نشریه علمی- پژوهشی

علوم و مهندسی آبخیزداری ایران
Iran-Watershed Management Science & Engineering

سال هشتم- شماره ۲۴- بهار ۱۳۹۳ Vol. 8, No. 24, Spring 2014

گزارش فنی

تحلیل منطقه‌ای سیلاب در حوزه‌های آبخیز استان مازندران با استفاده از تحلیل خوشه‌ای فازی

فرهاد فرسادنیا[۱]، پرویز حقیقت‌جو[۲]، زمان شامحمدی حیدری[۲] و علیرضا مقدم‌نیا[۳]

تاریخ دریافت : ۱۳۸۹/۱۰/۲۰ تاریخ پذیرش: ۱۳۹۰/۱۲/۲۳

چکیده

یکی از روش‌های تخمین چندک‌های سیلاب در حوضه‌های فاقد آمار استفاده از روش تحلیل فراوانی منطقه‌ای است. در روش‌های رایج منطقه‌ای کردن، براساس خصوصیات حوزه آبخیز مانند خصوصیات فیزیوگرافی حوزه، موقعیت جغرافیایی و آمار سیلاب در هر ایستگاه، مناطق همگن هیدرولوژیک ایجاد می‌شوند. البته اکثر حوزه‌های آبخیز فقط مقداری شباهت با سایر حوزه‌های آبخیز در منطقه دارند. بنابراین یک حوزه آبخیز نمی‌تواند کاملاً به یک گروه یا گروه دیگر اختصاص یابد. الگوریتم خوشه‌بندی فازی به یک حوزه آبخیز اجازه می‌دهد تا عضویت جزئی یا توزیعی از تمامی مناطق (گروه‌ها) تشخیص داده شده داشته باشد. در این مقاله کارایی الگوریتم خوشه بندی فازی برای منطقه ای کردن حوزه‌های آبخیز استان مازندران ارزیابی شده است و برای بهبود همگنی مناطقی که با این الگوریتم خوشه‌بندی شده‌اند و از لحاظ آماری همگن نیستند، راهکارهایی ارائه شده است. همچنین چندین شاخص ارزیابی کارایی خوشه بندی فازی در تعیین تعداد خوشه های بهینه نشان داده شده است. سپس آماره Z^{DIST} برای ایستگاه‌های منطقه با استفاده از زبان برنامه نویسی فرترن برای توزیع‌های ۳ پارامتری محاسبه شد و بهترین توزیع هر منطقه معرفی گردید. نتایج نشان داد که استان مازندران از سه منطقه همگن هیدرولوژیک تشکیل شده است و توزیع لجستیک تعمیم یافته با پارامترهای متفاوت برای هر یک از مناطق همگن انتخاب شد.

واژه‌های کلیدی: منطقه‌ای کردن؛ تحلیل خوشه‌ای فازی؛ گشتاورهای خطی؛ همگنی هیدرولوژیک؛ حوزه‌های آبخیز استان مازندران.

مقدمه

غالباً حوزه‌ها فاقد ایستگاه‌ها اندازه‌گیری هستند یا طول دوره آماری آن‌ها در مقایسه با دوره بازگشت مورد نیاز بسیار کوتاه می‌باشد. این امر مانع از استفاده روش‌های تحلیل فرآوانی در هر ایستگاه برای تخمین بزرگی سیلاب با دوره‌های بازگشت بالا می‌شود. برای حل این مشکل، هیدرولوژیست‌ها از روش‌های منطقه‌ای برای تفکیک حوزه‌های آبخیز موجود در یک منطقه به گروه‌های همگن استفاده می‌کنند.

تحلیل فراوانی منطقه‌ای سیلاب در طراحی و کاربری سیستم‌های منابع آب، نقشه کاربری اراضی و مدیریت، تشخیص بیمه سیل، محافظت از نواحی مسکونی و غیره مفید است. در روش‌های سنتی منطقه‌ای کردن، یک حوزه براساس عدم شباهت به حوزه‌های آبخیز دیگر منطقه طبقه‌بندی می‌شد، بدون توجه به اینکه قسمتی از خصوصیات اکثر حوزه‌ها مشابه آن خصوصیت در حوزه‌های آبخیز دیگر است. اما برعکس این روش‌ها، تحلیل خوشه‌ای فازی به یک حوزه آبخیز اجازه می‌دهد که در تمامی خوشه‌ها (گروه‌ها) عضویت داشته باشد.

در اکثر مطالعات منطقه‌ای کردن برای خوشه‌بندی حوزه‌های آبخیز به گروه‌های همگن هیدرولوژیک از تحلیل خوشه‌ای سخت (خوشه‌بندی k-میانگین، خوشه‌بندی سلسله مراتبی) استفاده شده است. به‌عنوان نمونه می‌توان به رائو و حامد [۱۰] رستمی و همکاران [۱۲] اشاره کرد. اما تلاش‌های کمی برای کشف پتانسیل خوشه‌بندی فازی برای دستیابی به مناطق همگن هیدرولوژیک صورت گرفته است. جینگی و هال [۸] روش‌های خوشه‌بندی وارد، شبکه عصبی و فازی را برای خوشه‌بندی حوزه رودخانه گان مینگ در جنوب شرقی چین بکار بردند. نتایج آن‌ها نشان داد که روش خوشه‌بندی فازی و خوشه‌بندی شبکه عصبی از روش‌های سنتی کاراتر هستند. رائو و اسرینیواس [۱۱] از خوشه‌بندی فازی برای طبقه‌بندی حوزه‌های آبخیز ایندیانا در ایالات متحده آمریکا استفاده کردند. آن‌ها از ۵ شاخص ارزیابی خوشه‌بندی برای تعیین تعداد خوشه بهینه استفاده کردند و در نهایت حوزه‌های آبخیز ایندیانا را به ۷ خوشه همگن تقسیم کردند.

در این مطالعه اولین هدف ارزیابی پتانسیل الگوریتم خوشه‌بندی فازی در دست‌یابی به مناطق همگن برای تحلیل فراوانی منطقه‌ای سیلاب است و حساسیت نتایج الگوریتم با تغییر در مقدار پارامتر فازی‌سازی نیز ارزیابی می‌شود. دومین هدف، ارزیابی چند شاخص

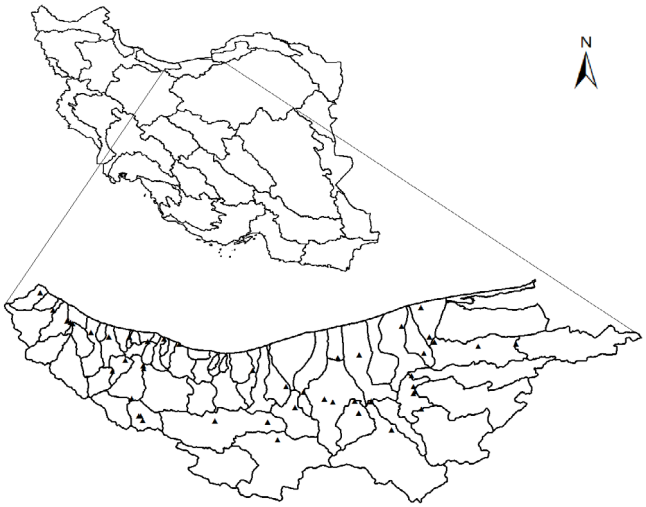
۱- دانشجوی دکتری آبیاری و زهکشی، دانشکده کشاورزی، دانشگاه فردوسی مشهد. (Email: farhadfarsad@ymail.com)
۲- استادیار گروه آبیاری و زهکشی، دانشکده کشاورزی، دانشگاه زابل.
۳- دانشیار هیدرولوژی، دانشکده منابع طبیعی، دانشگاه زابل.

خوشه‌بندی فازی در تعیین تعداد خوشه بهینه در الگوریتم خوشه‌بندی فازی است.

مواد و روش‌ها

منطقه مورد مطالعه

منطقه مورد مطالعه، استان مازندران در شمال ایران می‌باشد (شکل۱). ارتفاع این استان از زیر صفر در کناره‌های دریای خزر تا ۵۶۷۰ متر در قله دماوند متغیر است [٤]. میانگین بارندگی در غرب استان مازندران بیشتر از ۱۰۰۰ میلی‌متر و در شرق منطقه بیش از ۳۰۰ میلی‌متر است. قسمت مرطوب غربی ارتفاع کمی دارد اما بارندگی زیاد در آن باعث مناسب شدن آن منطقه برای فعالیت‌های کشاورزی شده است. برای تحلیل آماری ایستگاه‌های آب‌سنجی منطقه، از طریق آب منطقه‌ای استان مازندران آمار حداکثر دبی سالانه و خصوصیات حوزه‌های آبخیز استان مازندران تهیه شد.



شکل۱ – استان مازندران (منطقه مورد مطالعه)

خوشه بندی فازی

الگوریتم c–میانگین فازی (FCM)[1] توسط دون [۳] پیشنهاد گردید و توسط بزدک [۲] گسترش یافت. این الگوریتم، براساس بهینه‌سازی یک تابع هدف با تکرار است، که در تفکیک N حوزه آبخیز موجود در یک ناحیه به Cخوشهٔ فازی مفید است.

الگوریتم FCM با حداقل کردن تابع هدف زیر ماتریس خصوصیات ورودی X را به C زیر مجموعه (یا خوشه) تقسیم می‌کند [۱].

$$Minimize \quad J(U,V:X)=\sum_{i=1}^{C}\sum_{K=1}^{N}(u_k)^m d^2(X_K,V_i) \qquad (1)$$

U در معادله (۱) ماتریس افراز فازی است که شامل میزان تعلق هر بردار مشخصه تبدیل مقیاس شده در هر خوشه فازی می‌باشد؛ پارامتر $m\in[1,\infty]$ توان وزنی یا پارامتر فازی سازی است که درجه فازی بودن نتایج خوشه‌بندی را کنترل می‌کند و درجه همپوشانی بین خوشه‌ها است. $d^2(x_k,v_i)$ فاصله k- امین بردار خصوصیات تغییر مقیاس یافته (x_k) تا مرکز i–امین خوشه (v_i) است.

شاخص‌های ارزیابی خوشه‌بندی فازی

اگر چه در روش FCM تعداد خوشه‌ها از قبل مشخص شده است ولی در ابتدای کار تعداد خوشه‌ها برای هیدرولوژیست‌ها مشخص نیست و با روش سعی و خطا تعداد مناسب خوشه‌ها تعیین می‌شود. برای مشخص کردن تعداد درست خوشه‌ها توابع ارزیابی مختلفی تعریف شده است که می‌توان با استفاده از آن‌ها تعداد خوشه‌ها را مشخص کرد. در این مطالعه شاخص‌های ضریب افراز(PC)[2] [۱]، افراز بی‌نظمی(PE)[3](یا طبقه‌بندی بی‌نظمی) [۲]، عملکرد درجه فازی بودن FPI، طبقه‌بندی بی‌نظمی نرمالیزه(NCE) [۱۳]، تابع زی و بنی [١٤] و وون [۹] استفاده شده است. افراز بهینه مربوط به حداکثر مقدار PC (یا حداقل مقدار PE و FPI و NCE) است، که به حداقل همپوشانی بین خوشه‌ها اشاره دارد، همچنین مقدار حداقل تابع زی و بنی و وون به یک افراز خوب اشاره دارد، که به فشردگی و نیز جدایی خوب خوشه‌ها بستگی دارد.

آزمون گشتاورهای خطی[٦]

آزمون ناهمگونی[4]: مکان‌های ناجور را نسبت به کل گروه مشخص می‌کند. اگر مقدار D بزرگ‌تر از ۳ باشد ایستگاه ناجور بوده و از گروه حذف می‌شود.

آزمون ناهمگنی[5] منطقه‌ای: این آزمون شامل سه آماره H_1, H_2, H_3 است، اگر مقدار این آماره‌ها کم‌تر از یک باشد، منطقه همگن و اگر بین ۱ تا ۲ باشد، منطقه تا حدی همگن و اگر بزرگ‌تر از ۲ باشد، منطقه کاملاً ناهمگن است.

آزمون نکوئی برازش: به‌منظور انتخاب بهترین تابع توزیع منطقه‌ای انجام می‌شود و شامل محاسبه آماره Z^{DIST} است. تابع توزیع مناسب تابعی است که $|Z^{DIST}|$<۱/٦٤ باشد.

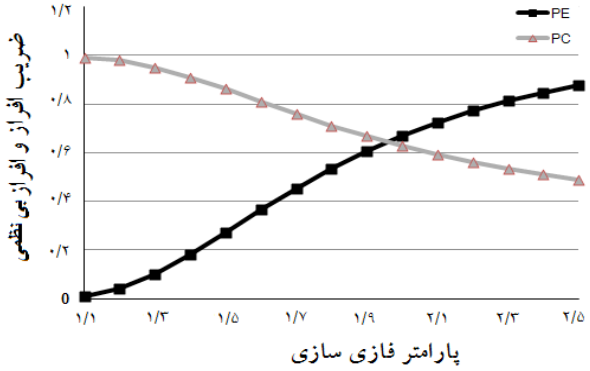
تجزیه و تحلیل نتایج

در این مقاله براساس ضرایب همبستگی محاسبه شده میان میانگین سیل سالانه و خصوصیات فیزیوگرافی و هواشناسی حوزه‌های آبخیز استان مازندران، ۵ خصوصیت طول آبراهه اصلی، مساحت حوزه، میانگین بارندگی سالانه، طول جغرافیایی و عرض جغرافیایی برای ورودی خوشه‌بندی منطقه انتخاب شدند.

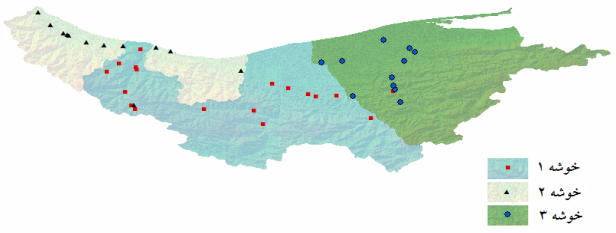
همگنی خوشه‌های (مناطق) به‌دست آمده از الگوریتم FCM توسط شاخص ناهمگنی هاسکینگ و والیس [۷] بررسی می‌شود.

1-Fuzzy c-Means
2-Partition coefficient
3- Partition entropy

4- Discordancy
5- Heterogeneity



شکل ۲- نمودار شاخص‌های ضریب افراز و افراز بی‌نظمی به ازاء مقادیر مختلف پارامتر فازی‌سازی و تعداد خوشه برابر ۳



شکل ۳ مناطق همگن هیدرولوژیک ایجاد شده با خوشه‌بندی فازی

جدول ۱- مقادیر آماره ناهمگنی قبل و بعد از تعدیل

شماره منطقه	N	قبل از تعدیل H_1	H_2	H_3	N	بعد از تعدیل H_1	H_2	H_3
۱	۲۴	۲/۵۲	۱/۵	۰/۴	۱۹	۰/۹۶	۰/۴۵	-۰/۱۹
۲	۱۳	۲/۴۶	۱/۷۵	۱/۲۴	۱۱	۰/۵	۰/۸۴	۰/۵۳
۳	۱۰	۰/۶۱	-۰/۳۹	-۰/۹۳	۱۲	۰/۸۶	-۰/۱۲	۰/۵۱

N: تعداد ایستگاه؛ H_i: آماره ناهمگنی

جدول ۲ - مقادیر آماره Z^{DIST} برای انتخاب بهترین توزیع منطقه‌ای

توزیع آماری	مقادیر آماره Z^{DIST} منطقه ۱	منطقه ۲	منطقه ۳
لوجستیک تعمیم یافته	۰/۷۲*	۰/۴۶*	۰/۳۹*
مقادیر حدی تعمیم یافته	۰/۴۵*	۱/۳*	۰/۸۷*
لوگ نرمال	۳۶/۲	۹۳/۱	۷۴/۱
پیرسون نوع ۳	۹۲/۳	۳/۳	۲۲/۳
پارتو تعمیم یافته	۶۷/۳	۱۴/۳	۵۱/۲

آماری منتخب توزیع‌های*

زمانی که تمام ۴۷ ایستگاه به‌عنوان یک منطقه مجزا مطرح شدند، منطقه بسیار ناهمگن بود (H_1=۲/۸۷، H_2=۱/۶۶، H_3=۰/۸۷). به‌طور کلی با افزایش تعداد خوشه‌ها، الگوریتم مناطقی که نسبتا همگن‌اند را آماده می‌کند، البته باید به این نکته توجه نمود که با افزایش مقدار c تعداد ایستگاه موجود در هر خوشه کاهش می‌یابد.

برای تعیین تعداد خوشه بهینه شاخص‌های ارزیابی خوشه‌بندی فازی به‌دست آمد. ضریب افراز (PC) به ازاء ۱/۹< m >۱/۴، c= ۳ را به‌عنوان تعداد خوشه بهینه معرفی می‌کند. افراز بی‌نظمی (PE) نیز به ازاء مقادیر فازی سازی ۱/۵ و ۱/۴ تعداد خوشه برابر ۳ را پیشنهاد می‌کند اما به ازاء مقادیر دیگر m، تعداد خوشه برابر ۲ را پیشنهاد می‌کند. به طور کلی، به ازاء بیشتر مقادیر m زمانی که تعداد خوشه برابر ۲ باشد PC حداکثر مقدار و PE حداقل مقدار را دارد که ناکارایی این دو شاخص را نشان می‌دهد. شاخص زی و بنی تعمیم یافته و شاخص وون برای تمام مقادیر m تعداد خوشه بهینه را برابر ۳ پیشنهاد می‌کند.

شاخص FPI نیز به ازاء مقادیر ۲/۴< m >۱/۶ تعداد خوشه‌ها را برابر ۳ پیشنهاد می‌کنند و شاخص NCE نیز در c= ۳ کاهش زیادی نشان می‌دهد. شاخص ناهمگنی هاسکینگ و والیس نیز به ازاء c= ۳ نشان داد که خوشه‌ها به همگنی نزدیک‌تر هستند. همان‌طور که در شکل (۲) مشاهده می‌شود در c= ۳ و m= ۳، ضریب افراز بیشترین مقدار و افراز بی‌نظمی کم‌ترین مقدار را داراست. به این ترتیب تعداد خوشه‌ها برابر ۳ و مقدار فازی سازی m=۲ برای مرحله بعدی تحلیل‌ها انتخاب می‌شود.

در تحلیل خوشه‌ای فازی، شناخت توزیع عضویت یک حوزه آبخیز درون یک منطقه فازی می‌تواند در تعدیل و بهبود همگنی مناطق مفید واقع شود. در این مطالعه، ابتدا ایستگاه‌هایی که توسط آماره ناجوری (D) هاسکینگ و والیس ناهمگون شناخته شدند، حذف گردیدند. دوماً، آماره ناهمگنی منطقه‌ای (H) با حذف و یا تعویض ایستگاه‌ها تعدیل شد. سوماً، ایستگاه‌های ناجور، که به میزان معنی داری از کم شدن آماره ناهمگنی منطقه‌ای (H) ممانعت می‌کنند، شناسایی شده و بعد از حصول اطمینان از زیاد نبودن عضویت فازی این ایستگاه‌ها، حذف شدند. در جدول ۱ مقادیر آماره ناهمگنی را قبل و بعد از تعدیل مناطق نشان می‌دهد، نتایج نشان می‌دهد که هر سه منطقه پس از تعدیل به طور قابل قبولی همگن‌اند (H_i<۱).

شکل ۳ مناطق همگن نهایی را پس از تعدیل نشان می‌دهد. به‌منظور انتخاب بهترین توزیع منطقه‌ای، آماره Z^{DIST} برای ایستگاه‌های هر منطقه با استفاده از برنامه نوشته شده به زبان فرترن [۵] برای توزیع‌های ۳ پارامتری محاسبه شد(جدول شماره ۲).

بحث و نتیجه‌گیری

در ذیل خلاصه‌ای از نتایج این پژوهش ارائه شده است.

۱) خوشه‌های حاصل از الگوریتم خوشه‌بندی فازی برای منطقه‌ای کردن به تغییرات میزان فازی سازی (m) برای یک مقدار مشخص از (c) حساس است. از این رو تعین مقدار بهینهٔ m برای منطقه‌ای کردن باید به دقت بررسی شود.

۲) شاخص‌های ارزیابی خوشه‌بندی فازی مانند ضریب افراز و افراز بی‌نظمی، که با داده‌ها ارتباط مستقیم کمی دارند، در تفکیک مناطق همگن هیدرولوژیک کارایی ندارند. صرف نظر از مقدار فازی‌سازی (m) انتخاب شده، اغلب هر دو شاخص تعداد خوشه را برابر ۲ پیشنهاد می‌کنند.

۳) شاخص‌های ارزیابی خوشه‌بندی فازی که همزمان میزان فازی سازی و ساختار داده‌ها را محاسبه می‌کنند (مانند شاخص‌های زی و بنی و وون)، کاملاً در تعیین بهترین افراز با الگوریتم خوشه‌بندی فازی موثراند.

٤) خوشه‌های فازی که از الگوریتم خوشه‌بندی فازی حاصل شدند همواره از لحاظ هیدرولوژیک همگن نیستند. تنها زمانی امکان همگنی این خوشه‌ها وجود دارد که تعدیل شوند.

منابع:

1- Bezdek, J.C. 1974. Numerical taxonomy with fuzzy sets. Mathematical Biology 1: 57–71.

2- Bezdek, J.C. 1981. Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press New York.

3- Dunn, J.C. 1974. A fuzzy relative of the ISODATA process and its use in detecting compact, well-separated clusters. Journal of Cybernetics. 3 (3): 32–57.

4- Farsadnia, F., and Moghaddamnia, A. 1392. Regional flood frequency analysis by self-organizing feature maps and fuzzy clustering approach. Iran-Water Resources Research. 9(3): 36 – 24. (In Persian)

5- Hosking, J.R.M. 2005. Fortran routines for use with the method of L moments. Version 3.04, IBM Research Division. T.J. Watson Research Center, Yorktown Heights, NY.

6- Hosking, J.R.M., and Wallis, J.R. 1997. Regional Frequency Analysis: An approach based on L-Moments. Cambridge University press, Cambridge.

7- Hosking, J.R.M., and Wallis, J.R. 1993. Some statistics useful in regional frequency analysis. Water Resources Research 29 (2), 271–281. 31(1): p. 251, 1995.

8- Jingyi, Z., and Hall, M.J. 2004. Regional flood frequency analysis for the Gan-Ming River basin in China. Journal of Hydrology 296, 98–117.

9- Kwon, S.H. 1998. Cluster validity index for fuzzy clustering. Electronics Letters. 34 (22): 2176–2177.

10- Rao, A.R., and Hamed, K.H. 1997. Regional frequency analysis of Wabash river flood data by L-moments. J. Hydrol. Eng. 2(4):169–179.

11- Rao, A.R., and Srinivas, V.V. 2006. Regionalization of watersheds by fuzzy cluster analysis. Journal of hydrology. 318: 57-79.

12- Rostami, R,. Sedghi, H., and Motamedi, A. 2009. Dez basin flood frequency analysis. Journal of Water Engineering. 61-70. (In Persian)

13- Roubens, M. 1982. Fuzzy clustering algorithms and their cluster validity. European Journal of Operational Research. 10(3): 294–301.

14- Xie, X.L., and Beni, G. 1991. A validity measure for fuzzy clustering. IEEE Transactions on Pattern Analysis and Machine Intelligence. 13 (8): 841–847.

نشریه علمی- پژوهشی
علوم و مهندسی آبخیزداری ایران
Iran-Watershed Management
Science & Engineering

Vol. 8, No. 24, Spring 2014
سال هشتم- شماره ۲۴- بهار ۱۳۹۳

Abstract (Technical Note)

Flood Regionalization of Mazandaran Province's Watersheds by Using Fuzzy Clustering Algorithm

F. Farsadnia[1], P. Haghighat Jou[2], Z. Shamohamadi Heidari[2] and A. Moghaddam Nia[3]

Received: 2011. 01. 10 Accepted: 2012. 03. 13

One of techniques for estimation of flood quantiles in ungauged watersheds is regional frequency analysis. In the common techniques of regionalization, homogeneous hydrological regions are created based on watershed characteristics such as physiographic characteristics of watershed, geographic location and flood statistics at each station. However, most watersheds only partially resemble other watersheds in a region. Therefore one watershed cannot be completely assigned to one group or another. The fuzzy clustering algorithm allows a watershed to have partial or distributed memberships of all the regions (groups) identified. In this paper, performance of fuzzy clustering algorithm has been evaluated for regionalization of Mazandaran province's watersheds, and some solutions have been represented to improve homogeneity of the regions that have been clustered by this algorithm and are not statistically homogeneous. In addition, several indices for performance evaluation of fuzzy clustering algorithm have been presented to determine the number of optimum clusters. Then, the Zdist statistic for the region's stations was calculated by FORTRAN programming language for the 3-parameter distributions and the best distribution was introduced for each region. The results showed that Mazandaran province has been composed of three hydrological homogeneous regions and generalized logistic distribution with different parameters was selected for each of the homogeneous regions.

Keywords: Regionalization; Fuzzy cluster analysis; Linier moment; Hydrologic homogeneity; Mazandaran Province's Watersheds.

1- Ph.D. Students of Irrigation and Drainage, Ferdowsi University of Mashhad. (Email: farhadfarsad@ymail.com)
2- Assistant Professors of Irrigation, Faculty of Agriculture, University of Zabol, Iran.
3- M.Sc. student of desert region management, University of Tehran, rshahbandari@ut.ac.ir