151-159.

Archive of SID