

بِسْمِ اللّٰهِ الرَّحْمٰنِ الرَّحِيْمِ



دانشگاه صنعتی شهرود

دانشکده مهندسی کشاورزی

گروه آب و خاک

ارزیابی عملکرد روش‌های تحلیل خوشهای در منطقه‌بندی حوضه‌های آبریز و تأثیر آن‌ها بر
تحلیل فراوانی منطقه‌ای سیلاب

دانشجو: علی آهنی

اساتید راهنما:

دکتر صمد امامقلی زاده

دکتر سید سعید موسوی ندوشنی

استاد مشاور:

دکتر خلیل اژدری

پایان نامه جهت اخذ درجه کارشناسی ارشد

شهریور ۱۳۹۲

دانشگاه صنعتی شهرود

دانشکده مهندسی کشاورزی

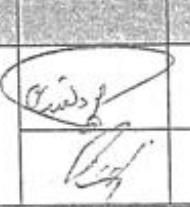
گروه آب و خاک

پایان نامه کارشناسی ارشد آقای علی آهنی

تحت عنوان: ارزیابی عملکرد روش‌های تحلیل خوش‌های در منطقه بندی حوضه‌های آبریز و تأثیر آن‌ها بر تحلیل فراوانی منطقه‌ای سیلاب

در تاریخ ۱۳۹۲/۶/۲۵ توسط کمیته تخصصی زیر جهت اخذ مدرک کارشناسی ارشد مهندسی کشاورزی- سازه‌های آبی مورد ارزیابی و با درجه عالی مورد پذیرش قرار گرفت.

اعضاء	اساتید مشاور	اعضاء	اساتید راهنمای
	نام و نام خانوادگی: دکتر خلیل ازدری		نام و نام خانوادگی: دکتر محمد امامقلی زاده نام و نام خانوادگی: دکتر سید سعید موسوی ندوشنی

اعضاء	نماینده تحصیلات تکمیلی	اعضاء	اساتید داور
	نام و نام خانوادگی: دکتر مسعود حکیمی تبار		نام و نام خانوادگی: دکتر مهدی دلتندی نام و نام خانوادگی: دکتر زهرا گنجهی نوروزی

تقدیم به خانواده‌ی عزیزم،

که وجودشان تجلی محبی پایان الی است.

تشکر و قدردانی

این پژوهش که نخستین پایان‌نامه‌ی رشته‌ی مهندسی کشاورزی- سازه‌های آبی در مقطع کارشناسی ارشد دانشگاه صنعتی شاهرود است، بی‌شک دچار کاستی‌هایی است که ناشی از کمبود دانش و پشتکار نگارنده‌ی آن است. با این حال امیدوارم با عنایت پروردگار یکتا، گامی هر چند ناچیز، اما صحیح در مسیر آشنایی بیشتر با دانش روز مهندسی آب برداشته باشیم.

بر خود وظیفه می‌دانم که مراتب سپاس خود را از دکتر صمد امامقلی‌زاده که مسئولیت راهنمایی این پایان‌نامه را با وجود مشکلات موجود در مسیر، با سعه‌ی صدر پذیرفتند، ابراز دارم. همچنین بدین وسیله از زحمات بی‌دریغ دکتر سید سعید موسوی ندوشنی که بخش عمده‌ی پژوهش حاضر، حاصل راهنمایی‌های آگاهانه و مشفقانه‌ی ایشان است، قدردانی می‌نمایم. در پایان سپاسگزاری خویش را از دکتر خلیل ازذری نیز که در زمان تهیه و تدوین این پایان‌نامه همواره از همکاری و مشورت‌های ایشان بهره بردم، اعلام می‌کنم.

امید است با عنایت خداوند متعال، پژوهشگران این مرز و بوم در مسیر رسیدن به مرزهای دانش هر روز پویاتر و استوارتر از پیش گام بردارند.

علی‌آهنی

شهریور ۱۳۹۲

تعهدنامه

اینجانب علی آهنی دانشجوی دوره کارشناسی ارشد رشته مهندسی کشاورزی- سازه‌های آبی دانشکده مهندسی کشاورزی دانشگاه صنعتی شاهرود نویسنده پایان نامه ارزیابی عملکرد روش‌های تحلیل خوش‌های در منطقه بندی حوضه‌های آبریز و تأثیر آن‌ها بر تحلیل فرآونی منطقه‌ای سیلاپ تحت راهنمایی دکتر صمد امامقلی‌زاده و دکتر سید سعید موسوی ندوشنی متعهد می‌شوم:

- تحقیقات در این پایان نامه توسط اینجانب انجام شده است و از صحت و اصالت برخوردار است.
- در استفاده از نتایج پژوهش‌های محققان دیگر به مرجع مورد استفاده استناد شده است.
- مطالب مندرج در پایان نامه تاکنون توسط خود یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی در هیچ جا ارائه نشده است.
- کلیه حقوق معنوی این اثر متعلق به دانشگاه صنعتی شاهرود می‌باشد و مقالات مستخرج با نام «دانشگاه صنعتی شاهرود» و یا «Shahrood University of Technology» به چاپ خواهد رسید.
- حقوق معنوی تمام افرادی که در به دست آمدن نتایج اصلی پایان نامه تأثیرگذار بوده‌اند در مقالات مستخرج از پایان نامه رعایت می‌گردد.
- در کلیه مراحل انجام این پایان نامه، در مواردی که از موجود زنده (یا بافت‌های آن‌ها) استفاده شده است، ضوابط و اصول اخلاقی رعایت شده است.
- در کلیه مراحل انجام این پایان نامه، در مواردی که به حوزه اطلاعات شخصی افراد دسترسی یافته یا استفاده شده است اصل رازداری، ضوابط و اصول اخلاق انسانی رعایت شده است.

تاریخ جلسه دفاع از پایان نامه: ۱۳۹۲/۶/۲۵

امضای دانشجو:

مالکیت نتایج و حق نشر

- کلیه حقوق معنوی این اثر و محصولات آن (مقالات مستخرج، کتاب، برنامه‌های رایانه‌ای، نرم‌افزارها و تجهیزات ساخته شده) متعلق به دانشگاه صنعتی شاهرود می‌باشد. این مطلب باید به نحو مقتضی در تولیدات علمی مربوطه ذکر شود.
- استفاده از اطلاعات و نتایج موجود در پایان نامه بدون ذکر مرجع مجاز نمی‌باشد.

* متن این صفحه نیز باید در ابتدای نسخه‌های تکثیر شده پایان نامه وجود داشته باشد.

چکیده

یکی از نخستین گام‌ها در فرآیند تحلیل فراوانی منطقه‌ای سیلاب، اختصاص ایستگاه‌های هیدرومتری مورد بررسی به مناطقی است که انتظار می‌رود ایستگاه‌های هیدرومتری موجود در آن‌ها دارای مکانیزم تولید سیلاب مشابه باشند. این فرآیند به عنوان منطقه‌بندی شناخته می‌شود. روش‌های منطقه‌بندی مبتنی بر تحلیل خوش‌های می‌توانند در تشخیص گروه‌های ایستگاه‌هایی که فرآیند تولید سیلاب مشابهی دارند، نقش مؤثری را ایفا کنند. در این پژوهش عملکرد انواع روش‌های تحلیل خوش‌های مانند خوش‌بندی ترکیبی، خوش‌بندی فازی با استفاده از الگوریتم فازی c-means و نیز خوش‌بندی با بهره‌گیری از نگاشتهای خودسازمانده کوهون در منطقه‌بندی حوضه‌ی آبریز سفیدرود بزرگ به منظور اجرای تحلیل فراوانی منطقه‌ای سیلاب مورد بررسی قرار گرفته است. همچنین ارزیابی وضعیت همگنی مناطق حاصل و اجرای تحلیل فراوانی منطقه‌ای با استفاده از الگوریتم گشتاورهای خطی انجام گرفته است. نتایج به دست آمده نشان داد که در میان الگوریتم‌های خوش‌بندی ترکیبی، ترکیب الگوریتم Ward با الگوریتم K-means می‌تواند مناسب‌ترین گزینه باشد. همچنین در مورد خوش‌بندی فازی با استفاده از الگوریتم فازی c-means نتایج حاصل حاکی از آن است که این الگوریتم در حوضه‌ی مورد مطالعه، از نظر تشکیل مناطق همگن و ارائه‌ی برآوردهای مناسب در تحلیل فراوانی منطقه‌ای سیلاب با استفاده از الگوریتم گشتاورهای خطی، عملکرد قابل قبولی دارد. افزون بر این، مشاهده شد که استفاده از خوش‌بندی فازی می‌تواند برآوردهای قابل اعتماد سیلاب را برای دوره‌های بازگشت طولانی‌تر امکان‌پذیر کند. در مورد نگاشتهای خودسازمانده مشاهده شد که مناطق مربوط به حالت ۳ منطقه‌ای حاصل از این روش، هم از نظر شاخص‌های صحت خوش‌بندی و هم از جهت همگنی مناطق تشکیل شده وضعیت مطلوبی را دارا هستند.

واژگان کلیدی: تحلیل فراوانی منطقه‌ای سیلاب، منطقه‌بندی، تحلیل خوش‌های، خوش‌بندی ترکیبی، خوش‌بندی فازی، نگاشتهای خودسازمانده، گشتاورهای خطی

فهرست مقالات مستخرج از پایان نامه

- «منطقه‌بندی حوضه آبریز سفیدرود بزرگ با استفاده از روش‌های خوش‌بندی ترکیبی»؛ هفتمین کنگره ملی مهندسی عمران، اردیبهشت ۱۳۹۲، دانشگاه سیستان و بلوچستان (پذیرش و ارائه شده در بخش مقالات شفاهی).
- «منطقه‌بندی فازی حوضه‌های آبریز با استفاده از الگوریتم *c-means* به منظور انجام تحلیل فراوانی منطقه‌ای سیلاب»؛ ارسال شده برای مجله علمی- پژوهشی مهندسی آبیاری و آب (در حال داوری) در اسفند ۱۳۹۱.
- «منطقه‌بندی حوضه‌های آبریز با استفاده از نگاشت‌های خودسازمانده»؛ ارسال شده برای مجله علمی- پژوهشی عمران مدرس (در حال داوری) در فروردین ۱۳۹۲.
- «تحلیل فراوانی منطقه‌ای سیلاب با استفاده از تحلیل خوش‌بندی ترکیبی و گشتاورهای خطی»؛ ارسال شده برای مجله علمی- پژوهشی مدیریت حوزه آبخیز (در حال داوری) در اردیبهشت ۱۳۹۲.

فهرست عناوین مطالب

۱	فصل اول: مقدمه
۲	۱-۱- تعریف مسئله و ضرورت تحقیق
۳	۲-۱- فرضیات تحقیق
۴	۳-۱- اهداف تحقیق
۵	۴-۱- سازماندهی متن پایان نامه
۶	فصل دوم: کلیات و مروری بر مطالعات پیشین
۷	۱-۲- منطقه‌بندی برای تحلیل فراوانی سیلاپ
۸	۲-۲- روش‌های منطقه‌بندی
۹	۳-۲- تحلیل خوش‌های در منطقه‌بندی
۹	۴-۲- ویژگی‌های مورد استفاده در منطقه‌بندی
۱۱	۵-۲- طبقه‌بندی الگوریتم‌های خوش‌بندی
۱۴	۶-۲- مراحل منطقه‌بندی به وسیله‌ی تحلیل خوش‌های
۱۶	فصل سوم: مواد و روش‌ها
۱۷	۱-۳- خوش‌بندی ترکیبی
۱۷	۱-۱-۳- معرفی تحلیل خوش‌های ترکیبی
۱۷	۲-۱-۳- طبقه‌بندی الگوریتم‌های خوش‌بندی سخت
۱۸	۱-۲-۱-۳- روش‌های خوش‌بندی سلسله مراتبی
۱۸	۱-۲-۱-الف- خوش‌بندی سلسله مراتبی متراکم کننده
۲۱	۱-۲-۱-ب- خوش‌بندی سلسله مراتبی تقسیم کننده
۲۲	۲-۱-۳- روش‌های خوش‌بندی افزایی

۲۴	۳-۲-۱-۳- خوشبندی ترکیبی
۲۵	۳-۱-۳- الگوریتم‌های خوشبندی ترکیبی و ارزیابی عملکرد آن
۲۵	۳-۱-۳-۱-۳- الگوریتم ترکیبی
۲۷	۳-۱-۳-۲- الگوریتم‌های تک پیوند، تمام پیوند و پیوند متوسط
۲۸	۳-۳-۱-۳- Ward
۲۹	۳-۱-۳-۴- شاخص صحت خوشبندی عرض silhouette
۳۰	۳-۲- خوشبندی فازی
۳۰	۳-۲-۱- معرفی خوشبندی فازی
۳۲	۳-۲-۲- الگوریتم فازی c-means
۳۲	۳-۲-۲-۱- توصیف الگوریتم
۳۸	۳-۲-۲-۲- اختصاص سایت‌های جدید به خوشبندی فازی
۳۹	۳-۲-۳- شاخص‌های صحت خوشبندی فازی
۴۴	۳-۳- خوشبندی با استفاده از نگاشتهای خودسازمانده
۴۴	۳-۳-۱- معرفی نگاشتهای مشخصه‌ی خود سازمانده کوهون
۴۷	۳-۳-۲- الگوریتم نگاشت مشخصه‌ی خود سازمانده کوهون
۵۲	۳-۴- تحلیل فراوانی منطقه‌ای سیلاب با استفاده از تئوری گشتاورهای خطی
۵۲	۳-۴-۱- معرفی و مروری بر مطالعات مربوط به گشتاورهای خطی
۵۳	۳-۴-۲- روش سیلاب نمایه دالریمپل
۵۴	۳-۴-۳- گشتاورهای خطی توزیع‌های احتمال
۵۵	۳-۴-۴- برآورد پارامترها با استفاده از گشتاورهای خطی
۵۶	۳-۴-۵- استفاده از گشتاورهای خطی در تحلیل فراوانی منطقه‌ای
۵۶	۳-۴-۵-۱- غربال کردن داده‌ها و شاخص ناجوری D

۵۸	۲-۵-۴-۳- تشخیص همگنی مناطق و استفاده از شاخص ناهمگنی H
۶۰	۳-۵-۴-۳- انتخاب یک توزیع فراوانی منطقه‌ای و شاخص نکویی برازش Z^{DIST}
۶۱	۴-۵-۴-۳- توزیع ویکبی
۶۲	۵-۵-۴-۳- الگوریتم گشتاورهای خطی منطقه‌ای
۶۳	۵-۵-۳- معرفی حوضه‌ی آبریز سفیدرود بزرگ
۶۶	۶-۳- داده‌های آماری سیلان
۶۸	۷-۳- زبان برنامه‌نویسی R
۷۰	فصل چهارم: محاسبات، نتایج و بحث
۷۱	۱-۴- مقدمه
۷۱	۲-۴- محاسبه‌ی گشتاورهای خطی و نسبت‌های گشتاور خطی
۷۳	۳-۴- شاخص D و شناسایی ایستگاه‌های ناجور
۷۴	۴-۴- ویژگی‌های مورد استفاده در خوشبندی و معیار فاصله
۷۴	۵-۴- منطقه‌بندی با استفاده از خوشبندی ترکیبی
۷۴	۱-۵-۴- خوشبندی با استفاده از الگوریتم‌های سلسله مراتبی
۷۵	۲-۵-۴- خوشبندی با استفاده از الگوریتم افزایی K-means
۷۸	۳-۵-۴- خوشبندی با استفاده از الگوریتم‌های خوشبندی ترکیبی
۷۸	۴-۵-۴- تعداد ایستگاه‌های اختصاص یافته به مناطق
۸۰	۵-۵-۴- کمینه‌سازی تابع هدف
۸۰	۶-۵-۴- شاخص صحت خوشبندی میانگین عرض silhouette
۸۱	۷-۵-۴- انتخاب الگوریتم خوشبندی بهینه
۸۴	۸-۵-۴- تحلیل فراوانی منطقه‌ای با استفاده از گشتاورهای خطی

۸۴	۱-۸-۵-۴- تعیین وضعیت همگنی مناطق و تعداد ایستگاههای هر منطقه
۹۰	۲-۸-۵-۴- شاخص نکویی برازش Z و انتخاب توزیع فراوانی منطقه‌ای سیلاب
۹۱	۳-۸-۵-۴- برآورد چندک‌های منطقه‌ای سیلاب و دبی سیلاب ایستگاهها
۹۶	۴-۶- خوشبندی با استفاده از الگوریتم فازی c-means
۹۶	۴-۶-۱- ایستگاهها مورد مطالعه و تعیین محدوده‌ی تعداد مناطق
۹۶	۴-۶-۲- تعیین مراکز اولیه‌ی خوشبندی
۹۷	۴-۶-۳- محدوده‌ی تغییرات μ
۹۷	۴-۶-۴- محاسبه‌ی مقادیر تابع هدف و بررسی تغییرات آن
۹۸	۴-۶-۵- محاسبه‌ی شاخص‌های صحت خوشبندی فازی
۱۰۳	۴-۶-۶- آستانه‌ی عضویت فازی و اختصاص ایستگاهها به خوشبندی
۱۰۹	۴-۶-۷- تحلیل فراوانی منطقه‌ای با استفاده از گشتاورهای خطی
۱۰۹	۴-۶-۷-۱- تعیین وضعیت همگنی مناطق و ایستگاههای موجود در هر منطقه
۱۱۵	۴-۶-۷-۲- شاخص نکویی برازش Z و انتخاب توزیع فراوانی منطقه‌ای سیلاب
۱۱۷	۴-۶-۷-۳- برآورد چندک‌های منطقه‌ای سیلاب و دبی سیلاب ایستگاهها
۱۲۶	۴-۷-۱- خوشبندی با استفاده از نگاشتهای خودسازمانده کوهونن
۱۲۶	۴-۷-۱-۱- ایستگاههای اختصاص یافته به خوشبندی و پیوستگی‌های آنها
۱۳۲	۴-۷-۲- شاخص میانگین عرض silhouette برای حالت‌های مختلف خوشبندی
۱۳۳	۴-۷-۳- بررسی وضعیت همگنی مناطق و انتخاب توزیع منطقه‌ای
۱۳۵	۴-۷-۴- برآورد چندک‌های منطقه‌ای سیلاب و دبی سیلاب ایستگاهها
۱۳۹	۴-۸- تحلیل فراوانی نقطه‌ای سیلاب
۱۴۱	۴-۹- مقایسه‌ی مقادیر برآوردها

۱۵۰	فصل پنجم: نتیجه‌گیری نهایی و پیشنهادات
۱۵۱	۷-۱- تحلیل نتایج و نتیجه‌گیری نهایی
۱۵۶	۷-۲- پیشنهادات
۱۵۸	پیوست الف: زبان برنامه‌نویسی R
۱۵۹	الف-۱- مقدمه
۱۶۰	الف-۲- توابع مورد استفاده در این مطالعه
۱۶۰	الف-۲-۱- ماتریس فاصله
۱۶۰	الف-۲-۲- خوشبندی سلسله‌مراتبی
۱۶۱	الف-۲-۳- خوشبندی با استفاده از الگوریتم K-means
۱۶۲	الف-۲-۴- شاخص عرض silhouette
۱۶۲	الف-۲-۵- گشتاورهای نمونه‌ی مجموعه‌های چندگانه‌ی داده‌ها
۱۶۳	الف-۲-۶- محاسبه‌ی شاخص‌ها و آماره‌های آزمون H , D و Z برای تحلیل فراوانی منطقه‌ای
۱۶۴	الف-۲-۷- میانگین (متوسط) وزنی منطقه‌ای گشتاورهای خطی
۱۶۴	الف-۲-۸- برازش یک توزیع فراوانی منطقه‌ای
۱۶۵	الف-۲-۹- چندک‌ها و توابع چندک یک توزیع فراوانی منطقه‌ای
۱۶۵	الف-۲-۱۰- رسم نمودار تابع
۱۶۵	الف-۲-۱۱- چندک‌ها و توابع چندک منحصر به فرد سایتها در یک منطقه
۱۶۶	الف-۲-۱۲- خوشبندی فازی با استفاده از الگوریتم c-means
۱۶۷	الف-۲-۱۳- شاخص‌های صحت خوشبندی فازی
۱۶۷	الف-۲-۱۴- نگاشتهای خودسازمانده کوهون
۱۶۹	منابع و مراجع

فهرست شکل‌ها

شکل (۱-۳) ترسیم تعریف فاصله (یا اختلاف) بین دو خوشه در الگوریتم‌های خوشبندی سلسله مراتبی متراکم‌کننده	۲۸
شکل (۲-۳) طرح نمادین نگاشت مشخصه‌ی خود سازمانده کوهونز	۴۸
شکل (۳-۳) طرح نمادین قطر (D) شبکه‌ی کوهونز SOFM؛ (الف) شبکه‌ی مربعی، (ب) شبکه‌ی مستطیلی	۵۱
شکل (۴-۳) نمایش محدوده‌ی جغرافیایی حوضه‌ی آبریز سفیدرود بزرگ	۶۴
شکل (۵-۳) نمایش موقعیت جغرافیایی ایستگاه‌های هیدرومتری مورد مطالعه در حوضه‌ی آبریز سفیدرود بزرگ	۶۵
شکل (۱-۴) نمایش تغییر اختلاف بین دو میانگین تجمعی متوالی شاخص میانگین عرض K-means در ۵۰ انتخاب تصادفی مراکز اولیه‌ی الگوریتم silhouette	۷۶
شکل (۲-۴) نمایش تغییر اختلاف بین دو میانگین تجمعی متوالی تابع هدف (F) در ۵۰ انتخاب تصادفی مراکز اولیه‌ی الگوریتم K-means	۷۷
شکل (۳-۴) تعداد مناطق شامل ۵ ایستگاه یا کمتر در حالات ۲ تا ۵ خوشه‌ای برای الگوریتم-های مختلف خوشبندی	۷۹
شکل (۴-۴) مقادیر تابع هدف (F) در حالات ۲ تا ۵ خوشه‌ای برای روش‌های مختلف خوشه-بندی	۸۲
شکل (۵-۴) مقادیر شاخص میانگین عرض silhouette در حالات ۲ تا ۵ خوشه‌ای برای روش‌های مختلف خوشبندی	۸۳
شکل (۶-۴) مقادیر شاخص‌های ناهمگنی H در مناطق حاصل از الگوریتم WAKM برای ایستگاه هیدرومتری	۸۵

- شکل (۷-۴) مقادیر شاخص‌های ناهمگنی H در مناطق حاصل از الگوریتم WAKM برای ۳۷
۸۵ ایستگاه هیدرومتری
- شکل (۸-۴) تعداد ایستگاه‌های هیدرومتری در هر یک از مناطق در حالات ۲ تا ۵ خوش‌های
۸۶ برای الگوریتم WAKM
- شکل (۹-۴) تعداد داده‌های آماری موجود در هر یک از مناطق در حالات ۲ تا ۵ خوش‌های
۸۷ برای الگوریتم WAKM
- شکل (۱۰-۴) توزیع جغرافیایی ایستگاه‌های هیدرومتری مورد مطالعه در مناطق حالت دو
۸۸ منطقه‌ای
- شکل (۱۱-۴) توزیع جغرافیایی ایستگاه‌های هیدرومتری در مناطق حالت سه منطقه‌ای
۸۹
- شکل (۱۲-۴) چندک‌های سیلان منطقه‌ای (q) منطقه‌ی اول در برابر احتمال عدم تجاوز
۹۲ متناظر $(F(x))$
- شکل (۱۳-۴) چندک‌های سیلان منطقه‌ای (q) منطقه‌ی دوم در برابر احتمال عدم تجاوز
۹۲ متناظر $(F(x))$
- شکل (۱۴-۴) مقادیر برآورد شده‌ی سیلان با دوره‌های بازگشت ۱۰، ۲۵، ۵۰ و ۱۰۰ سال برای
۹۴ ایستگاه‌های هیدرومتری موجود در منطقه‌ی اول
- شکل (۱۵-۴) مقادیر برآورد شده‌ی سیلان با دوره‌های بازگشت ۱۰، ۲۵ و ۵۰ سال برای
۹۵ ایستگاه‌های هیدرومتری موجود در منطقه‌ی دوم
- شکل (۱۶-۴) تغییرات مقدار تابع هدف الگوریتم خوشبندی c-means با تغییر در تعداد
۹۷ خوش‌ها و مقدار μ
- شکل (۱۷-۴) تغییرات مقدار ضریب افزای V_{PC} با تغییر مقدار μ در فاصله‌ی ۱/۱ تا ۳ برای
۹۹ تعداد ۲ تا ۵ خوش

شکل (۱۸-۴) تغییرات مقدار انتروپی افزای (V_{PE}) با تغییر مقدار μ در فاصله‌ی ۱/۱ تا ۳ برای

۹۹

تعداد ۲ تا ۵ خوشه

شکل (۱۹-۴) تغییرات مقدار شاخص عملکرد فازی (V_{FPI}) با تغییر مقدار μ در فاصله‌ی ۱/۱ تا

۱۰۰

تا ۳ برای ۲ تا ۵ خوشه

شکل (۲۰-۴) تغییرات مقدار شاخص فوکویاما و سوگنو (V_{FS}) با تغییر μ در فاصله‌ی ۱/۱ تا

۱۰۱

برای ۲ تا ۵ خوشه

شکل (۲۱-۴) تغییرات شاخص ژی-بنی توسعه یافته ($V_{XB,m}$) با تغییر μ در فاصله‌ی ۱/۱ تا

۱۰۲

برای ۲ تا ۵ خوشه

شکل (۲۲-۴) تغییرات مقدار شاخص کوان (V_K) با تغییر مقدار μ در فاصله‌ی ۱/۱ تا ۳ برای

۱۰۲

تعداد ۲ تا ۵ خوشه

شکل (۲۳-۴) تعداد ایستگاه‌های هیدرومتری متعلق به هر یک از خوشه‌ها در حالت دو

۱۰۵

خوشه‌ای

شکل (۲۴-۴) تعداد ایستگاه‌های هیدرومتری متعلق به هر یک از خوشه‌ها در حالت سه

۱۰۵

خوشه‌ای

شکل (۲۵-۴) تعداد ایستگاه‌های هیدرومتری متعلق به هر یک از خوشه‌ها در حالت چهار

۱۰۷

خوشه‌ای

شکل (۲۶-۴) تعداد ایستگاه‌های هیدرومتری متعلق به هر یک از خوشه‌ها در حالت پنج

۱۰۷

خوشه‌ای

شکل (۲۷-۴) حاصل جمع تعداد ایستگاه‌های هیدرومتری اختصاص یافته به کل خوشه‌ها در

۱۰۸

هر یک از حالات ۲ تا ۵ خوشه‌ای در محدوده‌ی $1.1 \leq \mu \leq 3.0$

شکل (۲۸-۴) مقادیر شاخص ناهمگنی H برای مناطق تشکیل شده در حالت دو خوشه‌ای به

۱۱۰

وسیله‌ی الگوریتم c-means

شکل (۲۹-۴) مقادیر شاخص ناهمگنی H_1 در مناطق تشکیل شده در حالت سه خوش‌های به ۱۱۰

وسیله‌ی الگوریتم c-means

شکل (۳۰-۴) مقادیر شاخص ناهمگنی H_1 برای مناطق تشکیل شده در حالت چهار خوش‌های ۱۱۲

به وسیله‌ی الگوریتم c-means

شکل (۳۱-۴) مقادیر شاخص ناهمگنی H_1 برای مناطق تشکیل شده در حالت پنج خوش‌های ۱۱۳

به وسیله‌ی الگوریتم c-means

شکل (۳۲-۴) تعداد داده‌های آماری سه حالت برگزیده ۱۱۴

شکل (۳۳-۴) چندک‌های سیلاب منطقه‌ای (q) در برابر احتمال عدم تجاوز متناظر ($F(x)$) ۱۱۷
برای حالت سه منطقه‌ای

شکل (۳۴-۴) چندک‌های سیلاب منطقه‌ای (q) در برابر احتمال عدم تجاوز متناظر ($F(x)$) ۱۱۸
برای حالت چهار منطقه‌ای

شکل (۳۵-۴) چندک‌های سیلاب منطقه‌ای (q) در برابر احتمال عدم تجاوز متناظر ($F(x)$) ۱۱۹
برای حالت پنج منطقه‌ای

شکل (۳۶-۴) مقادیر برآورد شده‌ی سیلاب با دوره‌ی بازگشت ۱۰ سال برای تمام ایستگاه‌های ۱۲۱
هیدرومتری مورد مطالعه در سه حالت برگزیده

شکل (۳۷-۴) مقادیر برآورد شده‌ی سیلاب با دوره‌ی بازگشت ۲۵ سال برای تمام ایستگاه‌های ۱۲۲
هیدرومتری مورد مطالعه در سه حالت برگزیده

شکل (۳۸-۴) مقادیر برآورد شده‌ی سیلاب با دوره‌ی بازگشت ۵۰ سال برای ۲۶ ایستگاه ۱۲۳
هیدرومتری در سه حالت برگزیده

شکل (۳۹-۴) مقادیر برآورد شده‌ی سیلاب با دوره‌ی بازگشت ۱۰۰ سال برای ایستگاه‌های ۱۲۴
هیدرومتری متعلق به منطقه‌ی اول حالت سه منطقه‌ای با $\mu = 1.8$

شکل (۴۰-۴) مقادیر برآورده سیلاب با دوره‌ی بازگشت ۱۰۰ سال برای ایستگاه‌های

۱۲۵

هیدرومتری متعلق به منطقه‌ی دوم حالت پنج منطقه‌ای با $\mu = 2.8$

۱۲۶

شکل (۴۱-۴) توزیع ایستگاه‌های هیدرومتری بین دو گره لایه‌ی خروجی

۱۲۷

شکل (۴۲-۴) وضعیت متوسط ویژگی‌های تجدید مقیاس شده دو گره لایه‌ی خروجی

۱۲۸

شکل (۴۳-۴) توزیع ایستگاه‌های هیدرومتری بین سه گره لایه‌ی خروجی

۱۲۸

شکل (۴۴-۴) وضعیت متوسط ویژگی‌های تجدید مقیاس شده سه گره لایه‌ی خروجی

۱۲۹

شکل (۴۵-۴) توزیع ایستگاه‌های هیدرومتری بین چهار گره لایه‌ی خروجی

۱۲۹

شکل (۴۶-۴) وضعیت متوسط ویژگی‌های تجدید مقیاس شده چهار گره لایه‌ی خروجی

۱۳۰

شکل (۴۷-۴) توزیع ایستگاه‌های هیدرومتری بین پنج گره لایه‌ی خروجی

۱۳۱

شکل (۴۸-۴) وضعیت متوسط ویژگی‌های تجدید مقیاس شده پنج گره لایه‌ی خروجی

۱۳۲

شکل (۴۹-۴) مقادیر متوسط شاخص صحت خوشبندی میانگین عرض silhouette برای

۱۳۲

حالات ۲ تا ۵ خوشبندی

شکل (۵۰-۴) مقادیر شاخص‌های ناهمگنی H برای مناطق تشکیل شده با استفاده از نگاشت-

۱۳۳

های خودسازمانده کوهونن برای ۳۷ ایستگاه هیدرومتری مورد مطالعه در حوضه‌ی آبریز

سفیدرود بزرگ

شکل (۵۱-۴) چندک‌های سیلاب منطقه‌ای (q) در برابر احتمال عدم تجاوز متناظر ($F(x)$)

۱۳۵

برای سه منطقه‌ی تشکیل شده با استفاده از نگاشتهای خودسازمانده

شکل (۵۲-۴) مقادیر برآورده سیلاب با دوره‌های بازگشت ۱۰ و ۲۵ سال برای ایستگاه-

۱۳۷

های هیدرومتری منطقه‌ی اول تشکیل شده به وسیله‌ی نگاشتهای خودسازمانده

شکل (۵۳-۴) مقادیر برآورده سیلاب با دوره‌های بازگشت ۱۰، ۲۵ و ۵۰ سال برای

۱۳۷

ایستگاه‌های هیدرومتری منطقه‌ی دوم تشکیل شده به وسیله‌ی نگاشتهای خودسازمانده

۱۳۸

شکل (۵۴-۴) مقادیر برآورده سیلاب با دوره‌های بازگشت ۱۰، ۲۵، ۵۰ و ۱۰۰ سال

برای ایستگاههای هیدرومتری منطقه‌ی سوم تشکیل شده به وسیله‌ی نگاشتهای خودسازمانده

شکل (۵۵-۴) مقادیر برآورده سیلاپ با استفاده از تحلیل فراوانی نقطه‌ای با دوره ۱۴۰

بازگشت ۱۰ سال برای ایستگاههای هیدرومتری حوضه‌ی آبریز سفیدرود بزرگ

شکل (۵۶-۴) مقادیر برآورده سیلاپ با استفاده از روش‌های تحلیل فراوانی منطقه‌ای با ۱۴۲

دوره بازگشت ۱۰ سال برای ایستگاههای هیدرومتری حوضه‌ی آبریز سفیدرود بزرگ

شکل (۵۷-۴) اختلاف نسبی بزرگ‌ترین و کوچک‌ترین برآورد حاصل از پنج گزینه‌ی نهایی ۱۴۳

تحلیل فراوانی منطقه‌ای برای هر یک از ۳۷ ایستگاه هیدرومتری مورد مطالعه

شکل (۵۸-۴) مقادیر برآورده دبی حداکثر سیلاپ با استفاده از تحلیل فراوانی نقطه‌ای و ۱۴۴ میانگین مقادیر برآورده دبی حداکثر سیلاپ با استفاده از پنج گزینه‌ی نهایی تحلیل

فراوانی منطقه‌ای برای دوره بازگشت ۱۰ سال

شکل (۵۹-۴) نسبی برآورده سیلاپ ۱۰ ساله با استفاده از تحلیل فراوانی نقطه‌ای و میانگین ۱۴۵

دبی‌های سیلاپ ۱۰ ساله برآورده شده به وسیله‌ی پنج گزینه‌ی تحلیل فراوانی منطقه‌ای

شکل (۶۰-۴) مقادیر میانگین قدر مطلق اختلاف‌های بین برآوردهای دبی سیلاپ با دوره ۱۴۷ بازگشت ۱۰ سال به دست آمده از هر یک از گزینه‌های نهایی تحلیل منطقه‌ای و برآوردهای

سیلاپ ۱۰ ساله در تحلیل نقطه‌ای

شکل (۶۱-۴) مقادیر میانگین قدر مطلق اختلاف‌های نسبی بین برآوردهای دبی سیلاپ با دوره بازگشت ۱۰ سال به دست آمده از هر یک از گزینه‌های نهایی تحلیل منطقه‌ای و برآوردهای ۱۴۷

های سیلاپ ۱۰ ساله در تحلیل نقطه‌ای

شکل (۶۲-۴) مقادیر میانگین اختلاف بین برآوردهای دبی سیلاپ با دوره بازگشت ۱۰ سال ۱۴۸ به دست آمده از هر یک از گزینه‌های نهایی تحلیل منطقه‌ای و برآوردهای سیلاپ ۱۰ ساله در

تحلیل نقطه‌ای

شکل (۶۳-۴) مقادیر میانگین اختلافهای نسبی بین برآوردهای دبی سیلاب با دوره بازگشت

۱۴۹ ۱۰ سال به دست آمده از هر یک از گزینه‌های نهایی تحلیل منطقه‌ای و برآوردهای سیلاب

۱۰ ساله در تحلیل نقطه‌ای

فهرست جداول

جدول (۱-۳) شاخص‌های اختلاف برای محاسبه فاصله‌ی بین مراکز خوش‌ها یا بردارهای مشخصه	۱۹
جدول (۲-۳) مقادیر بحرانی شاخص ناجوری D_i	۵۷
جدول (۳-۳) مشخصات جغرافیایی، فیزیوگرافیک و هواشناسی ایستگاه‌های هیدرومتری مورد مطالعه	۶۷
جدول (۴-۱) مقادیر گشتاور خطی مرتبه اول، ضریب تغییرات خطی، چولگی خطی و کشیدگی خطی محاسبه شده برای ایستگاه‌های هیدرومتری مورد مطالعه	۷۱
جدول (۴-۲) مقادیر محاسبه شده‌ی شاخص ناجوری D برای ایستگاه‌های هیدرومتری مورد مطالعه	۷۳
جدول (۴-۳) مقادیر میانگین تابع هدف، شاخص میانگین عرض silhouette و تعداد خوشه-	
های شامل ۵ ایستگاه هیدرومتری یا کمتر برای الگوریتم‌های مختلف خوشبندی در حالات ۲ تا ۵ خوش‌های	۸۴
جدول (۴-۴) مقدار شاخص نکوبی برازش Z برای توزیع‌های مورد آزمایش در هر یک از مناطق حالت دو منطقه‌ای	۹۰
جدول (۴-۵) مقادیر برآورده شده‌ی پارامترهای توزیع‌های برگزیده‌ی منطقه‌ای	۹۱
جدول (۴-۶) مقادیر شاخص نکوبی برازش Z در هر یک از مناطق سه حالت برگزیده‌ی منطقه‌بندی فازی	۱۱۵
جدول (۴-۷) مقادیر برآورده شده‌ی پارامترهای توزیع‌های منطقه‌ای برای سه حالت برگزیده‌ی منطقه‌بندی فازی	۱۱۶
جدول (۴-۸) مقادیر شاخص نکوبی برازش Z در هر یک از مناطق حالت ۳ منطقه‌ای ایجاد	۱۳۴

شده با استفاده از نگاشتهای خودسازمانده

جدول (۴-۹) مقادیر برآوردهی پارامترهای توزیع‌های منطقه‌ای در هر یک از مناطق حالت

۱۳۴

۳ منطقه‌ای ایجاد شده با استفاده از نگاشتهای خودسازمانده

فهرست علائم و نشانه‌های اختصاری به کار رفته در پایان‌نامه

y_i	بردار مشخصه‌ی i ام در یک فضای n بعدی مانند $y_i = [y_{i1}, \dots, y_{in}] \in R^n$
x_i	نامین بردار تجدید مقیاس شده در فضای ویژگی n بعدی ($x_i = [x_{i1}, \dots, x_{in}] \in R^n$)
y_{ij}	مقدار ویژگی j در بردار مشخصه‌ی n بعدی y_i
x_{ij}	مقدار ویژگی j در بردار مشخصه‌ی تجدید مقیاس شده‌ی n بعدی x_i
σ_j	انحراف معیار ویژگی j ام
F	مقدار تابع هدف الگوریتم خوشبندی K-means
$d(\cdot)$	یک شاخص فاصله، مانند شاخص فاصله‌ی اقلیدسی
W	تابع هدف در الگوریتم Ward
$s(i)$	شاخص صحت خوشبندی عرض silhouette
u_{ik}	درجه‌ی عضویت k امین بردار مشخصه‌ی تجدید مقیاس شده x_k در خوشبندی فازی i ام
J	مقدار تابع هدف الگوریتم خوشبندی فازی c-means
V_{PC}	ضریب افزایش
V_{PE}	انتروپی افزایش
V_{FPI}	شاخص عملکرد فازی
V_{NCE}	انتروپی دسته‌بندی نرمال شده
V_{FS}	شاخص فوکویاما و سوگنو
V_{XB}	شاخص ژئی-بنی
$V_{XB,m}$	شاخص ژئی-بنی توسعه یافته
V_K	شاخص کوان
w_j	بردار وزن بین گره خروجی j و گره‌های لایه‌ی ورودی

ω	گرهی خروجی برنده
$\eta(t)$	۲ پارامتر میزان یادگیرندگی برای تکرار عددی t
$h_{j,\omega}(t)$	تابع همسایگی حول گره برنده ω
$d_{\omega,j}$	فاصله‌ی بین گره برنده‌ی ω و گره j همسایه‌ی آن در لایه‌ی خروجی
r_ω	موقعیت گره برنده‌ی ω
r_j	موقعیت گره j
$\sigma(t)$	عرض مؤثر همسایگی توپولوژیک $h_{j,\omega}(t)$ در گام زمانی t
$M_{p,r,s}$	گشتاورهای وزنی احتمال
α_r	گشتاور وزنی احتمال $M_{1,0,r}$
β_r	گشتاور وزنی احتمال $M_{1,r,0}$
λ_r	گشتاور خطی مرتبه r
τ	ضریب تغییرات خطی (L -CV)
τ_r	نسبت گشتاور خطی مرتبه r
λ_1	گشتاور خطی مرتبه‌ی اول یا موقعیت خطی
λ_2	گشتاور خطی مرتبه‌ی دوم یا مقیاس خطی
τ_3	چولگی خطی
τ_4	کشیدگی خطی
b_r	برآوردگر ناریب β_r
l_r	برآوردگر ناریب λ_r
t	ضریب تغییرات خطی (L -CV) نمونه
t_r	نسبت گشتاور خطی مرتبه r نمونه
l_1	برآوردگر موقعیت خطی نمونه

l_2	برآورده‌گر مقیاس خطی نمونه
t_3	چولگی خطی نمونه
t_4	کشیدگی خطی نمونه
$\tilde{\alpha}_r$	برآورده‌گر احتمال تجربی r
$\tilde{\lambda}_r$	برآورده‌گر احتمال تجربی r
$\tilde{\tau}_r$	برآورده‌گر احتمال تجربی r
D_i	شاخص ناجوری سایت i
$t^{(i)}$	ضریب تغییرات خطی ($L\text{-}CV$) نمونه سایت i
$t_3^{(i)}$	چولگی خطی نمونه سایت i
$t_4^{(i)}$	کشیدگی خطی نمونه سایت i
t^R	ضریب تغییرات خطی میانگین (متوسط) منطقه‌ای
t_3^R	چولگی خطی میانگین (متوسط) منطقه‌ای
t_4^R	کشیدگی خطی میانگین (متوسط) منطقه‌ای
V	انحراف معیار وزنی $L\text{-}CV$ های نمونه‌ی درون‌سایتی
μ_V	میانگین N_{sim} مقدار V
σ_V	انحراف معیار N_{sim} مقدار V
H	شاخص ناهمگنی منطقه
H^*	شاخص ناهمگنی جایگزین
B_4	اریب t_4^R
σ_4	انحراف معیار t_4^R
Z^{DIST}	شاخص نکویی برازش توزیع ($DIST$)
μ_i	عامل مقیاس وابسته به سایت یا سیلاب نمایه

$\widehat{\mu}_l$	برآورده عامل مقیاس در سایت i
$Q_i(F)$	تابع چندک توزیع فراوانی در سایت i
$\widehat{Q}_i(F)$	برآورده $Q_i(F)$
$q(F)$	منحنی رشد (بی بعد) منطقه‌ای
$\widehat{q}(F)$	برآورده $q(F)$
GLO	نمانه‌ی اختصاری توزیع لجستیک تعمیم یافته
GEV	نمانه‌ی اختصاری توزیع مقادیر حدی تعمیم یافته
$PE3$	نمانه‌ی اختصاری توزیع پیرسون نوع ۳
GNO	نمانه‌ی اختصاری توزیع نرمال تعمیم یافته
GPA	نمانه‌ی اختصاری توزیع پرتوی تعمیم یافته
WAK	نمانه‌ی اختصاری توزیع ویکبی

فصل اول:

مقدمه

۱-۱- تعریف مسئله و ضرورت تحقیق

سیلاب‌ها خسارات مالی و جانی بسیاری را در بخش‌های مختلف دنیا به بار می‌آورند. طراحی سازه‌های کنترل سیلاب، مدیریت مخازن، ارزشیابی اقتصادی پروژه‌های حفاظت در برابر سیلاب، برنامه‌ریزی و مدیریت کاربری اراضی و ارزیابی بیمه‌ی سیلاب، همگی به آگاهی از بزرگی و تواتر سیلاب‌ها وابسته هستند. اغلب برآورد این اطلاعات به دلیل کمبود داده‌های ثبت شده‌ی سیلاب در نقطه‌ی مورد نظر، ساده نیست. در بسیاری از موارد مشاهده می‌شود که نقطه‌ی مورد نظر اساساً فاقد داده‌های کافی و مناسب برای اجرای تحلیل فراوانی نقطه‌ای و برآورد سیلاب است.

تحلیل فراوانی منطقه‌ای سیلاب، امکان برآورد توزیع فراوانی سیلاب در یک نقطه را از طریق به کارگیری داده‌های ثبت شده‌ی سیلاب که از زیرحوضه‌های مربوط به ایستگاههای هیدرومتری متعدد دیگر که از نظر فرآیند تولید سیلاب شبیه به حوضه‌ی آبریز نقطه‌ی مورد نظر هستند جمع آوری شده است را فراهم می‌کند. نخستین گام اساسی در اجرای تحلیل فراوانی منطقه‌ای، منطقه‌بندی است. روش تشخیص حوضه‌های مشابه برای جمع آوری اطلاعات جریان بیشینه به عنوان منطقه‌بندی شناخته می‌شود.

لزوم برآورد سیلاب برای نقاط فاقد داده‌های قابل اطمینان هیدرومتری در پروژه‌های مختلف تأمین منابع آب و کنترل سیلاب، استفاده از روش‌های مبتنی بر تحلیل فراوانی منطقه‌ای سیلاب را تا حدود زیادی اجتناب ناپذیر می‌نماید. بدین منظور استفاده از روش‌های منطقه‌بندی که بتوانند برآورد-های مناسبی از مقادیر سیلاب با دوره‌های بازگشت مختلف را فراهم کنند، امری مهم و ضروری است. از این رو بررسی عملکرد روش‌های مختلف منطقه‌بندی برای شناسایی نقاط قوت و ضعف هر یک از این روش‌ها و همچنین انتخاب روش بهینه به منظور استفاده در نواحی مختلف اهمیت ویژه‌ای دارد. روش‌های مبتنی بر تحلیل خوشه‌ای، دسته‌ای از روش‌های مورد استفاده در منطقه‌بندی هستند که بررسی عملکرد آن‌ها در منطقه‌بندی حوضه‌های آبریز می‌تواند میزان کارآمدی مناطق حاصل از

آن‌ها را در تحلیل فراوانی منطقه‌ای سیلاب مشخص نماید.

در طول چهار دهه‌ی اخیر پژوهش‌هایی در زمینه‌ی ارزیابی عملکرد روش‌های مختلف منطقه-بندی و شناسایی نقاط قوت و ضعف هر یک از آن‌ها انجام گرفته است و به طور مداوم یافته‌های جدیدی گزارش شده‌اند. روش‌های متعددی برای منطقه‌بندی توسعه یافته‌اند که شباهت بین سایتها را از طریق آزمودن ویژگی‌های حوضه مانند مشخصات فیزیوگرافیک، موقعیت جغرافیایی، ویژگی‌های هواشناسی، مشخصات زهکشی، معیارهای هواشناسی و آمارهای درون سایتی سیلاب جست و جو می‌کنند. یک گروه از این روش‌ها، روش‌های مبتنی بر تحلیل خوش‌های هستند.

در این پژوهش، عملکرد روش‌های تحلیل خوش‌های در منطقه‌بندی حوضه‌ی آبریز سفیدرود بزرگ و اثرات هر یک از این روش‌ها بر اجرای تحلیل فراوانی منطقه‌ای سیلاب در این حوضه مورد بررسی قرار گرفته است.

۱-۲- فرضیات تحقیق

روش‌های منطقه‌بندی مبتنی بر تحلیل خوش‌های می‌توانند در تشخیص گروه‌های حوضه‌هایی که فرآیند تولید سیلاب مشابهی دارند، نقش مؤثری را ایفا کنند. تحلیل خوش‌های نام گونه‌ای از روش‌های آماری چند متغیری است که به منظور تحقیق، تفسیر، و دسته بندی داده‌های موجود در گروه‌های مشابه یا خوش‌هایی است که می‌توانند همپوشانی داشته باشند یا نداشته باشند. نقاط داده‌ها در یک خوش باید تا حد امکان مشابه باشند و نقاط داده‌ها در خوش‌های مختلف باید حتی الامکان متفاوت باشند (رأو^۱ و سرینیواس^۲، ۲۰۰۸). پژوهش‌ها نشان می‌دهد قضاوت‌های ذهنی و تلاش مورد نیاز برای تعیین گروه‌های حوضه‌های آبریز با استفاده از روش‌های معمول، با به کارگیری شیوه‌های خوش‌بندی کارآمد به شکل چشم گیری کاهش می‌یابد. بنابراین انتظار می‌رود استفاده از این روش‌ها در منطقه-

¹ Rao A. R.

² Srinivas V. V.

بندی و تحلیل فراوانی منطقه‌ای سیلاب در حوضه‌ی آبریز سفیدرود بزرگ نیز بتواند نتایج مناسبی را فراهم کند.

بنا بر نتایج به دست آمده در مطالعات پیشین، کیفیت خوشبندی اجرا شده توسط روش‌های مختلف خوشبندی می‌تواند بر اساس معیارهایی چون اندازه‌ی مناطق تشکیل شده، مقادیر توابع هدف و شاخص‌های صحت خوشبندی مورد قضاوت قرار گیرد. همچنین به منظور شناخت داده‌ها یا ایستگاه‌های هیدرومتری که از نظر آماری با سایر ایستگاه‌ها به عنوان یک مجموعه‌ی کلی ناسازگار هستند، می‌توان از شاخص ناجوری^۱ D ، برای بررسی همگنی مناطق تشکیل شده از شاخص‌های ناهمگنی^۲ H و جهت سنجش نکویی برازش و انتخاب توزیع بهینه‌ی منطقه‌ای از شاخص نکویی برازش^۳ Z استفاده کرد.

۱-۳-۱- اهداف تحقیق

هدف از این مطالعه ارزیابی عملکرد روش‌های تحلیل خوشبندی در منطقه‌بندی حوضه‌های آبریز و به طور خاص، حوضه‌ی آبریز سفیدرود بزرگ است که به عنوان مطالعه‌ی موردی در نظر گرفته شده است. این ارزیابی می‌تواند منجر به شناخت هر چه بیشتر مزایا و معایب هر یک از این روش‌ها شده و در انتخاب روش بهینه برای اجرای منطقه‌بندی به منظور تحلیل فراوانی منطقه‌ای سیلاب در مطالعات آینده و پژوهش‌های عمرانی و زیست و محیطی مفید واقع شود. همچنین معرفی روش‌های جدیدتری چون منطقه‌بندی فازی و استفاده از شبکه‌های عصبی مصنوعی در منطقه‌بندی می‌تواند گشاینده‌ی راه‌هایی نوین در زمینه‌ی مطالعات منطقه‌بندی باشد.

از جمله دیگر اهداف این پژوهش، می‌توان به بررسی کاربرد شاخص‌های مختلف در سنجش کیفیت خوشبندی و ارتباط آن‌ها با مسائلی چون همگنی مناطق تشکیل شده اشاره کرد. افزون بر

¹ discordancy measure

² heterogeneity measure

³ goodness of fit measure

این، بررسی چگونگی تأثیر ویژگی‌های منتخب مورد استفاده در خوشبندی یا منطقه‌بندی نیز از اهداف جنبی این مطالعه است.

۱-۴- سازماندهی متن پایان‌نامه

پس از معرفی مسئله و ضرورت بررسی آن، فرضیات و اهداف تحقیق که در این فصل ارائه شد، در فصل دوم کلیات و مروری بر مطالعات پیشین ارائه شده است. در فصل سوم به مواد و روش‌های مورد استفاده در این مطالعه پرداخته شده است. فصل چهارم این مطالعه به بررسی محاسبات و ارائه نتایج و بحث در مورد آن‌ها اختصاص یافته است. در نهایت فصل پنجم دربرگیرنده نتیجه‌گیری نهایی و ارائه پیشنهاداتی به منظور استفاده در مطالعات آتی در این زمینه است.

شایان ذکر است که در نوشتار این پایان‌نامه، هر کجا واژه‌های ایستگاه یا سایت به تنها‌یی به کار رفته‌اند، مقصود ایستگاه هیدرومتری (با در نظر گرفتن مساحت سطح زهکشی آن) است. همچنین در مواردی که واژه‌ی حوضه به تنها‌یی ظاهر شده است، منظور همان حوضه‌ی آبریز است.

فصل دوم:

کلیات و مروری بر مطالعات پیشین

۲- منطقه‌بندی برای تحلیل فراوانی سیلاب

تعیین بزرگی محتمل و فراوانی این رویدادهای حدی هیدرولوژیک در طراحی سازه‌های مختلف کنترل جریان مانند گوره‌ها، کالورت‌ها، پل‌ها، آب بندها و سدها ضروری است. مدت زیادی است که در هیدرولوژی، روش‌های تحلیل فراوانی سیلاب به منظور مرتبط کردن بزرگی سیلاب‌ها با فراوانی رخداد آن‌ها مورد استفاده قرار می‌گیرند. در روش‌های سنتی تحلیل فراوانی اطلاعات هیدرولوژیک درون سایتی^۱ به صورت سری‌های شدت جریان بیشینه‌ی سالانه یا سری‌های اوج‌های فراتر از آستانه برای برآورد پتانسیل وقوع سیلاب در سایت مورد نظر نیاز است. برای مقابله با این وضعیت، هیدرولوژیست‌ها از روش‌های تحلیل فراوانی منطقه‌ای که مبتنی بر جمع آوری اطلاعات از حوضه‌های آبریز متعددی که فرآیند تولید سیلاب مشابه با سایت هدف دارند، استفاده می‌کنند. گروهی از سایت‌ها با همگنی کافی در فرآیندهای تولید سیلاب، یک منطقه‌ی همگن یا گروه ادغام شده برای تحلیل فراوانی منطقه‌ای سیلاب را تشکیل می‌دهند و روش تعیین مناطق همگن منطقه‌بندی نامیده می‌شود.

یک منطقه برای تحلیل فراوانی سیلاب با دوره‌ی بازگشت T سال باystsی حداقل به اندازه‌ی $5T$ آمار مقادیر اوج فراهم کند (رید^۲ و همکاران، ۱۹۹۹). انتخاب بین روش‌های تحلیل فراوانی درون سایتی و تحلیل فراوانی منطقه‌ای هم به طول آمار ثبت شده در سایت هدف و هم به دوره‌ی بازگشت هدف بستگی دارد. طراحی سازه‌های کنترل سیلاب نظیر کالورت‌ها، پل‌ها، مسیل‌های فاضلاب‌روی شهری، فرودگاه‌ها و سدهای کوچک ممکن است نیازمند برآوردهایی از چندک‌های سیلاب متناظر با فواصل تکرار ۵۰ تا ۱۰۰ سال باشند. طراحی گوره‌های اطراف شهرها و سدهای متوسط تا بزرگ ممکن است به برآوردهای چندک متناظر با فواصل تکرار ۱۰۰ تا ۲۰۰ سال احتیاج داشته باشد (چاو^۳ و همکاران، ۱۹۸۸، ص ۴۰۹). کانال‌های سیلاب‌رو برای رویدادهای سیلابی متناظر با ۵۰۰ سال یا حتی دوره‌ی بازگشت بزرگ‌تر طراحی می‌شوند. در گذشته روش‌های گوناگون منطقه‌بندی به منظور

¹ at-site

² Reed D. W.

³ Chow V. T.

شكل دادن مناطقی جهت تجمیع اطلاعات مورد نیاز برای برآورده قابل اعتماد چندکهای سیلاب متناظر با دوره‌ی بازگشت T سال توسعه یافته‌اند.

۲-۲- روش‌های منطقه‌بندی

مناطق مورد استفاده در تحلیل فراوانی منطقه‌ای سیلاب اغلب به گونه‌ای انتخاب می‌شوند که گروه‌هایی از حوضه‌های مجاور از نظر جغرافیایی بر اساس مرزهای سیاسی، اداری، یا فیزیوگرافیک به وجود آیند. با این حال، این روش مورد انتقاد است، چرا که طراحی مناطق با استفاده از این عوامل همگنی هیدرولوژیک را تضمین نمی‌کند. در نتیجه روش‌های متعددی برای منطقه‌بندی توسعه یافته‌اند که شباهت بین سایتها را از طریق آزمودن ویژگی‌های حوضه مانند مشخصات فیزیوگرافیک، موقعیت جغرافیایی و آماره‌های درون سایتی سیلاب جست و جو می‌کنند.

برخی از کارآمدترین روش‌های منطقه‌بندی عبارتند از: روش باقی‌مانده‌ها^۱، تحلیل همبستگی دوسویه^۲، منطقه‌ی تأثیر^۳، روش سلسله مراتبی^۴ و بسط آن در چارچوب منطقه‌ی تأثیر، و تحلیل خوشه‌ای^۵. هر روش منطقه‌بندی برتری‌ها و محدودیت‌های خاص خود را دارد. با این حال، به دلیل محدودیت‌های ناشی از کمبود داده‌ها و تأثیر قضاوت‌های ذهنی در انتخاب مشخصات، وزن‌ها، مقادیر آستانه و شاخص‌های فاصله، هیچ معیار قطعی که توسط آن برتری هر یک از روش‌های منطقه‌بندی بتواند به روشنی بیان شود، وجود ندارد.

به نظر می‌رسد، اخیراً افزایش توجه و آگاهی نسبت به استفاده از داده‌های هیدرولیماتیک نهادهای متعددی را به حرکت در جهت ایجاد بانک‌های داده‌های گونه‌ای از متغیرهایی که فرآیندهای هیدرولوژیک را تحت تأثیر قرار می‌دهند، سوق داده است. برای استفاده‌ی مؤثر از آرشیوهای داده‌ها در

¹ method of residuals (MOR)

² canonical correlation analysis (CCA)

³ region of influence (ROI)

⁴ hierarchical approach

⁵ cluster analysis

مطالعات منطقه‌بندی، نیاز به توسعه‌ی روش‌های بالقوه که برای تشخیص و تفسیر الگوهای ذاتی در داده‌های هیدرولوژیک مؤثر هستند، وجود دارد. به این منظور، الگوریتم‌های خوش‌بندی که در الگوهای تشخیصی در هر دو نوع مجموعه داده‌های بزرگ و کوچک امیدوار کننده ظاهر می‌شوند، مؤثر هستند (رائو و سرینیواس، ۲۰۰۸).

۳-۳- تحلیل خوش‌های در منطقه‌بندی

تحلیل خوش‌های نام گونه‌ای از روش‌های آماری چند متغیری است که به منظور تحقیق، تفسیر، و دسته‌بندی داده‌های موجود در گروه‌های مشابه یا خوش‌هایی است که می‌توانند هم‌پوشانی داشته باشند یا نداشته باشند. نقاط داده‌ها در یک خوش‌های مشابه باشند و نقاط داده‌ها در خوش‌های مختلف باید حتی امکان متفاوت باشند.

۴-۴- ویژگی‌های مورد استفاده در منطقه‌بندی

یک خوش‌های یک یا تعداد بیش‌تری بردار مشخصه تشکیل می‌شود. یک بردار مشخصه از ویژگی‌ها یا متغیرهای متعدد تشکیل می‌شود. ویژگی‌هایی که برای منطقه‌بندی حوضه‌های آبریز استفاده می‌شوند، عبارتند از: ویژگی‌های فیزیوگرافیک مانند سطح زهکشی، طول طویل‌ترین آبراه جریان، شیب آبراه جریان اصلی، شیب متوسط حوضه، شاخص ذخیره، نسبتی از سطح حوضه که با دریاچه‌ها، مخازن و مرداب‌ها پوشیده شده‌اند؛ ویژگی‌های پوشش خاک مانند پتانسیل نفوذ، کسر رطوبت مؤثر متوسط خاک و ضریب رواناب؛ ویژگی‌های مرتبط با الگوی کاربری زمین مانند نسبتی از سطح حوضه که با جنگل‌ها، زمین‌های کشاورزی، اراضی حومه‌ای و شهری پوشیده شده است؛ ویژگی‌های زهکشی حوضه مانند تراکم زهکشی؛ ویژگی‌های موقعیت جغرافیایی مانند طول و عرض و ارتفاع جغرافیایی ایستگاه‌های هیدرومتری و مرکز حوضه‌ای که سایت در آن قرار دارد؛ ویژگی‌های هواشناسی مانند

جهت رگبار در حوضه‌ها، میانگین سالانه‌ی تعداد روزهایی که دمای آن‌ها از یک مقدار مشخص کم‌تر یا بیش‌تر است (اوادا^۱ و همکاران، ۲۰۰۶)؛ ویژگی‌های زمین‌شناسی حوضه مانند نسبتی از سطح حوضه که با گونه‌های مختلف سنگ‌ها پوشیده شده است (ناتان^۲ و مکماهون^۳، ۱۹۹۰)؛ یک شاخص زمان واکنش حوضه، مانند زمان تأخیر یا زمان رسیدن به دبی اوج حوضه (پاتر^۴ و فاکنر^۵، ۱۹۸۷)؛ توصیف کننده‌های فصلی سیلاب^۶ مانند آماره‌های سویی (ماردیا^۷، ۱۹۷۲؛ فیشر^۸، ۱۹۹۳) و فراوانی نسبی رویداد سیلاب (بلک^۹ و وریتی^{۱۰}، ۱۹۹۷؛ لس^{۱۱}، ۲۰۰۰؛ کاندرلیک^{۱۲} و برن^{۱۳}، ۲۰۰۲b) کاندرلیک و همکاران، ۲۰۰۴a,b). آماره‌های سویی شامل شاخص‌هایی که زمان متوسط رویداد سیلاب‌ها و تغییرپذیری آن در حوضه و تأخیر میانگین بین بارندگی و سیلاب‌ها را توصیف می‌کنند، هستند (برن، ۱۹۹۷؛ کاندرلیک و برن، ۲۰۰۲a؛ کاندرلیک و همکاران، ۲۰۰۴a)؛ شاخص‌های شکل حوضه، مانند عامل شکل، ضریب فشردگی، نسبت (ضریب) کشیدگی، نسبت دایره‌ای نیز می‌توانند به عنوان ویژگی‌هایی برای شکل دادن مناطق جهت تحلیل فراوانی سیلاب مورد استفاده قرار گیرند.

همچنین در گذشته، آماره‌های سیلاب درون سایتی به عنوان ویژگی‌هایی برای منطقه‌بندی به کار می‌رفتند. برای نمونه می‌توان به میانگین، ضریب تغییرات و چولگی سری‌های سیلاب سالانه، برآورد احتمال تجربی رویداد سیلاب T ساله‌ی درون‌یابی شده از سری‌های سیلاب سالانه (برن، ۱۹۹۰)، و بزرگی سیلاب متناظر با یک دوره‌ی بازگشت T سال (تاکسکر^{۱۴}، ۱۹۸۰) اشاره کرد.

همگنی مناطق شکل گرفته به وسیله‌ی یک روش منطقه‌بندی، با استفاده از آماره‌های سیلاب

¹ Ouarda T. B. M. J.

² Nathan R. J.

³ McMahon T. A.

⁴ Potter K. W.

⁵ Faulkner E. B.

⁶ flood seasonality descriptors

⁷ Mardia K. V.

⁸ Fisher N. I.

⁹ Black A. R.

¹⁰ Werritty A.

¹¹ Lecce S. A.

¹² Cunderlik J. M.

¹³ Burn D. H.

¹⁴ Tasker G. D.

آزموده می‌شود. بنابراین استفاده از این آماره‌ها به عنوان ویژگی‌هایی برای شکل‌دهی مناطق توصیه نمی‌شود. یک پیامد استفاده از آماره‌های سیلاپ به عنوان ویژگی‌هایی برای تشکیل مناطق این است که مناطق حاصل ممکن است به صورت همگن ظاهر شوند، اما لزوماً برای تحلیل فراوانی منطقه‌ای سیلاپ مؤثر نباشند (برن و همکاران، ۱۹۹۷). افزون بر این استفاده از آماره‌های سیلاپ در تشکیل مناطق برای تحلیل فراوانی سیلاپ مانع استفاده از مناطق تشکیل شده به منظور برآورد چندک‌های سیلاپ در سایتها فاقد آمار هیدرومتری در ناحیه‌ی مورد مطالعه می‌شود. به طور مشابه شکل‌دهی مناطق نباید صرفاً بر اساس ویژگی‌های فیزیوگرافیک حوضه‌ها باشد، زیرا شباهت صرف در ویژگی‌های فیزیوگرافیک لزوماً بر شباهت واکنش هیدرولوژیک دلالت نمی‌کند. بنابراین در تشکیل مناطق، منطقی است که از چندین نوع ویژگی‌های اندازه‌گیری شده و برآورد شده در سایتها استفاده شود که این شاخص‌ها همبستگی بالایی با مقادیر سیلاپ متناظر خود ندارند (هاسکینگ^۱ و والیس^۲، ۱۹۹۷).

(۱۹۹۷)

۵-۲- طبقه‌بندی الگوریتم‌های خوشبندی

بیشتر الگوریتم‌های خوشبندی موجود می‌توانند در دو گروه طبقه‌بندی شوند: خوشبندی سلسه-مراتبی^۳ و خوشبندی افزایی^۴. روش‌های خوشبندی سلسه‌مراتبی، یک دنباله‌ی تو در تو از تقسیم-تقسیم‌بندی‌ها را ایجاد می‌کنند، در حالی که روش‌های افزایی یک تقسیم‌بندی منفرد از داده‌ها را برای بازیابی گروه‌بندی طبیعی در داده‌ها ارائه می‌کنند.

الگوریتم‌های خوشبندی سلسه‌مراتبی خود می‌توانند به دو زیر‌گروه تقسیم شوند: متراکم‌کننده^۵ کننده^۶ و تقسیم‌کننده^۷. برای یک مجموعه از N بردار مشخصه، روش‌های خوشبندی سلسه‌مراتبی

¹ Hosking J. R. M.

² Wallis J. R.

³ hierarchical clustering

⁴ partitional clustering

⁵ agglomerative

متراکم کننده با N خوشه‌ی تکی آغاز می‌شوند. خوشه‌های تکی، خوشه‌هایی هستند که تنها از یک بردار مشخصه تشکیل می‌شوند. یک شاخص فاصله مانند فاصله‌ی اقلیدسی برای ارزیابی تفاوت بین هر دو خوشه انتخاب می‌شود. خوشه‌هایی که کمترین تفاوت را با هم دارند، یافت شده و با هم ادغام می‌شوند. در این گام $N-2$ خوشه‌ی تکی و یک خوشه با دو بردار مشخصه ایجاد می‌شود. فرآیند تشخیص و ادغام دو خوشه‌ی نزدیک‌تر تا زمانی که تعداد مطلوبی از خوشه‌ها حاصل شود، تکرار می‌شود. از سوی دیگر، روش‌های سلسله‌مراتبی تقسیم کننده با یک خوشه‌ی واحد که از N بردار مشخصه تشکیل می‌شود، آغاز می‌شوند. بردار مشخصه‌ای که بیشترین اختلاف را با سایر بردارهای مشخصه‌ی خوشه دارد، مشخص شده و از خوشه جدا می‌شود تا یک خوشه‌ی مجزا را تشکیل دهد. سپس مقادیر اختلاف بردارهای مشخصه‌ی باقی مانده در خوشه‌ی اصلی آزموده می‌شوند تا چنانچه لزوم و امکان افزوده شدن هر یک از بردارهای مشخصه به خوشه‌ی مجزا وجود دارد، تشخیص داده شود. این مرحله خوشه‌ی اصلی را به دو بخش تقسیم می‌کند. فرآیند فوق در مرحله‌ی بعد بر روی خوشه‌ی بزرگ‌تر به اجرا در می‌آید. الگوریتم هنگامی که تعداد مطلوبی از خوشه‌ها حاصل شود، متوقف می‌گردد.

الگوریتم‌های خوشه‌بندی سلسله‌مراتبی متراکم کننده همواره خوشه‌ها را هم الحق کرده و خوشه‌های بزرگ‌تری ایجاد می‌کنند. در مقابل، الگوریتم‌های خوشه‌بندی سلسله‌مراتبی تقسیم کننده، همیشه خوشه‌ها را به خوشه‌های کوچک‌تر تقسیم می‌کنند. پیامد الگوریتم‌های خوشه‌بندی سلسله‌مراتبی این است که خوشه‌های حاصل معمولاً بهینه نیستند، زیرا بردارهای مشخصه‌ای که در مراحل ابتدایی به یک خوشه تعلق می‌گیرند، نمی‌توانند به خوشه‌ی دیگری منتقل شوند.

الگوریتم‌های خوشه‌بندی افزایی به یک حدس اولیه در مورد تعداد خوشه‌ها و مراکز خوشه‌ها نیازمند هستند. این روش‌ها می‌توانند بر اساس شیوه‌ی به کار رفته برای تعیین شرایط اولیه‌ی خوشه‌ها، معیار خوشه‌بندی، و نوع داده‌هایی که برای آن قابل کاربرد است، طبقه‌بندی شوند. در رایج‌ترین دسته‌ی این روش‌ها، مرکز خوشه به عنوان نماینده‌ی خوشه در نظر گرفته می‌شود.

¹ divisive

در حالی که روش‌های خوشبندی سلسله‌مراتبی تحت تأثیر تعیین شرایط اولیه قرار نمی‌گیرند، روش‌های خوشبندی افزایی از حدس‌های اولیه متأثر می‌شوند. روش‌های خوشبندی افزایی در حدی که بردارهای مشخصه بتوانند از یک خوش به خوش به دیگر برای کمینه‌سازی تابع هدف حرکت کنند، پویا هستند. در مقابل در روش‌های سلسله‌مراتبی، بردارهای مشخصه‌ی تعلق گرفته به یک خوش در مراحل ابتدایی، نمی‌توانند به خوش به دیگری انتقال یابند.

به منظور استفاده از نقاط قوت هر دو روش خوشبندی سخت سلسله‌مراتبی و افزایی و کاهش تأثیر نقاط ضعف این دو روش، با ترکیب این دو نوع روش، الگوریتم‌های خوشبندی ترکیبی^۱ حاصل می‌شوند. در الگوریتم خوشبندی ترکیبی، مراکز خوش‌های حاصل از یک الگوریتم خوشبندی سلسله‌مراتبی برای تعیین شرایط اولیه‌ی الگوریتم خوشبندی افزایی استفاده می‌شوند (رائو و سرینیواس، ۲۰۰۶a). رائو و سرینیواس (۲۰۰۶a)، عملکرد انواع الگوریتم‌های خوشبندی ترکیبی را در منطقه‌بندی ۲۴۵ سایت در ایالت ایندیانا مورد بررسی قرار دارند و ترکیب الگوریتم‌های Ward و K-means را بر اساس شاخص‌های صحت خوشبندی^۲ و توانایی کمینه‌سازی تابع هدف^۳، بهترین گزینه برای منطقه‌بندی ناحیه‌ی مورد مطالعه معرفی کردند.

الگوریتم‌های خوشبندی همچنین می‌توانند به صورت خوشبندی سخت^۴ و خوشبندی فازی^۵ طبقه‌بندی شوند. در خوشبندی سخت، هر بردار مشخصه به یکی از خوش‌ها با یک درجه‌ی عضویت برابر یک اختصاص داده می‌شود. این نوع خوشبندی بر اساس این فرض است که بردارهای مشخصه می‌توانند به خوش‌های بدون هم پوشانی با مرزهایی که به خوبی بین آن‌ها تعریف می‌شود، تقسیم شوند. روش‌های خوشبندی سخت در یک دسته‌بندی عمده به دو گروه روش‌های سلسله‌مراتبی و افزایی تقسیم می‌شوند که پیش‌تر به اختصار در مورد آن‌ها توضیح داده شد. این حالت برای گروه‌های فشرده و کاملاً مجزای داده‌ها طبیعی است. با این وجود، در بسیاری از موقعیت‌های واقعی، بردارهای

¹ hybrid clustering

² clustering validity measures

³ objective function

⁴ hard clustering

⁵ fuzzy clustering

مشخصه مشابهت‌های نسبی با خوشه‌های متعدد دارند و از این رو نمی‌توان یک بردار مشخصه را کاملاً به یک خوشه یا خوشه‌ی دیگر اختصاص داد. نظریه‌ی مجموعه‌ی فازی (زاده^۱، ۱۹۶۵) راهی طبیعی برای نشان دادن چنین موقعیتی است. خوشه‌بندی فازی به یک بردار مشخصه اجازه می‌دهد که همزمان با یک درجه‌ی عضویت مشخص در فاصله‌ی [۰,۱] به تمام خوشه‌ها تعلق داشته باشد. روش-های خوشه‌بندی فازی می‌توانند بر اساس راهبردی که برای تقسیم داده‌ها به کار می‌گیرند به دو گونه تقسیم می‌شوند: اول، گروهی که یک رابطه‌ی فازی را برای اجرای خوشه‌بندی فازی استفاده می‌کند؛ و دیگر گروهی که تابع هدف را برای خوشه‌بندی فازی به کار می‌گیرند. رائو و سرینیواس (۲۰۰۶b) الگوریتم خوشه‌بندی فازی c-means را به منظور منطقه‌بندی سایت‌های هیدرومتری ایالت ایندیانا مورد استفاده قرار دادند. همچنین صدری و برن (۲۰۱۱) از الگوریتم c-means برای منطقه‌بندی ۳۶ ایستگاه هیدرومتری در سه ایالت همچوار آلبرتا، ساسکاچوان و مانیتوبا به منظور تحلیل فراوانی منطقه‌ای خشکسالی استفاده کردند. آن‌ها روشی مبتنی بر به کار گیری گشتاورهای خطی چند متغیره را برای بررسی ناجوری سایت‌های مورد مطالعه و همگنی مناطق تشکیل شده پیشنهاد کردند.

در دهه‌ی اخیر، یک دسته‌ی خاص از شبکه‌های عصبی مصنوعی که نگاشت‌های مشخصه خود-سازمانده^۲ (SOFM) نام دارند، به عنوان تکنیک و ابزاری برای خوشه‌بندی به کار گرفته شده‌اند. با این حال، SOFM یک روش خوشه‌بندی نیست، زیرا خوشه‌های حاصل از یک SOFM به ندرت قابل تفسیر هستند (رائو و سرینیواس، ۲۰۰۸).

۶-۲- مراحل منطقه‌بندی به وسیله‌ی تحلیل خوشه‌ای

مراحل مختلف اجرای منطقه‌بندی حوضه‌های آبریز از طریق خوشه‌بندی عبارتند از:

- انتخاب ویژگی‌ها: هدف این مرحله، تحلیل داده‌های متغیرهای مختلف برای تعیین ویژگی-

¹ Zadeh L. A.

² self-organizing feature maps (SOFM)

های مؤثر بر واکنش سیلابی حوضه در منطقه‌ی مورد مطالعه است.

۲- آماده‌سازی بردارهای مشخصه^۱: داده‌های موجود برای هر ویژگی، برای خنثی کردن تفاوت‌ها در واریانس‌ها و بزرگی نسبی آن‌ها، تجدید مقیاس می‌شوند. تجدید مقیاس ممکن است تبدیل مقادیر ویژگی‌ها با استفاده از تابع تبدیل مناسب (مانند تبدیل لگاریتمی) و تقسیم مقادیر تبدیل یافته بر انحراف معیار، را به کار بگیرد. هر بردار مشخصه از ویژگی‌های تجدید مقیاس شده‌ی (بدون بعد) یک حوضه‌ی آبریز تشکیل می‌شود.

۳- شکل‌دهی خوش‌های: این مرحله شامل انتخاب یک الگوریتم خوش‌بندی برای بردارهای مشخصه‌ی آماده شده در مرحله‌ی ۲ در خوش‌های مجزا یا مداخل (دارای همپوشانی) است. حوضه‌های آبریزی که در یک خوش به وسیله‌ی بردارهای مشخصه معرفی می‌شوند، یک منطقه را برای تحلیل فراوانی سیلاب تشکیل می‌دهند. به طور کلی، شاخص فاصله و یک معیار خوش‌بندی، یک الگوریتم خوش‌بندی را مشخص و معرفی می‌کند.

۴- انتخاب تعداد مناطق: خوش‌های تشکیل شده در مرحله‌ی ۳ به صورت بصری و با استفاده از شاخص‌های صحتسنجی خوش برای تعیین تعداد بهینه‌ی مناطق تفسیر می‌شوند.

۵- آزمودن همگنی مناطق: مناطق تعیین شده در گام ۴ با استفاده از آزمون‌های همگنی آماری از نظر همگنی آزموده می‌شوند.

۶- اصلاح مناطق ناهمگن: مناطقی که بیشتر به حالت همگنی نزدیک هستند، به منظور بهبود وضعیت همگنیشان اصلاح می‌شوند.

۷- برآوردهای چندک‌های سیلاب: هدف این مرحله اجرای آزمون‌های نکویی برآذش منطقه‌ای برای تشخیص و برآذش یک توزیع فراوانی سیلاب مناسب بر داده‌های سیلاب سایتها در یک منطقه است. توزیع برآذش یافته، در نهایت برای به دست آوردن برآوردهای چندک سیلاب برای طراحی هیدرولوژیک استفاده می‌شوند (رائو و سرینیواس، ۲۰۰۸).

¹ feature vector

فصل سوم:

مواد و روش‌ها

۳-۱-۳- خوشبندی ترکیبی

۳-۱-۱- معرفی تحلیل خوشبندی ترکیبی

در الگوریتم‌های خوشبندی ترکیبی، خوشبندی‌ها از طریق ترکیب دو نوع الگوریتم خوشبندی سلسله مراتبی و خوشبندی افزایی، تشکیل می‌شوند.

الگوریتم‌های خوشبندی سخت، حوضه‌های آبریز را در ناحیه‌ی مورد نظر به خوشبندی‌های بدون همپوشانی تقسیم می‌کنند، به طوری که هر سایت در یکی از خوشبندی‌ها قرار می‌گیرد. چنانچه خوشبندی‌ها کاملاً مجزا باشند، نتایج مناسب خواهد بود. الگوریتم‌های خوشبندی سخت به صورت گستردگی در هیدرولوژی برای تحلیل منطقه‌ای استفاده می‌شوند.

۲-۱-۳- طبقه‌بندی الگوریتم‌های خوشبندی سخت

الگوریتم‌های خوشبندی سخت به صورت الگوریتم‌های خوشبندی سلسله مراتبی و افزایی طبقه-بندی می‌شوند. الگوریتم‌های خوشبندی سلسله مراتبی یک دنباله‌ی متوالی از تقسیمات را فراهم می‌کنند، در حالی که الگوریتم‌های خوشبندی افزایی یک افزای واحد از داده‌ها را ایجاد می‌کنند. الگوریتم‌های خوشبندی سلسله مراتبی به طور عمده می‌توانند به دو دسته طبقه‌بندی شوند: متراکم‌کننده و تقسیم‌کننده.

خوشبندی سلسله مراتبی متراکم‌کننده با خوشبندی‌ها تکی (خوشبندی‌های شامل یک بردار مشخصه) آغاز می‌شود و به تدریج با ادغام خوشبندی‌های کوچک‌تر و ایجاد خوشبندی‌های بزرگ‌تر پیش می‌رود. در سوی دیگر، خوشبندی سلسله مراتبی تقسیم‌کننده با یک خوشبندی بزرگ متشكل از تمام بردارهای مشخصه آغاز می‌شود و با تقسیم آن‌ها به خوشبندی‌های کوچک‌تر ادامه می‌یابد.

الگوریتم‌های خوشبندی افزایی به یک حدس اولیه در مورد تعداد خوشبندی‌ها و مراکز خوشبندی نیازمند هستند. آن‌ها می‌توانند بر اساس روش به کار رفته برای تعیین شرایط اولیه خوشبندی‌ها، معیار

خوشبندی، و نوع داده‌هایی که برای آن قابل استفاده هستند، طبقه‌بندی می‌شوند. الگوریتم K-means و الگوریتم‌های خوشبندی سلسله مراتبی متراکم کننده برای منطقه‌بندی در هیدرولوژی بیش از سایر روش‌ها به کار گرفته شده‌اند.

۱-۲-۱-۳- روش‌های خوشبندی سلسله مراتبی

۱-۲-۱-۳-الف- خوشبندی سلسله مراتبی متراکم کننده

برای مجموعه‌ای از N بردار مشخصه، روش‌های خوشبندی سلسله مراتبی متراکم کننده با N خوشه-ی تکی آغاز می‌شوند. یک معیار فاصله مانند مواردی که در جدول (۳-۱) نشان داده شده است برای ارزیابی اختلاف بین هر دو مرکز خوشه و یا بردار مشخصه انتخاب می‌شود. خوشه‌هایی که کمترین اختلاف را با یکدیگر دارند مشخص شده و به هم می‌پیوندند. این امر منجر به تشکیل $N-2$ خوشه‌ی تکی و یک خوشه شامل ۲ بردار مشخصه می‌شود. فرآیند تشخیص و ادغام دو خوشه‌ی نزدیک به هم آن قدر تکرار می‌شود تا تمام خوشه‌ها به هم بپیوندند و یک خوشه‌ی واحد را تشکیل دهند. به طور کلی تعداد خوشه‌های موجود پس از n مرتبه ادغام برابر $N-n$ است. کل فرآیند می‌تواند به صورت یک زنجیره تو در تو که dendrogram خوانده می‌شود و چگونگی ارتباط خوشه‌های تشکیل شده در مراحل مختلف فرآیند را نشان می‌دهد، ارائه شود (رائو و سرینیواس، ۲۰۰۸).

الگوریتم‌هایی که نماینده‌ی روش خوشبندی سلسله مراتبی متراکم کننده هستند، عبارتند از: Ward تک پیوند یا نزدیک‌ترین همسایه^۱، تمام پیوند یا دورترین همسایه^۲، پیوند متوسط، و الگوریتم کنند. که تفاوت آن‌ها در راهبردی است که برای تعریف نزدیک‌ترین همسایه به یک خوشه استفاده می‌کنند. در الگوریتم تک پیوند، فاصله‌ی بین دو خوشه برابر فاصله‌ی بین نزدیک‌ترین جفت از بردارهای مشخصه است که هر یک از آن‌ها به یکی از دو خوشه‌ی مورد نظر تعلق دارد. این الگوریتم به تشکیل

¹ nearest neighbor

² furthest neighbor

تعداد اندکی خوشه‌ی بزرگ تمایل دارد که به همراه آن‌ها خوشه‌های پرت کوچکی در حاشیه‌های فضای ویژگی‌ها باقی می‌مانند و بعید است که منجر به تشکیل مناطق مناسبی برای تحلیل فراوانی

منطقه‌ای سیلان شود (هاسکینگ و والیس، ۱۹۹۷؛ رائو و سرینیواس، ۲۰۰۶).

در الگوریتم تمام پیوند، فاصله‌ی بین دو خوشه برابر فاصله‌ی بین دورترین جفت از بردارهای مشخصه است که هر یک از آن‌ها به یکی از دو خوشه مورد نظر تعلق دارد. این الگوریتم تمایل به تشکیل خوشه‌های کوچک نزدیک به هم دارد. این روش معمولاً برای اعمال بر مجموعه‌های بزرگ داده‌ها مناسب نیست (رائو و سرینیواس، ۲۰۰۸).

جدول (۳-۱) شاخص‌های اختلاف برای محاسبه‌ی فاصله‌ی بین مراکز خوشه‌ها یا بردارهای مشخصه

معادله	شاخص فاصله
$\sqrt{\sum_{k=1}^n (x_{ik} - x_{jk})^2}$	اقلیدسی
$\sum_{k=1}^n (x_{ik} - x_{jk})^2$	اقلیدسی مربع
$\sum_{k=1}^n x_{ik} - x_{jk} $	منهتن
$\sum_{k=1}^n \frac{ x_{ik} - x_{jk} }{ x_{ik} + x_{jk} }$	کانبرا
$\max_{1 \leq k \leq n} x_{ik} - x_{jk} $	چبیشف
$1 - \frac{\sum_{k=1}^n x_{ik} x_{jk}}{\sqrt{\sum_{k=1}^n x_{ik}^2 \sum_{k=1}^n x_{jk}^2}}$	کسینوسی
$\left(\sum_{k=1}^n x_{ik} - x_{jk} ^t \right)^{1/t}$	مینکوفسکی

- تعداد ویژگی‌های x_{ik} : x_{ik} بردار مشخصه‌ی x_i در خوشه‌ی اول؛ x_{jk} : x_{jk} بردار مشخصه‌ی x_j در خوشه‌ی دوم؛ t معرف مرتبه‌ی فاصله‌ی مینکوفسکی است.

در الگوریتم پیوند متوسط، فاصله‌ی بین دو خوش به صورت فاصله‌ی متوسط بین آن‌ها تعریف می‌شود. روش‌های مختلفی برای محاسبه‌ی فاصله‌ی متوسط وجود دارد.

الگوریتم Ward (وارد^۱، ۱۹۶۳) روشی پرکاربرد برای مطالعات منطقه‌بندی در هیدرولوژی و کلیماتولوژی است (برای نمونه: ایکرمن^۲ و سینکلر^۳، ۱۹۸۶؛ ناتان و مکماهون ۱۹۹۰؛ هاسکینگ و والیس، ۱۹۹۷). این الگوریتم مبتنی بر این فرض است که اگر دو خوش به هم بپیونندند، تغییر در مقدار تابع هدف، تنها به رابطه‌ی بین این دو خوش وابسته است و به روابط با دیگر خوش‌ها بستگی ندارد. معادله‌ی حاکم و توضیح دقیق الگوریتم Ward در بخش‌های بعدی ارائه شده است.

در زمینه تحلیل فراوانی منطقه‌ای سیلاب، موزلی^۴ (۱۹۸۱) خوش‌بندی سلسله مراتبی متراکم کننده موجود در برنامه کامپیوتري BMDP2M (دیکسون^۵، ۱۹۷۵) برای منطقه‌بندی حوضه‌های آبریز در نیوزلند مورد استفاده قرار داد. تاسکر (۱۹۸۲) الگوریتم تمام پیوند معرفی شده توسط سوکال^۶ و اسنیث^۷ (۱۹۶۳) برای منطقه‌بندی حوضه‌های آریزونا در ایالات متحده به کار گرفت.

ناتان و مکماهون (۱۹۹۰) عملکرد الگوریتم‌های تک پیوند، تمام پیوند، پیوند متوسط و Ward از مجموعه‌ی روش‌های خوش‌بندی سلسله مراتبی متراکم کننده موجود در بسته‌ی آماری مطالعات اجتماعی^۸ (SPSS، ۱۹۸۸) را مقایسه کردند. شاخص‌های فاصله‌ی اقلیدسی^۹، اقلیدسی مربع^{۱۰}، منهتن^{۱۱}، چبیشف^{۱۲} و کسینوسی^{۱۳} در این مطالعه در نظر گرفته شدند. برن و همکاران (۱۹۹۷) الگوریتم خوش‌بندی سلسله مراتبی را برای منطقه‌بندی حوضه‌های آبریز کانادا به کار گرفتند. در

^۱ Ward J. H. Jr.

^۲ Acreman A. C.

^۳ Sinclair C. D.

^۴ Mosley M. P.

^۵ Dixon W. J.

^۶ Sokal R. R.

^۷ Sneath P. H. A.

^۸ Statistical Package for the Social Sciences

^۹ Euclidean

^{۱۰} Squared Euclidean

^{۱۱} Manhattan

^{۱۲} Chebychev

^{۱۳} Cosine

مطالعه‌ی آن‌ها از شاخص اختلافی که از تحقیق وبستر^۱ و باروف^۲ (۱۹۷۲) استخراج شده بود استفاده شد.

شاخص‌های فصلی حوضه که زمان متوسط رویداد وقایع سیلابی خوانده می‌شوند و ترتیب پدیده در هر ایستگاه هیدرومتری توسط برن و همکاران (۱۹۹۷) به عنوان ویژگی‌های مورد استفاده در شاخص اختلاف کانبرا^۳ مورد استفاده قرار گرفتند. شاخص‌های فصلی مبتنی بر وضعیت فصلی سیلاب ممکن است در زمانی که حوضه‌ها در ناحیه‌ی مورد نظر وضعیت فصلی سیلابی قوی را نشان نمی‌دهند یا چنانچه همه‌ی آن‌ها وضعیت فصلی سیلابی مشابهی داشته باشند، ویژگی‌های مفیدی نباشند.

۱-۲-۱-۳-ب- خوشبندی سلسله مراتبی تقسیم کننده

روش‌های خوشبندی سلسله مراتبی تقسیم کننده یا یک خوشه‌ی واحد که شامل هر N بردار مشخصه است، آغاز می‌شوند. در ادامه بردار مشخصه‌ای که بیشترین اختلاف را با دیگر بردارهای خوشه دارد مشخص شده و برای تشکیل یک گروه مجزا جدا می‌شود. سپس مقادیر اختلاف بردارهای مشخصه‌ی باقی مانده در خوشه‌ی اصلی آزموده می‌شوند تا چنانچه بردار دیگری باید به گروه مجزا اضافه شود، مشخص گردد. این مرحله، خوشه‌ی اصلی را به دو قسمت تقسیم می‌کند. این عملیات در مرحله‌ی بعد روی خوشه‌ی بزرگ‌تر اعمال می‌شود. این فرآیند تا زمانی که یک معیار توقف (مانند تعداد خوشه‌های مورد نظر) حاصل شود، ادامه می‌یابد. اگر هیچ معیار توقفی مشخص نشود، الگوریتم زمانی به پایان می‌رسد که تمامی خوشه‌های حاصل از تحلیل خوشه‌ای، خوشه‌های تکی باشند.
توصیف الگوریتم‌های خوشبندی تقسیم کننده می‌تواند در مطالعات انجام گرفته توسط مورتاف^۴ (۱۹۸۳) و گوئنوجه^۵ و همکاران (۱۹۹۱) یافت شود. ساوارسی^۶ و همکاران (۲۰۰۲) راهبردهای

¹ Webster R.

² Burrough P. A.

³ Canberra

⁴ Murtagh F.

⁵ Guenoche A.

⁶ Savaresi S. M.

انتخاب یک خوش برای جدا شدن در الگوریتم‌های خوشبندی تقسیم کننده را مورد بحث قرار دادند. عملکرد روش‌های خوشبندی تقسیم کننده همچنان در تحقیقات منطقه‌بندی در دست مطالعه است.

۳-۱-۲-۲- روش‌های خوشبندی افزایی

در روش‌های خوشبندی افزایی، تلاش می‌شود تا گروه‌بندی طبیعی موجود در داده‌ها از طریق یک K-medoids و K-means افزای واحد بازیابی شود. این دسته از روش‌ها خود به دو زیرگروه روش‌های تقسیم می‌شوند.

در روش K-means (بال^۱ و هال^۲، مک‌کوین^۳، ۱۹۶۷) هر خوش توسط مرکز خود که میانگین (وزنی یا غیروزنی) بردارهای مشخصه در آن خوش است، نماینده‌گی می‌شود. این روش به دلیل کارایی خود در خوشبندی مجموعه‌های بزرگ داده‌ها با ویژگی‌های عددی مشهور است. با این حال، این روش در خوشبندی داده‌های رسته‌ای دارای محدودیت‌هایی است (رالامبوندرینی^۴، ۱۹۹۵؛ هوانگ^۵ و انجی^۶، ۲۰۰۲). به علاوه، این روش نسبت به وجود داده‌های پرت حساس است.

در روش K-medoids میانه‌ی هر خوش به عنوان نماینده‌ی آن در نظر گرفته می‌شود. این کار دو مزیت دارد. نخست این که این روش می‌تواند برای هر دو نوع داده‌های عددی و رسته‌ای مورد استفاده قرار گیرد، و دوم این که انتخاب این نماینده‌ها توسط موقعیت بخش غالب نقاط داده‌های درون یک خوش تعیین می‌شوند و از این رو حساسیت کمتری نسبت به حضور داده‌های پرت دارد (رائو و سرینیواس، ۲۰۰۸). نمونه‌هایی از روش‌هایی که می‌توانند به عنوان زیرمجموعه‌های روش K-medoids در نظر گرفته شوند عبارتند از PAM^۷ (کافمن^۸ و روسيوف^۹، ۱۹۹۰)، CLARA^۱ (انجی و

^۱ Ball G.

^۲ Hall D.

^۳ MacQueen J.

^۴ Ralambondrainy H.

^۵ Huang Z.

^۶ Ng M. K.

^۷ Partitioning Around Medoids

^۸ Kaufman L.

^۹ Rousseeuw P.

هان^۲، ۱۹۹۴)، CLARANS^۳. در میان این روش‌ها PAM برای استفاده در مورد مجموعه‌های کوچک داده‌ها مؤثر است.

هوانگ (۱۹۹۷، ۱۹۹۸) با اصلاح الگوریتم K-means، الگوریتم K-modes را برای خوشبندی مجموعه‌های بزرگ داده‌های رسته‌ای پیشنهاد داد. هر خوشبندی توسط مد یا نمای خود معرفی می‌شود و روشی مبتنی بر فراوانی برای به هنگام کردن نماها در فرآیند خوشبندی به منظور کمینه‌سازی تابع هزینه‌ی خوشبندی به کار می‌رود.

الگوریتم‌هایی چون PAM و K-means برای مطالعات منطقه‌بندی در هیدرولوژی مناسب هستند. الگوریتم K-means و انواع اصلاح شده‌ی آن توسط ویلشاير^۴ (۱۹۸۶)، برن (۱۹۸۹)، باسکار^۵ و اوکانر^۶ (۱۹۸۹) و برن و گول^۷ (۲۰۰۰) برای اجرای تحلیل فراوانی منطقه‌ای سیلاب مورد استفاده قرار گرفته‌اند. برن (۱۹۸۹) الگوریتم خوشبندی K-means را برای تشخیص گروه‌بندی مناسب یک شبکه از ایستگاه‌های هیدرومتری در مانیتوبای جنوبی واقع در کانادا به کار گرفت. در این مطالعه آماره‌های سیلاب (ضریب تغییرات جریان‌های بیشینه، جریان میانگین سالانه تقسیم بر سطح زهکشی) و موقعیت جغرافیایی (طول و عرض جغرافیایی) حوضه‌ها (ایستگاه‌ها) به عنوان ویژگی‌های بردارهای مشخصه مورد استفاده قرار گرفتند. به طور سنتی، آماره‌های سیلاب نظری ضریب تغییرات به منظور آزمودن همگنی مناطق به دست آمده به کار می‌رond. استفاده از متغیرهای یکسان مربوط به سیلاب برای تشکیل مناطق و متعاقب آن برای ارزیابی همگنی مناطق حاصل، منجر به تشکیل مناطقی می‌شود که اگرچه همگن هستند اما ممکن است برای تحلیل فراوانی منطقه‌ای سیلاب مفید نباشند (برن و همکاران، ۱۹۹۷). اگر ویژگی‌های درون سایتی سیلاب به عنوان ویژگی‌های بردارهای مشخصه مورد استفاده قرار گیرند، باید به گونه‌ای اطمینان حاصل کرد که درجه‌ی همبستگی بالایی با

¹ Clustering Large Applications

² Han J.

³ Clustering Large Applications based on Randomized Search

⁴ Wiltshire S. E.

⁵ Bhaskar N. R.

⁶ O'Connor C. A.

⁷ Goel N. K.

چندک‌های سیلاب مورد نظر از خود نشان نمی‌دهند (رائو و سرینیواس، ۲۰۰۸). افزون بر این استفاده از آماره‌های سیلاب در یک شاخص اختلاف، امکان استفاده از مناطق به دست آمده را برای برآورده چندک‌های حدی سیلاب در سایت‌های فاقد تجهیزات هیدرومتری در منطقه‌ی مورد مطالعه را محدود می‌کند.

هنگامی که تحلیل خوش‌های مبتنی بر ویژگی‌های سایت‌ها باشد، آماره‌های درون سایتی به عنوان مبنای یک آزمون مستقل همگنی مناطق نهایی قابل استفاده هستند (هاسکینگ و والیس، ۱۹۹۷). برن و گول (۲۰۰۰) الگوریتم K-means را بر ویژگی‌های سایت‌های مجموعه‌ای از حوضه‌ها در هند به منظور حصول مناطقی برای اجرای تحلیل فراوانی منطقه‌ای سیلاب، اعمال کردند. ویژگی‌های مورد استفاده در این مطالعه شامل مساحت حوضه، طول و شیب آبراهه‌ی اصلی رودخانه بودند. اشکال استفاده‌ی صرف از ویژگی‌های فیزیوگرافیک برای تشکیل مناطق آن است که شباهت ویژگی‌های فیزیوگرافیک لزوماً بر شباهت واکنش هیدرولوژیک حوضه دلالت نمی‌کند (برن و همکاران، ۱۹۹۷).

۳-۲-۱-۳- خوش‌بندی ترکیبی

الگوریتم‌های خوش‌بندی سلسله مراتبی تحت تأثیر تعیین شرایط اولیه قرار نمی‌گیرند، در حالی که الگوریتم‌های خوش‌بندی افزایی به شدت تحت تأثیر حدس‌های اولیه درباره‌ی تعداد خوش‌ها، مراکز خوش‌ها و ... هستند. الگوریتم‌های خوش‌بندی افزایی در حدی که بردارهای مشخصه بتوانند از یک خوش‌به خوش‌ی دیگر برای کمینه‌سازی تابع هدف انتقال یابند، پویا هستند. در مقابل، در الگوریتم‌های خوش‌بندی سلسله مراتبی، بردارهای مشخصه‌ی اختصاص یافته به یک خوش‌ه در مراحل ابتدایی، نمی‌توانند از یک خوش‌به خوش‌ی دیگر انتقال یابند. مزیت‌های نسبی استفاده از هر دو الگوریتم خوش‌بندی سلسله مراتبی و تفکیکی، توسعه‌ی الگوریتم خوش‌بندی ترکیبی را به وجود آورد. در الگوریتم خوش‌بندی ترکیبی، مراکز خوش‌های حاصل از یک الگوریتم سلسله مراتبی برای تعیین مراکز اولیه‌ی الگوریتم خوش‌بندی افزایی مورد استفاده قرار می‌گیرند (رائو و سرینیواس، ۲۰۰۸).

۳-۱-۳- الگوریتم‌های خوشبندی ترکیبی و ارزیابی عملکرد آن

در این بخش، الگوریتم خوشبندی ترکیبی برای منطقه‌بندی حوضه‌ها ارائه می‌شود. در این روش، الگوریتم K-means (یک الگوریتم خوشبندی افزایی) برای تعیین کردن گروه‌های ایستگاه‌های هیدرومتری با اصلاح خوشبندی به دست آمده از الگوریتم خوشبندی سلسله مراتبی متراکم کننده‌ی تک پیوند، تمام پیوند، پیوند متوسط و الگوریتم Ward مورد استفاده قرار می‌گیرد. در ادامه، شاخص صحت خوشبندی که برای تعیین تقسیم‌بندی بهینه حوضه‌های فراهم شده توسط الگوریتم خوشبندی ترکیبی مفید است، توصیف شده است.

۳-۱-۳-۱- الگوریتم ترکیبی

اگر $\{y_i | i = 1, \dots, N\}$ یک مجموعه از N بردار مشخصه در یک فضای n بعدی (مانند $y_i = [y_{i1}, \dots, y_{in}] \in |R^n$) باشد که هر یک از آن‌ها یکی از N سایت را توصیف می‌کند و x_i امین بردار تجدید مقیاس شده در فضای ویژگی n بعدی ($x_i = [x_{i1}, \dots, x_{in}] \in |R^n$) به دست آمده به وسیله‌ی تجدید مقیاس y_i با استفاده از معادله‌ی (۳-۱) باشد، آنگاه

$$x_{ij} = \frac{w_j}{\sigma_j} [f(y_{ij})] \quad \text{for: } j = 1, \dots, n \quad (1-3)$$

که در آن $f(\cdot)$ نماینده‌یتابع تبدیل می‌باشد؛ y_i نشان دهنده‌ی مقدار ویژگی j در بردار مشخصه‌ی n بعدی y_i ؛ x_{ij} معرف مقدار تجدید مقیاس شده‌ی y_{ij} ؛ w_j وزن اختصاص یافته به ویژگی j است؛ σ_j انحراف معیار ویژگی j است. تجدید مقیاس ویژگی‌ها ممکن است به علت اختلافات موجود بین واریانس، اندازه‌ی نسبی و اهمیت آن‌ها لازم باشد.

خوشبندی شکل گرفته در گام $N-K$ از یک الگوریتم خوشبندی سلسله مراتبی متراکم کننده برای

تعیین شرایط اولیه‌ی الگوریتم K-means مورد استفاده قرار می‌گیرند (هارتیگان^۱ و وونگ^۲، ۱۹۷۹). الگوریتم K-means (KMA) یک روش عددی است که در آن بردارهای مشخصه از یک خوشه به خوشه‌ی دیگر برای کمینه‌سازی مقدار تابع هدف (F) تعریف شده در معادله‌ی (۲-۳) جابه‌جا می‌شوند.

$$F = \sum_{k=1}^K \sum_{j=1}^n \sum_{i=1}^{N_k} d^2(x_{ij}^k - x_{*j}^k) \quad (2-3)$$

که در آن K معرف تعداد خوشه‌ها است؛ N_k نماینده‌ی تعداد بردارهای مشخصه در خوشه‌ی k است؛ x_{ij}^k نشان دهنده‌ی مقدار تجدید مقیاس شده‌ی ویژگی j در بردار مشخصه‌ی i اختصاص یافته به خوشه‌ی k است؛ x_{*j}^k مقدار میانگین ویژگی j برای خوشه‌ی k است که به صورت معادله‌ی (۳-۳) محاسبه می‌شود.

$$x_{*j}^k = \frac{\sum_{i=1}^{N_k} x_{ij}^k}{N_k} \quad (3-3)$$

با کمینه‌سازی F در معادله‌ی (۲-۳)، فاصله‌ی هر بردار مشخصه از مرکز خوشه‌ای که به آن تعلق دارد، کمینه می‌شود. در مورد شکل یا اندازه‌ی کلی خوشه‌ها با استفاده از یک شاخص فاصله‌ی مناسب (d ، مانند فاصله‌ی اقلیدسی و ... می‌توان تا حدودی قضاوت کرد (رائو و سرینیواس، ۲۰۰۶a).

مقدار بهینه‌ی حاصل شده توسط تابع هدف F وابسته به مراکز خوشه‌های اولیه‌ی استفاده شده برای تعیین شرایط اولیه‌ی KMA است. از آنجا که هیچ روش خاصی برای تعیین شرایط اولیه‌ی مراکز خوشه‌ها برای منتهی شدن به یک مقدار حداقل کلی برای تابع هدف F ، به صورت قطعی تأیید نشده است، روش‌های متعدد تعیین شرایط اولیه مورد استفاده قرار می‌گیرند. ویلشایر (۱۹۸۶) به طور

¹ Hartigan J. A.

² Wong M. A.

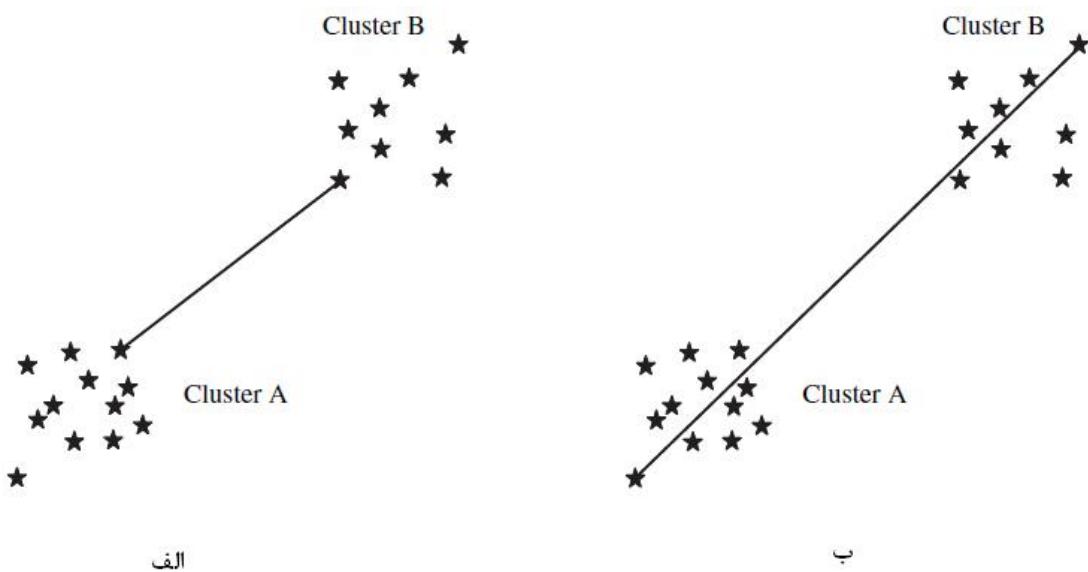
تصادفی داده‌ها را برای تعیین شرایط اولیه‌ی الگوریتم خوشبندی افزار کرد. باسکار و اوکانر (۱۹۸۹) مراکز خوشه‌های اولیه را به عنوان بردارهای مشخصه‌ای که حداقل توسط یک فاصله‌ی حداقل مشخص مجزا شده‌اند در نظر گرفتند.

هر بردار مشخصه به یک مرکز خوشه که در میان K خوشه به آن نزدیک‌تر است اختصاص می‌یابد. پس از اختصاص بردارها به مراکز K خوشه، مرکز هر یک از خوشه‌ها به هنگام شده و مقدار تابع هدف محاسبه می‌شود. این فرآیند یک مرتبه تکرار الگوریتم عددی K-means را کامل می‌کند. فرآیند اختصاص بردارها به نزدیک‌ترین مراکز خوشه‌ها و به هنگام کردن مراکز خوشه در هر یک از تکرارهای عددی متوالی انجام می‌شود. الگوریتم در یک نقطه، زمانی که تغییر در مقدار تابع هدف بین تکرارهای عددی متوالی به اندازه‌ی کافی کوچک شود، متوقف می‌شود (رائو و سرینیواس، ۲۰۰۶).

۳-۱-۳-۲- الگوریتم‌های تک پیوند، تمام پیوند و پیوند متوسط

الگوریتم‌ها با N خوشه‌ی تکی که هر یک شامل یک بردار مشخصه‌ی تجدید مقیاس شده هستند، آغاز می‌شوند. در میان N خوشه‌ی تکی، دو خوشه‌ی نزدیک‌تر x_i و x_j مشخص شده و برای تشکیل یک خوشه‌ی جدید $[x_i, x_j]$ به هم می‌پیوندند.

در الگوریتم تک پیوند، فاصله‌ی بین خوشه‌ی جدید $[x_i, x_j]$ و هر خوشه‌ی تکی x_k دیگر، کم‌تر یا مساوی فاصله‌ی بین x_i و x_k ، یا x_j و x_k است. در کل، فاصله‌ی بین دو خوشه‌ی غیرتکی از فواصل بین تمام زوج‌های ممکن بردارهای مشخصه در دو خوشه کوچک‌تر است (شکل ۱-۳). در سوی دیگر در الگوریتم تمام پیوند، فاصله‌ی بین خوشه‌ی جدید $[x_i, x_j]$ و هر خوشه‌ی تکی x_k بزرگ‌تر یا مساوی فواصل بین x_i و x_k ، یا x_j و x_k است. در کل، فاصله‌ی بین دو خوشه‌ی غیرتکی بزرگ‌تر از فواصل بین تمام جفت‌های بردارهای مشخصه‌ی ممکن در دو خوشه است (شکل ۱-۳). هر زوج شامل یک بردار مشخصه از هر یک از دو خوشه است (رائو و سرینیواس، ۲۰۰۸).



شکل (۳-۱) ترسیم تعریف فاصله (یا اختلاف) بین دو خوشه در الگوریتم‌های خوشه‌بندی سلسله مراتبی متراکم کننده (الف) تک پیوند، (ب) تمام پیوند. هر یک از ستاره‌ها نشان‌دهنده‌ی یک بردار مشخصه در یکی از خوشه‌ها است و خط اتصال بین ستاره‌ها، معرف فاصله‌ی بین آن‌ها است. (رائو و سرینیواس، ۲۰۰۸)

در هر مرحله، دو خوشه‌ی نزدیک‌تر شناسایی شده و به هم می‌پیوندند. در نتیجه، تعداد خوشه‌های موجود در هر مرحله، یک عدد کم‌تر خواهد شد. الگوریتم‌ها در مرحله‌ای که تعداد خوشه‌ها با مقدار معین K برابر شود، پایان می‌یابند (رائو و سرینیواس، ۲۰۰۸).

در الگوریتم پیوند متوسط، فاصله‌ی بین دو خوشه به صورت فاصله‌ی متوسط بین آن‌ها بیان می‌شود. این فاصله‌ی متوسط می‌تواند به شکل میانه‌ی فواصل بین بردارهای مشخصه یا میانگین‌های وزنی یا غیر وزنی فواصل بردارهای مشخصه‌ی دو خوشه تعریف شود.

۳-۳-۱-۳- Ward

تابع هدف W در الگوریتم Ward (وارد، ۱۹۶۳)، مطابق معادله‌ی (۴-۳) مجموع مربعات انحرافات بردارهای مشخصه از مراکز خوشه‌های متناظر آن‌ها را کمینه می‌کند.

$$W = \sum_{k=1}^K \sum_{j=1}^n \sum_{i=1}^{N_k} (x_{ij}^k - x_{*j}^k)^2 \quad (4-3)$$

الگوریتم Ward با خوشه‌های تکی آغاز می‌شود. در این مرحله، مراکز خوشه‌ها با بردارهای مشخصه‌ی موجود در خوشه یکسان هستند. بنابراین، مقدار تابع هدف برابر صفر است. در هر مرحله از تحلیل، ادغام هر جفت ممکن از خوشه‌ها مورد بررسی قرار گرفته و دو خوشه‌ای که به هم پیوستن آن‌ها به کمترین افزایش در W منجر می‌شود، به هم می‌پیوندند. تغییر در مقدار تابع هدف (W) در اثر ادغام، تنها به رابطه‌ی بین دو خوشه‌ی ادغام شده بستگی دارد و به روابط با سایر خوشه‌ها وابسته نیست (رأو و سرینیواس، ۲۰۰۸).

الگوریتم Ward در بازیابی ساختار خوشه مناسب است و تمایل به تشکیل خوشه‌های کروی با اندازه‌ی تقریباً مساوی دارد. این ویژگی الگوریتم Ward آن را برای تشخیص مناطق همگن برای منطقه‌بندی مفید می‌سازد. با این حال، مانند سایر روش‌های خوشه‌بندی سلسله مراتبی هیچ پیش زمینه‌ای در الگوریتم Ward برای جابه‌جایی بردارهای مشخصه‌ای که ممکن است در مرحله‌ی نخست در تحلیل به شکل نامناسبی دسته‌بندی شده باشند، وجود ندارد (رأو و سرینیواس، ۲۰۰۸).

۴-۳-۱-۳- شاخص صحت خوشه‌بندی عرض silhouette

شاخص‌های صحت خوشه‌بندی سخت به منظور تعیین تعداد بهینه‌ی خوشه‌ها در مجموعه‌ای از داده‌ها مورد استفاده قرار می‌گیرند. صحت خوشه‌بندی شامل روش ارزیابی نتایج یک الگوریتم خوشه‌بندی است.

عرض silhouette (روسیوف، ۱۹۸۷) برای یک بردار مشخصه، شاخصی از میزان تشابه آن بردار مشخصه با بردارهای مشخصه‌ی دیگر در خوشه‌ی خودش در مقایسه با بردارهای مشخصه در خوشه‌های دیگر است. عرض silhouette_(i) برای نامین بردار مشخصه در خوشه‌ی k به صورت

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (5-3)$$

تعریف می‌شود که در آن $a(i)$ فاصله‌ی میانگین از بردار مشخصه‌ی θ نسبت به تمام بردارهای مشخصه‌ی دیگر در خوشی k است؛ $b(i)$ حداقل فاصله‌ی میانگین از بردار مشخصه‌ی θ نسبت به تمام بردارهای مشخصه‌ی دیگر در خوشی j است. از این رابطه نتیجه می‌شود که $-1 \leq s(i) \leq 1$.

اگر $s(i)$ نزدیک به ۱ باشد، می‌توان این گونه استنتاج کرد که بردار مشخصه‌ی θ به یک خوشی مناسب اختصاص یافته است. از طرف دیگر، وقتی که $s(i)$ نزدیک به -۱ باشد، می‌توان این طور نتیجه‌گیری کرد که بردار مشخصه‌ی θ در خوشی مناسبی جای نگرفته است. هنگامی که $s(i)$ تقریباً برابر صفر است، این موضوع نشان‌دهنده‌ی آن است که بردار مشخصه‌ی θ به طور مساوی دور از دو خوشی قرار می‌گیرد. برای K خوشی موجود، عرض silhouette کلی، برابر میانگین عرض‌های silhouette برای تمام بردارهای مشخصه‌ی موجود در مجموعه‌ی داده‌ها است. افزای دارای بیشترین عرض silhouette متوسط کلی به عنوان افزای بهینه در نظر گرفته می‌شود (رائو و سرینیواس، ۲۰۰۸).

۲-۲-۳ - خوشبندی فازی

۱-۲-۳ - معرفی خوشبندی فازی

در فصل گذشته، منطقه‌بندی حوضه‌های آبریز با استفاده از تحلیل خوشهای ترکیبی مورد بحث قرار گرفت. تحلیل خوشهای ترکیبی یک روش خوشبندی سخت است. در منطقه‌بندی به وسیله‌ی خوشبندی سخت، یک حوضه بر اساس اختصاص یا عدم اختصاص به یک خوشه دسته‌بندی می‌شود. در واقعیت، اکثر سایتها دارای شباهت‌های جزئی به خوشهای متعدد هستند. از این رو نمی‌توان از اختصاص کامل یک سایت به یک خوشه یا خوشه‌ی دیگر را سخن گفت. در مقابل، خوشبندی فازی

به یک سایت اجازه‌ی داشتن عضویت‌های جزئی یا توزیعی در تمام خوش‌ها را می‌دهد. به عبارت دیگر، در خوش‌بندی فازی، یک سایت می‌تواند هم‌مان به بیش از یک خوش‌ اختصاص یابد. بنابراین این روش منجر به تعیین خوش‌هایی با مرزهای مبهم بین آن‌ها، در برابر خوش‌های مشخص با مرز-های آشکار در مورد خوش‌بندی سخت می‌شود. از این رو انتظار می‌رود روش خوش‌بندی فازی برای منطقه‌بندی که در این فصل مورد بحث قرار می‌گیرد، نسبت به خوش‌بندی سخت اطلاعات بیشتری را ارائه می‌دهد و از این جهت توصیف بهتری از شرایط واقعی را حاصل می‌کند.

نظریه‌ی مجموعه‌ی فازی (زاده، ۱۹۶۵) یک راه طبیعی برای توضیح موقعیت‌هایی است که بردارهای داده‌ها دارای شباهت‌های جزئی با خوش‌های متعدد هستند. خوش‌بندی فازی به یک بردار مشخصه اجازه می‌دهد که هم‌مان با یک درجه‌ی عضویت یا تعلق (اختصاص) مشخص در فاصله‌ی [۰,۱] به تمام خوش‌ها اختصاص یابد. روسپینی^۱ (۱۹۶۹، ۱۹۷۰) نخستین بار این ایده را مطرح کرد که توسط دان^۲ (۱۹۷۴) برای شکل دادن یک روش خوش‌بندی فازی مورد استفاده قرار گرفت.

در الگوریتم‌های خوش‌بندی فازی، اختصاص کامل یک نقطه‌ی داده به یک خوش‌ی موجود وجود ندارد. از این رو، این روش‌ها نیازمند فضای ذخیره‌ی حافظه‌ی بیشتری نسبت به الگوریتم‌های خوش-بندی سخت هستند که پیشرفت تجهیزات محاسباتی قدرتمند طی سه دهه‌ی گذشته، توسعه و استفاده از روش‌های خوش‌بندی فازی برای گونه‌هایی از کاربردها، شامل منطقه‌بندی را در پی داشته است (رائو و سرینیواس، ۲۰۰۸).

در میان روش‌های خوش‌بندی فازی موجود، الگوریتم فازی c-means (FCM) توسط بزدک^۳ (۱۹۸۱) ساده‌ترین و رایج‌ترین روش خوش‌بندی است. این روش، گسترشی از الگوریتم سخت K-means در چارچوب فازی است. الگوریتم FCM در گونه‌ای از زمینه‌ها شامل مهندسی کشاورزی، فضانوردی، شیمی، زمین‌شناسی، تحلیل تصویر، علوم پزشکی و تشخیص هدف کاربردهایی یافته

¹ Ruspini E. H.

² Dunn J. C.

³ Bezdek J. C.

است (بزدک، ۱۹۸۷).

در هیدرولوژی، پژوهشگران بسیاری تحلیل خوشهای را برای دسته بندی حوضه‌های آبریز در گروههایی که از نظر واکنش هیدرولوژیک همگن هستند، استفاده کرده‌اند، با این حال، تلاش‌های بسیار کمی برای جست و جوی پتانسیل خوشبندی فازی برای منطقه‌بندی انجام گرفته است. بارگایی^۱ و همکاران (۱۹۹۸) دو روش خوشبندی فازی، Iphigenie و ISODATA را برای منطقه‌بندی در نظر گرفتند. هال^۲ و مینز^۳ (۱۹۹۹) استفاده از الگوریتم فازی c-means را برای منطقه‌بندی با اعمال آن بر یک نمونه از ۱۰۱ سایت هیدرومتری از دو منطقه‌ی مشخص شده در گزارش مطالعات سیلاب انگلستان (NERC، ۱۹۷۵) آزمودند. این مطالعه، سطح حوضه‌ی زهکشی، طول آبراهه‌ی اصلی، شب آبراهه‌ی اصلی، بارندگی سالانه‌ی میانگین و اندیس خاک را به عنوان ویژگی‌هایی برای تحلیل در نظر گرفت.

۲-۲-۳- الگوریتم فازی c-means

در این بخش توصیفی از الگوریتم فازی c-means ارائه می‌شود. به دنبال آن، معیارهای مورد استفاده برای ارزیابی صحت خوشبندی حاصل از الگوریتم مورد بحث قرار می‌گیرند.

۲-۲-۱- توصیف الگوریتم

در ادبیات خوشبندی فازی، الگوریتم فازی C-Means (FCM) که توسط دان (۱۹۷۴) پیشنهاد شده و توسط بزدک (۱۹۸۱) گسترش یافته است، رایج است. این الگوریتم که مبتنی بر بهینه‌سازی عددی یک تابع هدف فازی است، برای تقسیم N سایت در یک منطقه به C خوشی فازی مفید است.

فرض کنید y_k ، معرف k امین بردار مشخصه که نماینده‌ی k امین حوضه در فضای مختصات n

¹ Bargaoui Z-K

² Hall M. J.

³ Minns A. W.

بعدی با محورهای مختصات با نامهای $(y_1, \dots, y_n) \in R^n$ مانند $y_k = [y_{1k}, \dots, y_{nk}]$ باشد که در آن y_{ik} مقدار ویژگی i در y_k را نشان می‌دهد. ویژگی‌های مختلفی می‌توانند برای منطقه‌بندی در نظر گرفته شوند. ویژگی‌های بردار مشخصه‌ی y_k به این صورت تجدید مقیاس می‌شوند:

$$x_{ik} = \frac{w_i}{\sigma_i} [f(y_{ik})] \quad \text{for} \quad 1 \leq i \leq n, \quad 1 \leq k \leq N \quad (6-3)$$

که در آن x_{ik} نشان دهنده‌ی مقدار تجدید مقیاس شده‌ی y_{ik} است؛ w_i وزن اختصاص یافته به ویژگی i است؛ σ_i معرف انحراف معیار ویژگی i است؛ $f(\cdot)$ نماینده‌ی تابع تبدیل و N نشان دهنده‌ی تعداد بردارهای مشخصه‌ی n بعدی است. تجدید مقیاس ویژگی‌ها به دلیل اختلافات موجود در واریانس، بزرگی نسبی و اهمیت آن‌ها ضروری است.

مجموعه‌ای از N بردار مشخصه‌ی می‌تواند به صورت یک ماتریس $n \times N$ داده‌ها مانند X نمایش داده شود.

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1N} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nN} \end{bmatrix} \quad (7-3)$$

افزون بر این، در نظر بگیرید $V = (v_1, v_c, \dots, v_c)$ نشان دهنده‌ی یک گروه c تایی از نمونه‌هایی باشد که هر یک از آن‌ها مرکز یکی از c خوش را توصیف می‌کند. الگوریتم FCM ماتریس X را از طریق کمینه‌سازی تابع هدفی که در ادامه معرفی می‌شود، به c زیرمجموعه (یا خوش) دارای هم-پوشانی تقسیم می‌کند. کمینه‌سازی

$$J(U, V; X) = \sum_{i=1}^c \sum_{k=1}^N (u_{ik})^\mu d^2(x_k, v_i) \quad (8-3)$$

تابع این قیود است،

$$\sum_{i=1}^c u_{ik} = 1 \quad \forall k \in \{1, \dots, N\} \quad (9-3)$$

$$0 < \sum_{k=1}^N u_{ik} < N \quad \forall i \in \{1, \dots, c\} \quad (10-3)$$

که در آن $u_{ik} \in [0,1]$ عضویت (یا درجه‌ی تعلق) k امین بردار مشخصه‌ی تجدید مقیاس شده x_k در خوشی فازی fam را نشان می‌دهد، U ماتریس افزای فازی است که شامل عضویت هر یک از بردارهای مشخصه‌ی تجدید مقیاس شده در هر خوشی فازی معادله‌ی (11-۳) است؛ پارامتر $\mu \in [1, \infty]$ مربوط به توان وزنی برای هر عضویت فازی است؛ $d^2(x_k, v_i)$ فاصله‌ی بین k امین بردار مشخصه‌ی تجدید مقیاس شده x_k تا مرکز خوشی fam است. هنگامی که نمونه‌های نقطه‌ای مورد استفاده قرار می‌گیرند، شکل عمومی معیار فاصله توسط معادله‌ی (12-۳) ارائه می‌شود،

$$U = \begin{bmatrix} u_{11} & \cdots & u_{1k} & \cdots & u_{1N} \\ \vdots & & \vdots & & \vdots \\ u_{i1} & \cdots & u_{ik} & \cdots & u_{iN} \\ \vdots & & \vdots & & \vdots \\ u_{c1} & \cdots & u_{ck} & \cdots & u_{cN} \end{bmatrix}_{c \times N} \quad (11-3)$$

$$d^2(x_k, v_i) = (x_k - v_i)^T A_i (x_k - v_i) \quad (12-3)$$

که در آن معیار (مقیاس) A_i یک ماتریس متقارن معین مثبت مربوط به خوشی fam است. برای برآورده فاصله‌ی اقلیدسی بین x_k و v_i ، $A_i = I$ است، که در آن I یک ماتریس واحد است. قید نخست یا معادله‌ی (9-۳) نیازمند این است که عضویت یک بردار مشخصه‌ی ورودی منتخب در کل c خوشی فازی باید در مجموع برابر ۱ باشد. اختصاص مقادیر عضویت بسیار کوچک به یک بردار مشخصه معنادار است، اگر آن بردار مشخصه نماینده‌ی حوضه‌ای باشد که واکنش هیدرولوژیک آن کاملاً متفاوت با واکنش هیدرولوژیک سایر حوضه‌های در نظر گرفته شده برای خوش بندی باشد. با این حال، قید نخست برای عضویت‌هایی از بردار مشخصه در تمام c خوشی که مجموعشان برابر ۱

نمی‌شود، اجازه نمی‌دهد تکرارهای عددی FCM به یک جواب همگرا شوند. از این رو در حل الگوریتم FCM، این امکان وجود دارد که سایت‌های مشخصی که در هیچ یک از مناطق تعیین شده جای نمی‌گیرند، باز هم مقادیر عضویت قابل توجهی در تمام خوش‌های خواسته باشند، به طوری که مجموع آن‌ها برابر یک شود. این موضوع می‌تواند همگنی خوش‌های خواسته را تحت تأثیر قرار دهد (رائو و سرینیواس، ۲۰۰۸). برای کم کردن اثر این مشکل، روش‌های اصلاح هاسکینگ و والیس (۱۹۹۷) که در ادامه‌ی این فصل به آن‌ها اشاره خواهد شد، می‌توانند مفید باشد.

در دهه‌ی گذشته، اصلاحات خاصی به منظور غلبه بر اثر نامطلوب مذکور قید نخست، برای الگوریتم FCM مرسوم پیشنهاد شده است (دیو^۱ و کریشناپورام^۲، ۱۹۹۷). با این وجود، این تکنیک‌های خوش‌بندی FCM اصلاح شده، هنوز به دنبال یافتن کاربردی در تحلیل فراوانی منطقه‌ای سیلان هستند و پژوهش در مورد مزیت آن‌ها نسبت به FCM مرسوم در منطقه‌بندی هنوز یک موضوع در دست تحقیق است.

قید دوم یا معادله‌ی (۳-۱۰) تضمین می‌کند که مجموع درجات عضویت در یک خوش‌ی فازی در N بردار مشخصه بین صفر تا N قرار می‌گیرد. اگر مجموع درجات عضویت در یک خوش‌ی فازی برابر صفر باشد، این مسئله دال بر این است که خوش‌های هیچ سایتی را شامل نمی‌شود. در مقابل، خوش‌ی فازی شامل تمام بردارهای مشخصه خواهد بود، اگر مجموع برابر N باشد (رائو و سرینیواس، ۲۰۰۶b).

μ در معادله‌ی (۳-۸) میزان فازی بودن خوش‌های را مشخص می‌کند. این پارامتر مقدار عضویت مشترک بین خوش‌های فازی را کنترل می‌کند. در $1 = \mu$ ، FCM در تئوری به جواب K-means می‌کند. برای عبارت دیگر، هنگامی که به ۱ میل کند، مقادیر عضویت به ۱ یا صفر میل سنتی همگرا می‌شود. به عبارت دیگر، خوش‌های $\mu \rightarrow 0$ می‌کنند. برای $\infty \rightarrow \mu$ ، بردارهای مشخصه به عضویت برابر در تمام C خوش‌های گرایش پیدا می‌کنند. بنابراین در کل، درجه‌ی عضویت k -امین بردار مشخصه‌ی تجدید مقیاس شده‌ی x_k در خوش‌ی فازی

¹ Dave R. N.

² Krishnapuram R.

نام، یعنی u_{ik} به $1/c$ میل می‌کند. افزایش در مقدار μ ، مشارکت مقادیر بزرگ (x_k, v_i) در تابع هدف را کاهش می‌دهد. به بیان دیگر، سایتهايی که ويژگی‌های آن‌ها نسبت به میانگین مشخصات خوشها (که توسط مراکز آن‌ها نشان داده می‌شود) تفاوت بیشتری دارد، سهم کمتری می‌یابند. در نتیجه، خوشها به در بر گرفتن سایتهاي بیشتر گرایش می‌یابند (رأو و سرینیواس، ۲۰۰۶b).

روش عددی الگوریتم FCM (بزدک، ۱۹۸۱) در ادامه به صورت خلاصه آمده است.

(۱) ماتریس اولیه افزار فازی U (یا ماتریس مراکز خوشهاي فازی V) با استفاده از یک تولید کننده اعداد تصادفی ایجاد می‌شود.

(۲) اگر الگوریتم FCM با ماتریس افزار فازی U آغاز شده است، عضويت‌های اولیه u_{ik}^{init} مربوط به x_k متعلق به خوشی i با استفاده از معادله (۱۳-۳) به منظور ارضای معادله (۹-۳) اصلاح می‌شود.

$$u_{ik} = \frac{u_{ik}^{init}}{\sum_{i=1}^c u_{ik}^{init}} \quad \text{for} \quad 1 \leq i \leq c, \quad 1 \leq k \leq N \quad (13-3)$$

اگر الگوریتم FCM با ماتریس مراکز خوشهاي فازی V (شامل c مرکز خوشهاي فازی $v_1^{init}, \dots, v_c^{init}$) آغاز شده است، عضويت‌های u_{ik} مربوط به x_k متعلق به خوشی i با استفاده از معادله (۱۵-۳) با جایگزین کردن v_i^{init} با v_i تعیین می‌شود.

(۳) مرکز فازی v_i برای $i = 1, 2, \dots, c$ به وسیله ای معادله (۱۴-۳) محاسبه می‌شود.

$$v_i = \frac{\sum_{k=1}^N (u_{ik})^\mu x_k}{\sum_{i=1}^c u_{ik}^{init}} \quad (14-3)$$

(۴) عضويت فازی u_{ik} با استفاده از معادله زیر به هنگام می‌شود.

$$u_{ik} = \frac{\left(\frac{1}{d^2(x_k, v_i)}\right)^{1/(\mu-1)}}{\sum_{i=1}^c \left(\frac{1}{d^2(x_k, v_i)}\right)^{1/(\mu-1)}} \quad \text{for } 1 \leq i \leq c, \quad 1 \leq k \leq N \quad (15-3)$$

گام‌های (۳) و (۴) تا زمانی که تغییر در مقادیر عضویت‌ها بین دو تکرار متوالی به قدر کافی کوچک شود، تکرار می‌شوند. در این نقطه، روش سنتی تحلیل خوش‌های فازی، غیر فازی کردن ماتریس افزای فازی U (نشان داده شده در معادله (۱۱-۳)) را برای اختصاص نهایی بردارهای مشخصه به خوش‌های توصیه می‌کند.

ماتریس افزای فازی می‌تواند با استفاده از روش حداکثر عضویت یا طبقه‌بندی کننده‌ی نزدیک‌ترین مرکز، غیر فازی یا سخت شود (راس^۱، ۱۹۹۵، ص ۳۹۸). در روش حداکثر عضویت مطابق معادله (۱۶-۳)، به بزرگ‌ترین مؤلفه در هر ستون U ، یک مقدار عضویت واحد اختصاص داده می‌شود و به تمام مؤلفه‌های دیگر آن ستون، مقدار عضویت صفر تعلق می‌گیرد. به عبارت دیگر، یک بردار مشخصه به خوش‌های تعلق می‌گیرد که بیشترین شباهت را به آن داشته باشد. از سوی دیگر در طبقه‌بندی کننده‌ی نزدیک‌ترین مرکز، هر یک از بردارهای مشخصه‌ی تجدید مقیاس شده، x_k به خوش‌های تعلق می‌گیرد که بنا بر معادله (۱۶-۳) نزدیک‌ترین فاصله را تا مرکز آن بر حسب فاصله‌ی اقلیدسی داشته باشد.

$$u_{jk} = \max_{1 \leq i \leq c} \{u_{ik}\} = 1; \quad u_{ik} = 0 \quad \text{for all } i \neq j \quad (16-3)$$

$$\begin{aligned} \text{If } d_{jk} &= \min_{1 \leq i \leq c} \{d_{ik}\} = \min_{1 \leq i \leq c} \|v_i - x_k\| \\ \text{then } u_{jk} &= 1; \quad u_{ik} = 0 \quad \text{for all } i \neq j \end{aligned} \quad (17-3)$$

در هیدرولوژی، هال و مینز (۱۹۹۹) هر دو معادله (۱۶-۳) و (۱۷-۳) را برای شکل دادن خوش‌های سخت در تحلیل خوش‌های فازی مورد استفاده قرار دادند. نتایج حاصل از پژوهش رأی و

^۱ Ross T. J.

سرینیواس (۲۰۰۶b) نشان داد که تلاش مورد نیاز برای تشکیل مناطق همگن برای تحلیل فراوانی منطقه‌ای سیلان با تشکیل خوشه‌های فازی نسبت به تشکیل خوشه‌های سخت با سخت کردن ماتریس افزای افزای به صورت چشمگیری کاهش می‌یابد.

یک خوشه‌ی فازی با اختصاص سایتهاست که عضویت آنها در خوشه از مقدار آستانه‌ی معینی تجاوز می‌کند به آن، شکل می‌گیرد. عموماً انتخاب یک مقدار آستانه برای تشکیل خوشه‌های فازی تا حدودی اختیاری و ذهنی است. در فازی‌ترین افزای، عضویت‌های یک بردار مشخصه در تمام خوشه‌ها برابر $c/1$ خواهد بود. از این رو مقدار $c/1$ به عنوان یک انتخاب قابل قبول برای عضویت فازی آستانه شناخته می‌شود. این چنین فرضیاتی ضروری هستند، اما همیشه نمی‌توانند رسیدن به بهترین نتیجه را تضمین کنند (راؤ و سرینیواس، ۲۰۰۸). در طول دو دهه‌ی گذشته، محققان معیارهای ابتکاری بسیاری را برای صحبت خوشبندی به منظور رسیدن به موضوع همگرایی توسعه داده‌اند. این موضوع در بخش ۳-۲-۳ مورد بحث قرار گرفته است.

۳-۲-۲-۲-۱- اختصاص سایتهاست جدید به خوشه‌های فازی

هنگامی که یک سایت جدید (دارای تجهیزات هیدرومتری یا فاقد آن) برای تحلیل فراوانی سیلان در نظر گرفته می‌شود، عضویت آن در تمام c خوشه (که به وسیله‌ی الگوریتم FCM مشخص می‌شود)، با استفاده از معادله‌ی (۱۵-۳) محاسبه می‌شود که در آن x_k با یک بردار مشخصه شامل ویژگی‌های تجدید مقیاس شده‌ی سایت جدید، x_{new} ، جایگزین می‌شود. به بیان ریاضی

$$u_i^{new} = \frac{\left(\frac{1}{d^2(x_{new}, v_i)} \right)^{1/(\mu-1)}}{\sum_{i=1}^c \left(\frac{1}{d^2(x_{new}, v_i)} \right)^{1/(\mu-1)}} \quad for \quad 1 \leq i \leq c \quad (18-3)$$

که در آن، u_i^{new} عضویت سایت جدید در خوشه‌ی فازی i است و نماینده‌ی مرکز خوشه‌ی فازی i ام

است. ویژگی‌های سایت جدید باید همانند ویژگی‌های استفاده شده برای تشکیل مناطق همگن باشد و باید با استفاده از معادله‌ی (۳-۶) تجدید مقیاس شده باشد. اگر سایت جدید فاقد تجهیزات هیدرومتری باشد، به تمام خوش‌هایی که عضویت آن سایت در آن‌ها بزرگ‌تر از مقدار آستانه‌ی پیش‌بینی شده مشخص عضویت فازی آستانه (مانند $c/1$) باشد، اختصاص می‌یابد. با این حال، اگر سایت جدید دارای تجهیزات هیدرومتری باشد، پس از اطمینان یافتن از این که افزودن سایت جدید به افزایش معناداری در ناهمگنی آماری خوش‌ها منجر نمی‌شود، به خوش‌هایی اختصاص می‌یابد که عضویتش در آن‌ها بزرگ‌تر از آستانه‌ی عضویت فازی باشد. همان‌گونه که در فصل اول اشاره شد، آماره‌های درون سایتی سیلاب نباید به عنوان ویژگی‌های مورد استفاده برای تشکیل مناطق به منظور تحلیل فراوانی سیلاب به کار گرفته شوند، زیرا آن‌ها به عنوان مبنای یک آزمون مستقل همگنی مناطق استفاده می‌شوند.

چندک سیلاب مطلوب برای یک سایت که بین دو یا تعداد بیشتری از مناطق مشترک است، می‌تواند با استفاده از میانگین وزنی مقادیر چندک‌های سیلاب برای آن سایت که بر اساس مناطق فازی برآورد شده است، محاسبه شود. وزن‌ها می‌توانند متناسب با درجه‌ی عضویت سایت در خوش‌های فازی اختصاص یابند (رائو و سرینیواس، ۲۰۰۸).

۳-۲-۳- شاخص‌های صحت خوش‌بندی فازی

ارزیابی صحت روشی برای ارزیابی و مقایسه‌ی خوش‌هایی به دست آمده از یک الگوریتم خوش‌بندی برای انتخاب‌های متفاوت پارامترها یا مقایسه‌ی خوش‌هایی حاصل از الگوریتم‌های خوش‌بندی مختلف است (بکر^۱ و جین^۲، ۱۹۸۱). در تحلیل خوش‌های فازی، ارزیابی صحت با استفاده از شاخص‌های صحت خوش‌بندی فازی^۳ اجرا می‌شود که متفاوت از تابع هدفی که با استفاده از الگوریتم خوش‌بندی فازی

¹ Backer E.

² Jain A. K.

³ cluster validity measures

بهینه می‌شود، در نظر گرفته می‌شوند.

معیارهایی که در ارزیابی و انتخاب خوش در نظر گرفته می‌شوند عبارتند از فشردگی و جدایی خوشها.

* فشردگی^۱: افزایش بهینه نیازمند آن است که اعضای هر خوش حتی الامکان به یکدیگر نزدیک باشند. یک شاخص رایج فشردگی واریانس است که باید کمینه شود. اگر تنها فشردگی به عنوان معیار صحت سنجی در نظر گرفته شود، آنگاه بهترین افزایش هنگامی به دست می‌آید که هر نقطه داده به عنوان یک خوشی مجزا در نظر گرفته شود.

* جدایی^۲: برای حصول افزایش بهینه لازم است که خوشها به صورت گستردگی در فضا جای گیرند. به عبارت دیگر، خوشها باید از یکدیگر دور باشند. اگر فقط جدایی بهینه به عنوان معیار صحت سنجی در نظر گرفته شود، آنگاه بهترین افزایش زمانی حاصل می‌شود که تمام نقاط داده‌ها در یک خوش جای بگیرند. در این مورد، فاصله‌ی جدایی تا نزدیک‌ترین خوش بی‌نهایت است.

یک افزایش فازی از مجموعه داده‌های $v_i (i = 1, 2, \dots, c)$ با $X = [x_k; k = 1, 2, \dots, N]$ که نشان دهنده‌ی مرکز هر خوش استند در نظر بگیرید. فرض کنید $u_{ik} (i = 1, 2, \dots, c; k = 1, 2, \dots, N)$ معرف عضویت فازی بردار مشخصه‌ی k در خوشی i است. در ادامه، شاخص‌های مختلف صحت خوشبندی که در ادبیات پژوهشی رایج هستند؛ به طور خلاصه توصیف می‌شوند. برخی شاخص‌های صحت خوشبندی تنها مقادیر عضویت یک افزایش فازی داده‌ها را مورد استفاده قرار می‌دهند. برخی نمونه‌ها عبارتند از ضرب افزایش^۳ (بزدک، ۱۹۷۴a,b)، انتروپی افزایش^۴ (بزدک، ۱۹۷۵)، نمای افزایش^۵ (ویندهام، ۱۹۸۱)، تابع داده یکنواخت^۶ (ویندهام، ۱۹۸۲). این موارد و شاخص‌های مشابه ممکن است قابل اعتماد نباشند، زیرا به هیچ یک از مشخصات داده‌ها ارتباطی ندارند.

¹ compactness

² separation

³ partition coefficient

⁴ partition entropy

⁵ partition exponent

⁶ Windham M. P.

⁷ uniform data functional

(۱) ضریب افزایش: این شاخص توسط بزدک (۱۹۷۴a,b) به منظور سنجش مقدار همپوشانی بین خوشه‌ها طرح شد.

$$V_{PC}(U) = \frac{1}{N} \sum_{i=1}^c \sum_{k=1}^N (u_{ik})^2 \quad (19-3)$$

(۲) انتروپی افزایش: بزدک (۱۹۸۱) انتروپی دسته بندی یک افزایش c فازی را به این صورت تعریف کرد:

$$V_{PE}(U) = -\frac{1}{N} \left[\sum_{i=1}^c \sum_{k=1}^N u_{ik} \log_a(u_{ik}) \right] \quad (20-3)$$

که در آن مبنای لگاریتمی $a \in (1, \infty)$ است. دو شاخص دیگر به نام‌های شاخص عملکرد فازی^۱ V_{FPI} ، و انتروپی دسته‌بندی نرمال شده^۲ V_{NCE} که توسط روبنز^۳ (۱۹۸۲) معرفی شدند، با استفاده از V_{PE} و V_{PC} به صورت زیر تعریف می‌شوند.

$$V_{FPI}(U) = 1 - \frac{c \times V_{PC}(U) - 1}{c - 1} \quad (21-3)$$

$$V_{NCE}(U) = \frac{V_{PE}(U)}{\log_a(c)} \quad (22-3)$$

افزایشیه متناظر با مقدار بیشینه‌ی V_{PC} (یا مقدار کمینه‌ی V_{PE} , V_{FPI} و V_{NCE}) است که بر حداقل همپوشانی بین خوشه‌ها دلالت می‌کند. دامنه‌ی تغییرات V_{PC} در محدوده‌ی $[1/c, 1]$ است، در حالی که V_{PE} در گستره‌ی $[0, \log_a(c)]$ جای می‌گیرد. در سوی دیگر، دامنه‌ی تغییرات V_{FPI} و V_{NCE} برابر $[0, 1]$ است. برای یک افزایش سخت، V_{PC} برابر ۱ است، در حالی که V_{PE} و V_{FPI} برابر V_{NCE} است.

¹ fuzziness performance index

² normalized classification entropy

³ Roubens M.

همگی مساوی صفر هستند.

هر گاه عضویت هر یک از بردارهای مشخصه در تمام خوشها برابر باشد ($u_{ik} = 1/c \quad \forall i, k$) که در فازی ترین افزار c خوشها روی می‌دهد، مقدار V_{PC} برابر $c/1$ و مساوی $\log_a(c)$ خواهد شد.

نقطه‌ی ضعف V_{NCE} ، V_{FPI} ، V_{PE} و V_{PC} این است که با هیچ یک از مشخصات داده‌ها ارتباط مستقیمی ندارند. در سال‌های اخیر، این شاخص‌های صحت در ادبیات هیدرولوژی مورد استفاده قرار گرفته‌اند (بارگایی و همکاران، ۱۹۹۸؛ هال و مینز، ۱۹۹۹؛ گولر^۱ و تاین^۲، ۲۰۰۰).

در حالی که V_{PC} با افزایش مقدار درجه‌ی فازی، تمایل به کاهش یکنواخت را نشان می‌دهد، V_{NCE} و V_{FPI} ، V_{PE} گرایش به افزایش یکنواخت را با افزایش مقدار این پارامتر به نمایش می‌گذارد. افزون بر این، V_{PC} و V_{PE} نسبت به مقدار درجه‌ی فازی در موقعی که $1 \rightarrow \mu$ و $\mu \rightarrow \infty$ حساس هستند (هاکیدی^۳ و همکاران، ۲۰۰۱، ص ۱۳۸).

(۳) شاخص فوکویاما و سوگنو^۴: فوکویاما^۵ و سوگنو^۶ (۱۹۸۹) یک شاخص صحت، V_{FS} ، را با بهره‌گیری از فشردگی و جدایی خوشها ارائه کردند. مقدار حداقل V_{FS} بیان‌گر یک افزار بهینه است که متناظر خوشها فشرده و کاملاً مجزا است.

$$V_{FS}(U, V: X) = \sum_{k=1}^N \sum_{i=1}^c (u_{ik})^\mu \|v_i - x_k\|_A^2 - \sum_{k=1}^N \sum_{i=1}^c (u_{ik})^\mu \|v_i - \bar{v}\|_A^2 \quad (۲۳-۳)$$

که در آن $\|\cdot\|$ معیار اقلیدسی است، \bar{v} بردار میانگین X معادله‌ی (۲۴-۳) است، A یک ماتریس متقارن معین مثبت و $\|X\|_A = \sqrt{X^T A X}$ یک معیار ضرب داخلی است. هنگامی که A برابر ماتریس واحد I باشد، شاخص فاصله‌ی $\|\cdot\|^2$ در معادله‌ی (۱۸-۳) محدود فاصله‌ی اقلیدسی خواهد بود.

¹ Güler C.

² Thine G. D.

³ Halkidi M.

⁴ Fukuyama and Sugeno index

⁵ Fukuyama Y.

⁶ Sugeno M.

$$\bar{v} = \frac{1}{N} \sum_{k=1}^N x_k \quad (24-3)$$

(۴) شاخص فاصله‌ی ژی-بنی^۱: شاخص پیشنهادی توسط ژی^۲ و بنی^۳ (۱۹۹۱) تابعی از مجموعه‌ی داده‌ها و مراکز خوش‌ها است. این شاخص به صورت نسبت فشردگی کل به جدایی یک افزار فازی با c خوش‌هه تعريف می‌شود (ژی و بنی، ۱۹۹۱، ص ۸۴۲). تابع صحت به این صورت تعريف می‌شود

$$V_{XB}(U, V: X) = \frac{\sum_{i=1}^c \sum_{k=1}^N (u_{ik})^2 \|v_i - x_k\|^2}{N \min_{i \neq k} \|v_i - v_k\|^2} \quad (25-3)$$

که در آن عبارت صورت کسر مجموع مجذورات انحراف فازی هر یک از بردارهای مشخصه x_k (که در آن عبارت صورت کسر مجموع مجذورات انحراف فازی هر یک از بردارهای مشخصه v_i است) از مرکز هر خوش‌هی فازی v_i ($i = 1, 2, \dots, c$) است. اندازه‌ی این عبارت با افزایش در فشردگی خوش‌ها کاهش می‌یابد. عبارت مخرج کسر، که حداقل جدایی بین مراکز خوش‌ها را می‌سنجد، برای خوش‌هایی که کاملاً از یکدیگر جدا هستند، مقدار بزرگ‌تری خواهد داشت. حداقل مقدار V_{XB} مبین یک افزار خوب است که متناظر با خوش‌هایی فشرده و کاملاً جدا از هم است. ژی و بنی (۱۹۹۱، ص ۸۴۳) جایگزینی $(u_{ik})^\mu$ با^۲ را در معادله‌ی (۲۵-۳) هنگامی که در معادله‌ی (۸-۳) $\mu \neq 2$ ، پیشنهاد کردند. پال^۴ و بزدک (۱۹۹۵، ص ۳۷۴) به این مورد به عنوان شاخص ژی-بنی توسعه یافته‌ی الگوریتم فازی c -means ($V_{XB,m}$) که بدین صورت حاصل می‌شود

$$V_{XB,m}(U, V: X) = \frac{\sum_{i=1}^c \sum_{k=1}^N (u_{ik})^\mu \|v_i - x_k\|^2}{N \min_{i \neq k} \|v_i - v_k\|^2} \quad (26-3)$$

شاره کردند. مقدار V_{XB} به صورت یکنواخت با افزایش تعداد خوش‌ها، کاهش می‌یابد. برای برطرف

¹ Xie-Beni validity measure

² Xie X. L.

³ Beni G.

⁴ Pal N. R.

کردن این مشکل، کوان^۱ (۱۹۹۸) یک شاخص جدید صحت خوشبندی V_K ارائه کرد که در صورت کسر دارای یک جمله‌ی اضافه است.

$$V_K(U, V; X) = \frac{\sum_{i=1}^c \sum_{k=1}^N (u_{ik})^2 \|v_i - x_k\|^2 + \frac{1}{C} \sum_{i=1}^c \|v_i - \bar{v}\|^2}{\min_{i \neq k} \|v_i - v_k\|^2} \quad (27-3)$$

صحت خوش اغلب برای تعیین تعداد بهینه‌ی خوش‌ها در یک مجموعه‌ی داده‌ها استفاده می‌شود (برای مثال گث^۲ و جوا^۳، ۱۹۸۹؛ ژی و بنی، ۱۹۹۱؛ تئودوریدیس^۴ و کوترباس^۵، ۱۹۹۹؛ هاکیدی و همکاران، ۲۰۰۱). این فرآیند نیازمند آن است که کلیه‌ی پارامترها الگوریتم خوشبندی به جز تعداد خوش‌ها، c ، ثابت نگاه داشته شوند. سپس پارامتر c از ۱ تا یک مقدار بیشینه C_{max} در افزایش‌های یک واحدی تغییر می‌کند. مقادیر یک شاخص صحت خوشبندی منتخب که برای خوش‌های به دست آمده برای هر انتخاب c محاسبه می‌شوند، به منظور تعیین تعداد بهینه‌ی خوش‌ها در یک مجموعه‌ی داده‌های مورد نظر مورد تحلیل قرار می‌گیرند (رأو و سرینیواس، ۲۰۰۸).

۳-۳-۳- خوشبندی با استفاده از نگاشت‌های خودسازمانده

۳-۱-۳- معرفی نگاشت‌های مشخصه‌ی خودسازمانده کوهون

در طول دو دهه‌ی اخیر مدل‌های مبتنی بر شبکه‌های عصبی مصنوعی^۶ به شکل گسترده‌ای توسعه یافته و به منظور شبیه رفتار نورون‌های بیولوژیک در مغز انسان مورد مطالعه قرار گرفته‌اند. غیرخطی بودن و انعطاف پذیری شبکه‌های عصبی مصنوعی آن‌ها را در گونه‌ای از کاربردهای علوم

¹ Kwon S. H.

² Gath I.

³ Geva A. V.

⁴ Theodoridis S.

⁵ Koutroubas K.

⁶ Artificial Neural Networks (ANN)

فیزیکی، شامل هیدرولوژی (گوینداراجو^۱ و رائو، ۲۰۰۰؛ رائو، ۲۰۰۰a,b, Neural Networks in Hydrology از این شبکه‌ها به نام نگاشت‌های مشخصه‌ی خود سازمانده کوهونن^۲ (SOFM) برای منطقه‌بندی حوضه‌های آبریز مورد بحث قرار گرفته است.

نگاشت‌های خودسازمانده که نگاشت‌های نگه دارنده‌ی توپولوژی^۳ هم نامیده می‌شوند، به گونه‌ای از شبکه‌های یاد گیرنده‌ی رقابتی تعلق دارند. آن‌ها مبتنی بر یاد گیرنده‌ی بدون نظارت هستند و این بدان معنا است که هیچ خروجی هدفی برای دسته‌بندی داده‌های موجود مورد نیاز نیست. این شبکه‌ها می‌کوشند تا به وسیله‌ی نگاشت داده‌های موجود بر یک نگاشت مشخصه، که همچنین می‌توان آن را به عنوان لایه‌ی خروجی یا فضای خروجی در نظر گرفت، ساختار توپولوژیکی در داده‌های ورودی بیابند (رائو و سرینیواس، ۲۰۰۸).

دو مدل مختلف از شبکه‌های عصبی خود سازمانده وجود دارد (سو^۴ و چانگ^۵، ۲۰۰۰؛ مدل ویلشاو-وندر مالزبرگ (ویلشاو^۶ و مالزبرگ^۷، ۱۹۷۶) و مدل کوهونن (کوهونن^۸، ۱۹۸۲). مدل نخست به طور ویژه برای نگاشت در جایی که ابعاد فضای ورودی با ابعاد فضای خروجی یکسان است کاربرد دارد، در حالی که مدل دوم قادر به تولید نگاشت‌هایی از فضاهای ورودی با تعداد ابعاد بالا به فضاهای خروجی با تعداد ابعاد پایین‌تر که به عنوان لایه‌ی کوهونن شناخته می‌شوند، است. تصویری از نگاشت‌های خود سازمانده، شامل پیشرفتهای اخیر و کاربردهای مهندسی آن‌ها در پژوهش‌های کوهونن و همکاران (۱۹۹۶)، اوبرمایر^۹ و ساینوسکی^{۱۰} (۲۰۰۱) یافت می‌شود.

SOFM (کوهونن، ۱۹۸۲) با لایه‌ی یک بعدی (1-D) کوهونن (که شبکه‌ی کوهونن خطی نیز

¹ Govindaraju R. S.

² Kohonen Self-Organizing Feature Maps (SOFM)

³ topology-preserving maps

⁴ Su M. C.

⁵ Chang H. T.

⁶ Wilshaw D. J.

⁷ Malsburg C. V.

⁸ Kohonen T.

⁹ Obermayer K.

¹⁰ Sejnowski T. J.

خوانده می‌شود) در مطالعات منطقه‌بندی مورد استفاده قرار گرفته است (هال و مینز، ۱۹۹۸؛ هال و همکاران، ۲۰۰۲؛ جینگی^۱ و هال، ۲۰۰۴). اما انتخاب تعداد گره‌ها در لایه‌ی کوهونن همچنان به صورت ذهنی باقی مانده است. هال و مینز (۱۹۹۸) استفاده از شبکه‌ی کوهونن را برای منطقه‌بندی با اعمال آن بر یک نمونه‌ی ۱۰۱ تایی از سایت‌های مجهز به تجهیزات هیدرومتری در جنوب غربی انگلیس و ولز در انگلستان مورد آزمایش قرار دادند. هال و همکاران (۲۰۰۲) یک SOFM یک بعدی (1-D) را برای سه مجموعه از جنوب غربی انگلیس و ولز، ولز و اسکاتلند و جزایر جاوه و سوماترا در اندونزی به کار گرفتند. این مطالعات در مورد صحت سنگی مناطق تشکیل شده با استفاده از شاخص‌های ناهمگنی، چیزی گزارش نکردند. با این حال، جینگی و هال (۲۰۰۴) شاخص‌های ناهمگنی هاسکینگ و والیس (۱۹۹۷) را برای ارزیابی همگنی مناطق مشخص شده به وسیله‌ی SOFM یک بعدی از ۸۶ ایستگاه هیدرومتری در استان‌های جیانگزی و فوجیان چین استفاده کردند. در حالت یاد گیرنده، SOFM بردارهای مشخصه را که نمایش دهنده خصوصیات حوضه‌ی آبریز هستند به گره‌های خروجی مختلف اختصاص می‌دهد. اگر گروه‌بندی‌های طبیعی خوش تعریف به صورت ذاتی در مجموعه‌ی داده‌ها وجود داشته باشند، بردارهای مشخصه حول گره‌های خروجی که کاملاً از یکدیگر جدا شده‌اند، مجتمع می‌شوند. این امر نشان می‌دهد که SOFM‌ها تعداد بهینه‌ی خوش‌های را به صورت خودکار تشخیص می‌دهند. این ویژگی SOFM‌ها یک مزیت نسبت به روش‌های خوش‌بندی سخت و فازی موردنظر بحث در بخش‌های قبل که تنها قادر به تقسیم‌بندی مجموعه‌ی داده‌های موجود در تعداد مشخصی از خوش‌های هستند، است (رأو و سرینیواس، ۲۰۰۸).

اگرچه در غیاب الگوهای قابل تمییز آشکار در داده‌های موجود، تفسیر خوش‌های از خروجی SOFM، صرف نظر از اندازه و ابعاد آن، به ندرت امکان پذیر است، با این حال، در چنین موقعیت‌هایی، SOFM‌ها می‌توانند به عنوان روشی مفید در میان الگوریتم‌های خوش‌بندی منظور شوند.

^۱ Jingyi Z.

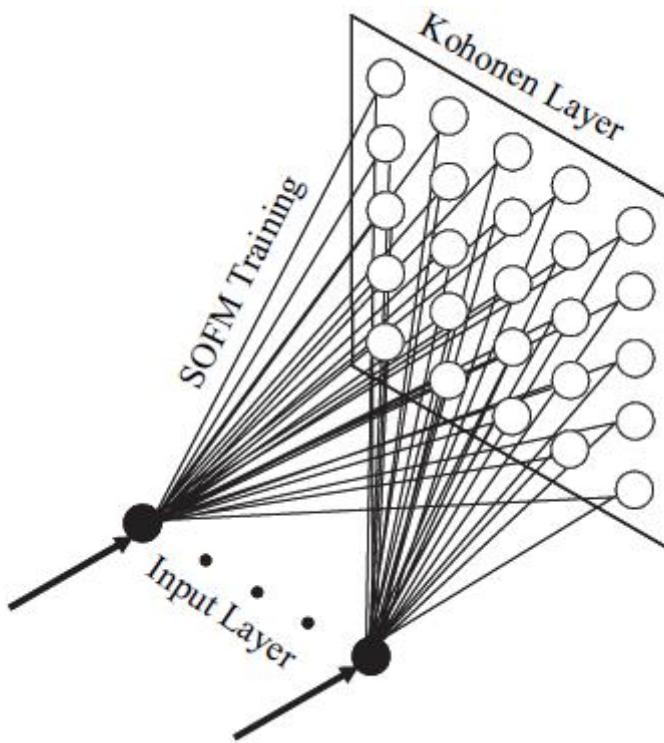
۳-۲-۳- الگوریتم نگاشت مشخصه‌ی خود سازمانده کوهونن

SOFM (کوهونن، ۱۹۸۲) یکی از پرکاربردترین شبکه‌های عصبی مصنوعی برای تشخیص ساختار توپولوژیک در داده‌های مورد بررسی است. این نوع شبکه در زمینه‌های تشخیص الگو، مدل‌سازی بیولوژیک، فشرده‌سازی داده‌ها، پردازش سیگنال و داده‌کاوی کاربردهایی پیدا کرده است (کوهونن، ۱۹۹۷). یک طرحواره برای SOFM در شکل (۲-۳) ارائه شده است. SOFM دارای یک لایه‌ی ورودی و یک لایه‌ی خروجی است که هر یک شامل تعدادی گره است. تعداد گره‌ها در لایه‌ی ورودی مساوی تعداد ویژگی‌های در نظر گرفته شده برای منطقه‌بندی است. هر گره در لایه‌ی ورودی به وسیله‌ی اتصالات سیناپتیک به تمام گره‌ها لایه‌ی خروجی متصل می‌شود. به همراه هر اتصال، یک توان یا وزن اتصال وجود دارد.

اگر y_k به عنوان k امین بردار مشخصه در فضای n بعدی با محورهای مختصاتی به نام‌های y_{ik} و $y_k = [y_{1k}, \dots, y_{nk}] \in R^n$ ، که y_{ik} معرف مقدار ویژگی i در k امین بردار مشخصه‌ی y_k است، در نظر گرفته شود، آنگاه ویژگی‌های بردار مشخصه‌ی y_k به صورت رابطه‌ی (۲-۳) بعدی y_k است، در نظر گرفته شود، آنگاه ویژگی‌های بردار مشخصه‌ی y_k به صورت رابطه‌ی (۲۸) تجدید مقیاس می‌شوند

$$x_{ik} = \frac{w_i}{\sigma_i} [f(y_{ik})] \quad \text{for } 1 \leq i \leq n, 1 \leq k \leq N \quad (28-3)$$

که در آن x_{ik} نشان‌دهنده‌ی مقدار تجدید مقیاس شده‌ی y_{ik} ، w_i وزن مربوط به ویژگی i مبتنی بر اهمیت نسبی آن است؛ σ_i معرف انحراف معیار ویژگی i است؛ $f(\cdot)$ تابع تبدیل تعیین شده را ارائه می‌کند؛ و N تعداد بردارهای مشخصه‌ی n بعدی را نشان می‌دهد. تجدید مقیاس ویژگی‌ها به دلیل تفاوت‌های موجود در واریانس، اندازه و اهمیت نسبی ضروری است.



شکل (۲-۳) طرح نمادین نگاشت مشخصه‌ی خود سازمانده کوهونن. لایه‌ی ورودی شامل ویژگی‌های تجدید مقیاس شده‌ی حوضه‌ی آبریز است. (راؤ و سرینیواس، ۲۰۰۸)

به طور کلی، k -امین بردار مشخصه‌ی تجدید مقیاس شده به صورت $x_k = [x_{1k}, \dots, x_{nk}]$ است. مجموعه‌ی N بردار مشخصه‌ی تجدید مقیاس شده می‌تواند به صورت ماتریس $N \times n$ داده‌های X نشان داده شود.

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1N} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nN} \end{bmatrix} \quad (29-3)$$

تعداد گره‌ها در لایه‌ی ورودی SOFM برابر ابعاد بردار مشخصه، یعنی n است. لایه‌ی خروجی، که لایه‌ی رقابتی یا کوهونن نیز خوانده می‌شود، دارای m گره است که در یک شبکه که معمولاً یک یا دو بعدی است، سازماندهی شده است (راؤ و سرینیواس، ۲۰۰۸). مقدار m می‌تواند به صورت حداقل تعداد مورد نظر برای تشکیل خوشها انتخاب شود (فاست، ۱۹۹۴). در زمینه‌ی منطقه‌بندی در هیدرولوژی، مقدار m عموماً به گونه‌ای انتخاب می‌شود که بزرگ‌تر از تعداد خوشها مورد انتظار

باشد. برای یک SOFM یک بعدی، هال و مینز (۱۹۹۹) مقدار m را برابر حداقل $2C_{Exp}$ انتخاب

کردند، در حالی که هال و همکاران (۲۰۰۲) مقدار آن را حداقل برابر $3C_{Exp}$ در نظر گرفتند.

الگوریتم SOFM به صورت زیر است:

(یک) وزن ابتدایی $\{w_{ij}; i = 1, \dots, n; j = 1, \dots, m\}$ اتصالات از n گره ورودی به m گره خروجی را به صورت تصادفی تعیین می‌شود. این وزن‌های تصادفی عموماً از محدوده‌ی همسانی

$w_j = \{w_{1j}, w_{2j}, \dots, w_{nj}\}$ از مقادیر به صورت مؤلفه‌های بردارهای ورودی انتخاب می‌شوند.

نشان دهنده‌ی بردار وزن بین گره خروجی j و گره‌های لایه‌ی ورودی در نظر گرفته می‌شود.

تکرار عددی را به صورت $t = 0$ تنظیم می‌گردد.

(دو) X با \hat{X} جایگزین می‌شود.

(سه) یک بردار ورودی x_k از \hat{X} (تصادفی بدون جایگزینی) ترسیم شده و فاصله‌ی آن از w_j با

استفاده از معیار اقلیدسی محاسبه می‌گردد. گرهی خروجی برنده ω را از طریق رابطه‌ی (۳۰-۳)

یافت می‌شود.

$$\omega = \arg \min_j \|x_k(t) - w_j(t)\| \quad j = 1, 2, \dots, m \quad (30-3)$$

(چهار) بردارهای وزن با استفاده از رابطه‌ی (۳۱-۳) به هنگام می‌شوند

$$w_j(t+1) = w_j(t) + \eta(t) h_{j,\omega}(t) [x_k(t) - w_j(t)] \quad (31-3)$$

که $\eta(t)$ پارامتر میزان یادگیرندگی برای تکرار عددی t است، و $h_{j,\omega}(t)$ یک تابع همسایگی

نامیده می‌شود. $\eta(t)$ برای کاهش یکنواخت با افزایش مقدار t به صورت رابطه‌ی (۳۲-۳) انتخاب

می‌شود

$$\eta(t) = \eta(0) \exp\left(-\frac{t}{\tau_1}\right) \quad (32-3)$$

که در آن $\eta(0)$ به صورتی انتخاب می‌شود که مقداری نزدیک به ۱، t را اختیار کند، τ_1 یک مقدار ثابت است که نوعاً برابر تعداد بیشینه تکرار عددی t_{max} تنظیم می‌شود (برای مثال ۱۰۰۰).

در رابطه‌ی (۳۱-۳)تابع همسایگی $h_{j,\omega}(t)$ که حول گره برنده ω متumerکز می‌شود، توسط رابطه‌ی (۳۳-۳) ارائه می‌شود

$$h_{j,\omega}(t) = \exp\left(-\frac{d_{\omega,j}^2}{2\sigma^2(t)}\right) \quad (33-3)$$

که در آن $d_{\omega,j}$ فاصله‌ی بین گره برنده‌ی ω و گره j همسایه‌ی آن در لایه‌ی خروجی است،

$$d_{\omega,j} = \|r_\omega - r_j\| \quad (34-3)$$

که در آن بردار گسسته‌ی r_ω موقعیت گره برنده‌ی ω و بردار گسسته‌ی r_j موقعیت گره j همسایه‌ی آن را تعریف می‌کند، که هر دو آن‌ها در فضای خروجی گسسته اندازه‌گیری می‌شوند. در رابطه‌ی (۳۳-۳)، پارامتر $\sigma(t)$ عرض مؤثر همسایگی توپولوژیک $h_{j,\omega}(t)$ در گام زمانی t است

$$\sigma(t) = \sigma(0) \exp\left(-\frac{t}{\tau_2}\right) \quad (35-3)$$

که برابر شعاع شبکه در لایه‌ی خروجی SOFM (شکل (۳-۳)) است؛ یک مقدار ثابت است که به وسیله‌ی رابطه‌ی (۳۶-۳) برآورد می‌شود.

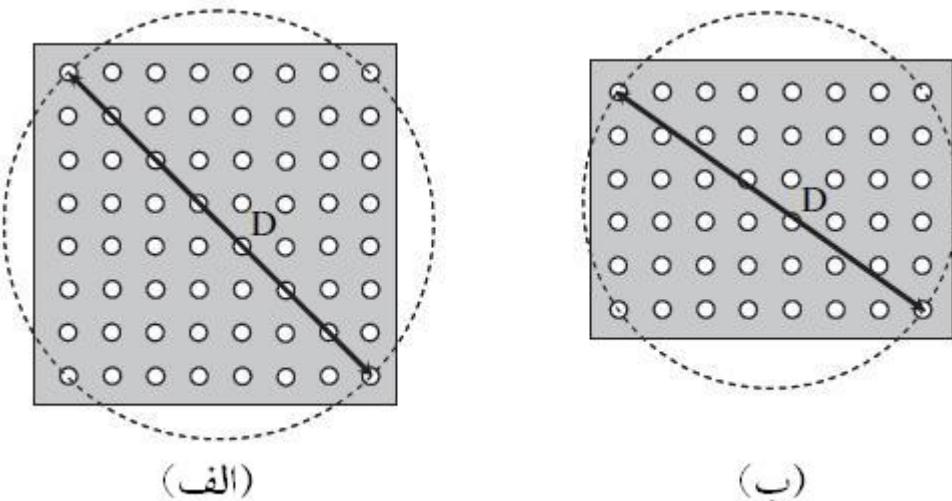
$$\tau_2 = \frac{t_{max}}{\ln \sigma(0)} \quad (36-3)$$

از روابط (۳۵-۳) و (۳۶-۳) برای $t = t_{max}$ ، $\sigma(t) = \sigma(0)$ است در حالی که برای $t = 0$ ، $\sigma(t) = 1$ است. به عبارت دیگر، در $t = 0$ تقریباً تمام گره‌ها در لایه‌ی خروجی متمرکز روی گره برنده به هنگام می‌شوند، در حالی که در پایان تکرارهای عددی، تنها تعداد اندکی گره‌های همسایه حول یک گره برنده به هنگام می‌شوند. افرون بر این، از رابطه‌ی (۳۳-۳) می‌توان نتیجه گرفت که $h_{j,\omega}$ با افزایش مقدار ω به صورت نمایی کاهش می‌یابد.

(پنج) اگر \hat{X} خالی است به گام (۴) رفته، در غیر این صورت به گام (۳) مراجعه می‌شود.

(شش) اگر $t \geq t_{max}$ باشد، گام (۷) آغاز می‌شود، در غیر این صورت t به $t + 1$ افزایش یافته و گام (۲) تا زمانی که تغییر قابل توجهی در نگاشت مشخصه دیده نشود، ادامه می‌یابد.

(هفت) به هر بردار ورودی x_k نام گره خروجی برنده‌ی j ($j = 1, \dots, m$) آن با استفاده از رابطه‌ی (۳۱-۳) اختصاص می‌یابد. تعداد m گره خروجی برنده $m \leq m$ به عنوان پیش الگو شناخته می‌شوند. توصیف دقیق الگوریتم SOFM در پژوهش هایکین^۱ (۲۰۰۳) موجود است.



شکل (۳-۳) طرح نمادین قطر (D) شبکه‌ی کوهونن SOFM؛ (الف) شبکه‌ی مربعی، (ب) شبکه‌ی مستطیلی. همسایگی نورون برنده در تکرار عددی $t = 0$ برای SOFM به عنوان شعاع شبکه انتخاب می‌شود که برابر $D/2$ است. (رائو و سرینیواس، ۲۰۰۸)

^۱ Haykin S.

۳-۴-۳- تحلیل فراوانی منطقه‌ای سیلاب با استفاده از تئوری گشتاورهای خطی

۳-۱-۴- معرفی و مروری بر مطالعات مربوط به گشتاورهای خطی

تحلیل فراوانی سیلاب برآورده است از این که یک سیلاب با بزرگی مشخص به طور متوسط هر چند وقت یک بار، اتفاق می‌افتد. برآورد فراوانی رویدادهای حدی اغلب از اهمیت زیادی برخوردار است. برای غلبه بر مشکل کمبود داده‌ها، روش‌های متعددی ابداع شده‌اند که از منابع جایگزین یا اضافه‌ی داده‌ها استفاده می‌کنند. یکی از این روش‌ها تحلیل فراوانی منطقه‌ای است. یک تحلیل منطقه‌ای مناسب به برآوردهایی از چندک منتهی می‌شود که به قدر کافی برای استفاده در بسیاری از کاربردهای عملی دقیق هستند.

یکی از ساده‌ترین روش‌های تحلیل فراوانی منطقه‌ای که مدت زمان زیادی است از آن استفاده می‌شود، روش سیلاب نمایه می‌باشد. دالریمپل^۱ (۱۹۶۰) برای اولین بار روش سیلاب نمایه در تحلیل فراوانی منطقه‌ای را معرفی کرد.

فرضیه‌ی کلیدی این روش این است که توزیع سیلاب‌ها در مکان‌های مختلف یک منطقه مشابه هستند، به جز برای پارامتر مقیاس یا سیلاب نمایه که انعکاس دهنده‌ی خواص بارندگی و رواناب در هر منطقه می‌باشند. سیلاب نمایه می‌تواند سیل میانگین باشد، اگرچه هر پارامتر محل از توزیع فراوانی می‌تواند مورد استفاده قرار گیرد.

داده‌های شدت جریان حداکثر سالانه، اگرچه معمولاً به وسیله‌ی توزیع گامبل مدل می‌شوند، اما اغلب چولگی بالاتری نسبت به مقدار متناسب با توزیع داشتند. این موضوع در اوایل دهه ۱۹۷۰ میلادی به شدت توجه هیدرولوژیست‌ها را به خود معطوف کرد. آماره‌های گشتاوری به صورت گسترده‌ای به عنوان مبنایی برای تشخیص و برآش توزیع‌های فراوانی به کار گرفته شدند، اما برای استفاده‌ی مؤثر از آن‌ها، به دانستن ویژگی‌های نمونه‌گیری آن‌ها در نمونه‌های کوچک نیاز بود. حجم

^۱ Dalrymple T.

زیادی تلاش‌های محاسباتی با استفاده از داده‌های شبیه سازی شده توسط والیس و همکاران (۱۹۷۴) انجام گرفت. این تلاش‌ها برخی از ویژگی‌های نامطلوب آماره‌های گشتاوری مانند اریب بالا و محدودیت جبری را آشکار کرد. دشواری برآورد به وسیله‌ی روش‌های رسمی مانند حداکثر درستنمایی یا روش گشتاورها، ابداع توزیع ویکبی^۱ و تمایل برای به دست آوردن برآوردهایی از پارامترهای ویکبی با نزدیک‌ترین حالت به آن‌ها، گرینوود^۲ و همکاران (۱۹۷۹) را به سوی ابداع گشتاورهای وزنی احتمال رهنمون کرد. گشتاورهای وزنی احتمال با عملکرد مناسب برای سایر توزیع‌ها یافت شدند، اما تفسیر آن‌ها دشوار بود (لندور^۳ و همکاران، ۱۹۷۹؛ هاسکینگ و والیس، ۱۹۸۷). استفاده‌ی نوین از روش سیلاب نمایه توسط والیس (۱۹۸۱، ۱۹۸۲) زمانی آغاز شد که او این روش را در ترکیب با گشتاورهای وزنی احتمال و توزیع ویکبی به عنوان روشی برای برآورد چندک‌ها در امتداد کران بالایی توزیع فراوانی به کار گرفت. هاسکینگ (۱۹۹۰) ترکیبات خطی معینی از گشتاورهای وزنی احتمال یافت. او آن‌ها را «گشتاورهای خطی» نامید، که به عنوان شاخص‌هایی از توزیع‌های فراوانی قابل تفسیر بودند و بدین ترتیب مبنایی برای یک تئوری قابل درک توصیف، تشخیص و برآورد توزیع‌ها شکل گرفت.

در نهایت هاسکینگ و والیس (۱۹۹۷) روشی را برای تحلیل فراوانی منطقه‌ای بر اساس گشتاورهای خطی ارائه کردند. در این روش شاخص‌هایی برای بررسی ناجوری آماری سایت‌ها، همگنی مناطق تشکیل شده و نکویی برازش و انتخاب توزیع‌های منطقه‌ای معرفی شدند.

۳-۴-۲- روش سیلاب نمایه دالریمپل

مراحل روش سیلاب نمایه دالریمپل که به صورت سنتی در آن از گشتاورهای معمولی استفاده می‌شود، به صورت خلاصه عبارتند از انتخاب دوره‌ی مشترک آماری، بازسازی آمار ناقص، تحلیل فراوانی

¹ Wakeby distribution

² Greenwood J. A.

³ Landwehr J. M.

در هر سایت با استفاده از عامل فراوانی، آزمون همگنی، تعیین رابطه‌ی دبی متوسط سیلاب با سایر ویژگی‌های حوضه، محاسبه‌ی دبی‌های بدون بعد و تعیین دبی سیلاب در سایت فاقد آمار احتمالی (مهدوی، ۱۳۸۶).

۳-۴-۳- گشتاورهای خطی توزیع‌های احتمال

برای یک متغیر تصادفی X با تابع چندک (u) ، موارد ویژه‌ی مفید گشتاورهای وزنی احتمال مرتبه r به صورت زیر تعریف می‌شوند.

$$\alpha_r = \int_0^1 x(u)(1-u)^r du , \beta_r = \int_0^1 x(u)u^r du \quad (37-3)$$

گشتاورهای خطی X و نسبت‌های مربوط به آن‌ها که توسط هاسکینگ (۱۹۹۰) تعریف شده‌اند را بر حسب گشتاورهای وزنی احتمال می‌توان به صورت

$$\begin{aligned} \lambda_1 &= \alpha_0 = \beta_0 \\ \lambda_2 &= \alpha_0 - 2\alpha_1 = 2\beta_1 - \beta_0 \\ \lambda_3 &= \alpha_0 - 6\alpha_1 + 6\alpha_2 = 6\beta_2 - 6\beta_1 + \beta_0 \\ \lambda_4 &= \alpha_0 - 12\alpha_1 + 30\alpha_2 - 20\alpha_3 = 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0 \end{aligned} \quad (38-3)$$

۹

$$\tau_r = \lambda_r / \lambda_2 , \quad r = 3,4, \dots \quad (39-3)$$

تعریف کرد که τ_3 را چولگی خطی^۱ و τ_4 را کشیدگی خطی^۲ می‌نامند. ضریب تغییرات خطی^۳ هم به صورت زیر تعریف می‌شود.

^۱ L-skewness

^۲ L-kurtosis

^۳ L-CV

$$\tau = \lambda_2 / \lambda_1 \quad (40-3)$$

برای یک نمونه‌ی n عضوی، گشتاورهای خطی به صورت

$$\begin{aligned} l_1 &= b_0 \\ l_2 &= 2b_1 - b_0 \\ l_3 &= 6b_2 - 6b_1 + b_0 \\ l_4 &= 20b_3 - 30b_2 + 12b_1 - b_0 \end{aligned} \quad (41-3)$$

تعريف می‌شوند که در آن‌ها b_r یک برآوردگر ناریب β_r است.

$$b_r = n^{-1} \binom{n-1}{r} \sum_{j=r+1}^n \binom{j-1}{r} X_{j:n} \quad (42-3)$$

همچنین نسبت‌های گشتاور خطی نمونه (برآوردگر طبیعی τ_r) به صورت

$$t_r = l_r / l_2 \quad (43-3)$$

و $L-CV$ نمونه (برآوردگر طبیعی τ) به شکل

$$t = l_2 / l_1 \quad (44-3)$$

تعريف می‌شوند (هاسکینگ و والیس، ۱۹۹۷).

۳-۴-۴-۴- برآورد پارامترها با استفاده از گشتاورهای خطی

هاسکینگ و والیس (۱۹۸۷) دریافتند که برای نمونه‌های کوچک و متوسط، روش گشتاورهای خطی

به برآوردهایی کارآمد و از نظر محاسباتی ساده از پارامترها و چندکها منتج می‌شود که اغلب کارآمدتر از روش حداکثر درستنمایی است و انتظار می‌رود این ویژگی‌ها در روش‌های سیلاب نمایه برای تحلیل فراوانی منطقه‌ای ادامه یابند.

۴-۳-۵- استفاده از گشتاورهای خطی در تحلیل فراوانی منطقه‌ای

بهره‌گیری از گشتاورهای خطی در تحلیل فراوانی منطقه‌ای با استفاده از روش سیلاب نمایه در ادامه چگونگی به کارگیری گشتاورهای خطی در استفاده از روش سیلاب نمایه تشریح شده است.

۴-۳-۵-۱- غربال کردن داده‌ها و شاخص ناجوری D

در این گام هدف تشخیص سایت‌هایی است که به صورت فاحشی با گروهی از سایت‌ها به عنوان یک مجموعه، ناجور هستند.

در یک مجموعه با N سایت، D_i به عنوان شاخص ناجوری سایت i مطابق رابطه (۴۵-۳) تعریف می‌شود،

$$D_i = \frac{1}{3} N(u_i - \bar{u})^T A^{-1}(u_i - \bar{u}) \quad (45-3)$$

که در آن

$$u_i = [t^{(i)} \quad t_3^{(i)} \quad t_4^{(i)}]^T, \quad \bar{u} = N^{-1} \sum_{i=1}^N u_i \quad (46-3)$$

$$A = \sum_{i=1}^N (u_i - \bar{u})(u_i - \bar{u})^T \quad (47-3)$$

است. چنانچه مقدار D_i یک سایت از مقدار بحرانی ارائه شده در جدول (۲-۳) تجاوز کند آن را ناجور در نظر می‌گیریم.

برای شاخص ناجوری D_i دو کاربرد را می‌توان در نظر گرفت. نخست در ابتدای تحلیل که شاخص ناجوری می‌تواند برای گروه بزرگی از ایستگاهها به کار رود که تمامی آن‌ها در چند منطقه‌ی بزرگ جغرافیایی قرار گرفته‌اند. مقصود این است که ایستگاه‌هایی که در داده‌های ایشان دارای خطاهای فاحش هستند از داده‌های دیگر جدا شده و به عنوان ایستگاه‌های ناجور مشخص می‌شوند. از این رو ایستگاه‌هایی که در این مرحله به عنوان ناجور مشخص شده‌اند، باید به دقت برای یافتن خطاهای روی داده در ثبت و نسخه‌برداری از داده‌ها یا منابع عدم اطمینان در داده‌ها مانند یک ایستگاه اندازه‌گیری که جایه‌جا شده است، یا تغییراتی که در طول زمان توسط انسان به توزیع فراوانی ایستگاه تحمیل شده‌اند، مورد بررسی قرار گیرند (هاسکینگ و والیس، ۱۹۹۷).

جدول (۲-۳) مقادیر بحرانی شاخص ناجوری D_i

مقدار بحرانی D_i	تعداد سایتها در منطقه
۱/۳۳۳	۵
۱/۶۴۸	۶
۱/۹۱۷	۷
۲/۱۴۰	۸
۲/۳۲۹	۹
۲/۴۹۱	۱۰
۲/۶۳۲	۱۱
۲/۷۵۷	۱۲
۲/۸۶۹	۱۳
۲/۹۷۱	۱۴
۳	≥ 15

پس از آن طی فرآیند تحلیل، هنگامی که مناطق همگن حداقل به صورت نسبی مشخص شدند، شاخص ناجوری می‌تواند برای هر ایستگاه در یک منطقه‌ی مورد نظر، محاسبه شود. چنانچه ایستگاهی در مقایسه با سایر ایستگاه‌های منطقه به عنوان یک کل واحد ناجور باشد، امکان انتقال آن ایستگاه به مناطق دیگر بایستی مورد تأمل قرار گیرد. با این حال این مسئله باید این نکته را مد نظر قرار داد که گشتاورهای خطی یک سایت ممکن است تنها بر حسب تصادف با سایر ایستگاه‌هایی که از نظر فیزیکی با آن مشابه هستند، متفاوت باشد. برای مثال، یک پدیده‌ی هواشناسی حدی اما محدود از نظر مکانی، ممکن است تنها بر روی تعداد کمی از ایستگاه‌ها در یک منطقه اثر بگذارد. در صورتی که چنین پدیده‌ای تقریباً به طور مساوی احتمال دارد که در آینده روی هر یک از ایستگاه‌ها اثر بگذارد، آنگاه رفتار با کل گروه ایستگاه‌ها به عنوان یک منطقه‌ی همگن، با وجود این که برخی ایستگاه‌ها ممکن است نسبت به منطقه‌ی کلی ناجور باشند، صحیح خواهد بود (هاسکینگ و والیس، ۱۹۹۷).

۴-۳-۵-۲- تشخیص همگنی مناطق و استفاده از شاخص ناهمگنی H

هدف تشکیل گروه‌هایی از سایت‌هایی است که به طور تقریبی شرایط همگنی را ارضا کنند. شاخص ناهمگنی تغییرات بین سایتی در گشتاورهای خطی نمونه برای گروه سایت‌ها را با آنچه که از یک منطقه‌ی همگن انتظار می‌رود، مقایسه می‌کند. برای ارزیابی همگنی مناطق از شاخص ناهمگنی H استفاده می‌شود که در روابط (۴۸-۳) معرفی شده است. توصیه می‌شود که اگر $1 < H$ باشد، منطقه به عنوان «دارای همگنی قابل قبول»، اگر $2 \leq H < 1$ باشد به عنوان «ناهمگن احتمالی» و اگر $2 \geq H$ باشد به عنوان «ناهمگن قطعی» در نظر گرفته شود.

برای منطقه‌ی پیشنهادی با N سایت که سایت i آن دارای طول رکورد n_i و نسبت‌های گشتاور خطی $t^{(i)}$ ، $t_3^{(i)}$ و $t_4^{(i)}$ است، $L-kurtosis$ ، $L-CV$ و $L-skewness$ میانگین وزنی منطقه‌ای به ترتیب با t^R ، t_3^R و t_4^R نشان داده می‌شوند (هاسکینگ و والیس، ۱۹۹۷).

با برآش یک توزیع کاپا^۱ بر نسبت‌های گشتاور خطی میانگین منطقه‌ای ۱، t^R ، t_3^R و t_4^R ، به یک مقدار بزرگ N_{sim} منطقه‌ی همگن فاقد همبستگی متوالی یا متقابل با N سایت که هریک دارای توزیع کاپا به عنوان توزیع فراوانی خود است، شبیه سازی می‌شود. سایتها طول آمار یکسانی با همتایان واقعی خود دارند. برای هر منطقه‌ی شبیه سازی شده، V محاسبه می‌شود. میانگین و انحراف معیار N_{sim} مقدار V محاسبه شده و به ترتیب μ_V و σ_V نامیده می‌شوند. سپس شاخص ناهمگنی H را محاسبه می‌شود (هاسکینگ و والیس، ۱۹۹۷).

$$\begin{aligned} t^R &= \sum_{i=1}^N n_i t^{(i)} \Bigg/ \sum_{i=1}^N n_i \\ t_r^R &= \sum_{i=1}^N n_i t_r^{(i)} \Bigg/ \sum_{i=1}^N n_i, \quad r = 3, 4, \dots \\ V &= \left\{ \sum_{i=1}^N n_i (t^{(i)} - t^R)^2 \Bigg/ \sum_{i=1}^N n_i \right\}^{1/2} \\ H &= \frac{(V - \mu_V)}{\sigma_V} \end{aligned} \tag{۴۸-۳}$$

ایجاد شاخص‌های ناهمگنی که در آنها V موجود در معادله (۴۹-۳) با شاخص‌های دیگری از پراکندگی بین سایتی گشتاورهای خطی نمونه جایگزین شده است، امکان پذیر است، یک شاخص بر مبنای $L-CV$ و چولگی خطی

$$V_2 = \sum_{i=1}^N n_i \left\{ (t^{(i)} - t^R)^2 + (t_3^{(i)} - t_3^R)^2 \right\}^{1/2} \Bigg/ \sum_{i=1}^N n_i \tag{۴۹-۳}$$

و شاخصی دیگر بر مبنای چولگی خطی و کشیدگی خطی (هاسکینگ و والیس، ۱۹۹۷).

^۱ Kappa distribution

$$V_3 = \sum_{i=1}^N n_i \left\{ \left(t_3^{(i)} - t_3^R \right)^2 + \left(t_4^{(i)} - t_4^R \right)^2 \right\}^{1/2} \Bigg/ \sum_{i=1}^N n_i \quad (50-3)$$

V_3 و V_2 فاصله‌ی متوسط وزنی از سایت تا میانگین وزنی گروه، به ترتیب روی گراف‌های t در مقابله با t_3 و t_4 در برابر t_3 هستند. هم برای داده‌های واقعی و هم برای مناطق شبیه سازی شده مصنوعی، آماره‌های H مبتنی بر V_2 و V_3 قدرت کمتری برای تشخیص تمایز بین مناطق همگن و ناهمگن دارند. آن‌ها حتی برای مناطق بسیار ناهمگن، به ندرت به مقادیر H بزرگ‌تر از ۲ منتهی می‌شوند. آماره‌ی H مبتنی بر V قدرت تشخیص تمایز بالاتری دارد. از این رو این طور قضاوت می‌شود که V به وضوح نسبت به V_2 و V_3 برای مقایسه‌ی بین سایتی نسبت‌های گشتاور خطی نمونه‌ی مورد نیاز برای روش سیلاپ نمایه، برتر است (هاسکینگ و والیس، ۱۹۹۷). به همین جهت در میان سه شاخص H_1 ، H_2 و H_3 که به ترتیب بر مبنای V ، V_2 و V_3 تعریف می‌شوند، بنا بر توصیه‌ی هاسکینگ و والیس (۱۹۹۷)، برای قضاوت در مورد وضعیت همگنی مناطق حاصل از تحلیل خوش‌های تأکید بیشتری نسبت به شاخص H_1 وجود دارد و دو شاخص دیگر از اهمیت کمتری برخوردار هستند.

۴-۳-۵-۳- انتخاب یک توزیع فراوانی منطقه‌ای و شاخص نکویی برازش Z^{DIST}

در تحلیل فراوانی منطقه‌ای یک توزیع فراوانی واحد که برآوردهای چندک دقیقی را برای هر سایت حاصل کند بر داده‌های به دست آمده از سایت‌های متعدد برازش داده می‌شود.

به منظور شناسایی چنین توزیعی می‌توان از یک شاخص نکویی برازش استفاده کرد. Z^{DIST} که در روابط (۳-۵) معرفی شده است شاخصی است که چنین کارکردی دارد. اگر $|Z^{DIST}| \leq 1.64$ می‌توانیم برازش را مناسب بدانیم. هنگامی که توزیع‌های متعددی به طور مناسب بر داده‌ها برازش می‌یابند، بهترین گزینه از میان آن‌ها توزیعی خواهد بود که نزدیک‌ترین مقدار Z به صفر را ارائه کند و بالاترین توانایی را در ارائه برآوردهای چندک خوب داشته باشد. چنانچه منطقه به صورت قابل قبولی همگن نباشد، استفاده از یک توزیع استوار نسبت به ناهمگنی معتل در توزیع‌های فراوانی

درون سایتی مانند توزیع‌های کاپا و ویکبی، می‌تواند مناسب باشد (هاسکینگ و والیس، ۱۹۹۷).

مجموعه‌ای از توزیع‌های سه پارامتری منطقی محتمل شامل لجستیک تعمیم یافته^۱، مقادیر حدی تعمیم یافته^۲، پارتوى تعمیم یافته^۳، نرمال تعمیم یافته^۴ و پیرسون تیپ III^۵ را بر نسبت‌های گشتاور خطی متوسط منطقه‌ای ۱، t_3^R و t_4^R برازش داده می‌شود. کشیدگی خطی توزیع برازش داده شده را با τ_4^{DIST} نشان داده می‌شوند که در آن $DIST$ می‌تواند GEV , GLO و... باشد.

برازش یک توزیع کاپا و عملیات شبیه سازی را درست مانند مرحله‌ی محاسبه‌ی شاخص ناهمگنی انجام می‌شود. برای m امین منطقه‌ی شبیه سازی شده، چولگی خطی متوسط منطقه‌ای $t_3^{[m]}$ و کشیدگی خطی متوسط منطقه‌ای $t_4^{[m]}$ محاسبه می‌شوند. در نهایت برای هر توزیع، شاخص نکویی برازش Z^{DIST} محاسبه می‌گردد (هاسکینگ و والیس، ۱۹۹۷).

$$\begin{aligned} Z^{DIST} &= (\tau_4^{DIST} - t_4^R + B_4) / \sigma_4 \\ B_4 &= N_{sim}^{-1} \sum_{m=1}^{N_{sim}} (t_4^{[m]} - t_4^R) \\ \sigma_4 &= \left[(N_{sim} - 1)^{-1} \left\{ \sum_{m=1}^{N_{sim}} (t_4^{[m]} - t_4^R)^2 - N_{sim} B_4^2 \right\} \right]^{1/2} \end{aligned} \quad (51-۳)$$

۴-۵-۴-۳- توزیع ویکبی

این توزیع یک توزیع پنج پارامتری با پارامترهای ξ , α , β , γ , δ است. $x(F)$ برای توزیع ویکبی به شکل زیر تعریف می‌شود.

$$x(F) = \xi + \frac{\alpha}{\beta} \{1 - (1 - F)^\beta\} - \frac{\gamma}{\delta} \{1 - (1 - F)^{-\delta}\} \quad (52-۳)$$

¹ Generalized Logistic (GLO)

² Generalized Extreme Values (GEV)

³ Generalized Pareto (GPA)

⁴ Generalized Normal (GNO)

⁵ Pierson type III (PE3)

پارامترهای توزیع ویکبی از اکثر توزیع‌های رایج بیشتر است، بنابراین نسبت به توزیع‌های رایج می‌تواند به دامنه‌ی وسیع‌تری از اشکال توزیعی دست یابد و برای مقادیر مناسب پارامترها، قادر است شکل‌های بسیاری از توزیع‌های چوله‌ی پرکاربرد را تقلید کند. هنگامی که $0 < \delta$ است، توزیع ویکبی یک امتداد بالایی طویل‌تر دارد و می‌تواند بر مجموعه‌های داده‌های شامل داده‌های پرت بالا انطباق یابد. توزیع ویکبی کران پایینی محدودی دارد که از نظر فیزیکی برای مشاهدات واقعی بسیار منطقی است. فرم صریح تابع چندک، شبیه سازی نمونه‌های تصادفی از توزیع ویکبی را ساده می‌کند. در تحلیل فراوانی منطقه‌ای چنانچه منطقه به صورت قابل قبولی همگن نباشد، استفاده از توزیع ویکبی، می‌تواند مناسب باشد (هاسکینگ و والیس، ۱۹۹۷).

۴-۵-۵- الگوریتم گشتاورهای خطی منطقه‌ای

برای منطقه‌ای دارای N سایت که سایت i دارای طول آمار n_i و داده‌های مشاهده شده‌ی Q_{ij} ، $j = 1, \dots, n_i$ است، $Q_i(F)$ را تابع چندک توزیع فراوانی در سایت i در نظر بگیرید. برای یک منطقه‌ی همگن داریم:

$$Q_i(F) = \mu_i q(F) \quad (53-3)$$

که در آن μ_i عامل مقیاس وابسته به سایت یا سیلاب نمایه است. $\hat{\mu}_i$ را برآورد عامل مقیاس در سایت i و داده‌های بازمقیاس بندی شده‌ی بدون بعد $q_{ij} = Q_{ij}/\hat{\mu}_i$ ، $j = 1, \dots, n_i$ نمایه را میانگین توزیع فراوانی در هر سایت در نظر بگیریم که به وسیله‌ی میانگین نمونه‌ی درون سایتی برای هر سایت برآورد می‌شود، آنگاه میانگین داده‌های بازمقیاس بندی شده برای هر سایت ۱ است و از این رو متوسط منطقه‌ای این میانگین‌ها برابر ۱ است. سایت i دارای میانگین نمونه‌ی $\bar{l}_i^{(i)}$ ، و نسبت‌های گشتاور خطی $t_4^{(i)}, t_3^{(i)}, t_2^{(i)}, \dots$ است. میانگین متوسط منطقه‌ای را برابر ۱ تنظیم می‌کنیم،

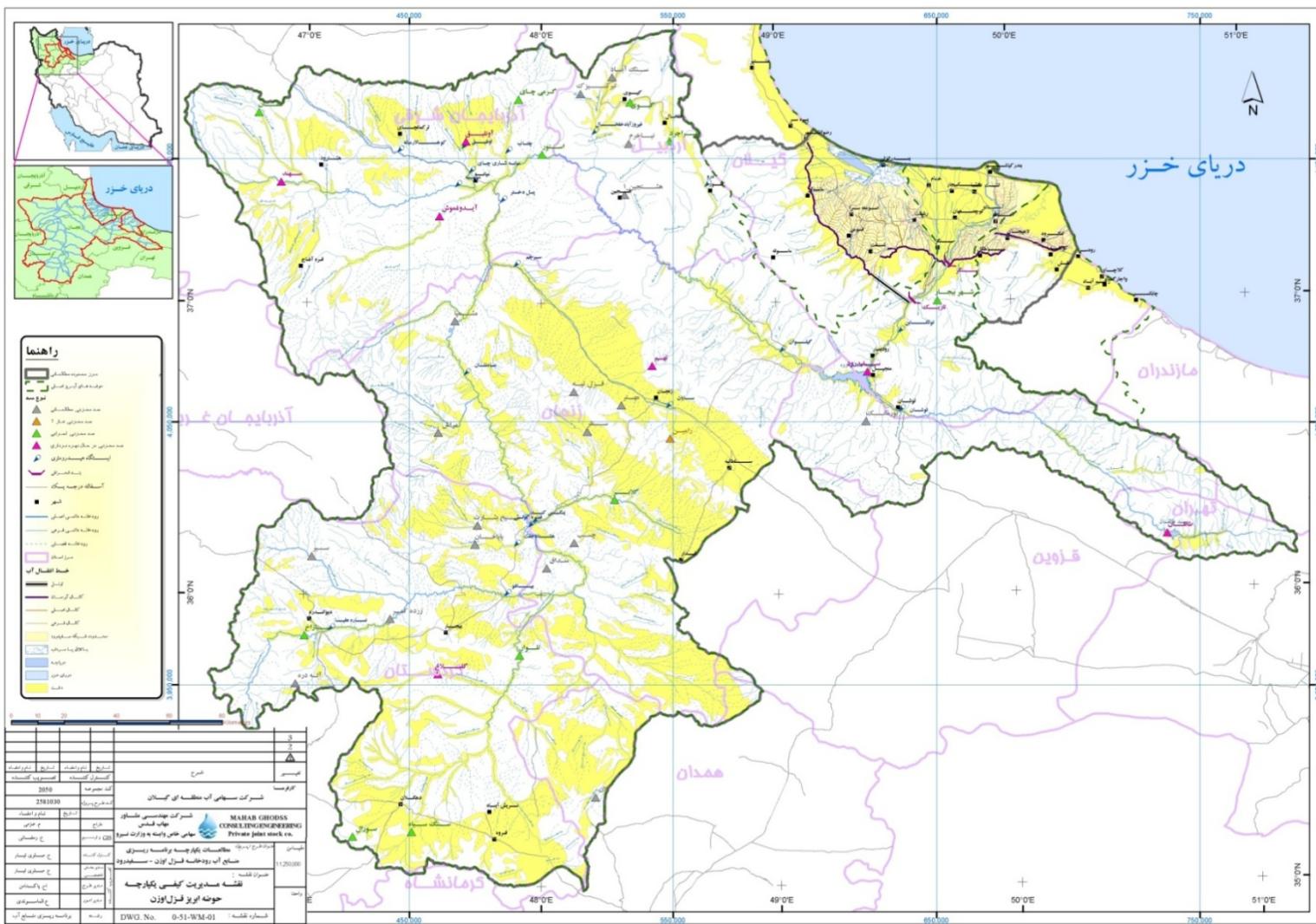
یعنی $l_1^R = 1$. توزیع را از طریق معادل قرار دادن نسبت‌های گشتاور خطی $\lambda_1, \tau_3, \tau_4, \dots$ ، با نسبت-های گشتاور خطی متوسط منطقه‌ای $t_1^R, t_3^R, t_4^R, \dots$ ، برازش می‌دهیم.تابع چندک توزیع فراوانی منطقه‌ای برازش داده شده را با (\hat{q}) نشان می‌دهیم که در واقع برآوردهای μ و q به دست می‌آیند. برآورد چندک با توزیع احتمال عدم تجاوز F به صورت زیر است (هاسکینگ و والیس، ۱۹۹۷).

$$\hat{Q}_i(F) = l_1^{(i)} \hat{q}(F) \quad (54-3)$$

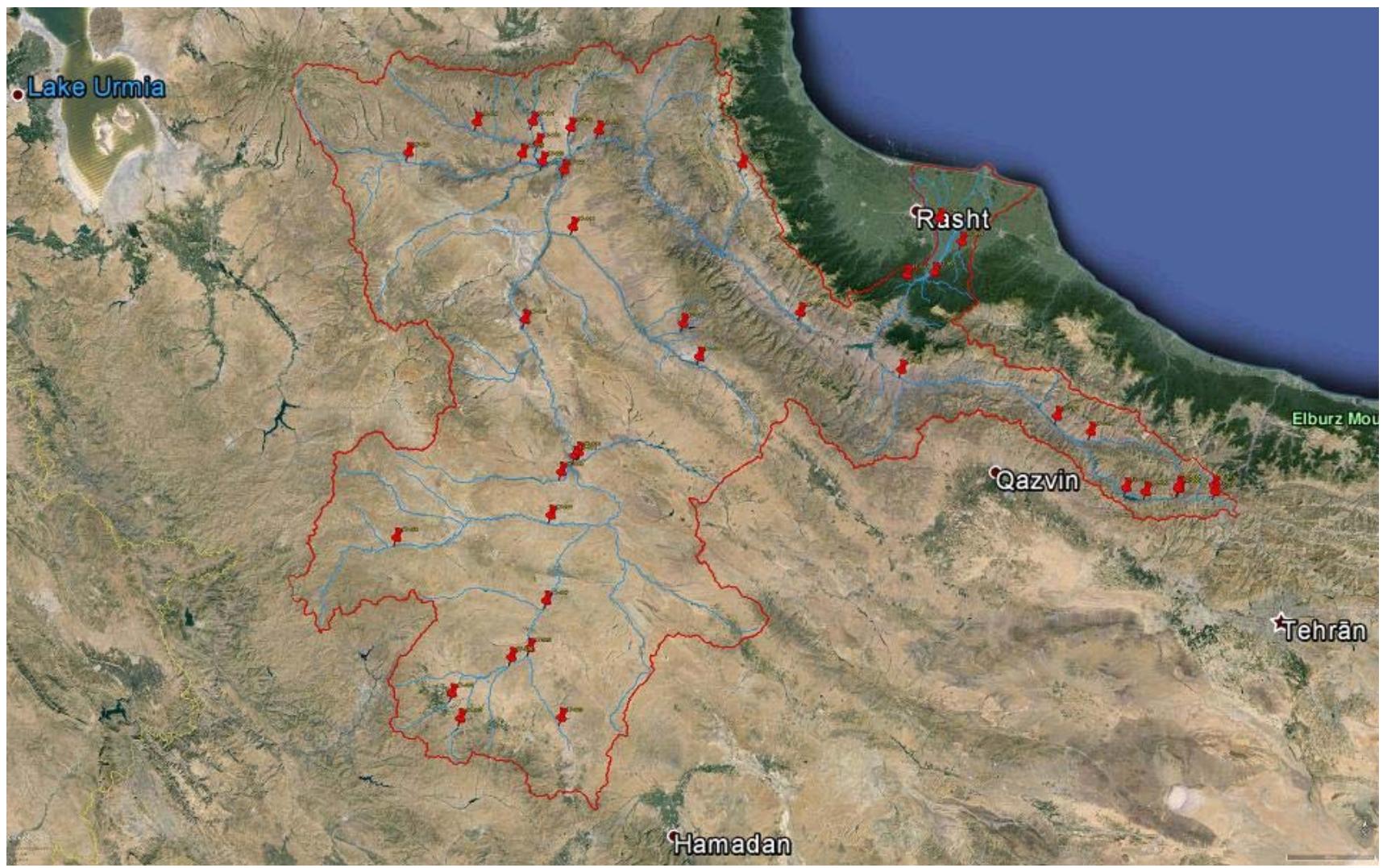
۳-۵- معرفی حوضه‌ی آبریز سفیدرود بزرگ

حوضه‌ی آبریز قزل اوزن یا سفیدرود بزرگ بعد از حوضه دریاچه‌ی نمک از نظر اراضی قابل آبیاری، بزرگ‌ترین حوضه‌ی کشور است. مساحت این حوضه که بزرگ‌ترین زیرحوضه‌ی درجه دو حوضه آبریز دریای خزر در ایران است، ۶۳۹۴۵ کیلومتر مربع می‌باشد و به خاطر وجود اقلیمهای متفاوت و منابع غنی آب و خاک از اهمیت خاصی برخوردار است.

این محدوده در محل تلاقی رشته کوههای البرز، زاگرس و مرکزی واقع شده و در تقسیم‌بندی‌های طرح جامع آب به عنوان حوضه‌ی سوم از منطقه‌ی اول مطالعاتی مشخص گردیده است. این حوضه‌ی آبریز از دو شاخه‌ی رودخانه‌ای اصلی به نامهای قزل اوزن و شاهروド تشکیل یافته است که در محل سد سفیدرود به هم می‌پیوندند و رودخانه‌ی سفیدرود را تشکیل می‌دهند. ذینفعان این حوضه شامل استان‌های آذربایجان شرقی، اردبیل، تهران، زنجان، قزوین، کردستان، گیلان و همدان می‌باشند. شکل (۴-۳) محدوده‌ی حوضه‌ی آبریز قزل اوزن را به همراه رودخانه‌های اصلی آن نمایش می‌دهد. جمعیت ساکن در محدوده‌ی این حوضه حدود ۲۰۶۳۵۸۵ نفر است. در سال ۱۳۹۰ تراکم نسبی جمعیت این حوضه، ۳۵ نفر در کیلومتر مربع بوده است.



شکل (۴-۳) نمایش محدوده جغرافیایی حوضه‌ی آبریز سفیدرود بزرگ (نقشه‌ی مدیریت کیفی یکپارچه حوضه‌ی آبریز قزل اوزن، شرکت مهندسی مشاور مهاب قدس)



شکل (۵-۳) نمایش موقعیت جغرافیایی ایستگاه‌های هیدرومتری مورد مطالعه در حوضه‌ی آبریز سفیدرود بزرگ

در وضع موجود شش سد در دست بهره‌برداری به نام‌های آیدوغموش، سهند، تهم، طالقان، گلبلاغ و سفیدرود در این حوضه وجود دارند و کل حجم آب مصرفی در حوضه‌ی آبریز در حدود ۴۹۰۰ میلیون متر مکعب در سال می‌باشد که از این مقدار در حدود ۳۷۰۰ میلیون متر مکعب از منابع سطحی و در حدود ۱۲۰۰ میلیون متر مکعب از منابع زیرزمینی (در فصول زراعی) تأمین می‌گردد. (شرکت سهامی آب منطقه‌ای گیلان، مطالعات یکپارچه‌ی برنامه‌ریزی منابع آب رودخانه قزل اوزن-سفیدرود، گزارش نیازهای آب شرب و صنعت محدوده‌ی مطالعاتی)

در تحقیق حاضر تعداد ۴۰ ایستگاه هیدرومتری از میان ایستگاه‌های هیدرومتری موجود در حوضه‌ی آبریز سفیدرود بزرگ به منظور استفاده در عملیات منطقه‌بندی و تحلیل فراوانی منطقه‌ای سیلاب مورد مطالعه قرار گرفته‌اند که مشخصات این ایستگاه‌ها در جدول (۳-۳) ارائه شده است. همچنین موقعیت جغرافیایی هر یک از ایستگاه‌های مورد مطالعه در شکل (۳-۵) نشان داده شده است.

برآورد سیلاب در تحلیل فراوانی منطقه‌ای تا زمانی قابل اعتماد است که تعداد سال‌های آماری موجود در ایستگاه‌های یک منطقه بزرگ‌تر یا مساوی پنج برابر دوره‌ی بازگشت مورد نظر جهت برآورد بزرگی سیلاب باشد (رید و همکاران، ۱۹۹۹). از این رو تعداد ایستگاه‌های موجود در هر منطقه و سال‌های آماری موجود برای هر یک از آن‌ها عاملی تعیین کننده در انتخاب تعداد خوشها است. در این تحقیق نتایج حاصل از اختصاص ایستگاه‌ها یا سایتها موجود در حوضه‌ی آبریز سفیدرود بزرگ به ۲ تا ۵ خوشه یا منطقه با استفاده از روش‌های مختلف تحلیل خوشها مورد ارزیابی قرار گرفته است.

۳-۶- داده‌های آماری سیلاب

در این مطالعه، برای اجرای محاسبات تحلیل فراوانی منطقه‌ای سیلاب با استفاده از الگوریتم گشتاور-

های خطی که پس از اجرای فرآیند منطقه‌بندی انجام می‌گیرد، از آمار دبی‌های حداکثر لحظه‌ای ایستگاه‌های هیدرومتری که توسط شرکت مدیریت منابع آب ایران ثبت و گردآوری شده است، استفاده شده است.

جدول (۳-۳) مشخصات جغرافیایی، فیزیوگرافیک و هواشناسی ایستگاه‌های هیدرومتری مورد مطالعه

کد ایستگاه	طول جغرافیایی (dd)	عرض جغرافیایی (dd)	سطح زهکشی (km ²)	بارندگی متوسط سالانه (mm)	ضریب رواناب	ارتفاع از سطح دریا (m)
۱۷-۰۰۱	۴۷/۸۵	۳۵/۶۷	۵۲۱۵	۲۷۸	۰/۳۱	۱۵۴۰
۱۷-۰۰۳	۴۷/۴۷	۳۵/۱۸	۲۴۷	۳۲۵	۰/۲۱	۱۸۳۰
۱۷-۰۰۷	۴۷/۸۵	۳۵/۶۷	۶۵۲۵	۲۶۰	۰/۱۵	۱۶۵۰
۱۷-۰۰۹	۴۷/۸۸	۳۶/۱۷	۱۸۲۵	۳۵۰	۰/۲۱	۱۴۵۰
۱۷-۰۱۱	۴۷/۹۵	۳۶/۲۲	۱۹۷۱۹	۳۵۰	۰/۲۰	۱۴۱۰
۱۷-۰۱۳	۴۷/۹۷	۳۶/۲۵	۲۴۸۴	۳۰۱	۰/۲۱	۱۴۲۰
۱۷-۰۱۵	۴۷/۶۷	۳۶/۷۵	۲۴۵۱۱	۳۷۵	۰/۱۱	۱۲۵۰
۱۷-۰۱۷	۴۸/۴۵	۳۶/۷۷	۱۷۲	۳۵۰	۰/۰۲	۱۷۰۰
۱۷-۰۱۹	۴۷/۸۸	۳۷/۱۲	۴۵۷۷	۳۱۰	۰/۱۰	۱۱۵۰
۱۷-۰۲۱	۴۷/۸۲	۳۷/۳۵	۳۳۰۰۴	۲۹۵	۰/۶۰	۱۰۸۰
۱۷-۰۲۳	۴۷/۷۲	۳۷/۳۸	۱۸۰۲	۲۸۵	۰/۳۳	۱۰۶۰
۱۷-۰۲۵	۴۷/۶۲	۳۷/۴۰	۳۵۲۶	۲۹۶	۰/۳۵	۱۱۰۰
۱۷-۰۲۷	۴۷/۳۸	۳۷/۰۲	۱۰۱۱	۳۳۶	۰/۳۰	۱۵۰۰
۱۷-۰۲۹	۴۷/۹۸	۳۷/۰۲	۴۲۰۷۲	۳۱۱	۰/۲۵	۱۰۰۰
۱۷-۰۳۰	۴۸/۵۲	۳۶/۶۳	۴۵۷۷	۳۰۰	۰/۳۱	۱۶۵۰
۱۷-۰۳۳	۴۹/۰۲	۳۶/۸۳	۴۹۳۰۰	۱۸۴	۰/۳۷	۳۲۰
۱۷-۰۳۵	۵۰/۷۳	۳۶/۱۷	۸۴۸	۴۵۶	*۱/۰۶	۱۷۷۰
۱۷-۰۳۷	۵۰/۶۲	۳۶/۱۸	۱۰۶۰	۴۷۸	۰/۸۰	۱۷۰۰
۱۷-۰۳۹	۵۰/۴۵	۳۶/۳۷	۶۹۵	۳۸۱	*۱/۲۱	۱۳۵۰
۱۷-۰۴۱	۴۹/۵۲	۳۶/۶۲	۴۸۸۲	۱۸۷	*۱/۱۳	۳۰۰
۱۷-۰۴۰	۴۹/۶۷	۳۷/۰۲	۲۳۳	۱۲۱۰	۰/۷۶	۴۰
۱۷-۰۴۴	۴۹/۸۰	۳۷/۱۳	۱۳۴	۱۴۰۰	۰/۹۲	۶۰
۱۷-۰۴۲	۴۷/۴۲	۳۵/۲۸	۲۳۸	۲۷۹	۰/۳۷	۱۸۲۰
۱۷-۰۴۵	۴۷/۶۵	۳۷/۰۳	۱۱۹	۳۰۰	۰/۳۸	۱۳۲۰
۱۷-۱۰۹	۴۷/۶۸	۳۷/۴۵	۲۲۲۰	۲۲۵	۰/۳۷	۱۱۰۰

ادامه‌ی جدول (۳-۳) مشخصات جغرافیایی، فیزیوگرافیک و هواشناسی ایستگاه‌های هیدرومتری مورد مطالعه

کد ایستگاه	طول جغرافیایی (dd)	عرض جغرافیایی (dd)	سطح زهکشی (km ²)	بارندگی متوسط سالانه (mm)	ضریب رواناب	ارتفاع از سطح دریا (m)
۱۷-۱۱۱	۴۹/۶۵	۳۷/۱۷	۸۲	۱۲۷۵	.۰/۶۰	۴۰
۱۷-۱۳۳	۴۸/۷۰	۳۷/۴۰	۱۵۳	۵۰۰	.۰/۳۱	۱۶۰۰
۱۷-۱۵۰	۴۷/۷۵	۳۵/۸۷	۱۵۶۸	۳۵۰	.۰/۰۸	۱۷۲۵
۱۷-۲۰۱	۵۰/۲۸	۳۶/۴۵	۲۴۴۵	۳۳۴	.۰/۴۷	۹۷۰
۱۷-۴۲۸	۴۷/۰۵	۳۷/۳۸	۱۴۶۹	۳۳۱	.۰/۰۳	۱۶۵۰
۱۷-۴۳۰	۴۷/۸۵	۳۷/۵۲	۶۷۰	۳۰۱	.۰/۳۵	۱۳۰۰
۱۷-۸۷۴	۵۱/۰۷	۳۶/۲۰	۴۱	۶۵۸	.۰/۹۳	۲۸۰۰
۱۷-۹۲۳	۴۷/۹۳	۳۵/۲۰	۱۰۵۰	۳۵۰	.۰/۰۳	۱۸۵۰
۱۷-۹۲۵	۴۷/۷۷	۳۵/۴۷	۲۰۱۳	۲۷۵	.۰/۰۶	۱۶۸۵
۱۷-۹۲۶	۴۷/۶۸	۳۵/۴۳	۲۴۶۵	۲۹۳	.۰/۲۲	۱۷۰۰
۱۷-۹۳۴	۵۰/۹۰	۳۶/۲۰	۶۴	۵۵۰	.۰/۲۷	۲۱۰۰
۱۷-۹۶۵	۵۰/۹۲	۳۶/۱۸	۹۸	۵۴۷	*۱/۳۲	۲۰۰۰
۱۷-۹۶۶	۵۰/۹۰	۳۶/۱۸	۴۲۸	۵۴۷	*۱/۱۴	۱۹۹۰
۱۷-۹۶۷	۴۹/۵۷	۳۷/۱۸	۲۹	۱۴۰۰	.۰/۶۸	۴۰

* مقادیر بالاتر از ۱ احتمالاً ناشی از وجود چشممه‌های کارستی در منطقه‌ی مربوط به ایستگاه مربوطه هستند. (سنتر مطالعات جامع احیاء و توسعه‌ی کشاورزی استان‌ها)

۷-۳- زبان برنامه‌نویسی R

در این پایان‌نامه بخش عمده‌ی عملیات مربوط به برنامه‌نویسی و محاسبات در فضای نرم‌افزاری نسخه-۲.15.3 R اجرا شده است. در این راه از برخی بسته‌های محاسباتی تهیه شده در محیط R، مانند بسته‌های kohonen، e1071، cluster و lmomRFA نیز استفاده شده است.

زبان R توسط دو نفر به نام‌های Ross Ihaka و Robert Gentleman در دانشکده‌ی آمار دانشگاه Auckland در سال ۱۹۹۶ تحت عنوان R & R یا به طور خلاصه R، پا به عرصه وجود گذاشت. اما اکنون یک تیم بین‌المللی نگهداری و بسط آن را به عهده دارد.

R یک زبان برنامه نویسی ریاضی شئ^۱ گرا است که بسیار شبیه S-plus (بسته‌ی نرم‌افزاری مشهور آماری) می‌باشد و برای محاسبات آماری طراحی شده است.

مزایای استفاده از زبان برنامه نویسی R به جای استفاده از نرم افزارهای مختلف آماری به شرح

زیر هستند :

- زبان R رایگان است و کدهای آن باز می‌باشد و روی سیستم‌های ویندوز (Windows)

- یونیکس (Unix)، لینوکس (Linux) و مکینتاش (Macintosh) اجرا می‌شود.

- زبان R دارای راهنمای داخلی خوبی است.

- زبان R دارای قابلیت‌های گرافیکی قابل ملاحظه‌ای است.

- آشنایی با این زبان به منزله آشنایی با زبان تجاری آماری S-plus است.

- زبان R، زبانی قوی است. یادگیری آن ساده است و دارای توابع پیش‌ساخته‌ی آماری فراوانی

- است. در ضمن بسته‌های (packages) بسیار زیادی به آن اضافه می‌شود، که آن‌ها هم رایگان

- هستند. امکان ساخت و استفاده از بسته‌ها نیز در محیط R وجود دارد.

- در این زبان به سادگی می‌توان توابع مورد نظر کاربر را ساخت.

- فراوانی ارتقاء این زبان قابل ملاحظه است.

- با استفاده از زبان R کنترل کاربر روی خروجی و توانایی بسط برنامه یا نرم افزار برای

- روش‌های جدید آماری و الگوریتم‌های مورد نظر کاربر افزایش می‌یابد و کاربر به روش‌های

- کلاسیک آماری محدود نمی‌شود. (موسوی ندوشنی، ۱۳۹۱)

مشخصات بسته‌های محاسباتی که در مراحل مختلف انجام کار به R اضافه شده‌اند و برخی

جزئیات دستوری برنامه‌های نوشته شده در بخش پیوست ارائه شده است.

¹ object

فصل چهارم:

محاسبات، نتایج و بحث

۱-۴- مقدمه

در این فصل عملکرد انواع روش‌های تحلیل خوش‌های که در فصول پیشین معرفی و تشریح شدند، در منطقه‌بندی حوضه‌ی آبریز سفیدرود بزرگ به عنوان موضوع مطالعه‌ی موردی، بررسی شده است.

۲-۴- محاسبه‌ی گشتاورهای خطی و نسبت‌های گشتاور خطی

با استفاده از داده‌های آماری سیلاب که در بخش (۳-۶) مورد اشاره قرار گرفت، مقادیر گشتاورهای خطی و نسبت‌های گشتاور خطی مربوط به هر ایستگاه محاسبه شد.

برای داده‌های آماری مربوط به هر ایستگاه، مقادیر گشتاورهای خطی مرتبه اول تا چهارم، یعنی l_1 , l_2 , l_3 و l_4 محاسبه شدند. همچنین مقادیر مربوط به نسبت‌های گشتاور خطی t , t_3 و t_4 یعنی ضریب تغییرات خطی، چولگی خطی و کشیدگی خطی نیز مورد محاسبه قرار گرفتند.

در جدول (۴-۱) مقادیر گشتاور خطی مرتبه اول هر ایستگاه به همراه ضریب تغییرات خطی، چولگی خطی و کشیدگی خطی ارائه شده است. این موارد در محاسبه‌ی شاخص ناجوری D , شبیه-سازی مناطق همگن، محاسبه‌ی شاخص‌های ناهمگنی H , محاسبه‌ی شاخص نکویی برازش Z و برآورد پارامترهای توزیع و چندک‌های سیلاب کاربرد دارند.

جدول (۴-۱) مقادیر گشتاور خطی مرتبه اول، ضریب تغییرات خطی، چولگی خطی و کشیدگی خطی محاسبه شده برای ایستگاه‌های هیدرومتری مورد مطالعه

کد ایستگاه	مرتبه اول (l_1)	گشتاور خطی خطی (t)	ضریب تغییرات خطی (t_3)	کشیدگی خطی (t_4)
۱۷-۰۰۱	۲۰۹/۶۵	۰/۴۲۸	۰/۳۲۸	۰/۲۰۱
۱۷-۰۰۳	۲۷/۳۲	۰/۵۱۴	۰/۳۷۵	۰/۱۸۳
۱۷-۰۰۷	۱۱۳/۵۴	۰/۴۸۸	۰/۳۳۷	۰/۱۱۸
۱۷-۰۰۹	۶۰/۰۱	۰/۵۰۹	۰/۴۰۲	۰/۲۵۸
۱۷-۰۱۱	۳۲۶/۲۷	۰/۴۶۵	۰/۳۲۲	۰/۱۹۰
۱۷-۰۱۳	۶۵/۸۰	۰/۵۱۹	۰/۴۷۲	۰/۲۸۲

ادامه جدول (۱-۴) مقادیر گشتاور خطی مرتبه اول، ضریب تغییرات خطی، چولگی خطی و کشیدگی خطی محاسبه شده برای ایستگاه‌های هیدرومتری مورد مطالعه

کد ایستگاه	مرتبه اول (l_1)	گشتاور خطی	ضریب تغییرات خطی (t)	چولگی خطی	کشیدگی خطی (t_4)
۱۷-۰۱۵	۲۹۸/۳۰	۰/۴۳۲	۰/۲۰۴	۰/۰۳۲	
۱۷-۰۱۷	۱۹/۹۱	۰/۴۸۱	۰/۲۷۲	۰/۱۳۳	
۱۷-۰۱۹	۱۱۴/۸۰	۰/۵۰۷	۰/۴۰۲	۰/۲۱۶	
۱۷-۰۲۱	۴۶۱/۹۱	۰/۴۲۸	۰/۳۲۴	۰/۲۳۲	
۱۷-۰۲۳	۷۸/۵۱	۰/۴۷۵	۰/۲۷۱	۰/۱۱۸	
۱۷-۰۳۳	۷۴۵/۹۱	۰/۳۵۹	۰/۱۶۶	۰/۱۳۲	
۱۷-۰۳۵	۱۲۲/۰۲۳	۰/۲۴۵	۰/۱۲۶	۰/۱۶۷	
۱۷-۰۳۷	۱۴۸/۹۹	۰/۲۷۰	۰/۳۹۰	۰/۲۱۰	
۱۷-۰۳۹	۱۱۷/۰۷	۰/۳۵۲	۰/۲۴۸	۰/۱۸۷	
۱۷-۰۴۱	۲۴۱/۷۶	۰/۲۸۳	۰/۰۶۴	۰/۰۹۸	
۱۷-۰۵۰	۱۳/۳۱	۰/۳۱۴	۰/۱۴۵	۰/۰۵۷	
۱۷-۰۵۱	۱۲۳/۱۵	۰/۲۹۸	۰/۰۶۴	۰/۱۲۲	
۱۷-۰۵۵	۱۳۸/۳۸	۰/۳۰۲	۰/۲۱۷	۰/۱۹۹	
۱۷-۰۸۳	۱۳/۹۹	۰/۴۶۸	۰/۲۳۸	۰/۰۶۹	
۱۷-۱۰۱	۸/۵۸	۰/۳۹۲	۰/۲۷۸	۰/۲۰۳	
۱۷-۱۰۹	۸۶/۲۵	۰/۳۹۱	۰/۳۹۱	۰/۲۳۷	
۱۷-۱۱۱	۶۹/۶۸	۰/۲۲۵	۰/۰۴۵	-۰/۰۲۸	
۱۷-۱۲۳	۱۳/۷۶	۰/۳۸۸	۰/۱۳۷	۰/۰۱۲	
۱۷-۱۵۰	۱۲۰/۸۵	۰/۰۴۰	۰/۳۵۷	۰/۰۷۸	
۱۷-۲۰۱	۲۱۷/۵۲	۰/۳۴۱	۰/۱۶۹	۰/۱۱۰	
۱۷-۴۲۸	۱۲۴/۸۱	۰/۳۳۵	۰/۱۰۷	-۰/۰۳۰	
۱۷-۴۳۰	۶۵/۵۱	۰/۴۶۰	۰/۵۱۸	۰/۳۶۰	
۱۷-۸۷۴	۷/۰۳	۰/۳۱۹	۰/۰۹۹	۰/۰۴۸	
۱۷-۹۲۳	۱۰/۰۹	۰/۵۰۶	۰/۱۶۷	-۰/۰۱۱	
۱۷-۹۲۵	۲۹/۰۸	۰/۵۳۸	۰/۴۲۰	۰/۲۴۷	
۱۷-۹۲۶	۱۲۶/۰۸	۰/۶۸۶	۰/۵۸۸	۰/۴۱۲	
۱۷-۹۳۴	۱۷/۷۴	۰/۳۱۲	۰/۰۷۹	۰/۰۷۲	
۱۷-۹۶۵	۲۰/۸۳	۰/۳۸۹	۰/۳۹۵	۰/۳۱۰	
۱۷-۹۶۶	۶۲/۵۰	۰/۲۶۱	۰/۰۷۷	۰/۰۷۵	
۱۷-۹۶۷	۴۲/۵۶	۰/۲۴۲	۰/۳۲۰	۰/۲۷۷	

۴-۳-۴- شاخص D و شناسایی ایستگاههای ناجور

برای شناسایی ایستگاههایی که از نظر آماری با سایر ایستگاههای مورد مطالعه به شکل قابل توجهی اختلاف دارند یا در اصطلاح ناجور هستند، از شاخص ناجوری D که پیشتر در بخش (۱-۵-۴-۳) معرفی شد، استفاده شده است. احتمال وجود خطأ در مراحل ثبت داده‌ها، نسخه‌برداری و ... در مورد داده‌های چنین ایستگاههایی به شکل قابل توجهی بیش از سایر ایستگاههای مورد مطالعه است. مقادیر محاسبه شده شاخص D برای ۴۰ ایستگاه مورد مطالعه در جدول (۲-۴) درج شده است. بنا بر نتایج ارائه شده در جدول (۲-۴)، مقدار شاخص ناجوری D در ۲ ایستگاه، از مقدار مجاز این شاخص که برابر ۳ می‌باشد، بیشتر است. این دو ایستگاه که با کدهای ۱۷-۰۳۷ و ۱۷-۹۲۶ مشخص می‌شوند، از ادامه‌ی فرآیند مطالعه کنار گذاشته شدند و عملیات منطقه‌بندی و تحلیل فراوانی منطقه‌ای سیلان با استفاده از داده‌های مربوط به ۳۸ ایستگاه باقی‌مانده انجام گرفت.

جدول (۲-۴) مقادیر محاسبه شده شاخص ناجوری D برای ایستگاههای هیدرومتری مورد مطالعه

D	کد ایستگاه	D	کد ایستگاه	D	کد ایستگاه	D	کد ایستگاه
۱/۲۵	۱۷-۴۲۸	۰/۴۶	۱۷-۰۵۰	۰/۳۰	۱۷-۰۲۳	۰/۰۸	۱۷-۰۰۱
۱/۶۰	۱۷-۴۳۰	۱/۶۶	۱۷-۰۵۱	۱/۰۹	۱۷-۰۲۵	۰/۳۷	۱۷-۰۰۳
۰/۲۱	۱۷-۸۷۴	۰/۶۳	۱۷-۰۵۵	۰/۳۳	۱۷-۰۲۷	۰/۵۴	۱۷-۰۰۷
۱/۸۵	۱۷-۹۲۳	۰/۵۲	۱۷-۰۸۳	۰/۹۰	۱۷-۰۲۹	۰/۵۹	۱۷-۰۰۹
۰/۶۸	۱۷-۹۲۵	۰/۱۸	۱۷-۱۰۱	۲/۶۰	۱۷-۰۳۰	۰/۱۶	۱۷-۰۱۱
*۴/۳۶	۱۷-۹۲۶	۰/۷۳	۱۷-۱۰۹	۰/۴۰	۱۷-۰۳۳	۰/۷۵	۱۷-۰۱۳
۰/۷۶	۱۷-۹۳۴	۱/۸۲	۱۷-۱۱۱	۱/۲۱	۱۷-۰۳۵	۰/۶۷	۱۷-۰۱۵
۱/۰۰	۱۷-۹۶۵	۰/۶۷	۱۷-۱۳۳	*۴/۰۶	۱۷-۰۳۷	۰/۳۷	۱۷-۰۱۷
۰/۷۸	۱۷-۹۶۶	۱/۵۰	۱۷-۱۵۰	۰/۲۲	۱۷-۰۳۹	۰/۳۶	۱۷-۰۱۹
۲/۴۱	۱۷-۹۶۷	۰/۲۰	۱۷-۲۰۱	۱/۱۵	۱۷-۰۴۱	۰/۲۹	۱۷-۰۲۱

* مقادیر بالاتر از حد مجاز (بزرگ‌تر از ۳)

۴-۴- ویژگی‌های مورد استفاده در خوشبندی و معیار فاصله

در این تحقیق، از میان ویژگی‌های مؤثر بر واکنش سیلابی سایت‌های مورد نظر در حوضه، با توجه به آمار و اطلاعات در دسترس و نتایج مطالعات پیشین، طول و عرض جغرافیایی به عنوان ویژگی‌های جغرافیایی، مساحت سطح زهکشی و ضریب رواناب به عنوان ویژگی‌های فیزیوگرافیک، ارتفاع از سطح دریا به عنوان یک ویژگی جغرافیایی و فیزیوگرافیک و متوسط بارندگی سالانه به عنوان یک ویژگی هواشناسی جهت استفاده در عملیات خوشبندی انتخاب شدند. مقادیر هر یک از این ویژگی‌ها برای هر ایستگاه، در جدول (۳-۳) ارائه شده است. در میان ویژگی‌های منتخب در این مطالعه، تنها مساحت سطح زهکشی سایت‌ها با لگاریتم طبیعی آن‌ها جایگزین شد و سپس داده‌های مربوط به هر گروه از ویژگی‌ها بر انحراف معیار داده‌های آن گروه تقسیم شدند. جایگزینی مقادیر مساحت سطح زهکشی ایستگاه‌ها با لگاریتم طبیعی آن‌ها به منظور کاهش اثر اختلاف‌های بعضًا بسیار بزرگ بین مساحت‌های مربوط به سایت‌های مختلف انجام گرفت که این روش در مطالعات پیشین نیز، تجربه و توصیه شده است (رانو و سرینیواس، ۲۰۰۶a, b).

همچنین به منظور سنجش فاصله یا شباهت و تفاوت بین بردارهای مشخصه مربوط به ایستگاه‌ها یا سایت‌ها مختلف، در این مطالعه از معیار فاصله‌ی اقلیدسی که در بخش (۳-۱-۱-الف) معرفی شد، استفاده شده است.

۴-۵- منطقه‌بندی با استفاده از خوشبندی ترکیبی

۴-۵-۱- خوشبندی با استفاده از الگوریتم‌های سلسله مراتبی

برای اجرای منطقه‌بندی حوضه‌ی آبریز سفیدرود بزرگ با استفاده از الگوریتم‌های خوشبندی سلسله مراتبی، چهار الگوریتم متراکم کننده‌ی تک پیوند، تمام پیوند، پیوند متوسط و Ward مورد استفاده

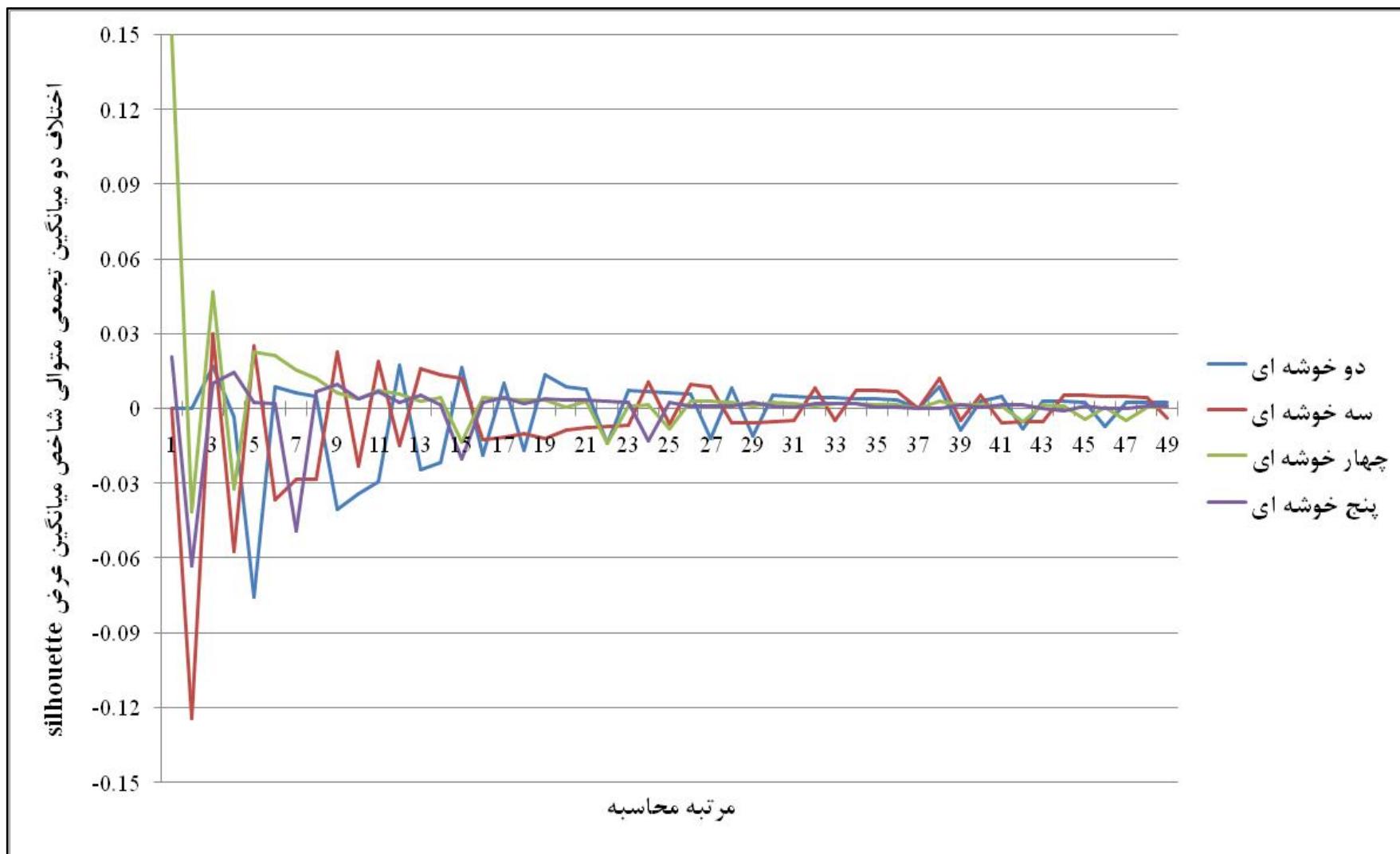
قرار گرفتند. کاربرد هر یک از این الگوریتم‌ها در منطقه‌بندی حوضه‌ی آبریز سفیدرود بزرگ در حالت‌های ۲ تا ۵ خوش‌های مورد مطالعه قرار گرفت. تعداد ایستگاه‌های اختصاص یافته به هر یک از مناطق تشکیل شده در حالات مختلف، همچنین مقدار تابع هدف و شاخص صحت خوش‌های بزرگ در حالت silhouettes برای هر حالت محاسبه شد.

۴-۵-۲- خوش‌های بزرگ در حالت silhouettes با استفاده از الگوریتم افزایی K-means

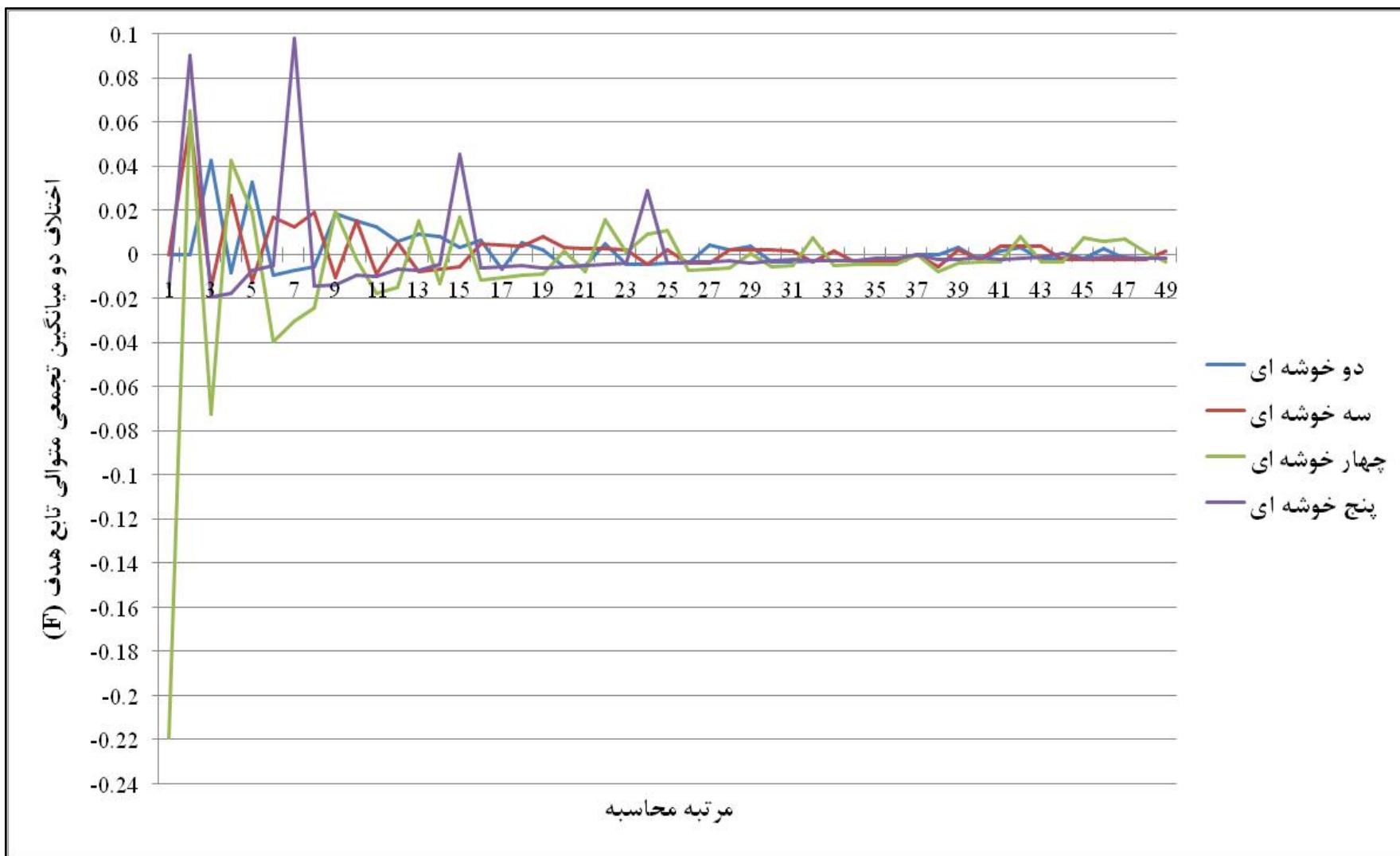
از میان الگوریتم‌های خوش‌های بزرگ در حالت silhouettes، الگوریتم K-means با توجه عملکرد مناسب آن در مطالعات منطقه‌بندی، برای منطقه‌بندی حوضه‌ی آبریز سفیدرود بزرگ در حالت‌های ۲ تا ۵ منطبقه‌ای مورد استفاده قرار گرفت.

از آنجا که عملکرد این الگوریتم تا حدودی تحت تأثیر انتخاب مراکز اولیه‌ی خوش‌های بزرگ در ابتدای فرآیند خوش‌های بزرگ است، لذا اجرای الگوریتم خوش‌های بزرگ در حالت silhouettes با یک انتخاب منحصر به فرد از ایستگاه‌ها به عنوان مراکز اولیه‌ی خوش‌های بزرگ در حالت silhouettes ممکن است معیار مناسبی برای قضاوت در مورد عملکرد این روش نباشد.

برای برطرف کردن این مشکل، انتخاب مراکز اولیه‌ی خوش‌های بزرگ در حالت silhouettes به صورت انتخاب اعداد تصادفی در محدوده‌ی تعداد ایستگاه‌های مورد استفاده در عملیات منطقه‌بندی، که هر عدد معرف یکی از ایستگاه‌های موجود در حوضه است، به دفعات متعدد انجام گرفت. این فرآیند مطابق شکل‌های (۱-۴) و (۲-۴) تا جایی ادامه یافت که اختلاف میانگین تجمعی تابع هدف و میانگین شاخص صحت خوش‌های بزرگ در مراتب متوالی اجرای الگوریتم به مقداری کمتر از ۱٪ کاهش یافت. میانگین مقادیر تابع هدف و شاخص صحت خوش‌های بزرگ در حالت silhouettes با ۵۰ انتخاب تصادفی مراکز اولیه، به عنوان معیاری برای قضاوت در مورد عملکرد الگوریتم K-means در منطقه‌بندی حوضه‌ی آبریز سفیدرود بزرگ مورد استفاده قرار گرفت. تعداد تکرار فرآیند عددی الگوریتم K-means برای هر بار انتخاب مراکز اولیه‌ی خوش‌های بزرگ در حالت silhouettes، برابر ۱۰۰۰ در نظر گرفته شد.



شکل (۱-۴) نمایش تغییر اختلاف بین دو میانگین تجمعی متوالی شاخص میانگین عرض silhouette در ۵۰ انتخاب تصادفی مراکز اولیه‌ی الگوریتم K-means



شکل (۲-۴) نمایش تغییر اختلاف بین دو میانگین تجمعی متوالی تابع هدف (F) در ۵۰ انتخاب تصادفی مراکز اولیه‌ی الگوریتم K-means

۴-۵-۳- خوشبندی با استفاده از الگوریتم‌های خوشبندی ترکیبی

برای منطقه‌بندی حوضه‌ی سفیدرود با استفاده از الگوریتم‌های خوشبندی ترکیبی، مراکز خوشبندی تشکیل شده با استفاده از الگوریتم‌های خوشبندی سلسه مراتبی تک پیوند، تمام پیوند، پیوند متوسط و Ward در حالت‌های ۲ تا ۵ منطقه‌ای، به عنوان مراکز اولیه‌ی خوشبندی در الگوریتم خوشبندی افزایی K-means مورد استفاده قرار گرفتند. تعداد دفعات تکرار فرآیند عددی الگوریتم K-means برابر ۱۰۰۰ مرتبه تنظیم شد.

در هر حالت مقدار تابع هدف و شاخص میانگین عرض silhouette برای خوشبندی انجام گرفته، محاسبه و تعداد ایستگاه‌های تعلق یافته به هر منطقه نیز مشخص شد.

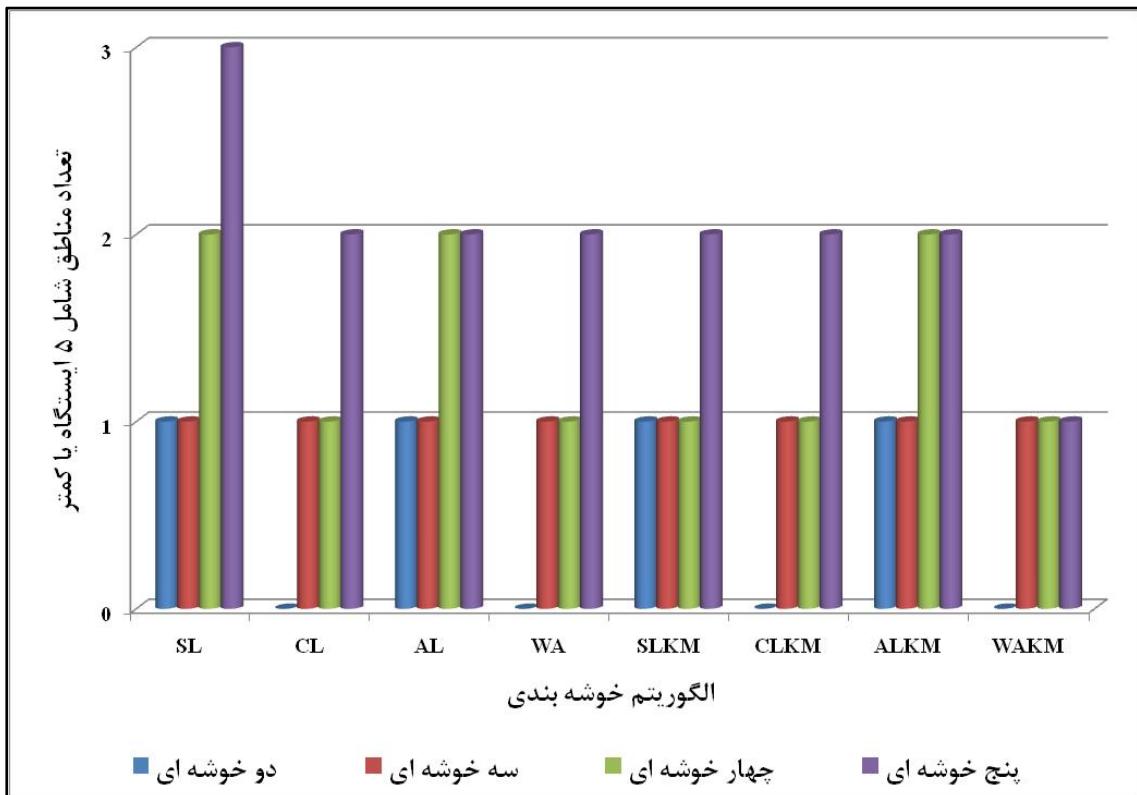
۴-۵-۴- تعداد ایستگاه‌های اختصاص یافته به مناطق

هر چه تعداد ایستگاه‌های موجود در یک منطقه و به تبع آن داده‌های آماری سیلاب مربوط به آن بیشتر باشد، امکان برآوردهای دوره‌های بازگشت طولانی‌تری فراهم خواهد شد. از طرفی آن چنان که پیشتر گفته شد، برآوردهای سیلاب در تحلیل فراوانی منطقه‌ای تا زمانی قابل اعتماد است که تعداد سال‌های آماری موجود در ایستگاه‌های یک منطقه بزرگ‌تر یا مساوی پنج برابر دوره‌ی بازگشت مورد نظر جهت برآوردهای بزرگی سیلاب باشد این بدان معناست که اگر تمام ایستگاه‌های موجود در یک منطقه از نظر طول داده‌های آماری یکسان باشند، زمانی تحلیل فراوانی منطقه‌ای می‌تواند به صورتی قابل اعتماد برای برآوردهای سیلاب با دوره‌ی بازگشت طولانی‌تر از تعداد سال‌های آماری هر ایستگاه مورد استفاده قرار گیرد، که تعداد ایستگاه‌های موجود در منطقه بیش از ۵ ایستگاه باشد.

این معیار اگرچه یک معیار دقیق و قطعی نیست، اما تا حدودی می‌تواند نشان‌دهنده‌ی کیفیت پرآکنش ایستگاه‌های موجود در مناطق به منظور اجرای تحلیل فراوانی منطقه‌ای باشد. در واقع هر چه تعداد مناطقی که شامل تعداد اندکی از ایستگاه‌ها هستند کمتر باشد و به طور کلی مناطق تشکیل

شده از نظر تعداد داده‌های آماری موجود در آن‌ها به یکدیگر نزدیک‌تر باشند، وضعیت مطلوب‌تر خواهد بود. البته چنانچه اجرای تحلیل فراوانی منطقه‌ای و برآورد سیلاپ برای ایستگاه یا ایستگاه‌های خاصی مد نظر باشد، آنچه بیش‌تر اهمیت می‌یابد افزایش داده‌های آماری مربوط به مناطق دربر-گیرنده‌ی آن ایستگاه یا ایستگاه‌ها است و نه کلیه‌ی مناطق.

در شکل (۳-۴)، تعداد مناطق شامل ۵ منطقه یا کمتر برای هر یک از روش‌های خوشبندی سلسله مراتبی و ترکیبی در هر کدام از حالات ۲ تا ۵ منطقه‌ای به نمایش در آمده است. همان گونه که در شکل (۳-۴) مشاهده می‌شود، الگوریتم ترکیبی WAKM که در واقع ترکیب الگوریتم سلسله مراتبی Ward و الگوریتم افزایی K-means است، در تمامی حالات از نظر برآورده کردن این معیار، بهترین وضعیت را دارد.



شکل (۳-۴) تعداد مناطق شامل ۵ ایستگاه یا کمتر در حالات ۲ تا ۵ خوشبندی برای الگوریتم‌های مختلف خوشبندی. WAKM، ALKM، CLKM، WA، AL، CL، SL، SLKM، Ward، K-means، ترکیب تمام پیوند و ترکیب پیوند متوسط و K-means، ترکیب تک پیوند و K-means، ترکیب تک پیوند و K-means، ترکیب تک پیوند و Ward، ترکیب تک پیوند و Ward هستند.

۴-۵-۵- کمینه‌سازی تابع هدف

به منظور بررسی توانایی هر یک از الگوریتم‌های خوشبندی در کمینه‌سازی تابع هدف معرفی شده در بخش (۳-۱-۳)، مقدار نهایی تابع هدف برای هر یک از الگوریتم‌های خوشبندی شامل الگوریتم-های سلسله مراتبی و الگوریتم‌های ترکیبی در حالت‌های ۲ تا ۵ منطقه‌ای محاسبه شد. همچنین مقدار میانگین تابع هدف در ۵۰ انتخاب تصادفی مراکز اولیه‌ی خوشبها برای اجرای الگوریتم K-means که در هر انتخاب فرآیند عددی این الگوریتم به تعداد ۱۰۰۰ مرتبه تکرار شده است، محاسبه شد. شکل (۴-۴) مقادیر تابع هدف را برای هر کدام از روش‌ها و در هر یک از حالات نشان می‌دهد. همان گونه که در شکل (۴-۴) مشاهده می‌شود، در میان روش‌های مورد بررسی دو الگوریتم ترکیبی WAKM و CLKM در حالت‌های ۲ تا ۴ منطقه‌ای، بهترین عملکرد را در زمینه‌ی کمینه-سازی تابع هدف از خود نشان می‌دهند. در حالت ۵ خوشبها میانگین تابع هدف اجرای الگوریتم K-means با ۵۰ انتخاب تصادفی مراکز اولیه، کمترین مقدار را اختیار کرده و پس از آن الگوریتم CLKM در این زمینه موفق ظاهر شده است. در این حالت، الگوریتم CLKM با تفاوت بسیار اندکی نسبت به WAKM در جای سوم قرار گرفته است.

۴-۶- شاخص صحت خوشبندی میانگین عرض silhouette

برای ارزیابی کیفیت خوشبندی اجرا شده با استفاده از هر یک از روش‌های خوشبندی و همچنین انتخاب تعداد بهینه‌ی خوشبها، مقدار شاخص میانگین عرض silhouette که در بخش (۳-۱-۳) توصیف شد، برای هر یک از الگوریتم‌های خوشبندی سلسله مراتبی و ترکیبی که پیش‌تر معرفی شدند، در حالت‌های ۲ تا ۵ خوشبها محاسبه شد. همچنین میانگین این شاخص برای ۵۰ انتخاب تصادفی مراکز اولیه‌ی الگوریتم K-means در هر یک از حالت‌های ۲ تا ۵ منطقه‌ای مورد محاسبه قرار گرفت. در شکل (۵-۴) مقادیر محاسبه شده‌ی این شاخص برای هر حالت از اجرای هر کدام از

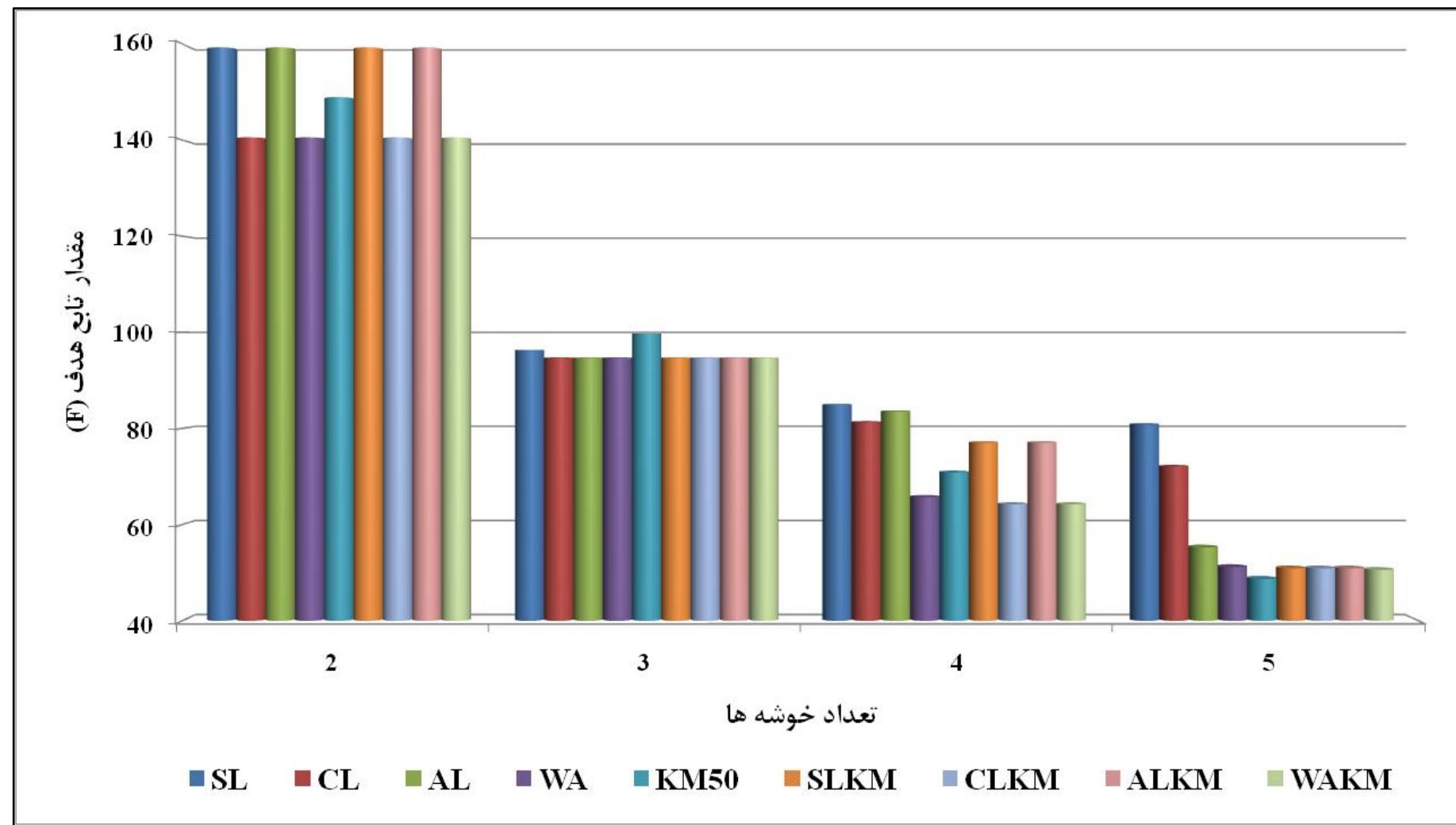
الگوریتم‌های خوشبندی به نمایش در آمده است.

آن چنان که در شکل (۴-۵) دیده می‌شود، در حالت ۲ خوشبایی، الگوریتم‌های AL، SLK و ALKM از نظر مقدار شاخص میانگین عرض silhouette بهترین وضعیت را دارا هستند. در حالت ۳ خوشبایی به جز الگوریتم SL و همچنین میانگین شاخص صحت خوشبندی ۵۰ انتخاب تصادفی مراکز اولیه‌ی K-means که مقادیر کمتر و نامناسب‌تری را اختیار می‌کنند، سایر روش‌ها وضعیت یکسانی دارند. در حالت ۴ خوشبایی، دو الگوریتم ترکیبی CLKM و WAKM بالاترین مقادیر شاخص صحت خوشبندی را دارا هستند. در حالت ۵ خوشبایی بیشترین مقدار میانگین شاخص صحت خوشبندی برای اجرای الگوریتم K-means با ۵۰ انتخاب تصادفی مراکز اولیه است و پس از آن الگوریتم‌های Ward و WAKM بالاترین مقادیر را دارا هستند.

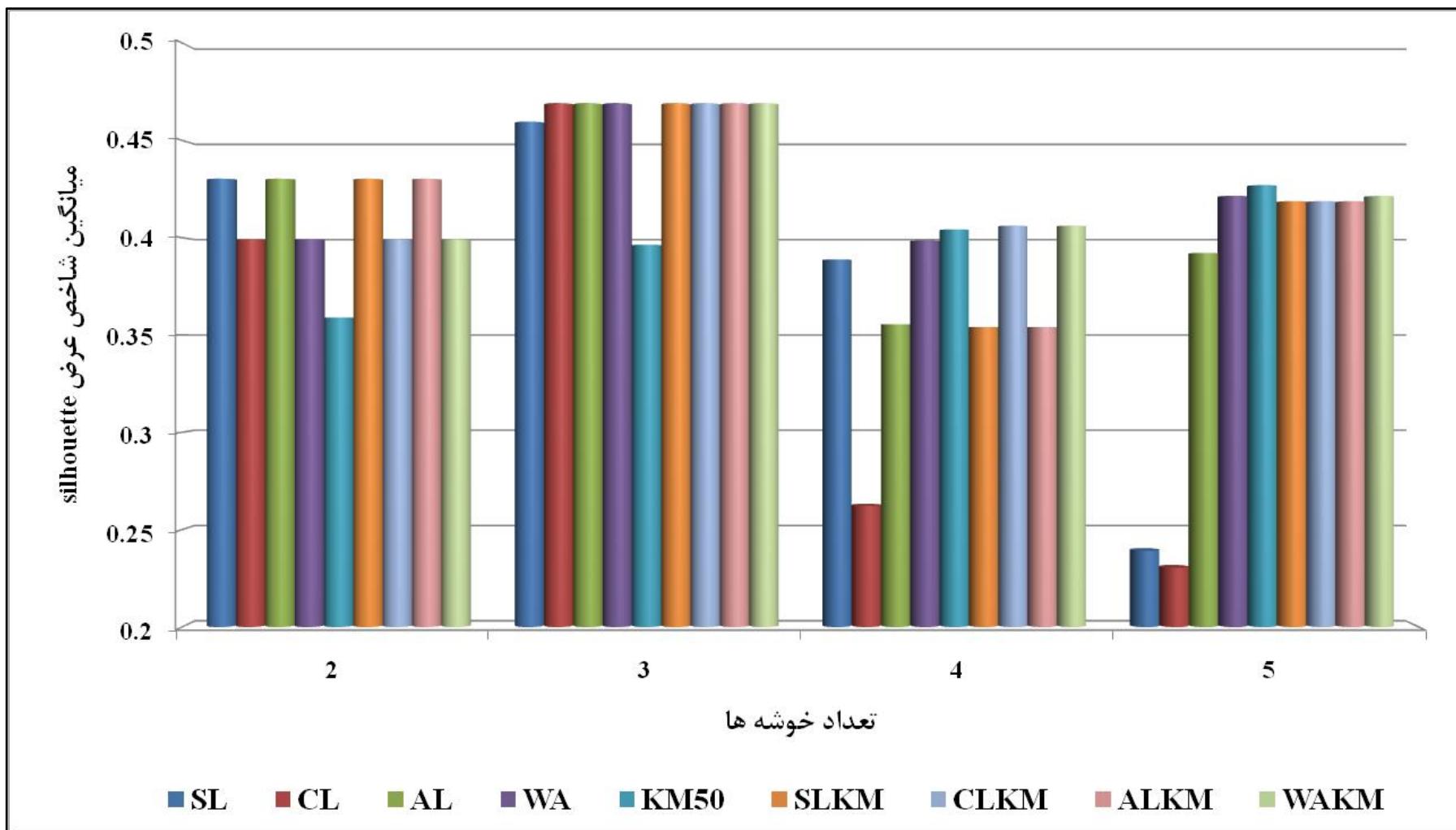
۴-۵-۷- انتخاب الگوریتم خوشبندی بهینه

در ستون‌های اول و دوم جدول (۳-۴) مقادیر میانگین تابع هدف و شاخص صحت خوشبندی حالت‌های ۲ تا ۵ خوشبایی برای هر یک از الگوریتم‌های خوشبندی مورد بررسی ارائه شده است. همچنین برای هر یک از گزینه‌ها مجموع تعداد خوشبایی که در حالت‌های ۲ تا ۵ خوشبایی شامل ۵ یا تعداد کمتری ایستگاه هستند، در ستون سوم جدول (۳-۴) آمده است. بر اساس نتایج مندرج در این جدول، الگوریتم ترکیبی WAKM در هر سه بخش نسبت به سایر روش‌ها برتری نسبی دارد. بنابراین از بین روش‌هایی که تا به اینجا مورد بررسی قرار گرفتند این الگوریتم به عنوان روش بهینه برای اجرای منطقه‌بندی حوضه‌ی سفیدرود بزرگ و تحلیل فراوانی منطقه‌ای سیلاب در این حوضه انتخاب شد.

از نظر تعداد خوشبایی، حالت سه خوشبایی که در آن شاخص صحت خوشبندی در کلیه‌ی روش‌ها، بیشترین مقدار خود را اختیار می‌کند، می‌تواند به عنوان حالت بهینه انتخاب شود. شایان ذکر است، از آنجا که مقدار تابع هدف همواره با افزایش تعداد خوشبایی کاهش می‌یابد، در تعیین تعداد خوشبایی عامل تعیین کننده‌ای نیست.



شکل (۴-۴) مقادیر تابع هدف (F) در حالات ۲ تا ۵ خوشه‌ای برای روش‌های مختلف خوشه‌بندی. WAKM، ALKM، CLKM، SLKM، WA، AL، CL، SL و WAKM به ترتیب معرف الگوریتم‌های تک پیوند، تمام پیوند، پیوند متوسط، Ward، و ترکیب آن‌ها با الگوریتم K-means هستند. KM50 معرف ۵ انتخاب تصادفی مراکز اولیه برای الگوریتم K-means است.



شکل (۴-۵) مقادیر شاخص میانگین عرض silhouette در حالات ۲ تا ۵ خوشه‌ای برای روش‌های مختلف خوشه‌بندی. K-means الگوریتم‌های تک پیوند، تمام پیوند، پیوند متوسط، Ward، و ترکیب آن‌ها با الگوریتم KM50 معرف است. انتخاب تصادفی مراکز اولیه برای الگوریتم K-means معرف WAKM، CLKM، SLKM، WA، AL، CL، SL، ALKM و WAKM است.

جدول (۳-۴) مقادیر میانگین تابع هدف، شاخص میانگین عرض silhouette و تعداد خوشه‌های شامل ۵ ایستگاه هیدرومتری یا کمتر برای الگوریتم‌های مختلف خوشبندی در حالات ۲ تا ۵ خوشه‌ای

الگوریتم	میانگین تابع هدف (F)	میانگین شاخص عرض silhouette	تعداد خوشه‌های شامل ۵ ایستگاه یا کمتر
SL	۱۰۵/۳۴	۰/۳۷۹	۷
CL	۹۷/۱۷	۰/۳۴۰	۴
AL	۹۸/۲۱	۰/۴۱۲	۶
WA	۸۸/۰۵	۰/۴۲۲	۴
KM50	۹۲/۰۶	۰/۳۹۷	*
SLKM	۹۵/۴۹	۰/۴۱۸	۵
CLKM	۸۷/۶۰	۰/۴۲۳	۴
ALKM	۹۵/۴۹	۰/۴۱۸	۶
WAKM	۸۷/۵۲	۰/۴۲۴	۳

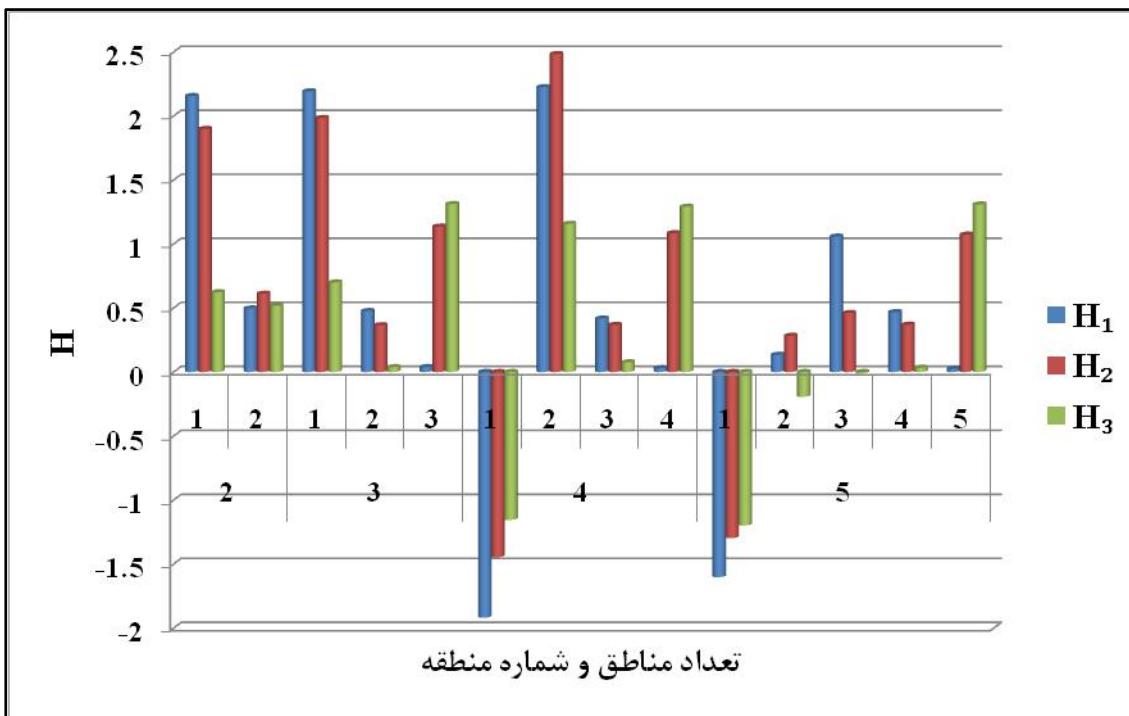
* با توجه به این که تعداد چنین خوشه‌هایی برای این گزینه در دفعات مختلف اجرای الگوریتم K-means متفاوت است، امکان محاسبه‌ی این مورد به صورت تعداد مجموع وجود ندارد.

۴-۵-۴-۸-۵-۴- تحلیل فراوانی منطقه‌ای با استفاده از گشتاورهای خطی

۴-۵-۴-۸-۱- تعیین وضعیت همگنی مناطق و تعداد ایستگاه‌های هر منطقه

پس از انتخاب الگوریتم ترکیبی WAKM به عنوان روش بهینه برای منطقه‌بندی حوضه‌ی آبریز سفیدرود بزرگ، مقادیر شاخص‌های ناهمگنی H برای مناطق تشکیل شده در حالتهای ۲ تا ۵ منطقه‌ای محاسبه شد که نتایج به دست آمده، در شکل (۶-۴) نشان داده شده است.

اگر بنا بر توصیه‌ی هاسکینگ و والیس (۱۹۹۷)، برای قضاوت در مورد وضعیت همگنی مناطق حاصل از تحلیل خوشه‌ای تنها شاخص H_1 را ملاک قرار دهیم، مشاهده می‌شود که در هر یک از حالتهای ۲ تا ۵ منطقه‌ای، مقدار شاخص ناهمگنی H_1 برای یکی از مناطق تشکیل شده بیشتر از ۱ بوده و در نتیجه آن منطقه ناهمگن است. از این رو چگونگی و میزان تأثیر حضور یا عدم حضور هر یک از ایستگاه‌های موجود در منطقه‌ی ناهمگن بر تغییرات شاخص ناهمگنی H_1 در هر یک از حالات ۲ تا ۵ منطقه‌ای مورد بررسی قرار گرفت و مشخص شد که با حذف ایستگاه با کد شناسه‌ی ۰۴۱-۰۱۷، مطابق شکل (۷-۴) در کلیه‌ی حالتهای مقدار شاخص H_1 در منطقه‌ی مورد نظر کاهش می‌یابد.



شکل (۶-۴) مقادیر شاخص‌های ناهمگنی H در مناطق حاصل از الگوریتم WAKM برای ۳۸ ایستگاه هیدرومتری



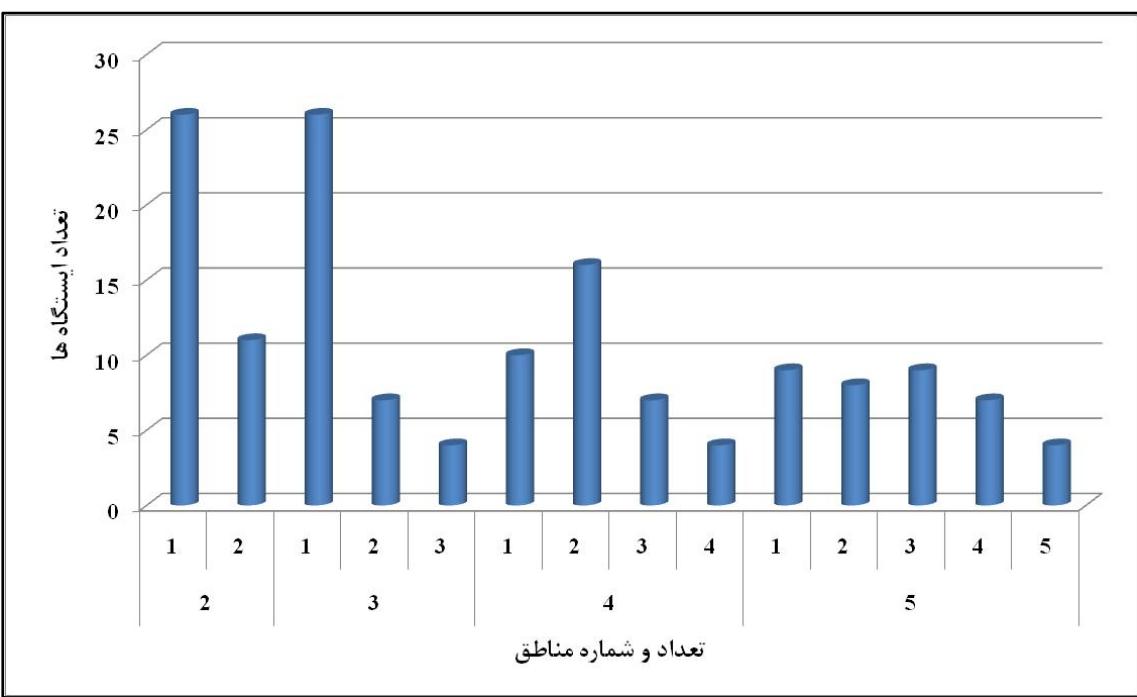
شکل (۷-۴) مقادیر شاخص‌های ناهمگنی H در مناطق حاصل از الگوریتم WAKM برای ۳۷ ایستگاه هیدرومتری

همان گونه که در شکل (۷-۴) مشاهده می‌شود، پس از حذف ایستگاه ۱۷-۰۴۱ و اجرای دوباره-منطقه‌بندی، در حالت‌های ۲، ۳ و ۴ منطقه‌ای، مقدار شاخص H_1 به کمتر از ۱ کاهش یافته و شرط

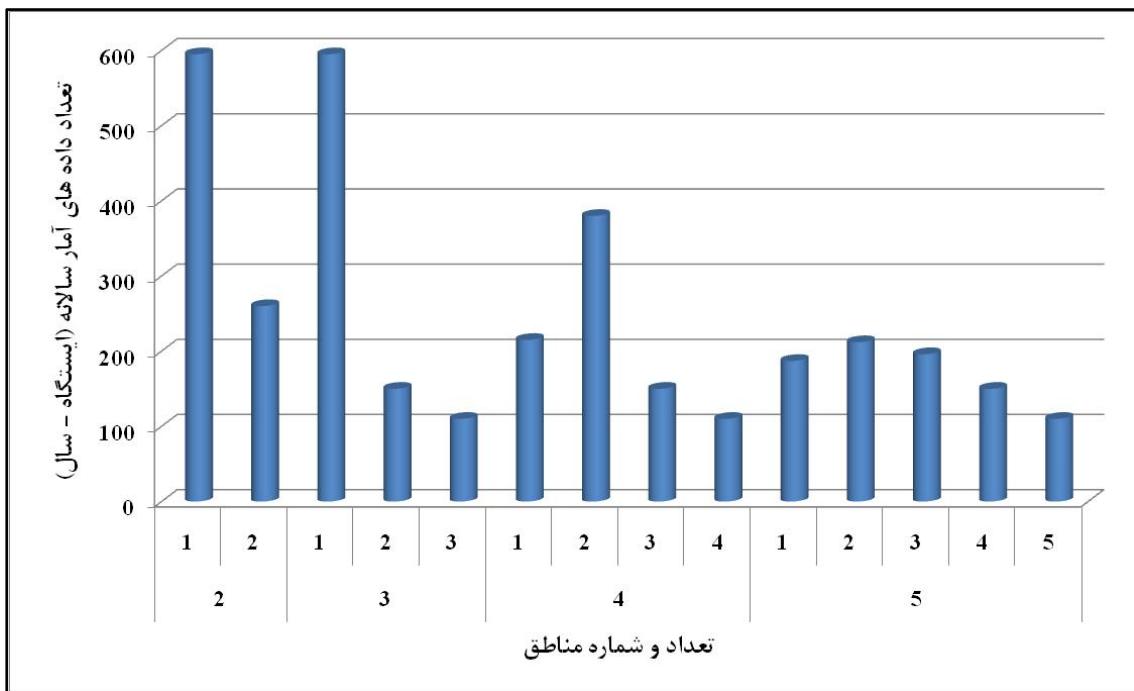
همگنی برقرار می‌شود. البته چنانچه شرط همگنی را کمتر از ۱ بودن مقدار هر سه شاخص H_1 , H_2 و H_3 در نظر بگیریم، همگنی کامل تنها در حالت دو منطقه‌ای برقرار است.

موضوع دیگر این است که هر چه تعداد مناطق کمتر باشد، به طور متوسط تعداد ایستگاه‌های موجود در هر منطقه و داده‌های آماری مربوط به آن منطقه افزایش می‌یابد و از این رو امکان برآورد چندک‌های سیلاب با دوره‌های بازگشت طولانی‌تر فراهم می‌شود. لذا از این حیث نیز می‌توان حالت دو منطقه‌ای را به عنوان بهترین گزینه در نظر گرفت. در شکل‌های (۴-۸) و (۴-۹) تعداد ایستگاه‌های موجود در هر منطقه و تعداد داده‌های آماری متناظر با آن نشان داده شده است.

با توجه به موارد ذکر شده، حالت دو منطقه‌ای از نظر مقادیر شاخص‌های ناهمگنی و تعداد داده‌های آماری موجود در مناطق به عنوان بهترین حالت شناسایی شد. حالت سه منطقه‌ای نیز که از نظر شاخص‌های ناهمگنی وضعیت قابل قبولی داشته و همچنین از نظر مقادیر شاخص صحت خوشبندی به عنوان بهترین گزینه شناسایی شد، به همراه حالت دو منطقه‌ای برای انتخاب بهترین گزینه جهت اجرای تحلیل فراوانی منطقه‌ای مورد بررسی‌های بیشتر قرار گرفت.

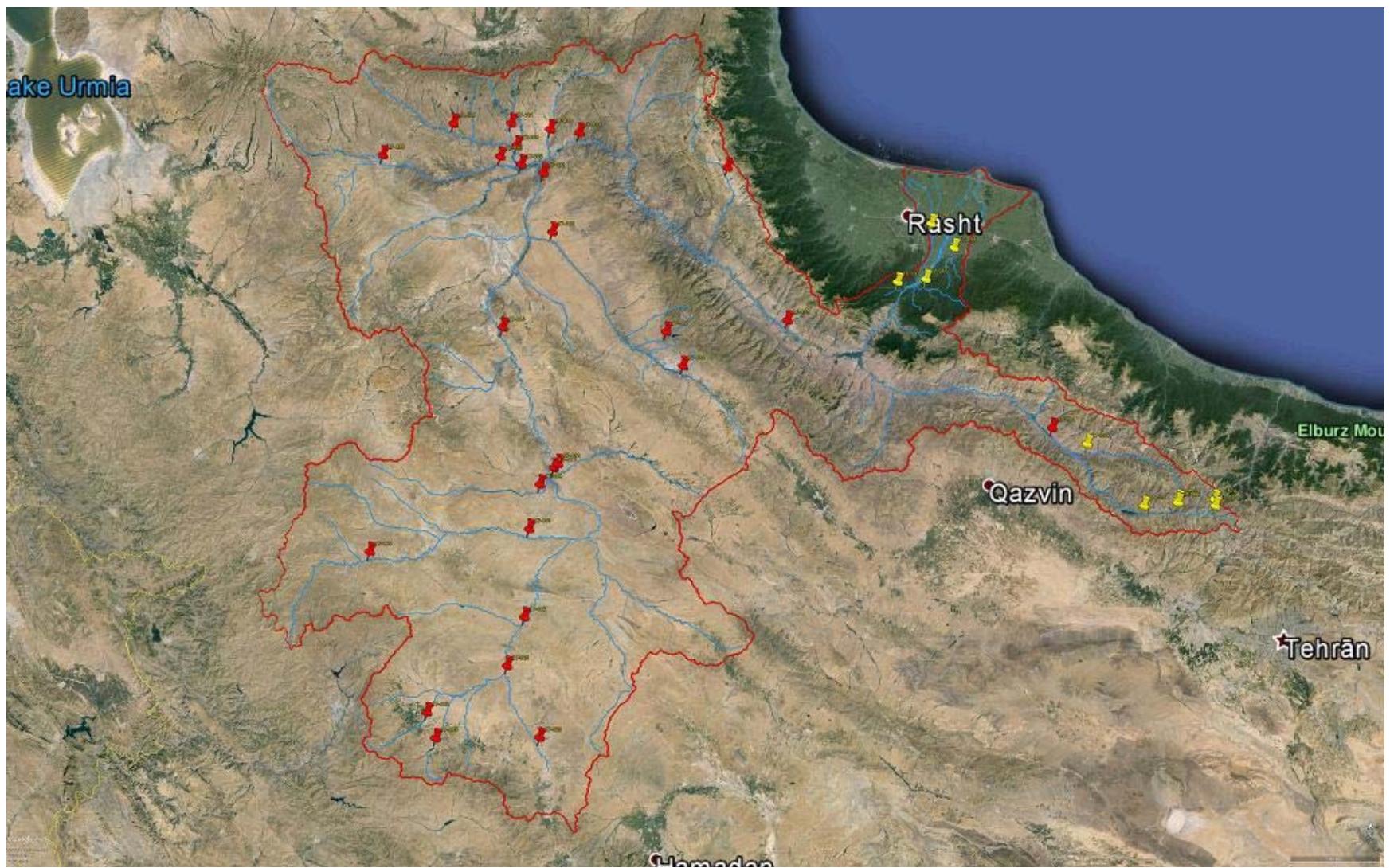


شکل (۴-۸) تعداد ایستگاه‌های هیدرومتری در هر یک از مناطق در حالات ۲ تا ۵ خوشبندی برای الگوریتم WAKM

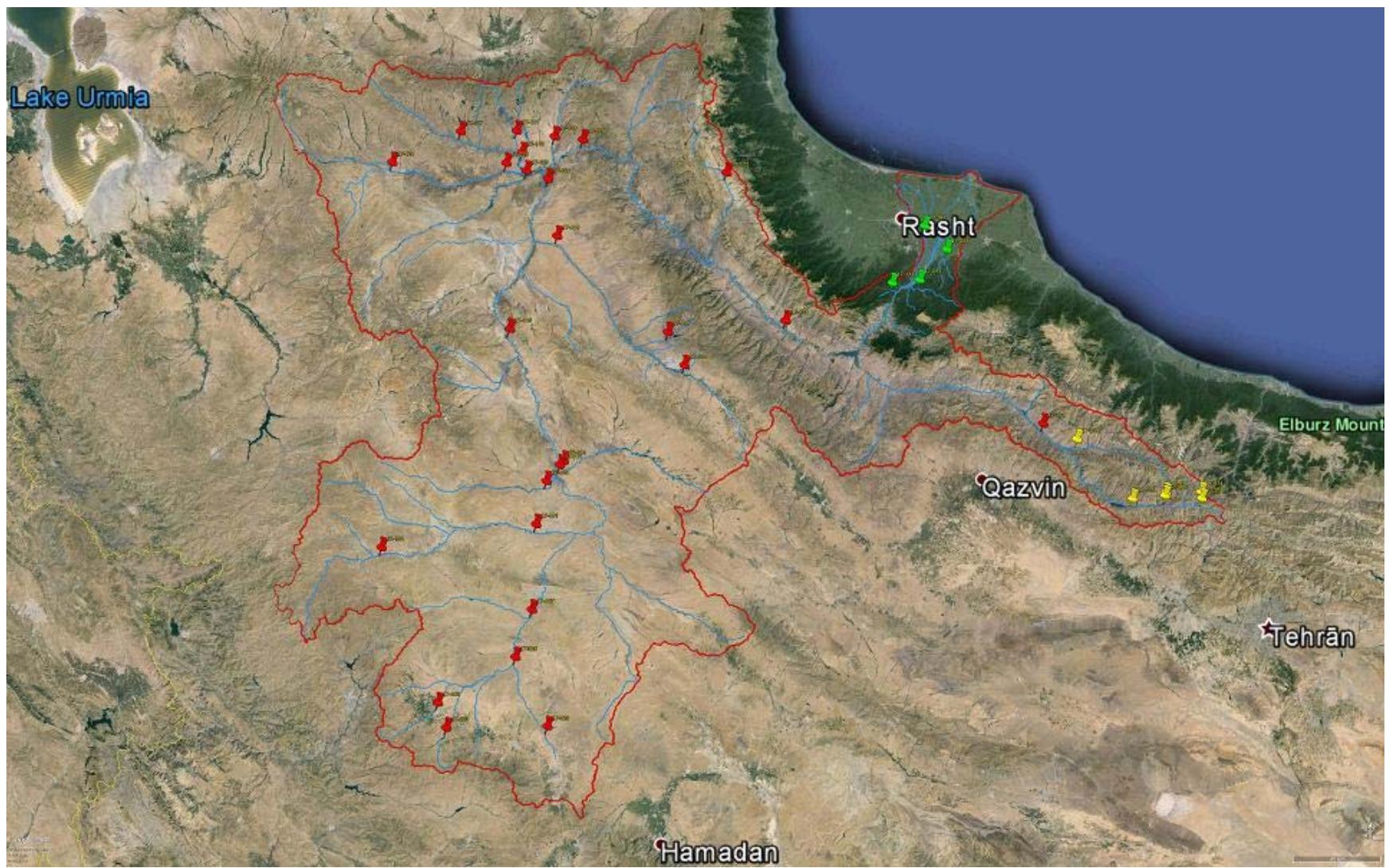


شکل (۹-۴) تعداد داده‌های آماری موجود در هر یک از مناطق در حالات ۲ تا ۵ خوش‌های برای الگوریتم WAKM

شکل‌های (۱۰-۴) و (۱۱-۴) چگونگی توزیع ایستگاه‌های مورد مطالعه در مناطق تشکیل شده را نمایش می‌دهند. بررسی نحوه‌ی توزیع ایستگاه‌ها در مناطق تشکیل شده در دو حالت مورد بررسی نشان می‌دهد که منطقه‌ی اول تشکیل شده در حالت‌های دو و سه منطقه‌ای، یک منطقه‌ی یکسان شامل ۲۶ ایستگاه است. اما منطقه‌ی دوم مربوط به حالت دو منطقه‌ای که شامل ۱۱ ایستگاه است، در حالت سه منطقه‌ای خود به دو منطقه، یکی شامل ۷ ایستگاه و دیگری شامل ۴ ایستگاه تقسیم می‌شود. این مسئله خود به نوعی بیانگر برتری حالت دو منطقه‌ای نسبت به حالت سه منطقه‌ای است. زیرا با تبدیل منطقه‌ی دوم حالت دو منطقه‌ای به مناطق دوم و سوم حالت سه منطقه‌ای، نه تنها تعداد ایستگاه‌ها و داده‌های آماری موجود در این مناطق نسبت به حالت پیشین کاهش می‌یابد، بلکه آن چنان که پیش‌تر در شکل (۸-۴) نشان داده شد، همگنی مناطق نیز تا حدودی دچار افت می‌شود. در نهایت با توجه وضعیت مناسب حالت دو منطقه‌ای از نظر همگنی مناطق و میزان داده‌های آماری موجود در آن‌ها که امکان برآورد سیلاب برای دوره‌های بازگشت بزرگ‌تر را فراهم می‌کند، حالت دو منطقه‌ای مربوط به الگوریتم WAKM برای اجرای تحلیل فراوانی منطقه‌ای انتخاب شد.



شکل (۱۰-۴) توزیع جغرافیایی ایستگاه‌های هیدرومتری مورد مطالعه در مناطق حالت دو منطقه‌ای؛ ایستگاه‌های مناطق اول و دوم به ترتیب با رنگ‌های قرمز و زرد مشخص شده‌اند.



شکل (۱۱-۴) توزیع جغرافیایی ایستگاه‌های هیدرومتری در مناطق حالت سه منطقه‌ای؛ ایستگاه‌های مناطق اول، دوم و سوم به ترتیب با رنگ‌های قرمز، زرد و سبز مشخص شده‌اند.

۴-۵-۲-۸- شاخص نکویی برازش Z و انتخاب توزیع فراوانی منطقه‌ای سیلاب

در این پژوهش شاخص نکویی برازش Z که در بخش (۳-۵-۴) معرفی شد، به منظور انتخاب توزیع فراوانی منطقه‌ای بهینه برای هر یک از مناطق تشکیل شده مورد استفاده قرار گرفت. بدین منظور برای هر منطقه از میان توزیع‌هایی که شرط نکویی برازش $Z^{DIST} \leq 1.64$ در مورد آن‌ها برقرار بود، توزیعی که نزدیک‌ترین مقدار به صفر را اختیار می‌کرد به عنوان توزیع منطقه‌ای سیلاب انتخاب شد. نکویی برازش پنج توزیع آماری لجستیک نعمیم یافته، مقادیر حدی تعمیم یافته، نرمال تعمیم یافته، پیرسون تیپ III و پارتولی تعمیم یافته مورد بررسی قرار گرفت تا از میان این توزیع‌ها، بهترین گزینه بر اساس شاخص نکویی برازش Z به عنوان توزیع فراوانی منطقه‌ای سیلاب برگزیده شود. هر یک از توزیع‌های دو پارامتری شناخته شده نیز، معمولاً حالت خاص از این توزیع‌های سه پارامتری هستند. در جدول (۴-۴) مقدار شاخص نکویی برازش Z محاسبه شده برای هر یک از توزیع‌های سه پارامتری مورد آزمایش، ارائه شده است.

همان گونه که در جدول (۴-۴) مشاهده می‌شود، در حالت دو منطقه‌ای برای منطقه‌ی اول توزیع پیرسون تیپ III و برای منطقه‌ی دوم توزیع مقادیر حدی تعمیم یافته به عنوان توزیع منطقه‌ای انتخاب شدند. البته آن چنان که مشاهده می‌شود، برای منطقه‌ی اول توزیع پارتولی تعمیم یافته و برای منطقه‌ی دوم توزیع‌های نرمال تعمیم یافته، پیرسون تیپ III و لجستیک تعمیم یافته نیز در محدوده‌ی مجاز شاخص نکویی برازش Z برای انتخاب به عنوان توزیع منطقه‌ای قرار گرفتند.

جدول (۴-۴) مقدار شاخص نکویی برازش Z برای توزیع‌های مورد آزمایش در هر یک از مناطق حالت دو منطقه‌ای

توزيع فراوانی					
شماره منطقه	لجستیک تعمیم یافته	مقادیر حدی تعمیم یافته	نرمال تعمیم یافته	پیرسون تیپ III	پارتولی تعمیم یافته
۱	۴/۲۱	۲/۸۰	۱/۷۴	*-۰/۱۲	-۱/۰۳
۲	۱/۵۰	*۰/۰۷	-۰/۱۷	-۰/۶۵	-۳/۳۰

*توزیع منطقه‌ای برگزیده

۴-۵-۳-۸- برآورد چندکهای منطقه‌ای سیلاب و دبی سیلاب ایستگاهها

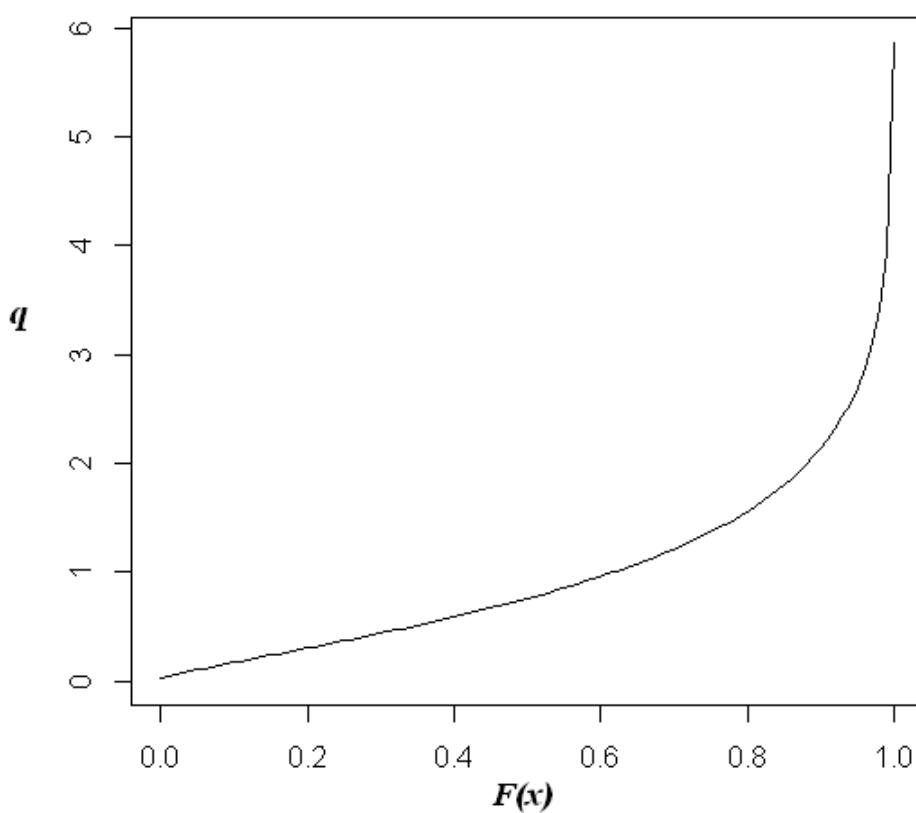
پس از انتخاب توزیع فراوانی منطقه‌ای سیلاب برای هر یک از مناطق، برآورد پارامترهای توزیع مربوط به هر منطقه با استفاده از تئوری گشتاورهای خطی انجام گرفت. مقادیر برآورد شده‌ی پارامترهای توزیع پیرسون تیپ III برای منطقه‌ای اول و توزیع مقادیر حدی تعمیم یافته برای منطقه‌ای دوم در جدول (۴-۵) آمده است.

پس از برآورد پارامترهای مربوط به هر یک از توزیع‌های منطقه‌ای منتخب، مقادیر چندکهای سیلاب منطقه‌ای به ازای مقادیر مختلف احتمال عدم تجاوز محاسبه شدند. منحنی‌های مربوط به مقادیر چندکهای سیلاب منطقه‌ای (q) یا همان دبی بدون بعد سیلاب منطقه‌ای، در برابر احتمال عدم تجاوز متناظر ($F(x)$) خود در شکل‌های (۱۲-۴) و (۱۳-۴) نمایش داده شده‌اند.

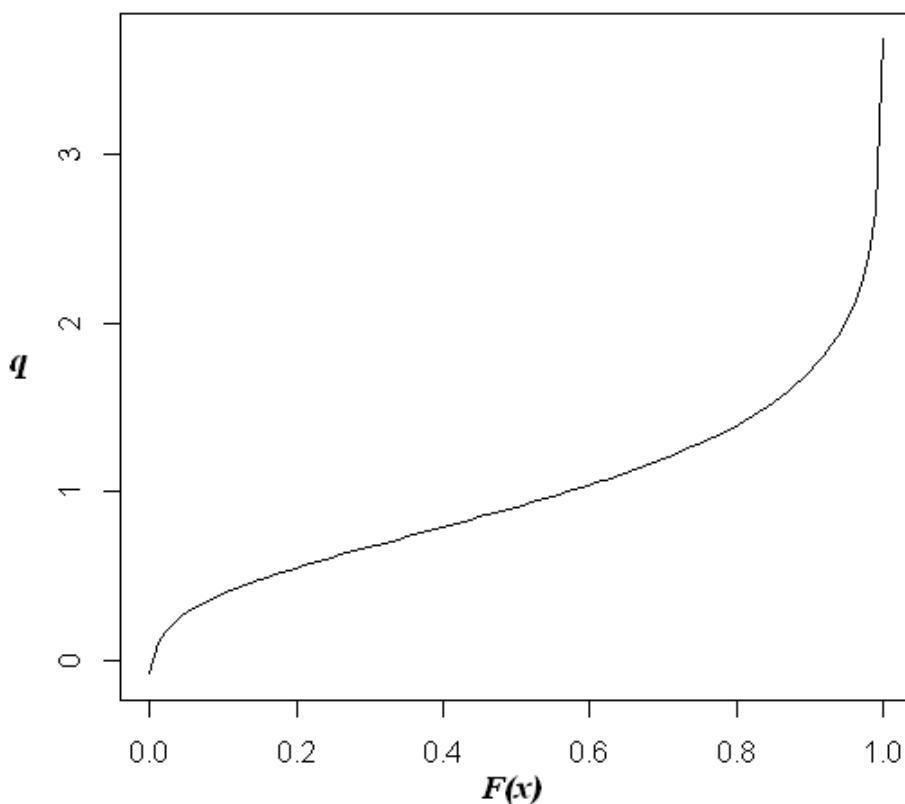
برای هر سایت یا ایستگاه، توزیع فراوانی منطقه‌ای با استفاده از عاملی که شاخص سیلاب یا سیلاب نمایه نامیده می‌شود، تبدیل به توزیع فراوانی ویژه‌ی هر سایت یا ایستگاه می‌شود. در واقع با ضرب عامل سیلاب نمایه‌ی هر ایستگاه در چندکهای منطقه‌ای سیلاب با دوره‌های بازگشت مختلف، برآورد سیلاب با دوره‌های بازگشت موردنظر انجام می‌شود. معمولاً از میانگین آمار سالانه‌ی حداکثر دبی لحظه‌ای سیلاب هر سایت به عنوان عامل سیلاب نمایه می‌شود. در این مطالعه نیز، همین عامل به عنوان عامل سیلاب نمایه انتخاب شده و برای برآورد سیلاب در ایستگاه‌ها مورد استفاده قرار گرفت.

جدول (۴-۵) مقادیر برآورد شده‌ی پارامترهای توزیع‌های برگزیده‌ی منطقه‌ای

شماره منطقه	توزیع منطقه‌ای	پارامترهای توزیع	برآورد مقدار پارامتر
۱	μ		
۰/۸۶۷۶	σ	پیرسون تیپ III	۱
۱/۷۶۷۴	γ		
۰/۷۵۳۶	ζ		
۰/۴۲۹۵	α	مقادیر حدی تعمیم یافته	۲
۰/۰۰۳۷	k		



شکل (۱۲-۴) چندک‌های سیلاب منطقه‌ای (q) (بدون بعد) منطقه‌ی اول در برابر احتمال عدم تجاوز متناظر ($F(x)$)



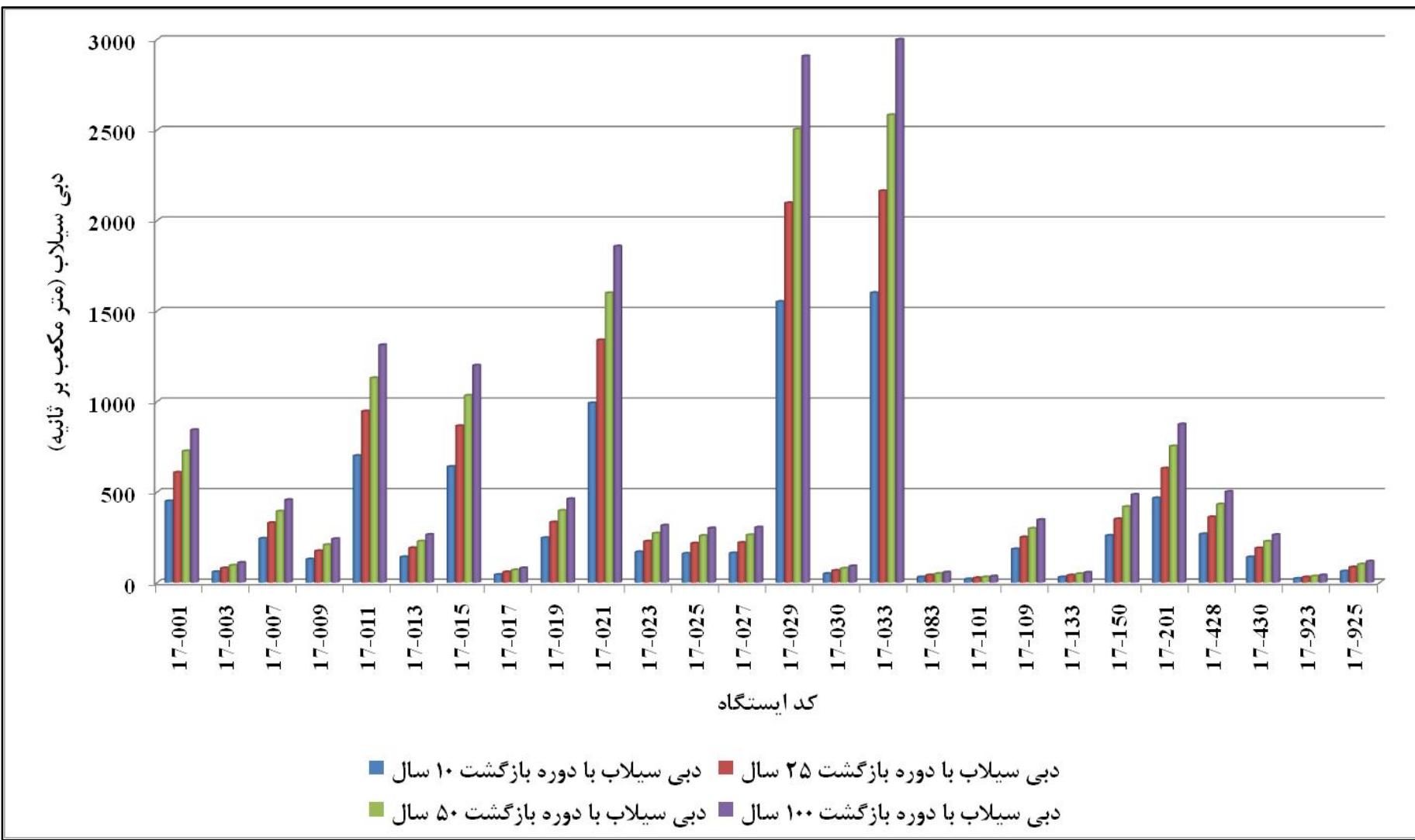
شکل (۱۳-۴) چندک‌های سیلاب منطقه‌ای (q) (بدون بعد) منطقه‌ی دوم در برابر احتمال عدم تجاوز متناظر ($F(x)$)

همان طور که پیش تر گفته شد، برآورد سیلاب در تحلیل فراوانی منطقه‌ای تا زمانی قابل اعتماد است که تعداد سال‌های آماری موجود در ایستگاه‌های یک منطقه بزرگ‌تر یا مساوی پنج برابر دوره‌ی بازگشت مورد نظر جهت برآورد بزرگی سیلاب باشد (رید و همکاران، ۱۹۹۹). بنابراین با توجه به مجموع داده‌های آماری موجود در منطقه‌ی اول حدود ۶۰۰ ایستگاه- سال و داده‌های آماری موجود در منطقه‌ی دوم حدود ۲۵۰ ایستگاه- سال است، برآورد سیلاب برای منطقه‌ی اول تا دوره‌ی بازگشت حدود ۱۲۰ سال و برای منطقه‌ی دوم تا دوره‌ی بازگشت حدود ۵۰ سال قابل اعتماد است. لذا بر اساس سیلاب‌های معیار در زمینه‌ی مطالعات هیدرولوژی، برآورد سیلاب در ایستگاه‌های منطقه‌ی اول تا دوره‌ی بازگشت ۱۰۰ سال و در منطقه‌ی دوم تا دوره‌ی بازگشت ۵۰ سال انجام شد.

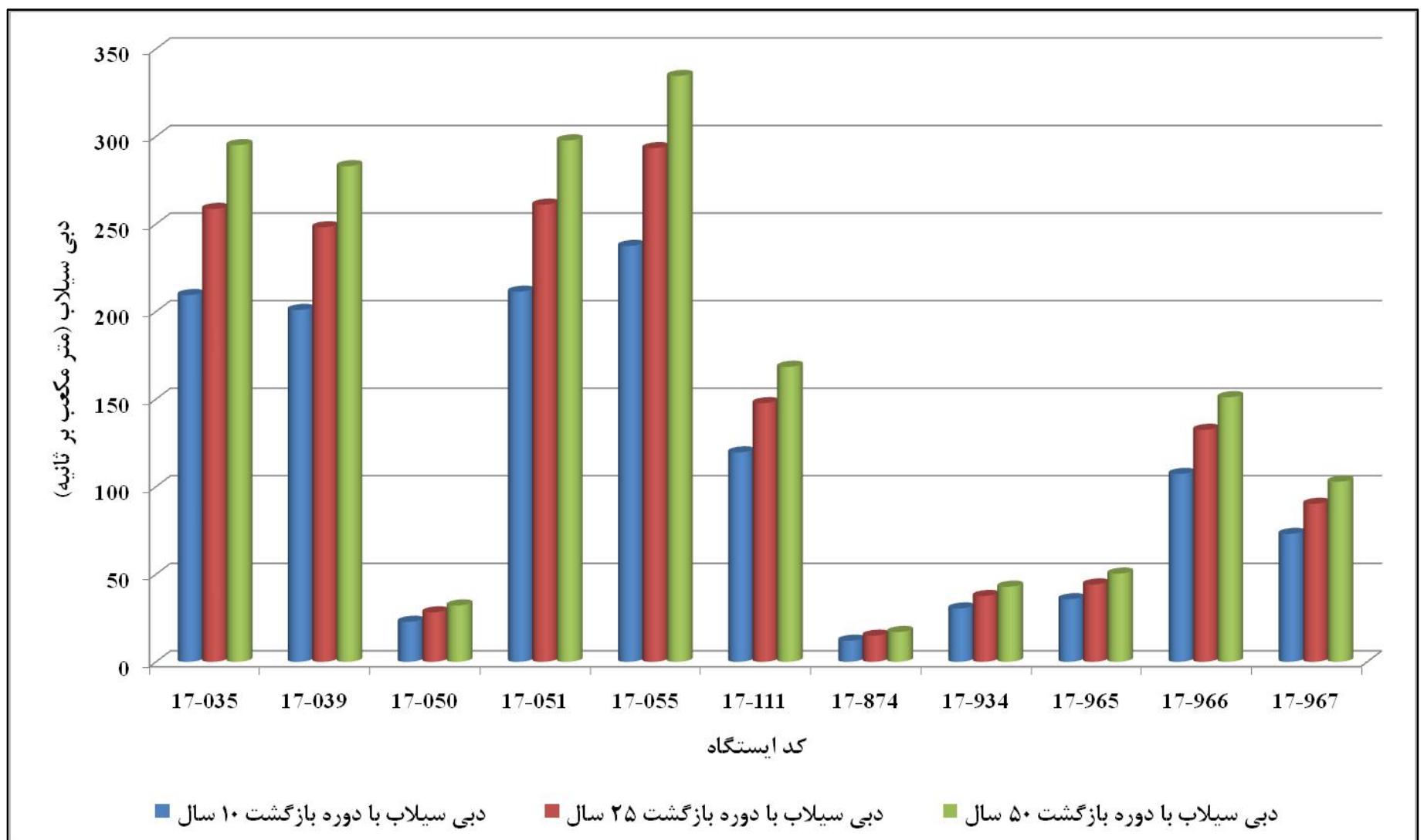
در شکل (۱۴-۴) مقادیر برآورد شده‌ی سیلاب با دوره‌های بازگشت ۱۰، ۲۵، ۵۰ و ۱۰۰ سال برای ایستگاه‌های موجود در منطقه‌ی اول نشان داده شده است. همچنین در شکل (۱۵-۴) مقادیر برآورد شده‌ی سیلاب با دوره‌های بازگشت ۱۰، ۲۵ و ۵۰ سال به نمایش در آمده است.

همان طور که در شکل (۱۴-۴) مشاهده می‌شود، کمترین مقدار دبی سیلاب برآورد شده با دوره‌های بازگشت ۱۰، ۲۵، ۵۰ و ۱۰۰ سال در منطقه‌ی اول به ترتیب برابر $18/4$ ، $24/88$ ، $24/71$ و $34/51$ متر مکعب بر ثانیه و مربوط به ایستگاه ۱۷-۱۰۱ می‌باشد، در حالی که بیشترین مقادیر دبی سیلاب با دوره‌های بازگشت مذکور به ترتیب مساوی $1599/97$ ، $2162/34$ ، $2582/3$ و $2999/05$ متر مکعب بر ثانیه بوده و به ایستگاه ۱۷-۰۳۳ تعلق دارد.

همچنین بنا بر نتایج مندرج در شکل (۱۵-۴)، کوچک‌ترین مقدار دبی برآورد شده‌ی سیلاب با دوره‌های بازگشت ۱۰، ۲۵ و ۵۰ سال در منطقه‌ی دوم به ایستگاه ۱۷-۸۷۴ اختصاص داشته و به ترتیب معادل $12/07$ ، $14/9$ و 17 متر مکعب بر ثانیه می‌باشد. در مقابل بزرگ‌ترین مقادیر برآورد شده‌ی دبی سیلاب با دوره‌های بازگشت مورد نظر در این منطقه، مربوط به ایستگاه ۱۷-۰۵۵ به ترتیب با مقادیر $237/48$ ، $237/28$ ، $293/28$ و $334/54$ متر مکعب بر ثانیه است.



شکل (۱۴-۴) مقادیر برآورد شدهی سیلاب با دوره‌های بازگشت ۱۰، ۲۵، ۵۰ و ۱۰۰ سال برای ایستگاه‌های هیدرومتری موجود در منطقه‌ی اول



شکل (۱۵-۴) مقدار برآورد شدهی سیلاب با دوره های بازگشت ۱۰، ۲۵ و ۵۰ سال برای ایستگاه های هیدرومتری موجود در منطقه دوم

۴-۶-۴- خوشبندی با استفاده از الگوریتم فازی c-means

۴-۱-۶- ایستگاه‌ها مورد مطالعه و تعیین محدوده تعداد مناطق

همان طور که در بخش (۳-۴) اشاره شد، در میان ۴۰ ایستگاه مورد مطالعه، مقدار شاخص ناجوری D در ۲ ایستگاه، از مقدار مجاز این شاخص بیشتر بوده و لذا این دو ایستگاه که با کدهای ۳۷-۰۳۷ و ۹۲۶-۱۷ مشخص می‌شوند، از ادامه‌ی فرآیند مطالعه کنار گذاشته شدند.

آن چنان که پیش‌تر در بخش (۳-۲-۱) ذکر شد، هنگامی که $1 = \mu$ باشد، پاسخ الگوریتم خوشبندی فازی c-means در تئوری به جواب K-means سنتی همگرا می‌شود. از این روش مشاهده شد که ایستگاه ۴۱-۰۱۷ که در منطقه‌بندی با استفاده از الگوریتم‌های ترکیبی موجب افزایش شاخص‌های ناهمگنی و ایجاد ناهمگنی در مناطق تشکیل شده می‌شد، در اینجا نیز اثری مشابه از خود به جای می‌گذارد. بنابراین این ایستگاه از ادامه‌ی فرآیند محاسبه و تحلیل کنار گذاشته شد و مراحل خوشبندی با ۳۷ ایستگاه باقی‌مانده ادامه یافت.

در ادامه، اختصاص ۳۷ ایستگاه مورد بررسی به ۲ تا ۵ منطقه با استفاده از الگوریتم خوشبندی فازی c-means مورد مطالعه قرار گرفت.

۴-۶-۴-۲- تعیین مراکز اولیه‌ی خوشها

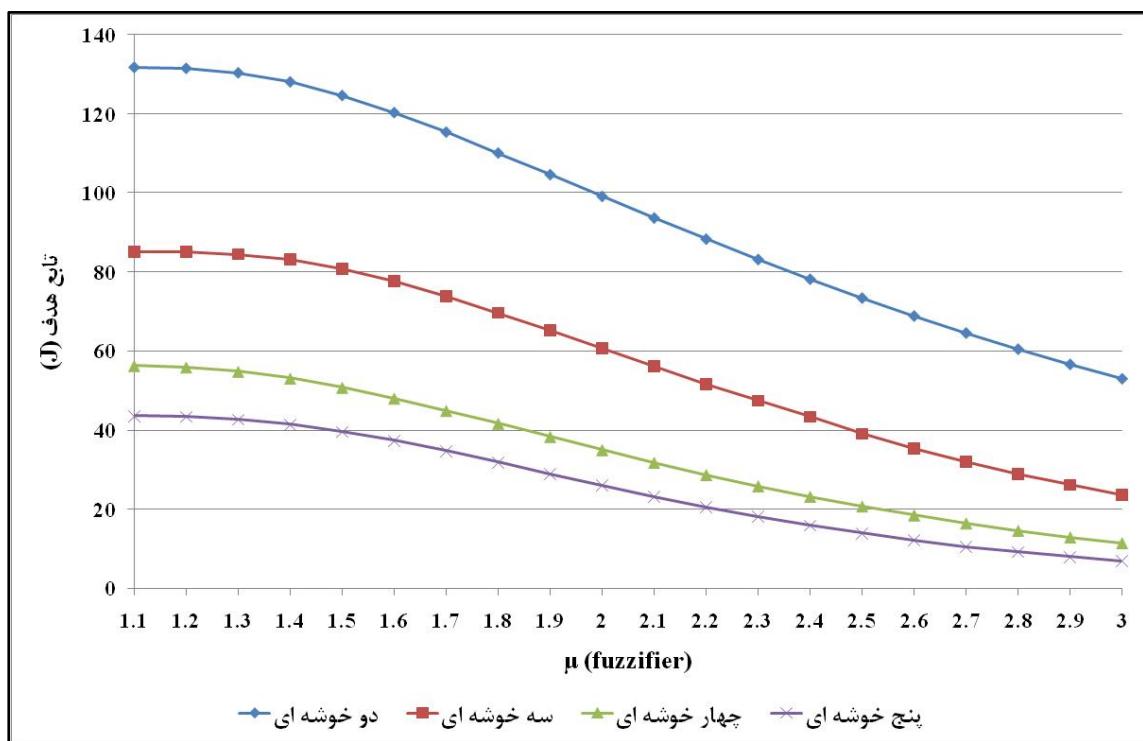
با توجه به عملکرد قابل قبول الگوریتم Ward در تعیین مراکز اولیه‌ی خوشها برای استفاده در الگوریتم K-means که در بررسی منطقه‌بندی حوضه با استفاده از روش‌های خوشبندی ترکیبی مشاهده شد، در این بخش نیز برای تعیین مراکز اولیه‌ی مورد استفاده برای الگوریتم c-means از مراکز خوشها حاصل شده به وسیله‌ی الگوریتم Ward استفاده شد. تعداد دفعات تکرار الگوریتم عددی c-means برابر ۱۰۰۰ مرتبه در نظر گرفته شد.

۴-۶-۳- محدوده‌ی تغییرات μ

پال و بزدک (۱۹۹۵، ص. ۳۷۰) به این نکته اشاره کردند که الگوریتم خوشبندی فازی c-means محدوده‌ی تغییرات $2.5 \leq \mu \leq 1.5$ بهترین عملکرد را از خود نشان می‌دهد. در این مطالعه، حساسیت نتایج حاصل از کاربرد الگوریتم c-means نسبت به تغییر مقدار μ در محدوده‌ی ۱/۱ تا ۳ با گام‌هایی معادل ۰/۱ مورد آزمایش قرار گرفته است.

۴-۶-۴- محاسبه‌ی مقادیر تابع هدف و بررسی تغییرات آن

مقدار تابع هدف برای تعداد ۲ تا ۵ خوش و در محدوده‌ی تغییرات $1.1 \leq \mu \leq 3.0$ محاسبه شده و مورد بررسی قرار گرفت. در شکل (۱۶-۴) مقادیر تابع هدف و نحوی تغییرات آن برای موارد و محدوده‌های ذکر شده نشان داده شده است.



شکل (۱۶-۴) تغییرات مقدار تابع هدف الگوریتم خوشبندی c-means با تغییر در تعداد خوشها و مقدار μ

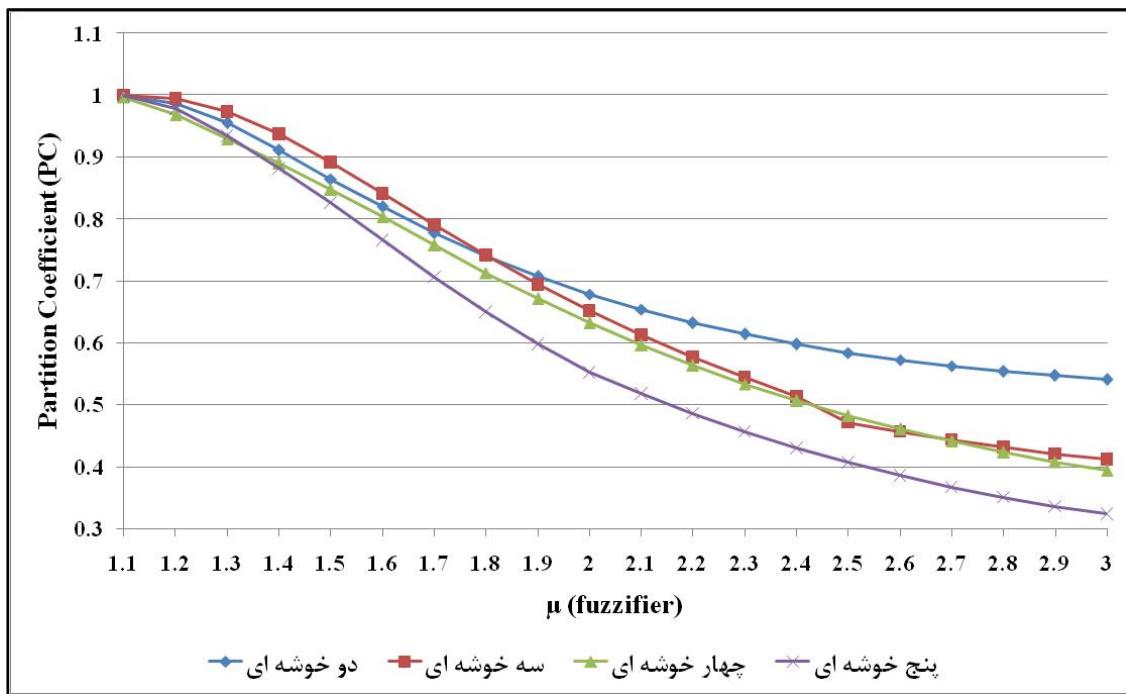
بررسی نتایج منعکس شده در شکل (۱۶-۴) بیانگر دو نکته است. نخست این که برای یک مقدار ثابت μ , مقدار تابع هدف با افزایش تعداد خوشها کاهش می‌یابد، و دوم این که برای تعداد ثابتی از خوشها با افزایش مقدار μ , مقدار تابع هدف کاهش می‌یابد. به لحاظ نظری، هنگامی که تمام بردار-های مشخصه یا ایستگاهها به یک خوش تعلق یابند، مقدار تابع هدف بیشینه خواهد بود و در مقابل زمانی که هر بردار مشخصه یا ایستگاه تشکیل یک خوشی مجزا با درجهی عضویت ۱ در آن را بدهد، مقدار تابع هدف برابر صفر خواهد بود.

۴-۵-۶- محاسبه‌ی شاخص‌های صحت خوشبندی فازی

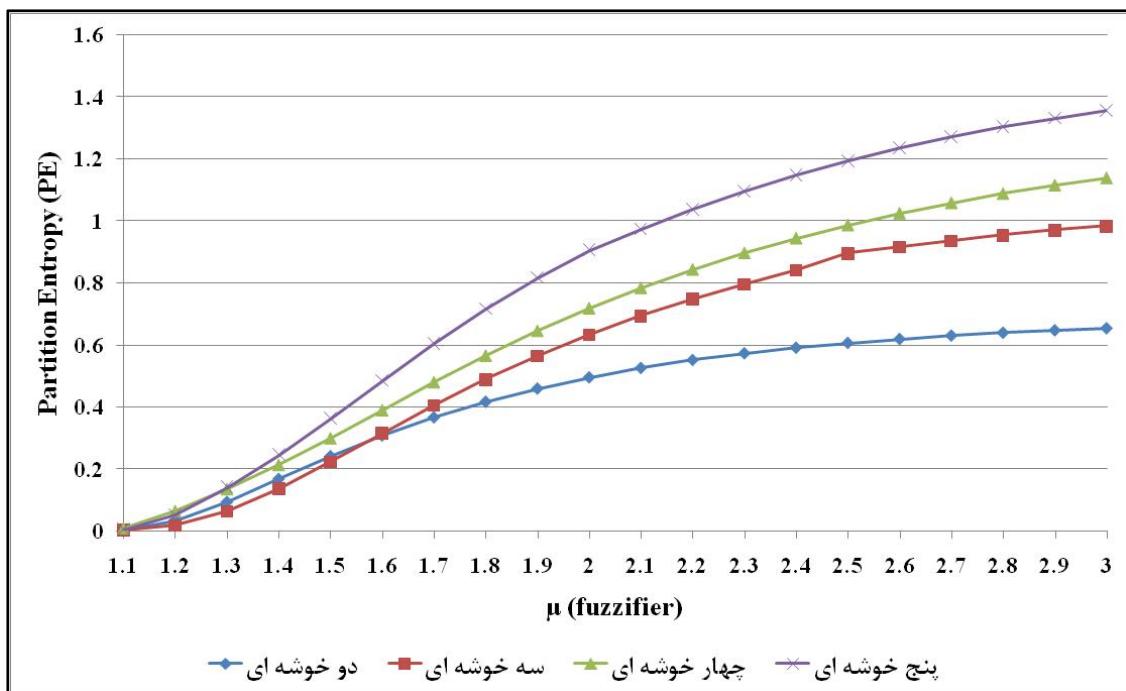
در این مطالعه، شش شاخص صحت خوشبندی فازی، شامل شاخص فوکویاما و سوگنو (V_{FS}), ضریب افزار (V_{PC}), انتروپی افزار (V_{PE}), شاخص عملکرد فازی (V_{FPI}), شاخص ژی-بنی توسعه یافته (V_K) و شاخص کوان ($V_{XB,m}$) محاسبه شده و مورد بررسی قرار گرفتند. توصیف این شاخص‌ها پیش‌تر در بخش (۳-۲-۳) آمده است. هر یک از این شاخص‌ها برای تعداد ۲ تا ۵ خوش و در محدوده‌ی تغییرات $3.0 \leq \mu \leq 1.1$ مورد محاسبه قرار گرفتند.

در شکل‌های (۱۷-۴) تا (۲۲-۴) مقادیر محاسبه شده و نحوه‌ی تغییرات شاخص ضریب افزار (V_{PC}), شاخص انتروپی افزار (V_{PE}), شاخص عملکرد فازی (V_{FPI}) و شاخص فوکویاما و سوگنو (V_{FS}), با تغییر مقدار μ در محدوده‌ی $1/1$ تا 3 برای تعداد ۲ تا ۵ خوش به نمایش در آمده است. همان گونه که در شکل (۱۷-۴) مشاهده می‌شود، در محدوده‌ی $1.8 \leq \mu \leq 1.1$, مقدار محاسبه شده‌ی ضریب افزار (V_{PC}) برای حالت سه خوش‌های بیش از سایر حالات است و در گستره‌ی $3.0 \leq \mu \leq 1.9$, این حالت دو خوش‌های است که بالاترین مقادیر را اختیار می‌کند.

بنا بر نتایج مندرج در شکل (۱۸-۴)، در فاصله‌ی $1.5 \leq \mu \leq 1.1$, شاخص انتروپی افزار برای تعداد ۳ خوش کمترین مقادیر خود را دارد، اما در فاصله‌ی $3.0 \leq \mu \leq 1.6$, این حالت دو خوش‌های است که پایین‌ترین مقادیر محاسبه شده‌ی این شاخص را به خود اختصاص می‌دهد.

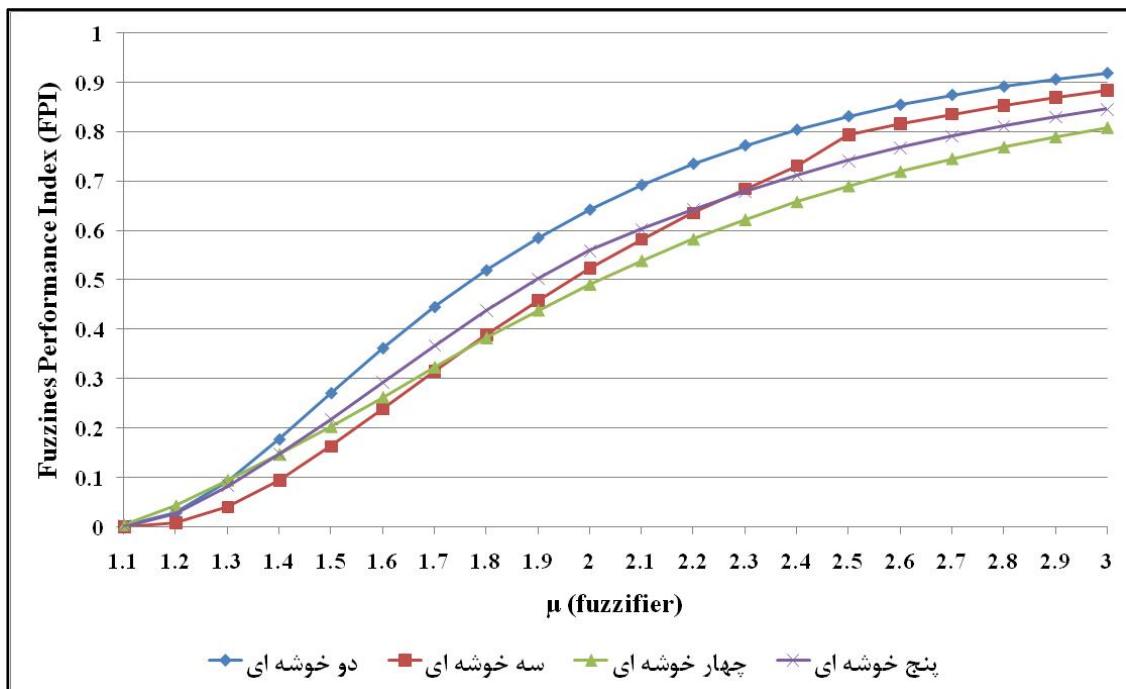


شکل (۱۷-۴) تغییرات مقدار ضریب افزای (V_{PC}) با تغییر مقدار μ در فاصله‌ی ۱/۱ تا ۳ برای تعداد ۲ تا ۵ خوشه



شکل (۱۸-۴) تغییرات مقدار انتروپی افزای (V_{PE}) با تغییر مقدار μ در فاصله‌ی ۱/۱ تا ۳ برای تعداد ۲ تا ۵ خوشه

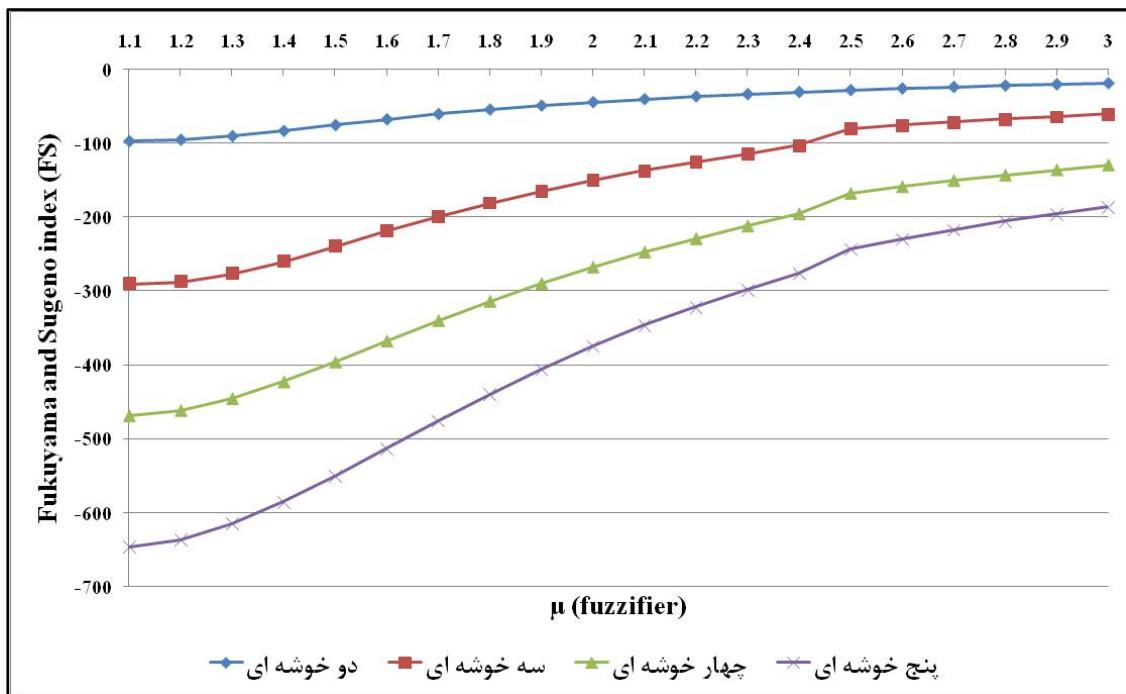
مطابق شکل (۱۹-۴)، در محدوده‌ی $1.1 \leq \mu \leq 1.7$ کمترین مقادیر شاخص عملکرد فازی (V_{FPI}) در حالت سه خوشه‌ای ظاهر می‌شود، اما در گستره‌ی $1.8 \leq \mu \leq 3.0$ ، کوچکترین مقادیر شاخص مذکور به حالت چهار خوشه‌ای تعلق می‌گیرد.



شکل (۴-۱۹) تغییرات مقدار شاخص عملکرد فازی (V_{FPI}) با تغییر مقدار μ در فاصله‌ی ۱/۱ تا ۳ برای ۲ تا ۵ خوشه

در مورد سه شکل اخیر، لازم به ذکر است که افزار بهینه متناظر با مقدار بیشینه‌ی ضربی افزار (V_{PC}) یا مقدار کمینه‌ی انتروپی افزار (V_{PE}) و شاخص عملکرد فازی (V_{FPI}) است. نکته‌ای که در شکل (۴-۱۷) مشاهده می‌شود این است که به ازای تعداد ثابتی از خوشها همواره با افزایش مقدار μ ، مقدار ضربی افزار کاهش می‌یابد. همچنین با توجه به شکل‌های (۱۸-۴) و (۱۹-۴) می‌توان دریافت که در صورت ثابت ماندن تعداد خوشها، با افزایش مقدار μ ، منحنی تغییرات انتروپی افزار و شاخص عملکرد فازی، حالت اکیداً صعودی خواهند داشت. از این رو می‌توان گفت که انتخاب مقدار بهینه‌ی μ برای تعداد مشخصی از خوشها بر مبنای مقادیر محاسبه شده‌ی این سه شاخص امکان پذیر نیست.

تغییرات مقدار شاخص فوکویاما و سوگنو (V_{FS}) به ازای تغییر مقدار μ در فاصله‌ی ۱/۱ تا ۳ برای ۲ تا ۵ خوشه در شکل (۲۰-۴) نشان داده شده است. برای مقدار μ یکسان، مقدار شاخص فوکویاما و سوگنو برای حالت پنج خوشه‌ای کمترین مقدار و برای حالت دو خوشه‌ای بیشترین مقدار را دارد.

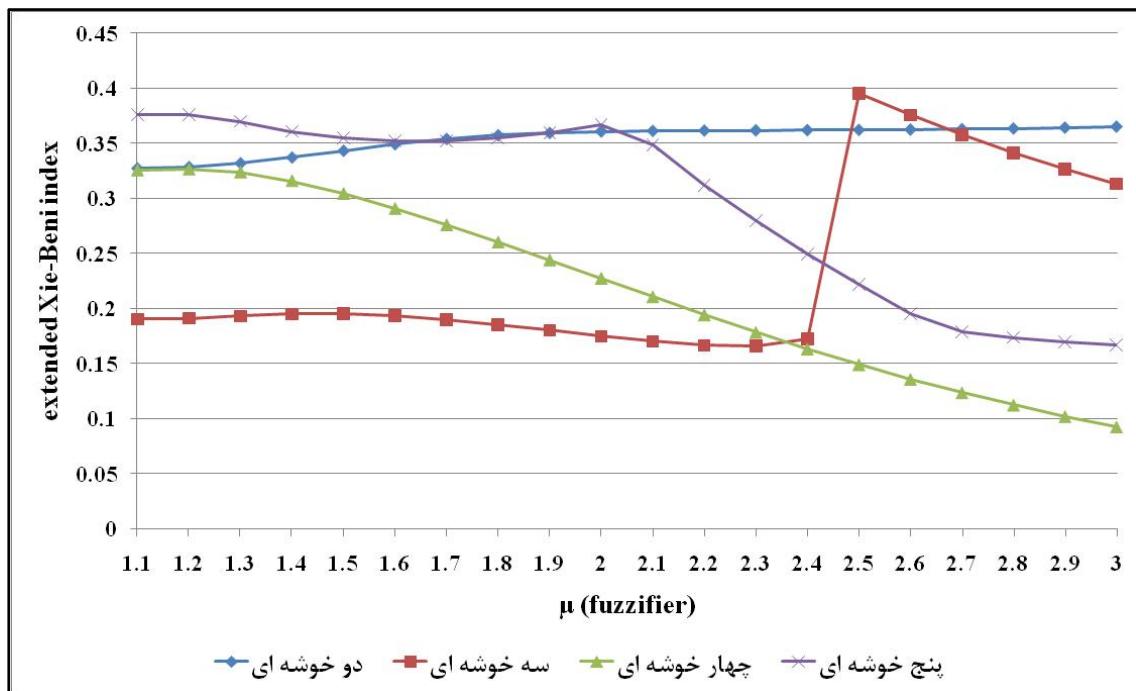


شکل (۲۰-۴) تغییرات مقدار شاخص فوکویاما و سوگنو (V_{FS}) با تغییر μ در فاصله‌ی ۱/۱ تا ۳ برای ۲ تا ۵ خوشه

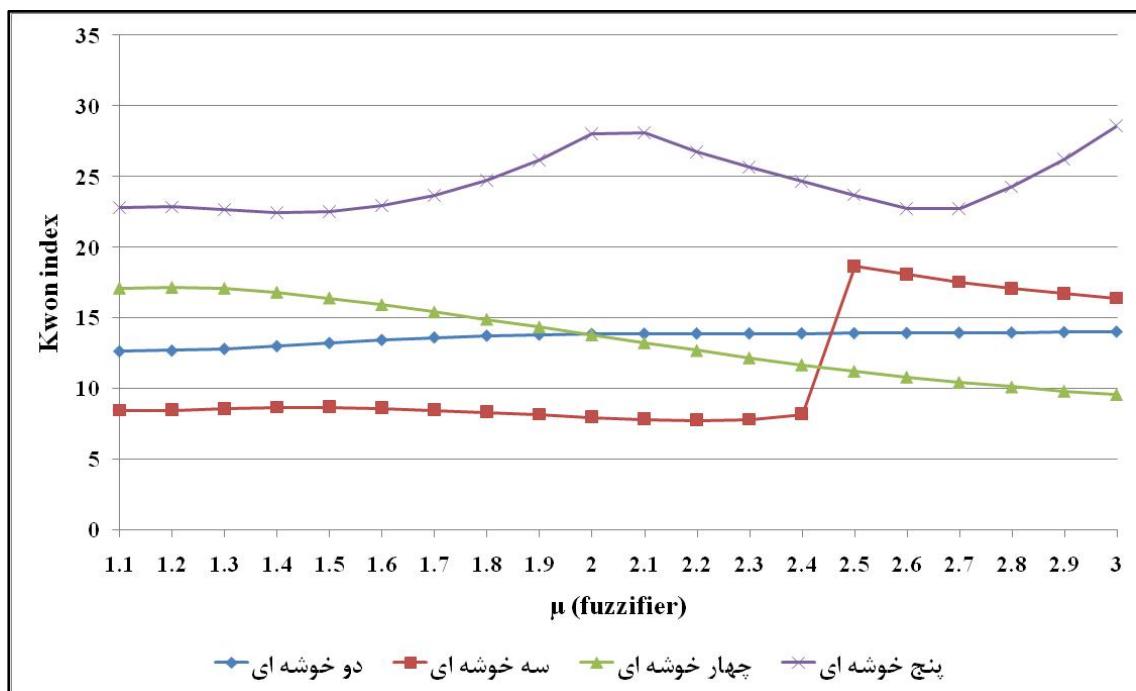
همان طور که در شکل دیده می‌شود، مقدار این شاخص با افزایش تعداد خوشه‌ها به صورت یکنواخت کاهش می‌یابد و از این رو به نظر می‌رسد با توجه به عملکرد مشابه این شاخص در مطالعات مشابه پیشین، شاخص مذکور برای تعیین تعداد بهینه‌ی خوشه‌ها برای منطقه‌بندی حوضه‌ی آبریز مورد مطالعه کارایی چندانی ندارد.

در شکل (۲۱-۴) نمودار تغییرات مقدار شاخص ژی-بنی توسعه یافته به نمایش در آمد است. آن چنان که در شکل مشاهده می‌شود، در فاصله‌ی $2.3 \leq \mu \leq 1.1$ کمترین مقادیر شاخص مذکور که متناظر با افزار بهینه هستند به حالت سه خوشه‌ای تعلق دارند، اما در بازه‌ی $3.0 \leq \mu \leq 2.4$ این حالت جای خود را به حالت چهار خوشه‌ای می‌دهد.

مطابق شکل (۲۲-۴) که نشان دهنده‌ی تغییرات مقدار شاخص کوان به ازای تغییرات μ برای تعداد ۲ تا ۵ خوشه است، کمترین مقدار این شاخص در بازه‌های $2.4 \leq \mu \leq 1.1$ و $2.5 \leq \mu \leq 3.0$ به ترتیب مربوط به حالت‌های سه خوشه‌ای و چهار خوشه‌ای است. همچنین مشاهده می‌شود که این شاخص در حالت پنج خوشه‌ای همواره مقادیر بزرگ‌تری را نسبت به دیگر حالت‌ها اختیار می‌کند.



شکل (۲۱-۴) تغییرات شاخص ژی-بنی توسعه یافته ($V_{XB,m}$) با تغییر μ در فاصله‌ی ۱/۱ تا ۳ برای ۲ تا ۵ خوش



شکل (۲۲-۴) تغییرات مقدار شاخص کوان (V_K) با تغییر مقدار μ در فاصله‌ی ۱/۱ تا ۳ برای ۲ تا ۵ خوش

نکته‌ی مشترکی که در شکل‌های (۱۷-۴) تا (۲۲-۴) جلب توجه می‌کند، جهش قابل توجهی است که در مقدار شاخص‌های صحت خوشبندی در حالت سه خوشه‌ای و با افزایش مقدار μ از ۲/۴

به ۲/۵ اتفاق می‌افتد. این نکته در مورد تمامی شاخص‌ها باشد و ضعف مشاهده می‌شود، اما مقدار جهش مذکور در مورد شاخص‌های ذی-بنی توسعه یافته و کوان چشمگیرتر از سایر شاخص‌ها است. البته تغییر مقدار μ تنها علت ظهور این جهش نیست، بلکه آن چه که تأثیر بیشتری در اتفاق افتادن این جهش دارد، تغییر نسبتاً قابل توجه مؤلفه‌های بردار مشخصه‌ی مرکز خوش‌های دوم است. این تغییر قابل توجه به ویژه در محدوده‌ی تکرارهای ۱۰۰ تا ۱۵۰ مشاهده می‌شود و پس از تکرار مرتبه‌ی ۲۰۰ تغییر خاصی در مورد مراکز خوش‌های مشاهده نمی‌شود.

۴-۶-۶- آستانه‌ی عضویت فازی و اختصاص ایستگاه‌ها به خوش‌های

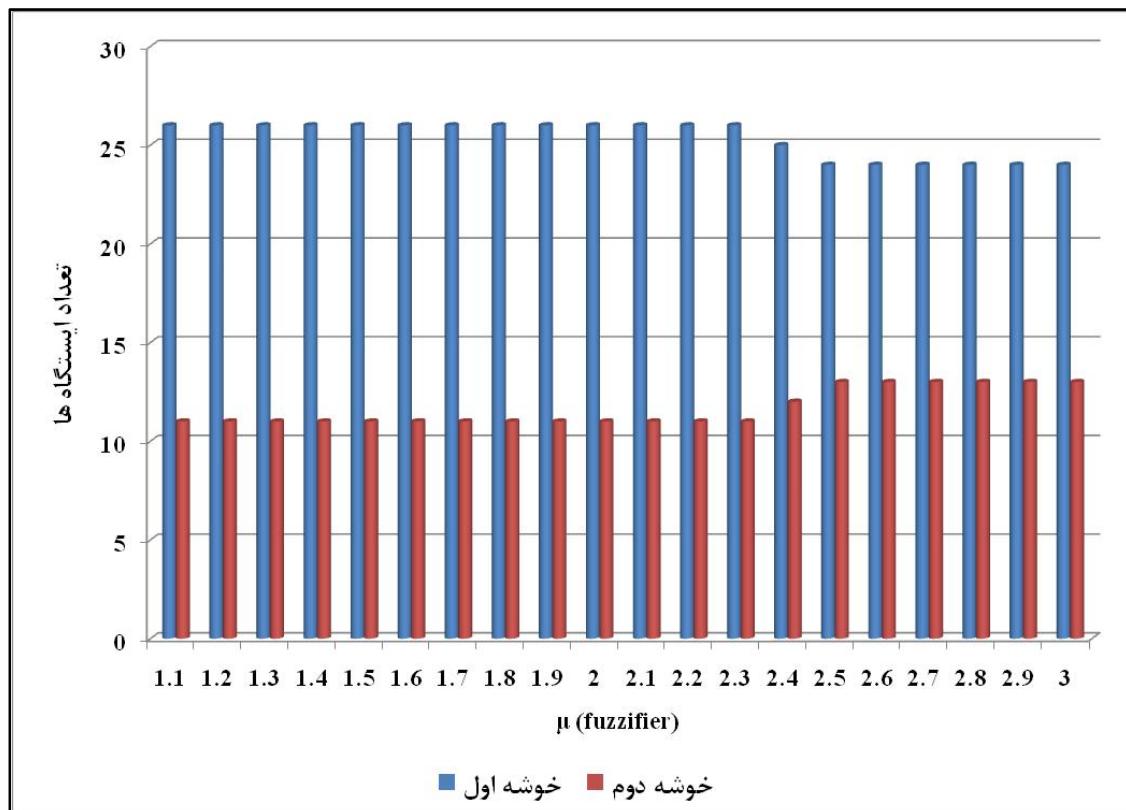
اختصاص بردارهای مشخصه‌ی مربوط به سایت‌های مورد مطالعه به خوش‌های، همان گونه که در بخش (۱-۲-۳) تشریح شد، می‌تواند به دو صورت سخت و فازی اجرا شود. در حالت سخت هر بردار مشخصه تنها به یک خوش‌های تعلق می‌گیرد. این خوش‌های می‌تواند خوش‌های باشد که بالاترین درجه‌ی عضویت بردار مشخصه مربوط به آن خوش‌های مرکز آن کمترین فاصله را با بردار مشخصه‌ی مورد نظر دارد. در حالت اختصاص فازی، سایت‌هایی که عضویت آن‌ها در یک خوش‌های از مقدار آستانه‌ی معینی تجاوز می‌کند، به آن خوش‌های اختصاص می‌یابند. در فازی‌ترین افزار، عضویت‌های یک بردار مشخصه در تمام خوش‌های بردار $c/1$ خواهد بود. از این رو مقدار $c/1$ به عنوان یک انتخاب قابل قبول برای عضویت فازی آستانه شناخته می‌شود.

مطالعات پیشین نشان می‌دهد که نتایج مربوط به اختصاص به صورت سخت، خوش‌هایی هستند که تا حدود زیادی مشابه خوش‌های تولید شده توسط روش‌های خوش‌بندی سخت مانند K-means هستند. همچنین در صورت استفاده از این روش، مزیت عمده‌ی استفاده از الگوریتم‌های خوش‌بندی فازی که مشارکت بردارهای مشخصه در بیش از یک خوش‌های افزایش تعداد داده‌های آماری در هر منطقه به منظور تحلیل فراوانی منطقه‌ای سیلاب است، نادیده گرفته می‌شود. از این رو، در این مطالعه، روش اختصاص فازی بردارهای مشخصه به خوش‌های مورد استفاده قرار گرفت و بدین منظور

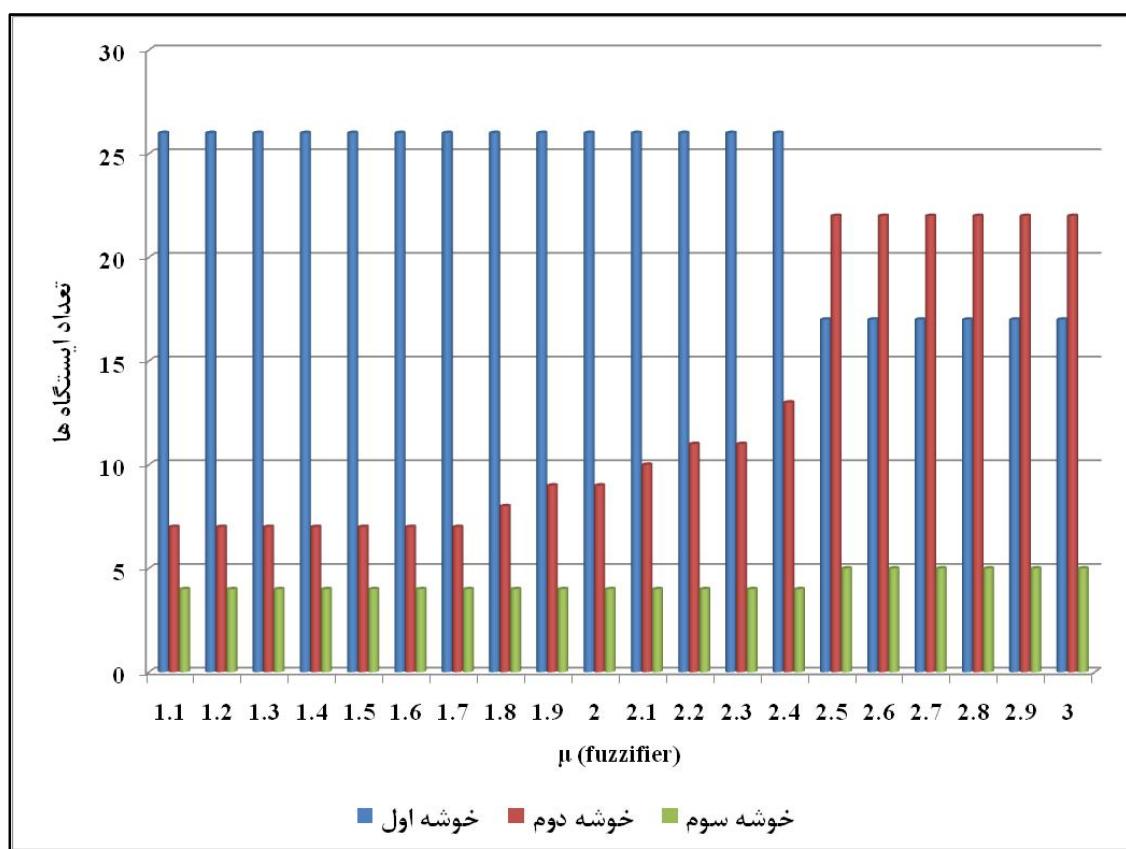
مقدار $c/1$ به عنوان یک انتخاب قابل قبول برای عضویت فازی آستانه برگزیده شد. با توجه به تغییر تعداد خوشهای مورد مطالعه (c) بین ۲ تا ۵، در این پژوهش مقدار $c/1$ برای حالت‌های ۲ تا ۵ خوشهای به ترتیب برابر $0/25$ ، $0/24$ و $0/2$ انتخاب شد.

شکل (۲۳-۴) تعداد ایستگاه‌ها یا سایت‌های اختصاص یافته به هر یک از خوشهای در حالت دو خوشهای را نشان می‌دهد. با توجه به این که مقدار آستانه‌ی عضویت برای این حالت برابر $0/5$ تعیین شده است و از آنجا که مجموع درجات عضویت بردار مشخصه مربوط به یک سایت همواره برابر یک می‌باشد، هیچ ایستگاهی نمی‌تواند همزمان به بیش از یک خوشه تعلق یابد، از این رو تعداد کل ایستگاه‌های اختصاص یافته به دو خوشه نیز همواره برابر تعداد کل ایستگاه‌های مورد مطالعه یعنی همان ۳۷ ایستگاه است. تنها در فاصله‌ی $2.5 \leq \mu \leq 2.3$ تعداد ایستگاه‌های متعلق به خوشه اول ۲ واحد کاهش یافته و به تعداد ایستگاه‌های متعلق به خوشه دوم ۲ واحد افزوده شده است.

شکل (۲۴-۴) مربوط به نمایش تعداد ایستگاه‌های موجود در خوشهای تشکیل شده در حالت سه خوشهای است. همان گونه که در شکل مشاهده می‌شود تعداد ایستگاه‌های تعلق یافته به خوشه‌های اول و سوم در فاصله‌ی $2.4 \leq \mu \leq 1.1$ ثابت باقی مانده است و تعداد ایستگاه‌های اختصاص یافته به خوشه دوم از ۷ به ۱۳ افزایش یافته است. اما هنگامی که مقدار μ از $2/4$ به $2/5$ افزایش می‌یابد، یک کاهش ناگهانی و قابل توجه در تعداد ایستگاه‌های موجود در خوشه اول و یک افزایش ناگهانی و چشمگیر در تعداد ایستگاه‌های موجود در خوشه دوم رخ می‌دهد. این تغییرات ناگهانی متناظر با همان جهش قابل توجهی است که در زمان افزایش مقدار μ از $2/4$ به $2/5$ در مورد شاخص‌های صحت خوشبندی فازی مشاهده شد. همزمان با این تغییرات در خوشهای اول و دوم، تعداد سایت‌های اختصاص یافته به خوشه سوم نیز یک واحد افزایش می‌یابد. البته برآیند کل این تغییرات، تنها یک واحد افزایش در تعداد کل ایستگاه‌های تعلق یافته به هر سه خوشه است. در ادامه نیز وضعیت ایستگاه‌های اختصاص یافته به هر یک از خوشهای در محدوده $3.0 \leq \mu \leq 2.5$ به همین صورت و بدون تغییر باقی مانده است.



شکل (۲۳-۴) تعداد ایستگاه‌های هیدرومتری متعلق به هر یک از خوشی‌ها در حالت دو خوشی‌ای

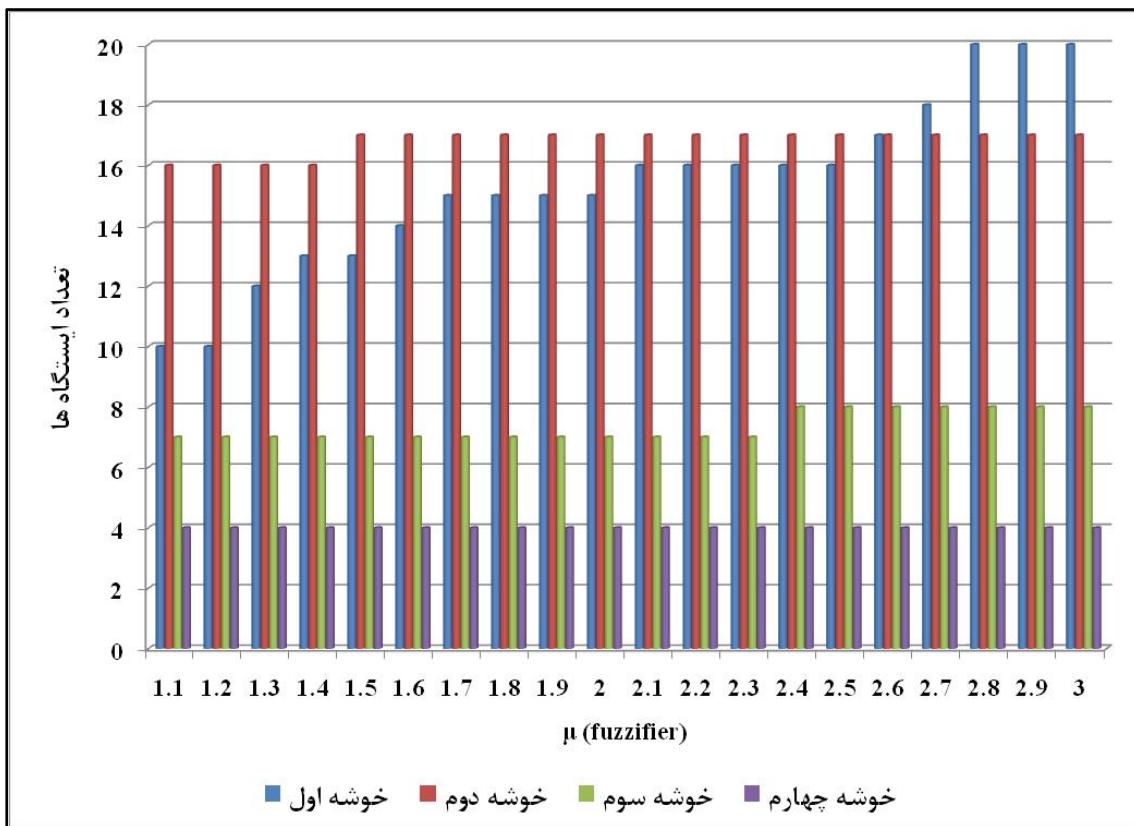


شکل (۲۴-۴) تعداد ایستگاه‌های هیدرومتری متعلق به هر یک از خوشی‌ها در حالت سه خوشی‌ای

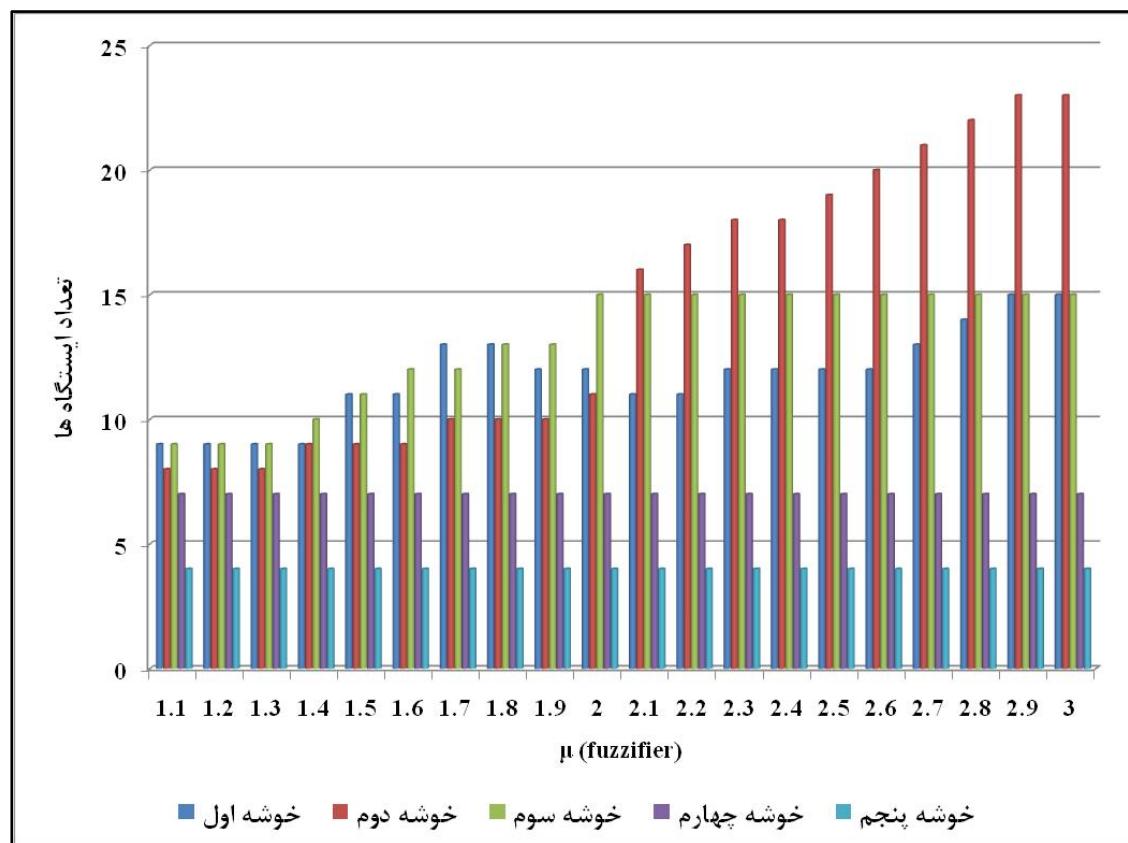
شکل (۲۵-۴) نشان دهندهی تعداد سایت‌های اختصاص یافته به هر یک از خوش‌های تشکیل شده در حالت چهار خوش‌های است. مطابق شکل، تعداد سایت‌های تعلق گرفته به خوش‌هی اول در محدوده‌ی مورد بررسی تقریباً به طور مداوم افزایش یافته است، به طوری که تعداد این سایت‌ها که در $\mu = 1.1$ برابر ۱۰ بوده است، در $\mu = 3.0$ به ۲۰ افزایش یافته است. تعداد ایستگاه‌های متعلق به خوش‌هی دوم در سراسر گستره‌ی مورد مطالعه، تنها در یک مورد، در زمان افزایش مقدار μ از $1/4$ به $1/5$ ، از ۱۶ ایستگاه به ۱۷ ایستگاه افزایش یافته است. خوش‌هی سوم نیز همانند خوش‌هی دوم تنها با یک مورد افزایش تعداد ایستگاه‌ها مواجه است که افزایش تعداد ایستگاه‌ها از ۷ عدد به ۸ عدد در هنگامی است که مقدار μ از $2/3$ به $2/4$ افزایش یافته است. در مورد خوش‌هی چهارم همان‌طور که در شکل مشاهده می‌شود، تعداد سایت‌های اختصاص یافته به آن همواره برابر ۴ است.

در شکل (۲۶-۴) چگونگی تغییرات تعداد ایستگاه‌های اختصاص یافته به هر یک از خوش‌های تشکیل شده در حالت پنج خوش‌های به نمایش در آمده است. تعداد ایستگاه‌های متعلق به خوش‌هی اول پس از ثابت ماندن در محدوده‌ی $1.4 \leq \mu \leq 1.8$ ، در فاصله‌ی $1.1 \leq \mu \leq 1.4$ ، روندی صعودی را در پیش گرفته است. سپس در گستره‌ی $2.1 \leq \mu \leq 2.8$ نزول کرده و پس از آن مجدداً افزایش یافته است. تعداد سایت‌های متعلق به این خوش‌ه که در ابتدا برابر ۹ عدد بوده است، در انتهای 15 عدد رسیده است. اما تعداد سایت‌های اختصاص یافته به خوش‌هی دوم از ابتدا تا انتهای روندی صعودی را دنبال کرده است. تعداد سایت‌ها متعلق به این خوش‌ه در زمان افزایش μ از ۲ به $2/1$ ، افزایشی ۵ واحدی را نشان داده و پس از آن نیز با سرعت قابل توجهی به رشد خود ادامه داده است. تعداد ایستگاه‌های متعلق به خوش‌هی سوم در فاصله‌ی $2.0 \leq \mu \leq 2.4$ از ۹ به 15 افزایش یافته و پس از آن ثابت مانده است. مطابق شکل، تعداد ایستگاه‌های اختصاص یافته به خوش‌های چهارم و پنجم در طول بررسی‌های انجام شده، تغییری نیافته و به ترتیب برابر ۷ و ۴ باقی مانده است.

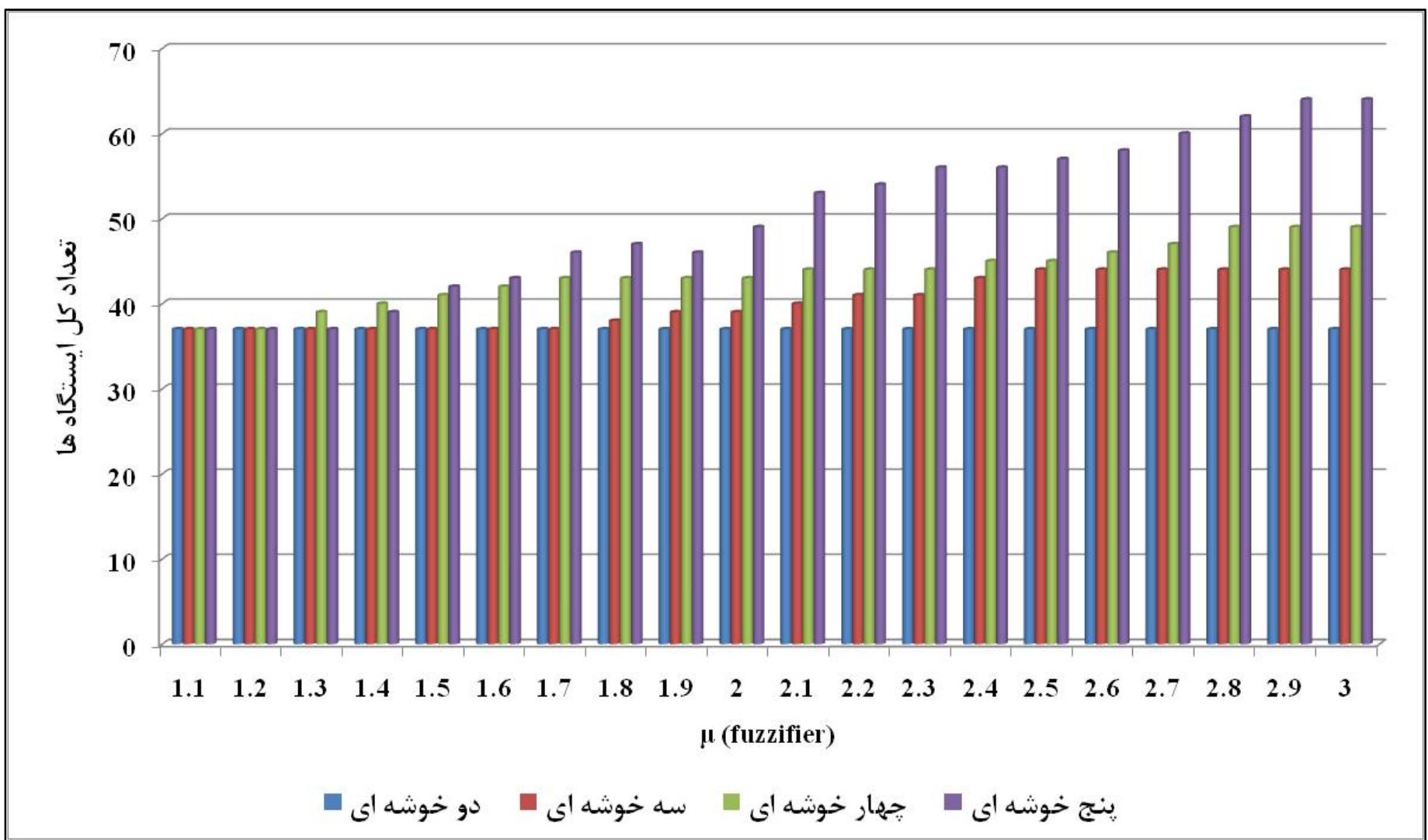
در شکل (۲۷-۴) حاصل جمع تعداد ایستگاه‌های اختصاص یافته به کل خوش‌های در هر یک از حالات ۲ تا ۵ خوش‌های در محدوده‌ی $1.1 \leq \mu \leq 3.0$ نشان داده شده است.



شکل (۲۵-۴) تعداد ایستگاه‌های هیدرومتری متعلق به هر یک از خوشه‌ها در حالت چهار خوشه‌ای



شکل (۲۶-۴) تعداد ایستگاه‌های هیدرومتری متعلق به هر یک از خوشه‌ها در حالت پنج خوشه‌ای



شکل (۲۷-۴) حاصل جمع تعداد ایستگاه های هیدرومتری اختصاص یافته به کل خوش ها در هر یک از حالات ۲ تا ۵ خوش های در محدوده $1.1 \leq \mu \leq 3.0$

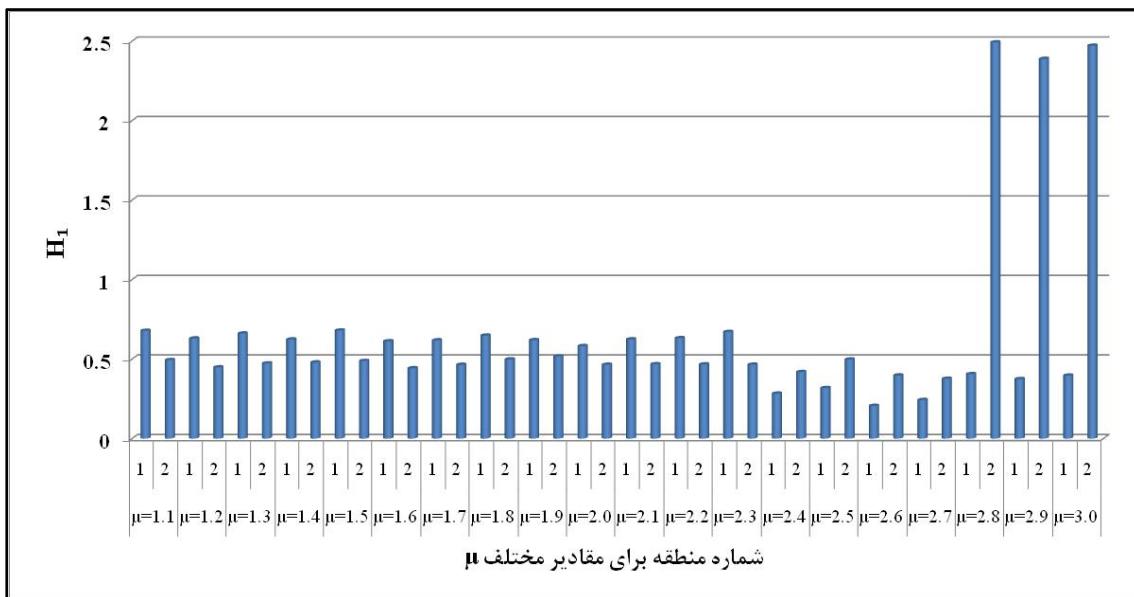
همان طور که در شکل (۴-۲۷) مشاهده می‌شود، حاصل جمع تعداد ایستگاه‌های اختصاص یافته به کل خوش‌ها در حالت دو خوش‌های از ابتدا تا انتهای بدون تغییر و برابر ۳۷ باقی مانده است. حاصل جمع تعداد ایستگاه‌های تعلق گرفته به سه خوش در حالت سه خوش‌های در محدوده‌ی $\mu \leq 1.7$ از ۳۷ به ۴۲ افزایش یافته و پس از آن تا انتهای ثابت مانده است. مجموع تعداد ایستگاه‌های جای گرفته در خوش‌های حالت چهار خوش‌های در فاصله‌ی مورد مطالعه از ۳۷ به ۴۹ رسیده است. در حالت پنج خوش‌های، افزایش حاصل جمع تعداد ایستگاه‌های اختصاص یافته به پنج خوش بسیار چشمگیر است، به طوری که حاصل جمعی که در ابتدای بررسی برابر ۳۷ است، در انتهای محدوده‌ی مورد بررسی به ۶۴ افزایش یافته است.

۴-۶-۷-۱- تحلیل فراوانی منطقه‌ای با استفاده از گشتاورهای خطی

۴-۶-۷-۲- تعیین وضعیت همگنی مناطق و ایستگاه‌های موجود در هر منطقه

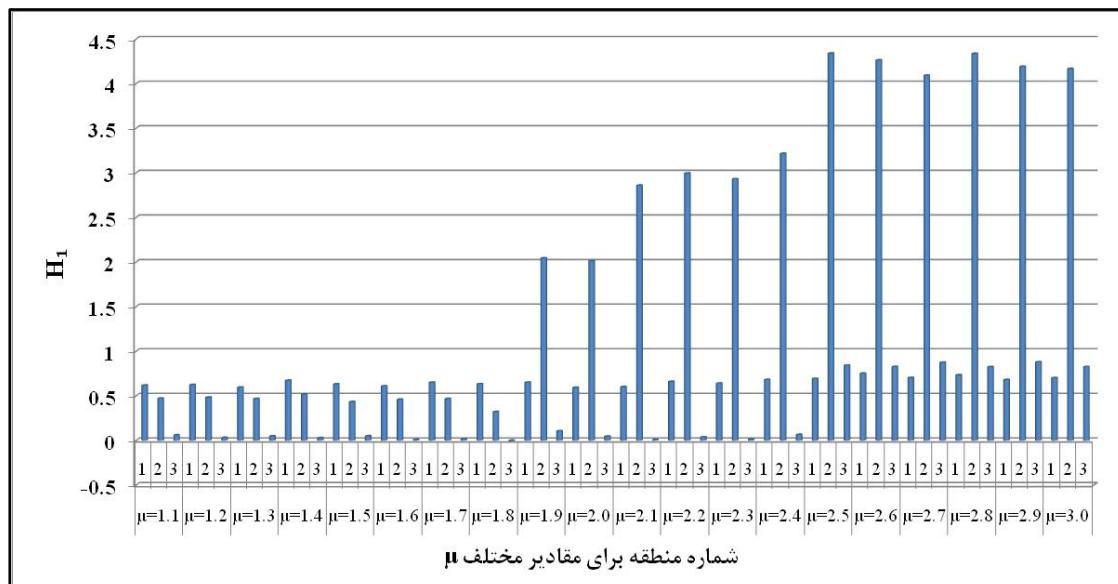
پس از اجرای خوشبندی با استفاده از الگوریتم c-means، وضعیت همگنی مناطق تشکیل شده بر اساس شاخص ناهمگنی H_1 که استفاده از آن برای ارزیابی مناطق حاصل از استفاده از روش‌های خوشبندی توسط هاسکینگ و والیس (۱۹۹۷) توصیه شده است، مورد بررسی قرار گرفت.

در شکل (۴-۲۸) مقادیر شاخص ناهمگنی H_1 برای مناطق حالت دو خوش‌های در فاصله‌ی $\mu \leq 1.1$ مشاهده می‌شود که بر اساس آن، تمامی مناطق تشکیل شده در این حالت در حد فاصل $2.7 \leq \mu \leq 3.0$ همگن هستند، اما در فاصله‌ی $3.0 \leq \mu \leq 2.8$ ، هر بار یکی از دو منطقه‌ی تشکیل شده در وضعیت ناهمگن قرار می‌گیرد. این اتفاق صرفاً به دلیل انتقال ایستگاه از منطقه‌ی اول به منطقه‌ی دوم روی می‌دهد. در حالی که در $\mu = 2.7$ درجه‌ی عضویت این ایستگاه در منطقه‌ی اول برابر 50.5% و در منطقه‌ی دوم برابر 49.5% است، این مقادیر برای $\mu = 2.8$ در منطقه‌ی اول به 49.9% و در منطقه‌ی دوم به 50.1% تغییر می‌کند.



شکل (۲۸-۴) مقادیر شاخص ناهمگنی H_1 برای مناطق تشکیل شده در حالت دو خوشهای به وسیله‌ی الگوریتم c-means

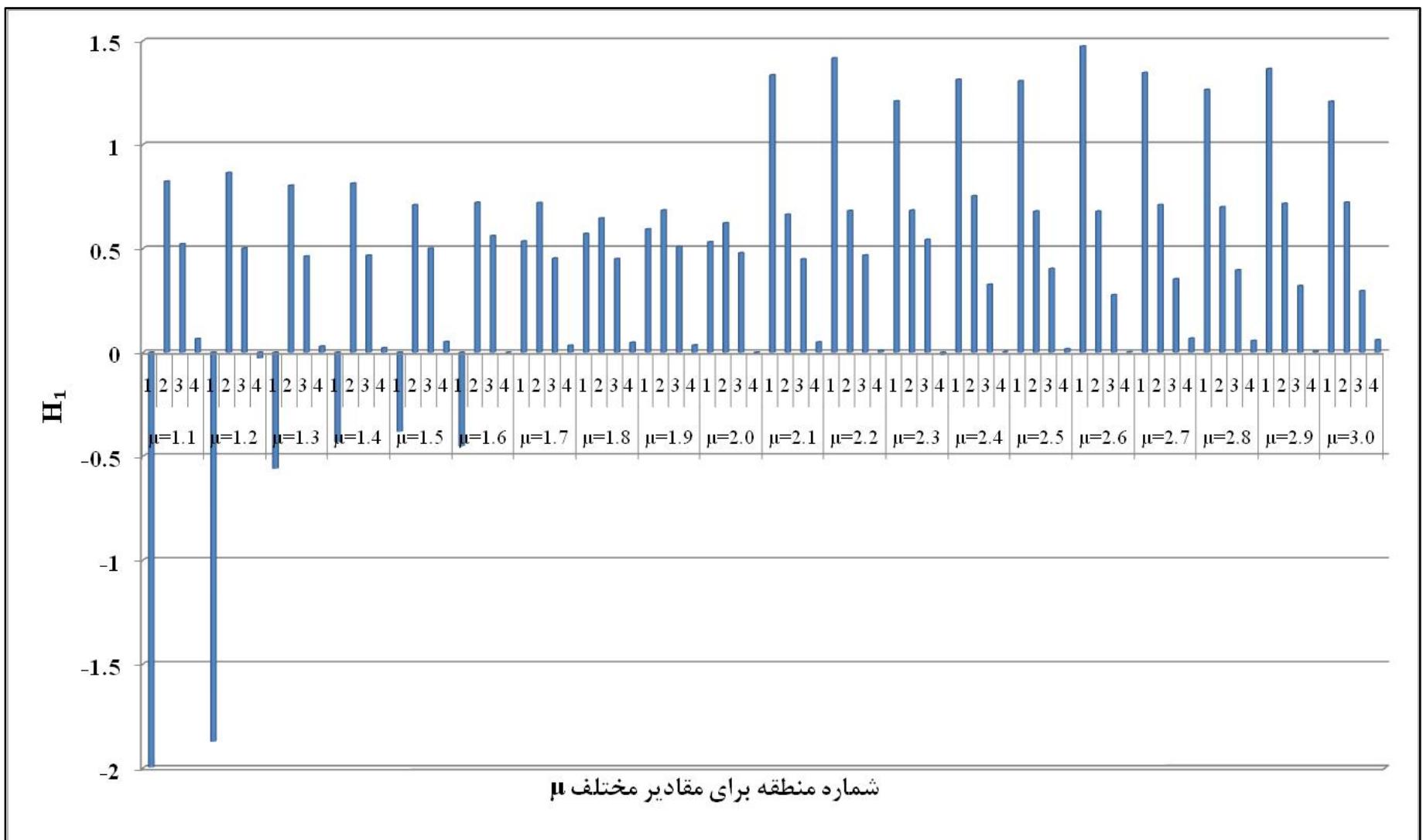
شکل (۲۹-۴) که مربوط به مقادیر شاخص ناهمگنی H_1 برای مناطق حالت سه خوشهای در فاصله‌ی $1.1 \leq \mu \leq 3.0$ می‌باشد، بیانگر همگنی تمامی مناطق تشکیل شده در محدوده‌ی $1.8 \leq \mu \leq 1.1$ است. این شکل همچنین نشان می‌دهد که در فاصله‌ی $3.0 \leq \mu \leq 1.9$ مقدار شاخص ناهمگنی H_1 برای منطقه‌ی دوم به قدری افزایش می‌یابد که در محدوده‌ی ناهمگنی قرار می‌گیرد.



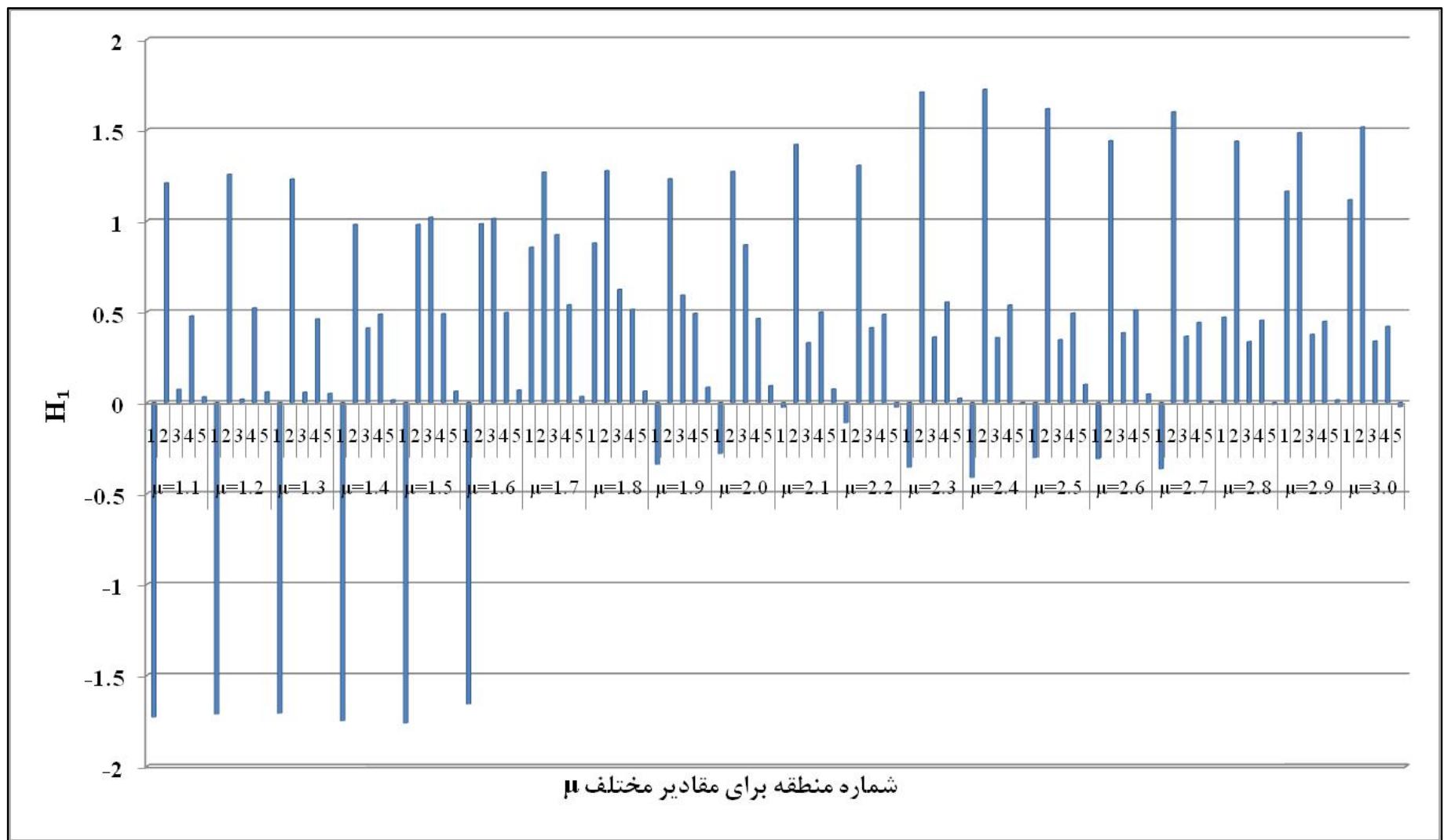
شکل (۲۹-۴) مقادیر شاخص ناهمگنی H_1 در مناطق تشکیل شده در حالت سه خوشهای به وسیله‌ی الگوریتم c-means

شکل (۳۰-۴) مربوط به نمایش مقادیر شاخص ناهمنگی H_1 برای مناطق تشکیل شده در حالت چهار خوشهای در فاصله‌ی $\mu \leq 3.0$ است. آن گونه که در شکل دیده می‌شود، در فاصله‌ی $\mu \leq 2.0$ ، مقدار شاخص ناهمنگی برای هیچ یک از مناطق از مقدار ۱ تجاوز نکرده است و تمام مناطق در وضعیت همگن قرار دارند. اما در فاصله‌ی $\mu \leq 3.0$ ، مقدار شاخص H_1 برای منطقه‌ی اول از ۱ فراتر رفته و این منطقه را در وضعیت احتمالاً ناهمنگن قرار داده است. با این حال به نظر می‌رسد مقدار این ناهمنگی چندان زیاد نیست و امکان برطرف کردن آن با استفاده از روش‌های معرفی شده توسط هاسکینگ و والیس (۱۹۹۷) برای رفع ناهمنگی مناطق وجود دارد.

شکل (۳۱-۴) نشان دهنده‌ی مقادیر شاخص ناهمنگی H_1 برای مناطق تشکیل شده در حالت پنج خوشهای در محدوده‌ی $\mu \leq 3.0$ است. مطابق شکل، در محدوده‌ی $\mu \leq 1.3$ منطقه‌ی دوم در وضعیت احتمالاً ناهمنگن قرار گرفته است و سایر مناطق همگن هستند. در $\mu = 1.4$ مقدار شاخص ناهمنگی برای منطقه‌ی دوم کاهش یافته و منطقه در حالت همگن قرار گرفته است. هم‌زمان مقدار شاخص ناهمنگی به را منطقه‌ی سوم نیز تا حدودی افزایش یافته است. به طور کلی در این حالا هر پنج منطقه، همگن هستند. در $\mu = 1.5$ و $\mu = 1.6$ افزایش مقدار شاخص ناهمنگی در منطقه‌ی سوم، سبب شده که مقدار این شاخص به مقدار بسیار اندک از ۱ فراتر رود و منطقه در وضعیت احتمالاً ناهمنگن قرار گیرد. در این دو حالت سایر مناطق همچنان وضعیت همگن خود را حفظ کرده‌اند. در فاصله‌ی $\mu \leq 2.8$ ، منطقه‌ی سوم همچنان در وضعیت احتمالاً ناهمنگن قرار داشته و سایر مناطق همگن هستند. در حالت‌های $\mu = 2.9$ و $\mu = 3.0$ مقدار شاخص ناهمنگی برای منطقه‌ی دوم مجدداً افزایش یافته تا این منطقه نیز در کنار منطقه‌ی سوم در وضعیت احتمالاً ناهمنگن قرار گیرد. در این دو حالت، سه منطقه‌ی دیگر همچنان همگنی خود را حفظ کرده‌اند. در مورد مقادیر منفی شاخص ناهمنگی، این نکته شایان ذکر است که این مقادیر تأثیری در ناهمنگی مناطق ندارند، اما در صورتی که مقدار منفی از -۲ کوچک‌تر باشد، احتمال وجود همبستگی معکوس بین داده‌های مورد مطالعه وجود دارد.

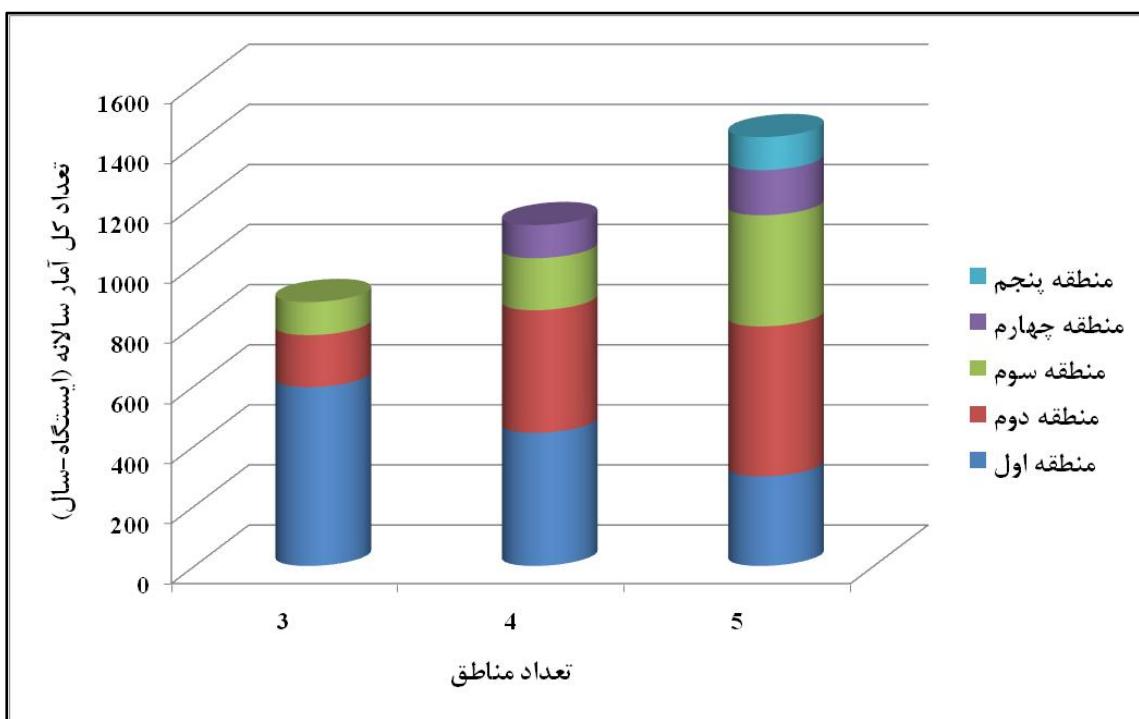


شکل (۴-۳۰) مقادیر شاخص ناهمگنی H_1 برای مناطق تشکیل شده در حالت چهار خوشهای به وسیله‌ی الگوریتم c-means



شکل (۳۱-۴) مقادیر شاخص ناهمنگی H_1 برای مناطق تشکیل شده در حالت پنج خوشهای به وسیله‌ی الگوریتم c-means

از میان کل حالت‌های بررسی شده، سه حالت برای اجرای مراحل بعدی تحلیل فراوانی منطقه‌ای سیلان برگزیده شدند. نخستین حالت، حالت سه خوش‌های با 1.8μ است. در این حالت هر سه منطقه‌ی تشکیل شده همگن هستند، اما افزایش قابل توجهی در تعداد حاصل جمع ایستگاه‌های عضو در سه منطقه دیده نمی‌شود و تنها با یک واحد افزایش از ۳۷ به ۳۸ می‌رسد. در واقع تنها یکی از ایستگاه‌ها همزمان در دو منطقه عضویت دارد و سایر ایستگاه‌ها هر کدام تنها عضو یک منطقه هستند. حالت‌های برگزیده‌ی دوم و سوم، حالت‌های چهار و پنج خوش‌های هر دو با 2.8μ هستند. حاصل جمع تعداد ایستگاه‌های عضو در کل مناطق این دو حالت به ترتیب برابر ۴۹ و ۶۲ ایستگاه است. در هر دو حالت اخیر یک منطقه‌ی ناهمگن وجود دارد که ناهمگنی آن با حذف ایستگاه ۹۳۴-۱۷ از منطقه‌ی ناهمگن، برطرف می‌شود و مقادیر شاخص ناهمگنی برای منطقه‌ی ناهمگن این دو حالت به ترتیب به $0/76$ و $0/95$ کاهش می‌یابد. البته از آنجا که این ایستگاه در مناطق دیگر نیز عضویت دارد، امکان برآورد چندک‌های سیلان برای آن همچنان وجود خواهد داشت. در شکل (۳۲-۴) تعداد داده‌های آماری موجود در هر یک از مناطق این سه حالت برگزیده نشان داده شده است.



شکل (۳۲-۴) تعداد داده‌های آماری سه حالت برگزیده

۴-۶-۷-۲- شاخص نکویی برازش Z و انتخاب توزیع فراوانی منطقه‌ای سیلاب

برای انتخاب توزیع منطقه‌ای بهینه برای هر یک از مناطق سه حالت برگزیده، از میان توزیع‌هایی که شرط $|Z^{DIST}| \leq 1.64$ در مورد آن‌ها برقرار بود، توزیعی که نزدیک‌ترین مقدار به صفر را اختیار می‌کرد به عنوان توزیع منطقه‌ای سیلاب انتخاب شد. بدین منظور توزیع‌های لجستیک تعیین یافته، مقادیر حدی تعیین یافته، نرمال تعیین یافته، پیرسون تیپ III و پارتوي تعیین یافته مورد ازمون قرار گرفتند. در جدول (۶-۴) مقدار شاخص نکویی برازش Z محاسبه شده برای هر یک از توزیع‌های سه پارامتری مورد آزمایش برای هر یک از مناطق سه حالت برگزیده و در جدول (۷-۴) مقادیر پارامتر-های برآورد شده با استفاده از گشتاورهای خطی برای هر یک از توزیع‌های منطقه‌ای منتخب، ارائه شده است. لازم به ذکر است که مقدار شاخص Z برای مناطقی که ناهمگن بوده و با حذف ایستگاه ۱۷-۹۳۴ همگن شده‌اند، مجدداً و پس از حذف این ایستگاه محاسبه شده است.

جدول (۶-۴) مقادیر شاخص نکویی برازش Z در هر یک از مناطق سه حالت برگزیده منطقه‌بندی فازی

Z						
توزیع فراوانی						تعداد مناطق
پارتوي	پیرسون	نرمال تعیین	مقادیر حدی	لجدستیک	تعیین یافته	شماره منطقه و مقدار μ
-۱/۰۶	*-۰/۱۲	۱/۸۲	۲/۹۳	۴/۳۹	۱	منطقه ۳
-۲/۵۱	-۰/۳۸	*۰/۰۶	۰/۲۴	۱/۴۷	۲	
-۲/۵۳	-۰/۵۷	-۰/۲۸	*-۰/۲۰	۰/۸۷	۳	
-۰/۹۳	*۰/۰۱	۱/۵۸	۲/۴۷	۳/۷۴	۱	منطقه ۴
-۱/۲۵	*۰/۰۶	۱/۴۴	۲/۲۲	۳/۵۵	۲	
-۲/۵۰	-۰/۳۱	*۰/۱۵	۰/۳۳	۱/۵۹	۳	
-۲/۵۴	-۰/۵۴	-۰/۲۴	*-۰/۱۵	۰/۹۵	۴	
-۰/۳۸	*۰/۱۳	۱/۴۹	۲/۲۷	۳/۲۳	۱	منطقه ۵
-۰/۸۴	*۰/۳۹	۲/۰۲	۲/۹۱	۴/۳۱	۲	
-۱/۴۵	*۰/۰۲	۱/۳۲	۲/۰۴	۳/۴۱	۳	
-۲/۳۵	-۰/۴۳	*-۰/۰۳	۰/۱۴	۱/۲۴	۴	
-۲/۵۷	-۰/۶۰	-۰/۳۱	*-۰/۲۳	۰/۸۵	۵	

*توزیع منطقه‌ای برگزیده

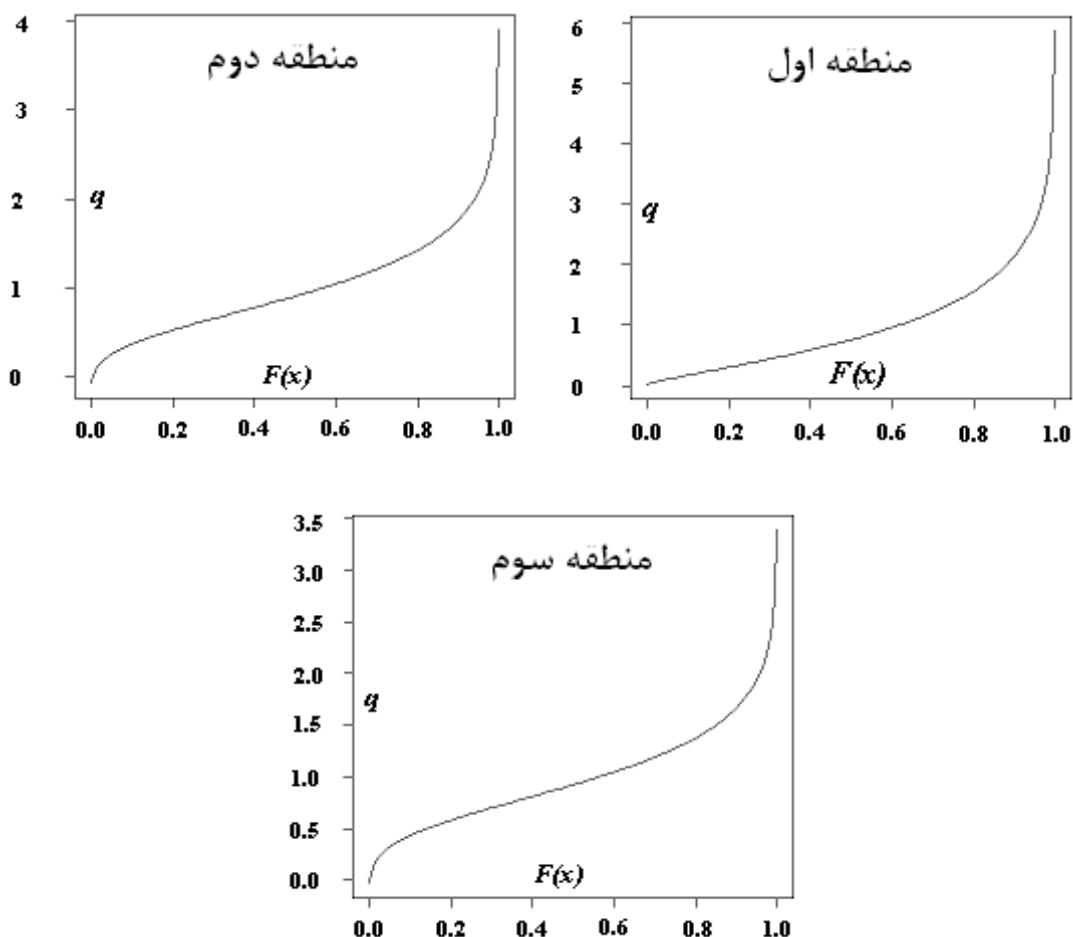
جدول (۷-۴) مقادیر برآورده شده پارامترهای توزیع‌های منطقه‌ای برای سه حالت برگزیده منطقه‌بندی فازی

تعداد مناطق و مقدار μ	شماره منطقه	توزیع منطقه‌ای	پارامترها	برآورد پارامتر
۱	۱	پیرسون تیپ III	μ	۱/۰۰
۳	۲	نرمال تعمیم یافته	σ	۰/۸۶۷۶
۴	۲	مقادیر حدی تعمیم یافته	γ	۱/۷۶۷۴
۴	۳	مقادیر حدی تعمیم یافته	ξ	۰/۹۰۱۳
۴	۳	مقادیر حدی تعمیم یافته	α	۰/۵۲۸۹
۴	۳	مقادیر حدی تعمیم یافته	k	-۰/۳۶۱۲
۴	۲	پیرسون تیپ III	μ	۰/۷۷۳۳
۴	۲	پیرسون تیپ III	σ	۰/۴۰۷۸
۴	۲	پیرسون تیپ III	γ	۰/۰۲۲۱
۴	۲	پیرسون تیپ III	μ	۱/۰۰
۴	۲	پیرسون تیپ III	σ	۰/۸۷۵۰
۴	۲	پیرسون تیپ III	γ	۱/۷۲۸۵
۴	۳	نرمال تعمیم یافته	μ	۰/۸۱۳۰
۴	۳	نرمال تعمیم یافته	σ	۱/۶۱۳۸
۴	۳	نرمال تعمیم یافته	γ	۰/۹۰۱۳
۴	۴	مقادیر حدی تعمیم یافته	ξ	۰/۵۲۸۹
۴	۴	مقادیر حدی تعمیم یافته	α	-۰/۳۶۱۲
۴	۴	مقادیر حدی تعمیم یافته	k	۰/۷۷۳۳
۴	۴	مقادیر حدی تعمیم یافته	μ	۰/۴۰۷۸
۴	۴	مقادیر حدی تعمیم یافته	σ	۰/۰۲۲۱
۴	۴	مقادیر حدی تعمیم یافته	γ	۱/۰۰
۴	۱	پیرسون تیپ III	μ	۰/۹۰۶۲
۴	۱	پیرسون تیپ III	σ	۱/۸۱۵۸
۴	۱	پیرسون تیپ III	γ	۱/۰۰
۴	۲	پیرسون تیپ III	μ	۰/۸۵۶۲
۴	۲	پیرسون تیپ III	σ	۱/۶۶۸۶
۴	۲	پیرسون تیپ III	γ	۱/۰۰
۵	۳	پیرسون تیپ III	μ	۰/۷۹۵۰
۵	۳	پیرسون تیپ III	σ	۱/۵۶۲۸
۵	۳	پیرسون تیپ III	γ	۰/۹۰۲۰
۵	۴	نرمال تعمیم یافته	ξ	۰/۵۲۱۹
۵	۴	نرمال تعمیم یافته	α	-۰/۳۶۳۲
۵	۴	نرمال تعمیم یافته	k	۰/۷۷۳۳
۵	۴	مقادیر حدی تعمیم یافته	μ	۰/۴۰۷۸
۵	۴	مقادیر حدی تعمیم یافته	σ	۰/۰۲۲۱

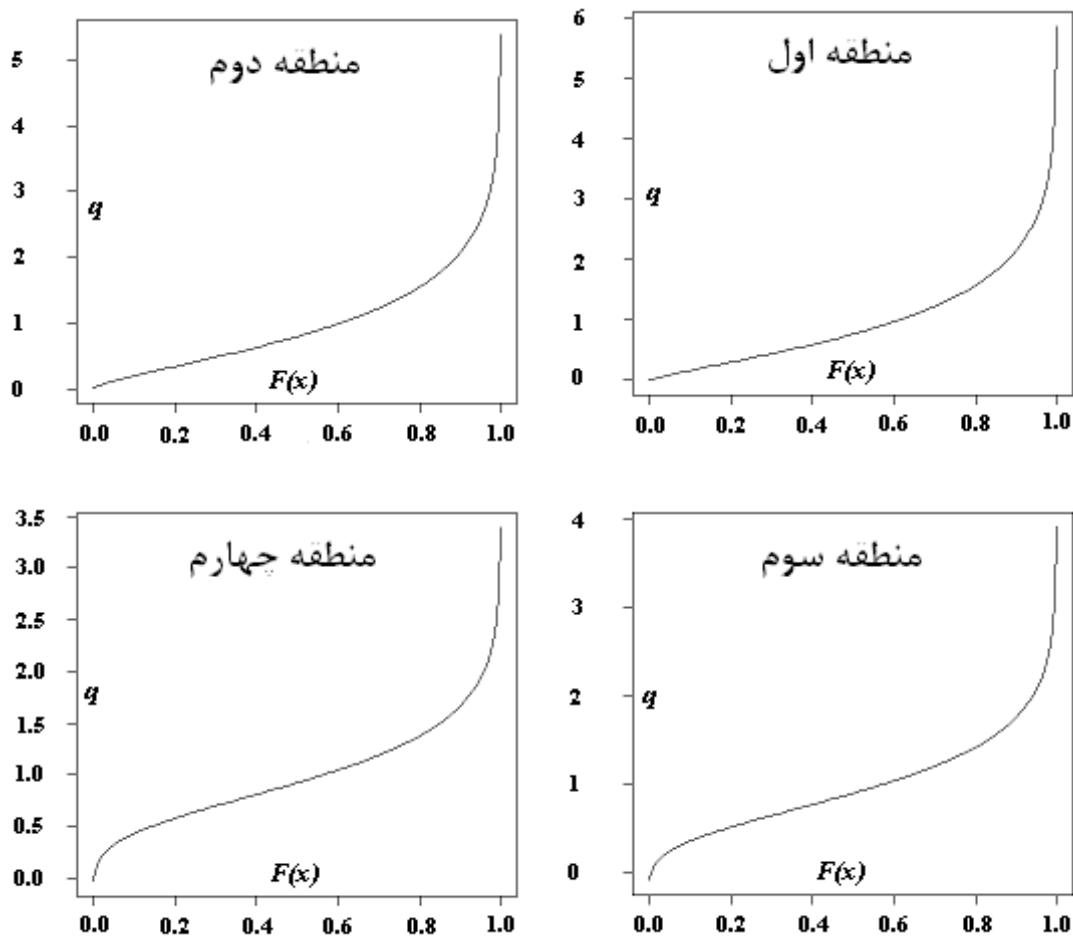
همان طور که از جداول (۶-۴) و (۷-۴) می‌توان دریافت، مناطق دوم و سوم حالت سه منطقه‌ای منتخب، همان مناطق سوم و چهارم حالت چهار منطقه‌ای برگزیده هستند. مناطق چهارم و پنجم حالت پنج منطقه‌ای نیز تا حد زیادی مشابه همین دو منطقه هستند.

۴-۶-۷-۳-برآورد چندک‌های منطقه‌ای سیلاب و دبی سیلاب ایستگاهها

پس از انتخاب توزیع بهینه برای هر منطقه و برآورد پارامترهای آن، چندک‌های منطقه‌ای سیلاب برای هر یک از مناطق محاسبه شد. در شکل‌های (۳۳-۴)، (۳۴-۴) و (۳۵-۴) چندک‌های منطقه‌ای محاسبه شده برای مناطق حالت‌های سه، چهار و پنج منطقه‌ای برگزیده، نشان داده شده است.



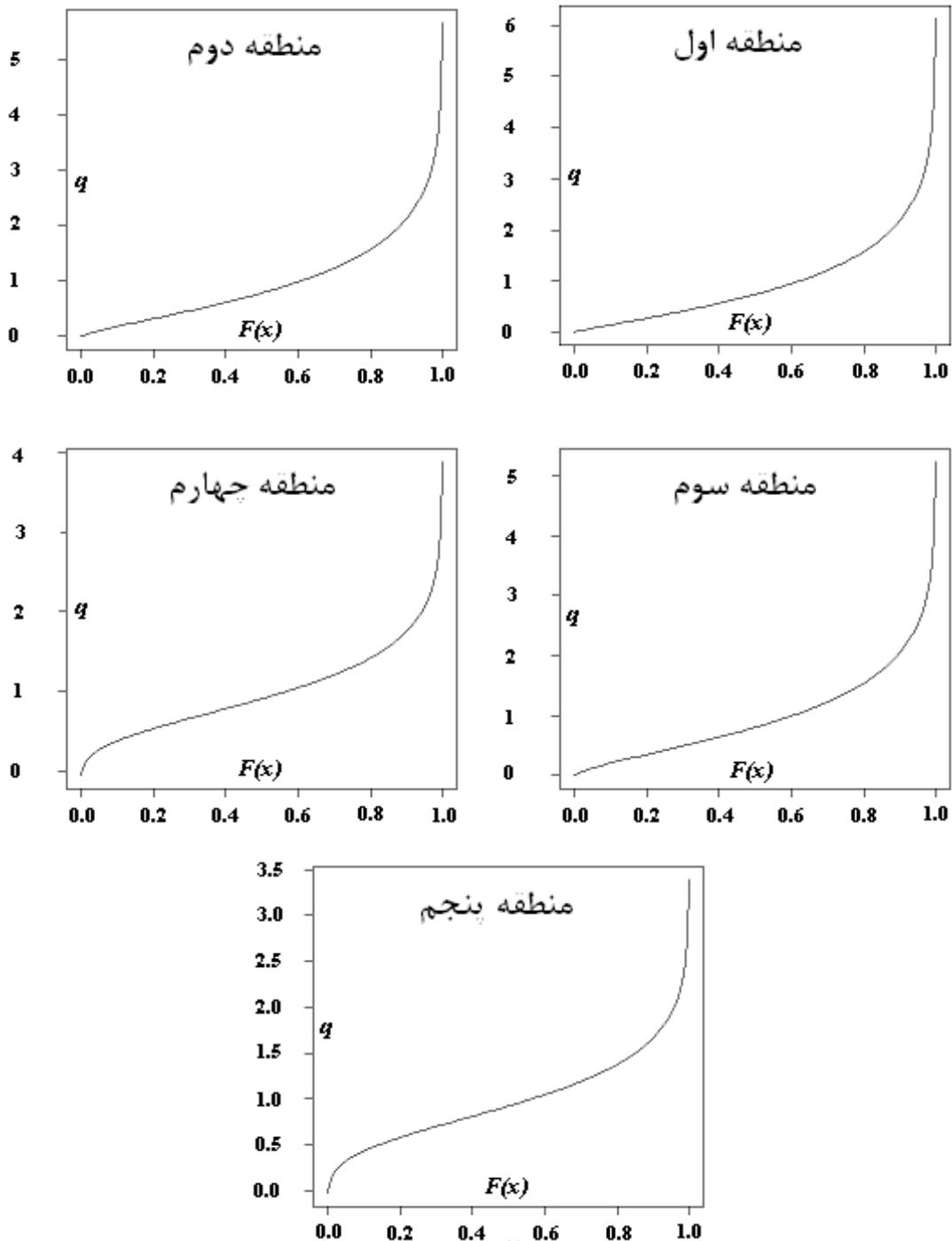
شکل (۳۳-۴) چندک‌های سیلاب منطقه‌ای (q) در برابر احتمال عدم تجاوز متناظر ($F(x)$) برای حالت سه منطقه‌ای



شکل (۳۴-۴) چندک‌های سیلاب منطقه‌ای (q) در برابر احتمال عدم تجاوز متناظر ($F(x)$) برای حالت چهار منطقه‌ای

برای هر سایت یا ایستگاه، توزیع فراوانی منطقه‌ای با استفاده از عاملی که شاخص سیلاب یا سیلاب نمایه نامیده می‌شود، تبدیل به توزیع فراوانی ویژه‌ی هر سایت یا ایستگاه می‌شود. در واقع با ضرب عامل سیلاب نمایه‌ی هر ایستگاه که در این مطالعه برابر میانگین آمار سالانه‌ی حداکثر دبی لحظه‌ای سیلاب هر سایت در چندک‌های منطقه‌ای سیلاب با دوره‌های بازگشت مختلف در نظر گرفته شده است، برآورد سیلاب با دوره‌های بازگشت مورد نظر انجام می‌شود.

در شکل‌های (۳۶-۴)، (۳۷-۴)، (۳۸-۴)، (۳۹-۴) و (۴۰-۴) مقادیر برآورد شده‌ی حداکثر سیلاب با دوره‌های بازگشت ۱۰، ۲۵، ۵۰ و ۱۰۰ سال برای ایستگاه‌های مورد مطالعه در سه حالت برگزیده نشان داده شده است.



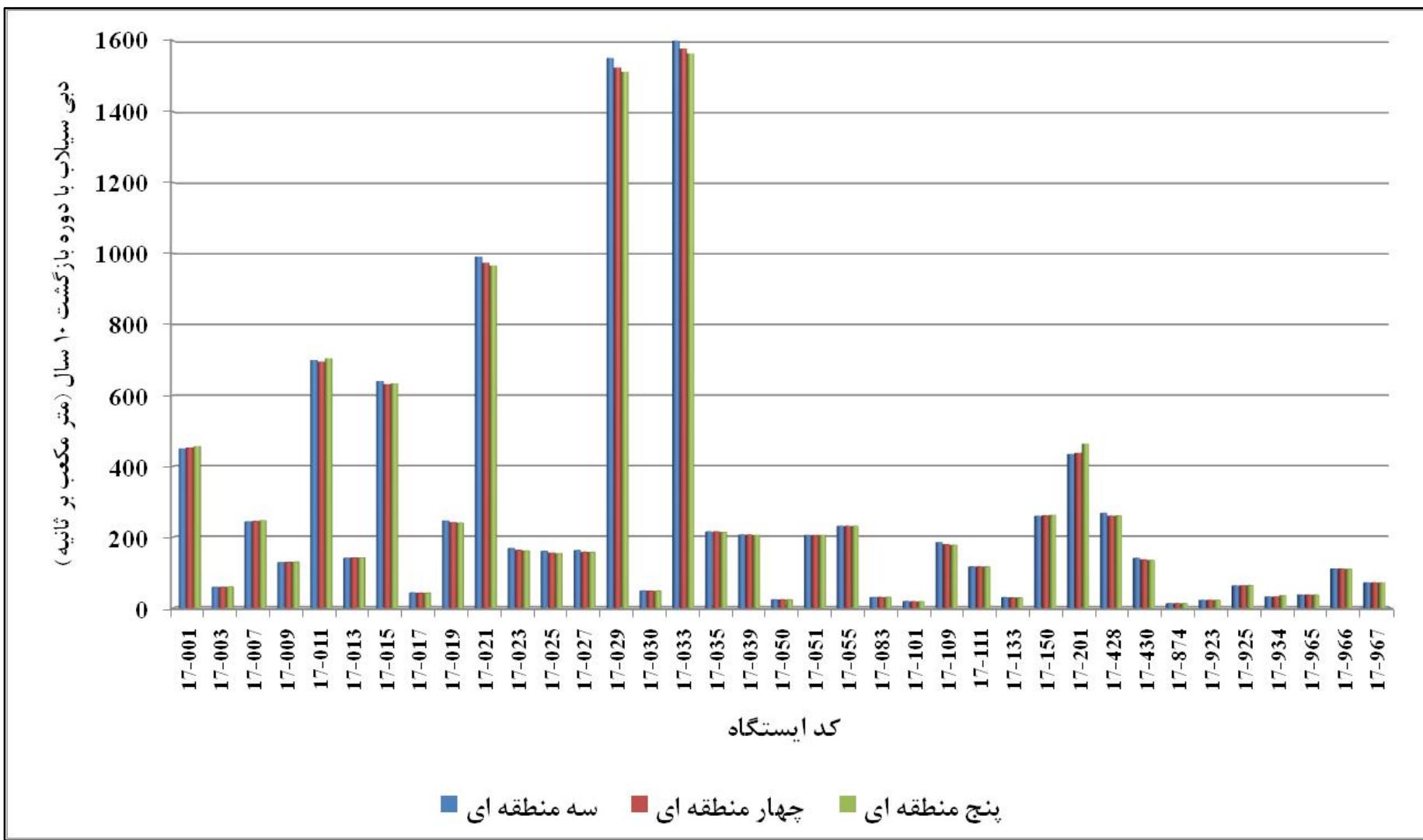
شکل (۴-۳۵) چندک‌های سیلاپ منطقه‌ای (q) در برابر احتمال عدم تجاوز متناظر ($F(x)$) برای حالت پنج منطقه‌ای

برآورد نهایی دبی سیلاپ با دوره‌ی بازگشت مشخص برای سایت‌هایی که به بیش از یک منطقه اختصاص یافته‌اند، با استفاده از رابطه‌ی (۱-۴) محاسبه شده است:

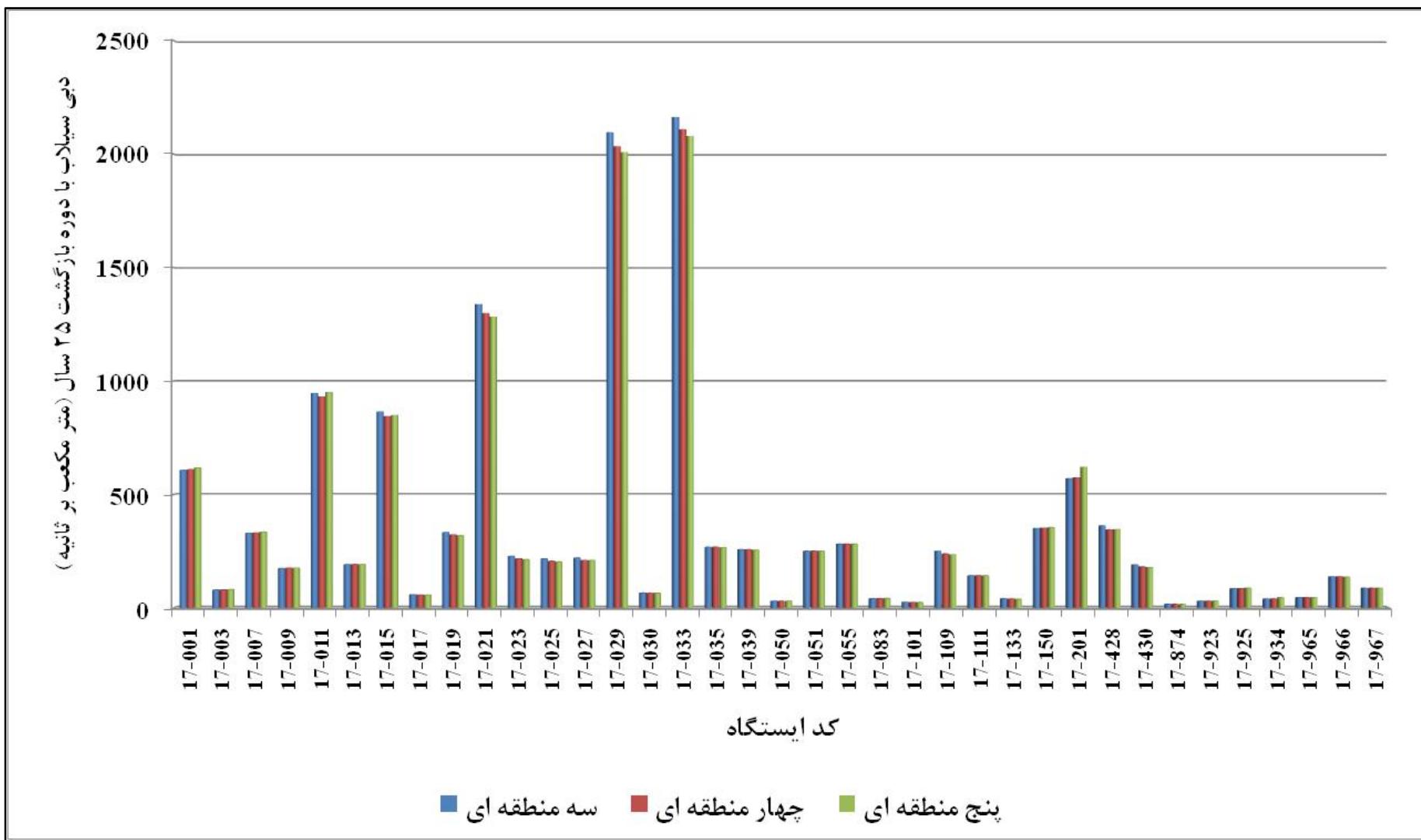
$$Q_T = \sum_{m=1}^t \left(\frac{u_m}{\sum_{i=1}^t u_i} \right) l_1 q_{m,T} \quad (1-4)$$

که در آن t تعداد مناطقی است که سایت مورد نظر به آن‌ها اختصاص یافته است. u_m و u_i در رابطه-ی (۱-۴) به ترتیب درجه‌ی عضویت سایت مورد نظر در منطقه‌ی m و i هستند. l_1 در رابطه‌ی مذکور، گشتاور خطی مرتبه اول یا میانگین داده‌های سیلاب سایت مورد نظر است که به عنوان عامل سیلاب نمایه‌ی ایستگاه انتخاب شده است. همچنین $q_{m,T}$ در رابطه‌ی فوق چندک منطقه‌ای سیلاب با دوره بازگشت T در m این منطقه‌ای است که سایت مورد نظر به آن اختصاص یافته است. در نهایت Q_T دبی نهایی برآورد شده برای سیلاب با دوره بازگشت T سال در سایت مورد نظر است.

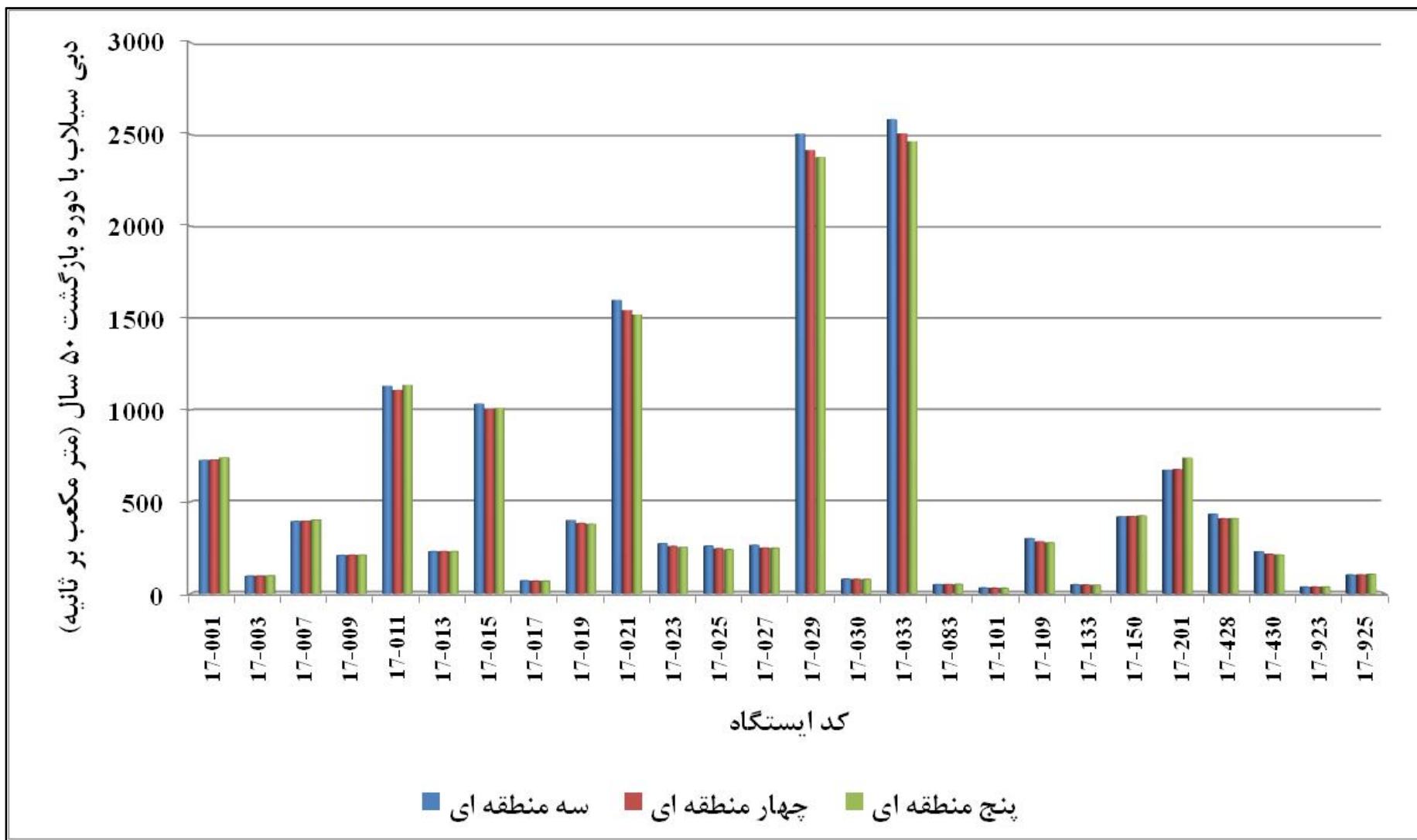
با توجه به تعداد داده‌های آماری موجود در مناطق، برآورد سیلاب با دوره‌های بازگشت ۵۰ و ۱۰۰ سال، تنها برای ایستگاه‌هایی امکان پذیر است که در مناطقی با داده‌هایی بیش از ۵ برابر تعداد این سال‌ها، یعنی به ترتیب ۲۵ و ۵۰۰ سال- ایستگاه واقع شده باشند. در مواردی که ایستگاهی به بیش از یک منطقه اختصاص یافته باشد، برای برآورد سیلاب با دوره بازگشت مورد نظر در آن ایستگاه، از میانگین وزنی برآوردهای چندک سیلاب آن ایستگاه در مناطقی که شرط مذکور در مورد کفایت داده‌ها را داشته باشند، استفاده می‌شود. با در نظر گرفتن این مسئله، برآورد سیلاب با دوره‌ی بازگشت ۵۰ سال برای تعداد ۲۶ ایستگاه در حالت‌های سه، چهار و پنج منطقه‌ای منتخب صورت گرفته است. همچنین برآورد سیلاب با دوره‌ی بازگشت ۱۰۰ سال، تنها برای منطقه‌ی اول حالت سه منطقه‌ای و منطقه‌ی دوم حالت پنج منطقه‌ای انجام گرفته است. همان گونه که در شکل‌ها دیده می‌شود بزرگ‌ترین مقادیر دبی سیلاب برآورد شده با دوره‌های بازگشت ۱۰، ۲۵ و ۵۰ سال به ترتیب مربوط به ایستگاه‌های ۳۳-۰۷، ۲۹-۰۷ و ۲۱-۰۷ است که در هر سه مورد بالاترین مقدار برآورد شده به حالت سه منطقه‌ای و پس از آن به حالت‌های چهار و پنج منطقه‌ای اختصاص دارد. اما این مسئله در مورد ایستگاه‌های دیگر عمومیت مطلق ندارد. در مورد مقادیر دبی سیلاب برآورد شده با دوره‌ی بازگشت ۱۰۰ سال نیز، همان سه ایستگاه بالاترین مقادیر را اختیار می‌کنند.



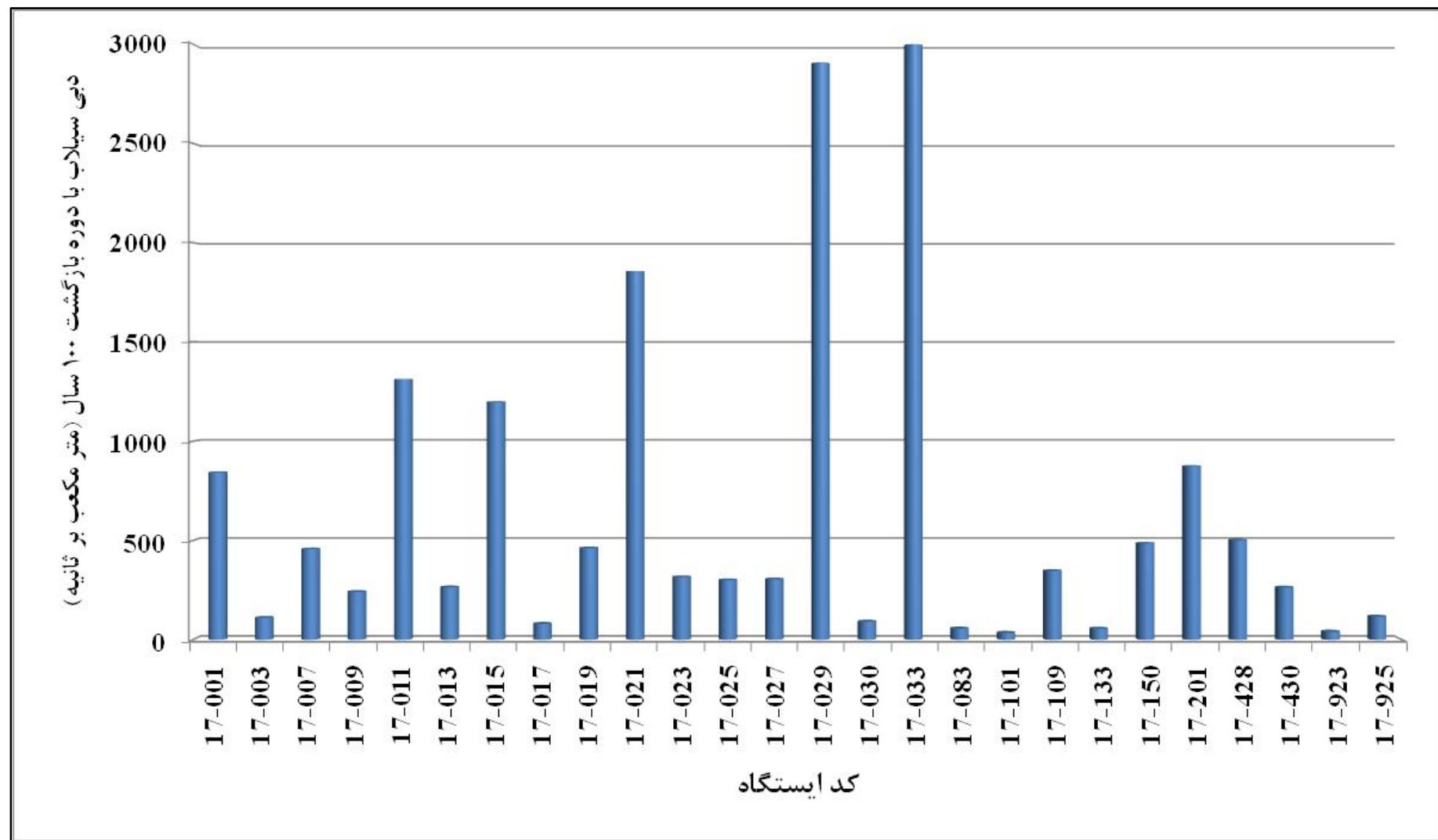
شکل (۳۶-۴) مقادیر برآورد شدهی سیلاب با دورهی بازگشت ۱۰ سال برای تمام ایستگاه‌های هیدرومتری مورد مطالعه در سه حالت برگزیده

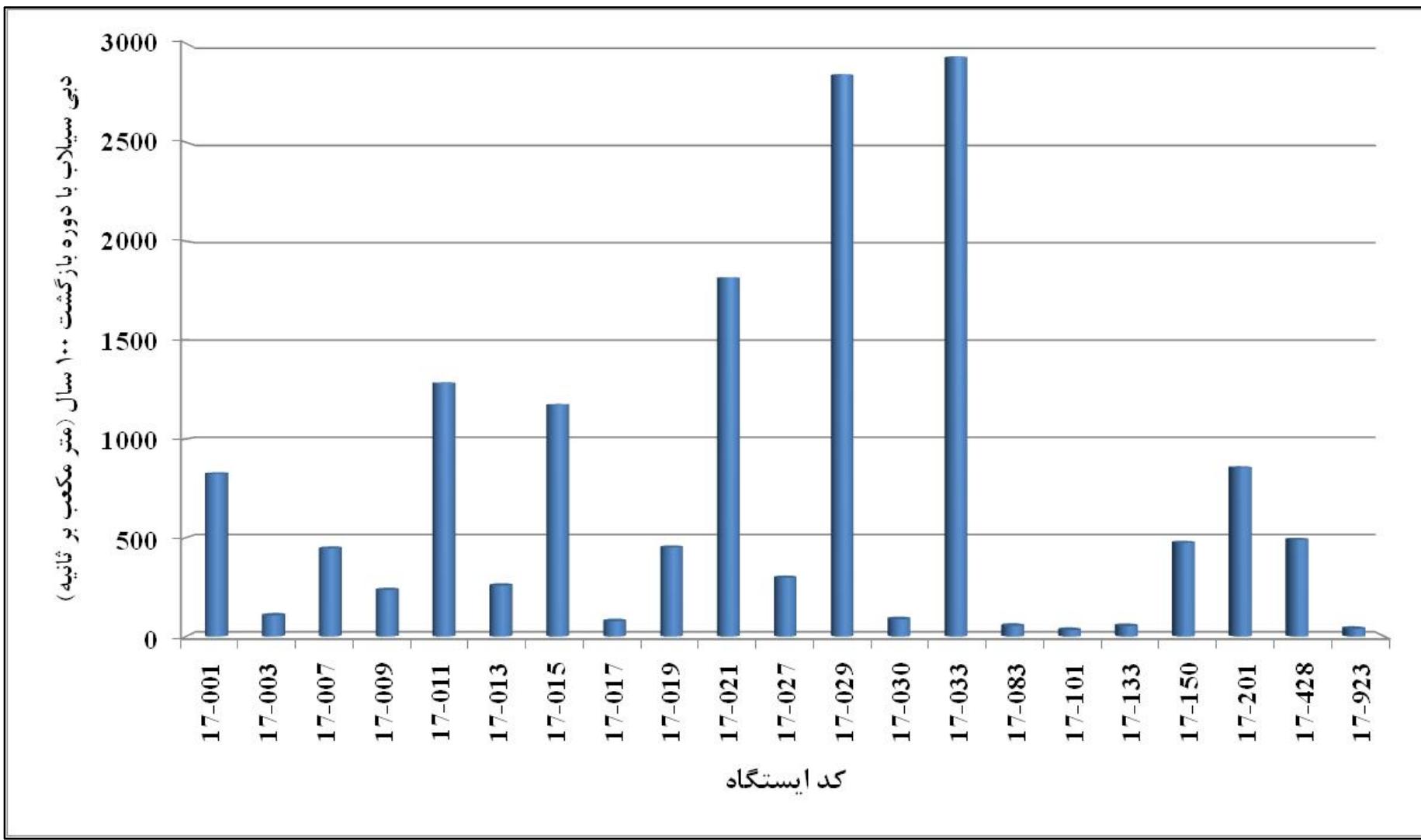


شکل (۳۷-۴) مقادیر برآورد شدهی سیلاب با دورهی بازگشت ۲۵ سال برای تمام ایستگاه‌های هیدرومتری مورد مطالعه در سه حالت برگزیده



شکل (۳۸-۴) مقادیر برآورد شدهی سیلاب با دورهی بازگشت ۵۰ سال برای ۲۶ ایستگاه هیدرومتری در سه حالت برگزیده



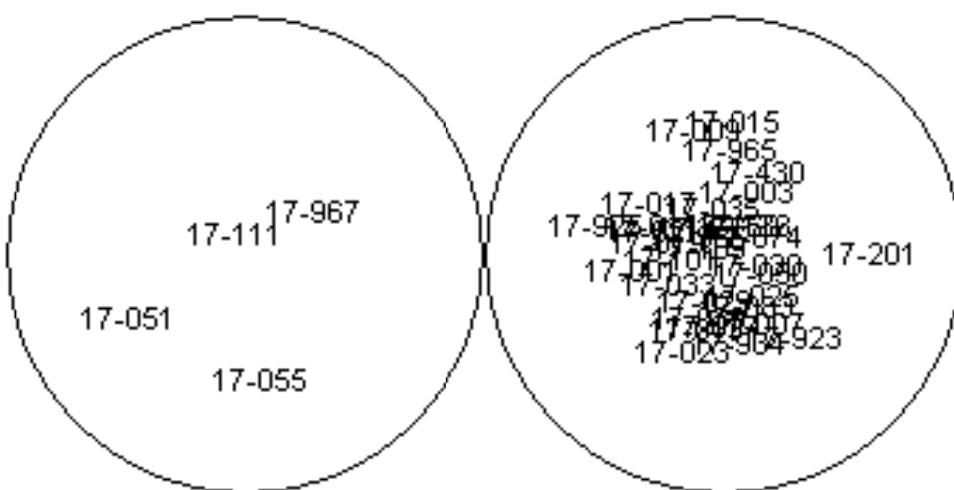


۴-۷-۴- خوشبندی با استفاده از نگاشتهای خودسازمانده کوهون

۴-۷-۱- ایستگاههای اختصاص یافته به خوشها و ویژگی‌های آن‌ها

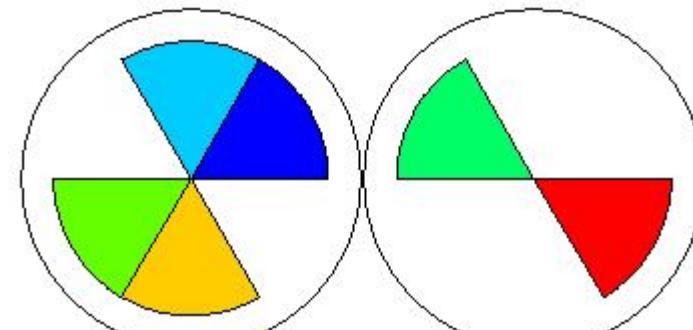
استفاده از نگاشتهای خودسازمانده کوهون، روش دیگری است که در این مطالعه عملکرد آن در زمینه‌ی منطقه‌بندی حوضه‌ی آبریز سفیدرود مورد بررسی قرار گرفته است. از مجموع ۴۰ ایستگاه ابتدایی انتخاب شده برای مطالعه، ۳ ایستگاهی که در مراحل گذشته به دلیل بالا بودن مقادیر شاخص ناجوری یا افزایش ناهمگنی مناطق کنار گذاشته شده بودند، در استفاده از این روش نیز به دلیل مشابه از ادامه‌ی فرآیند محاسبات و تحلیل حذف شدند. بنابراین خوشبندی ۳۷ ایستگاه هیدرومتری باقی‌مانده در چهار مرحله و با استفاده از شبکه‌های کوهون با ۲ تا ۵ گره در لایه‌ی خروجی برای حالت‌های ۲ تا ۵ منطقه‌ای اجرا شد.

در اولین مرحله برای لایه‌ی خروجی تنها ۲ گره در نظر گرفته شد. در این حالت تعداد ۴ ایستگاه در منطقه‌ی نخست (سمت چپ) و تعداد ۳۳ ایستگاه در منطقه‌ی دوم (سمت راست) جای گرفتند. شکل (۴-۴) نحوی اختصاص ایستگاههای مورد بررسی به دو منطقه‌ی مذکور را نشان می‌دهد.



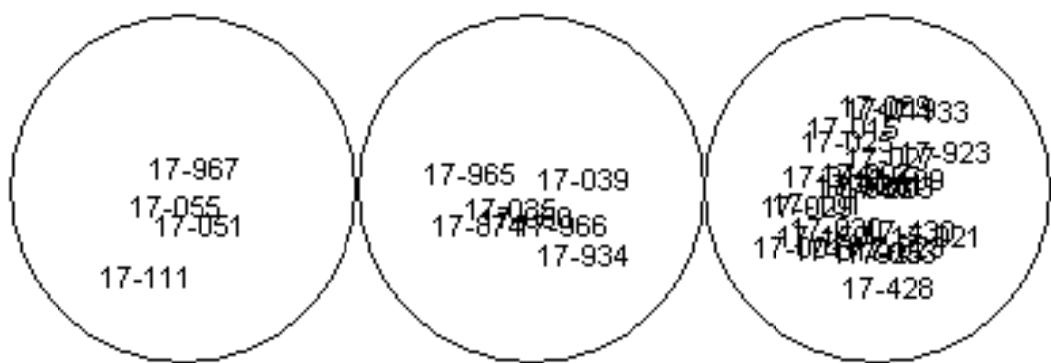
شکل (۴-۴) توزیع ایستگاههای هیدرومتری بین دو گره لایه‌ی خروجی. شماره‌ها معرف کد ایستگاههای مورد مطالعه هستند. (عدد ۱۷ کد حوضه است).

بر اساس شکل (۴۲-۴) که نشان دهندهی مقادیر متوسط ویژگی‌های تجدید مقیاس شده‌ی ایستگاه‌های اختصاص یافته به هر گره است، در این حالت ایستگاه‌هایی به گره اول اختصاص یافته‌اند که دارای مقادیر تجدید مقیاس شده‌ی طول و عرض جغرافیایی، بارندگی متوسط و ضریب رواناب بزرگ هستند و در مقابل، مقادیر تجدید مقیاس شده‌ی سایر ویژگی‌های آن‌ها منفی است. اما گره دوم در برگیرندهی ایستگاه‌هایی است که مقادیر تجدید مقیاس شده‌ی ارتفاع از سطح دریا و لگاریتم مساحت بزرگی داشته و دارای مقادیر تجدید مقیاس شده‌ی طول و عرض جغرافیایی، بارندگی متوسط و ضریب رواناب منفی هستند. برای حالت سه منطقه‌ای، شبکه‌ی کوهونن با لایه‌ی خروجی دارای ۳ گره در نظر گرفته شد که چگونگی توزیع ایستگاه‌ها در میان این سه گره در شکل (۴۳-۴) مشاهده می‌شود. در این حالت سه منطقه با تعداد ۴، ۷ و ۲۶ ایستگاه تشکیل شدند.

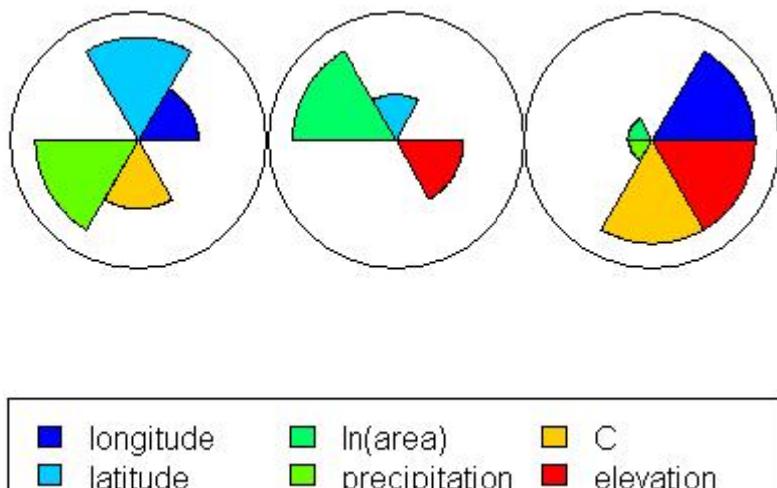


■	longitude	■	ln(area)	■	C
■	latitude	■	precipitation	■	elevation

شکل (۴۲-۴) وضعیت متوسط ویژگی‌های تجدید مقیاس شده‌ی دو گره لایه‌ی خروجی



شکل (۴۳-۴) توزیع ایستگاههای هیدرومتری بین سه گره لایه‌ی خروجی. شماره‌ها معرف کد ایستگاههای مورد مطالعه هستند. (عدد ۱۷ کد حوضه است).

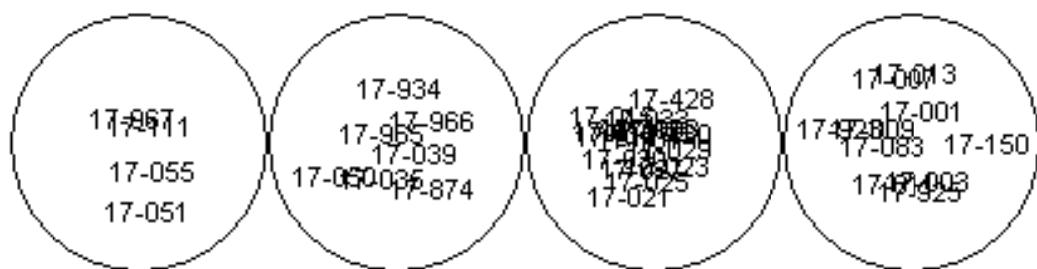


شکل (۴۴-۴) وضعیت متوسط ویژگی‌های تجدید مقیاس شده سه گره لایه‌ی خروجی

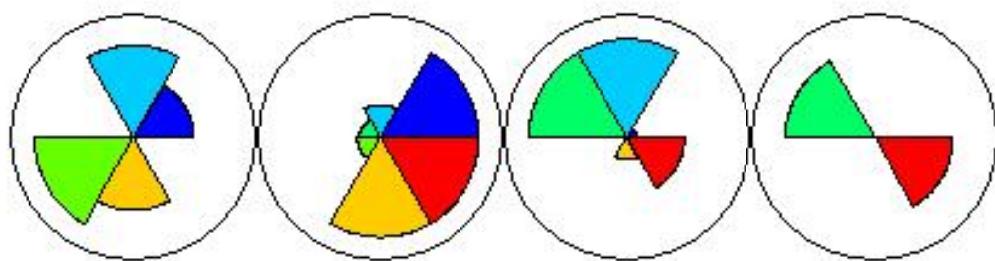
شکل (۴۴-۴) نیز وضعیت سه گره لایه‌ی خروجی را از نظر ویژگی‌های تجدید مقیاس شده نشان می‌دهد. مطابق شکل، گره اول به طور متوسط شامل مقادیر تجدید مقیاس شده عرض جغرافیایی و بارندگی متوسط سالانه‌ی بزرگ، طول جغرافیایی و ضریب رواناب تجدید مقیاس شده متوسط و لگاریتم طبیعی مساحت و تراز ارتفاع تجدید مقیاس شده منفی است. گره دوم به طور میانگین دربرگیرنده‌ی ایستگاههایی با مقدار تجدید مقیاس شده لگاریتم طبیعی مساحت بزرگ، ارتفاع متوسط، عرض جغرافیایی تجدید مقیاس شده نسبتاً کوچک و طول جغرافیایی، بارندگی متوسط و ضریب رواناب تجدید مقیاس شده منفی است. گره سوم حاوی ایستگاههایی با طول جغرافیایی،

ارتفاع و ضریب رواناب تجدید مقیاس شده‌ی بزرگ، بارندگی متوسط سالانه و لگاریتم طبیعی مساحت تجدید مقیاس شده‌ی کوچک و عرض جغرافیایی تجدید مقیاس شده‌ی منفی است. استقرار ایستگاه‌ها در گره‌های لایه‌ی خروجی شبکه‌ی کوهون برای حالت ۴ منطقه‌ای در شکل (۴۵-۴) مشاهده می‌شود. در این حالت، ۴ گره لایه‌ی خروجی شبکه‌ی عصبی کوهون به ترتیب محتوی ۱۰، ۷، ۴ و ۱۶ ایستگاه هستند.

چگونگی وضعیت مقدادیر ویژگی‌های تجدید مقیاس شده برای هر یک از ۴ گره لایه‌ی خروجی نیز در شکل (۴۶-۴) مشاهده می‌شود.



شکل (۴۵-۴) توزیع ایستگاه‌های هیدرومتری بین چهار گره لایه‌ی خروجی. شماره‌ها معرف کد ایستگاه‌های مورد مطالعه هستند. (عدد ۱۷ کد حوضه است).



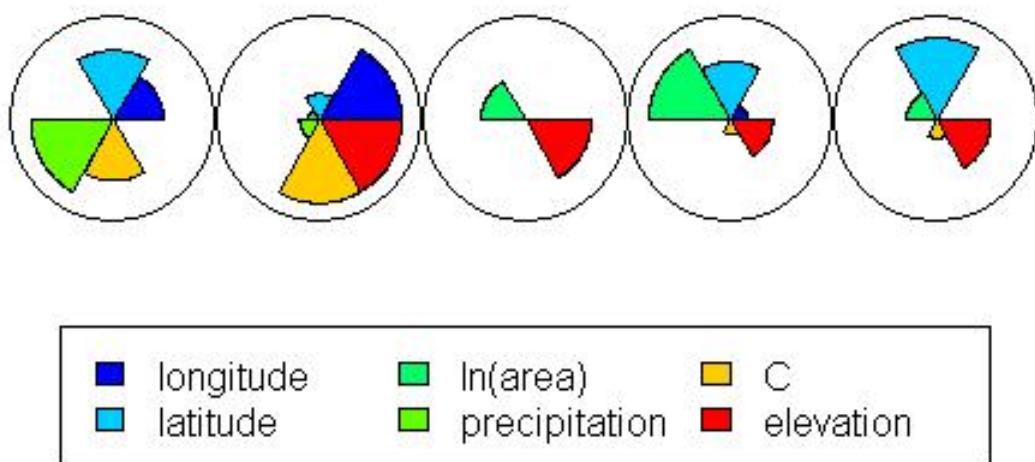
longitude	ln(area)	C
latitude	precipitation	elevation

شکل (۴۶-۴) وضعیت متوسط ویژگی‌های تجدید مقیاس شده‌ی چهار گره لایه‌ی خروجی

گره اول به طور متوسط شامل مقادیر تجدید مقیاس شده‌ی عرض جغرافیایی و بارندگی متوسط سالانه‌ی بزرگ، طول جغرافیایی و ضریب رواناب تجدید مقیاس شده‌ی متوسط و لگاریتم طبیعی مساحت و تراز ارتفاع تجدید مقیاس شده‌ی منفی است. گره دوم مربوط به ایستگاه‌هایی با طول جغرافیایی، ارتفاع و ضریب رواناب تجدید مقیاس شده‌ی بزرگ و در مقابل، عرض جغرافیایی، بارندگی متوسط سالانه و لگاریتم طبیعی مساحت تجدید مقیاس شده‌ی کوچک است. گره سوم در برگیرنده‌ی ایستگاه‌هایی است که مقادیر تجدید مقیاس شده‌ی عرض جغرافیایی و لگاریتم طبیعی مساحت بزرگی داشته و دارای مقدار تجدید مقیاس شده‌ی ارتفاع متوسط و مقادیر تجدید مقیاس شده‌ی ضریب رواناب و طول جغرافیایی بسیار کوچک هستند. همچنین مقدار تجدید مقیاس شده‌ی متوسط بارندگی سالانه برای ایستگاه‌های مربوط به این گره به طور میانگین منفی است. سایت‌های اختصاص یافته به گره چهارم، اغلب دارای مقادیر تجدید مقیاس شده‌ی بارندگی متوسط سالانه و ارتفاع متوسط تا بزرگ هستند و مقادیر تجدید مقیاس شده‌ی سایر ویژگی‌های آن‌ها به طور متوسط منفی است. لایه‌ی خروجی شبکه‌ی کوهونن یک بعدی برای حالت پنج منطقه‌ای، مطابق شکل (۴۷-۴) شامل ۵ گره به ترتیب با تعداد ۴، ۷، ۹، ۸ و ۹ ایستگاه است. وضعیت متوسط ویژگی‌های تجدید مقیاس شده‌ی ایستگاه‌های توزیع شده در میان این پنج گره، نیز به صورت شکل (۴۸-۴) است.



شکل (۴۷-۴) توزیع ایستگاه‌های هیدرومتری بین پنج گره لایه‌ی خروجی. شماره‌ها معرف کد ایستگاه‌های مورد مطالعه هستند. (عدد ۱۷ کد حوضه است).



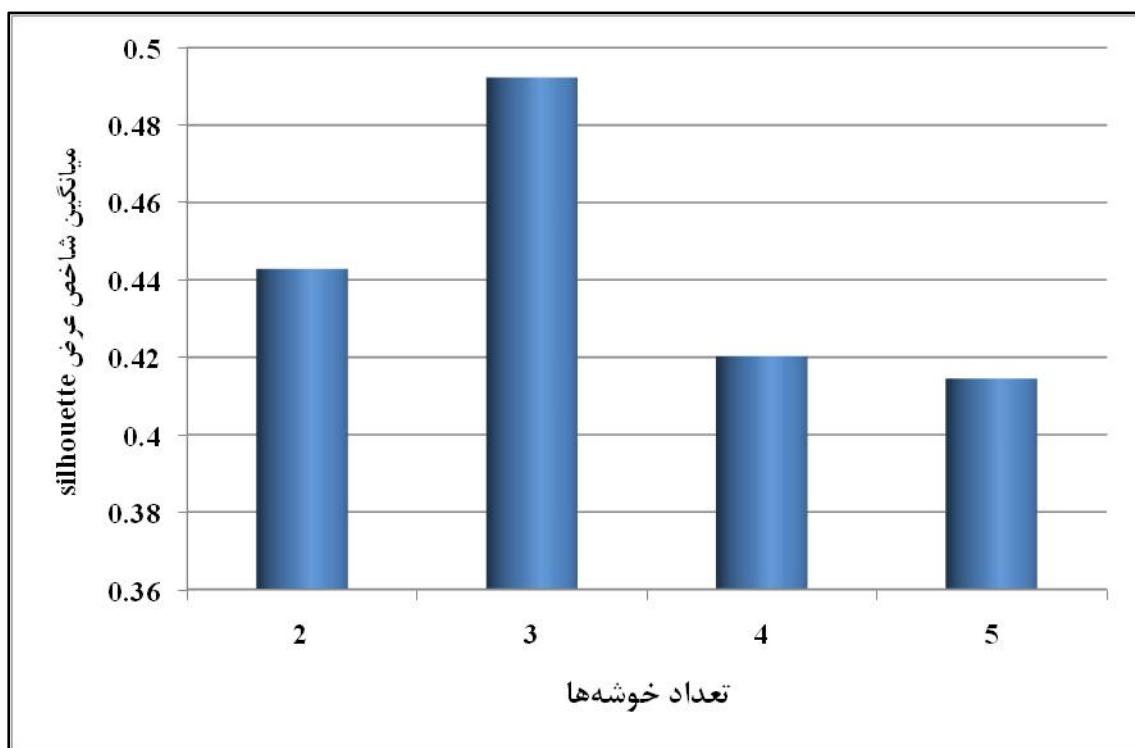
شکل (۴۸-۴) وضعیت متوسط ویژگی‌های تجدید مقیاس شده‌ی پنج گره لایه‌ی خروجی

مطابق شکل، گره اول نیز به طور متوسط شامل مقادیر تجدید مقیاس شده‌ی عرض جغرافیایی و بارندگی متوسط سالانه‌ی بزرگ، طول جغرافیایی و ضریب رواناب تجدید مقیاس شده‌ی متوسط و لگاریتم طبیعی مساحت و تراز ارتفاع تجدید مقیاس شده‌ی منفی است. گره دوم بیشتر مربوط به ایستگاه‌هایی با طول جغرافیایی، ارتفاع و ضریب رواناب تجدید مقیاس شده‌ی بزرگ، و از سوی دیگر عرض جغرافیایی، بارندگی متوسط سالانه و لگاریتم طبیعی مساحت تجدید مقیاس شده‌ی کوچک است. سایت‌های متعلق به گره سوم، اغلب دارای مقادیر تجدید مقیاس شده‌ی بارندگی متوسط سالانه نسبتاً کوچک یا متوسط و ارتفاع متوسط هستند و مقادیر تجدید مقیاس شده‌ی سایر ویژگی‌های آن‌ها به طور متوسط منفی است. گره چهارم به طور میانگین دربرگیرنده‌ی ایستگاه‌هایی با مقدار تجدید مقیاس شده‌ی لگاریتم طبیعی مساحت بزرگ، عرض جغرافیایی تجدید مقیاس شده‌ی متوسط، ارتفاع نسبتاً کوچک یا متوسط و طول جغرافیایی و ضریب رواناب تجدید مقیاس شده‌ی کوچک است. مقدار تجدید مقیاس شده‌ی بارندگی متوسط سالانه‌ی ایستگاه‌های موجود در این گره نیز به طور متوسط منفی است. گره پنجم به طور متوسط شامل مقادیر تجدید مقیاس شده‌ی عرض جغرافیایی بزرگ و ارتفاع متوسط، لگاریتم طبیعی مساحت و ضریب رواناب تجدید مقیاس شده‌ی کوچک و طول جغرافیایی و بارندگی متوسط سالانه‌ی تجدید مقیاس شده‌ی منفی است.

۴-۷-۲- شاخص میانگین عرض silhouette برای حالت‌های مختلف خوشبندی

در ادامه با محاسبه‌ی شاخص صحت خوشبندی میانگین عرض silhouette برای حالت‌های ۲ تا ۵ خوشبندی مشخص شد که مقدار این شاخص در حالت سه خوشبندی بیش از سایر حالت‌ها است و لذا بر این اساس، تعداد ۳ منطقه را می‌توان، تعداد بهینه از نظر وضعیت خوشبندی ایستگاه‌های مورد بررسی دانست. در هر چهار حالت خوشبندی، مقدار میانگین این شاخص، عددی بزرگ‌تر از ۰/۴ را اختیار می‌کند که با توجه به دامنه‌ی تعریف $1 \leq s(i) \leq 1$ ، مقادیر نسبتاً مناسبی برای شاخص صحت خوشبندی هستند.

شکل (۴-۴۹) مقادیر میانگین محاسبه شده شاخص میانگین عرض silhouette برای تعداد خوشبندی مختلف را نشان می‌دهد. پس از حالت سه خوشبندی، بر اساس شاخص صحت خوشبندی میانگین عرض silhouette، به ترتیب تعداد ۲، ۴ و ۵ خوشبندی توانند تعداد بهینه‌ی خوشبندی تشکیل شده با استفاده از نگاشتهای خودسازمانده باشند.

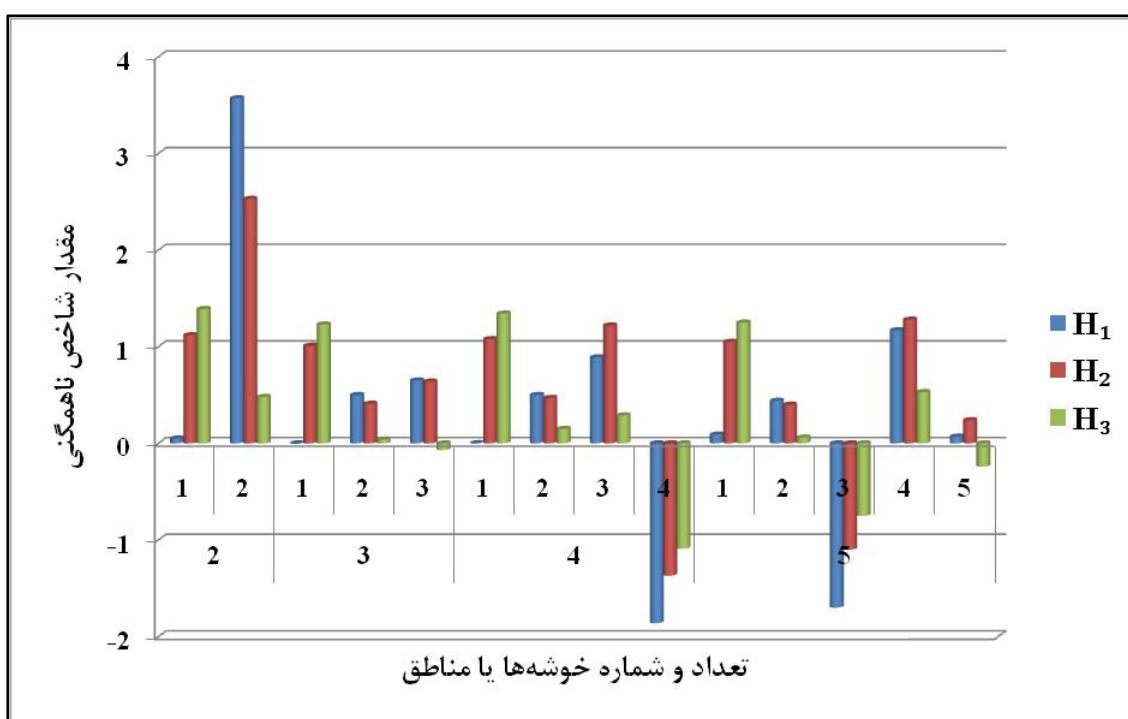


شکل (۴-۴۹) مقادیر متوسط شاخص صحت خوشبندی میانگین عرض silhouette برای حالت‌های ۲ تا ۵ خوشبندی

۴-۷-۳- بررسی وضعیت همگنی مناطق و انتخاب توزیع منطقه‌ای

پس از تشکیل مناطق با استفاده از نگاشتهای خودسازمانده، مقادیر شاخص‌های ناهمگنی H برای مناطق تشکیل شده در حالت‌های ۲ تا ۵ منطقه‌ای محاسبه شد که نتایج به دست آمده، در شکل (۴-۵۰) نشان داده شده است.

اگر بنا بر توصیه‌ی هاسکینگ و والیس (۱۹۹۷)، برای قضاوت در مورد وضعیت همگنی مناطق حاصل از تحلیل خوش‌های تنها شاخص H_1 را ملاک قرار دهیم، مشاهده می‌شود که در حالت‌های ۳ و ۴ منطقه‌ای، تمام مناطق تشکیل شده همگن هستند، اما در هر یک از حالت‌های ۲ و ۵ منطقه‌ای برای یک منطقه، شاخص ناهمگنی H_1 بیشتر از ۱ بوده و در نتیجه آن منطقه ناهمگن است. اگر برای همگنی یک منطقه، مقادیر هر سه شاخص H_1 , H_2 و H_3 در نظر گرفته شود، مشاهده می‌شود که وضعیت همگنی حالت ۳ منطقه‌ای نسبت به حالت ۴ منطقه‌ای نیز مناسب‌تر است.



شکل (۴-۵۰) مقادیر شاخص‌های ناهمگنی H برای مناطق تشکیل شده با استفاده از نگاشتهای خودسازمانده کوهون برای ۳۷ ایستگاه هیدرومتری مورد مطالعه در حوضه‌ی آبریز سفیدرود بزرگ

با توجه به این که حالت ۳ خوشبازی هم بر اساس شاخص صحت خوشبندی میانگین عرض silhouette، و هم از نظر همگنی مناطق تشکیل شده، نسبت به سایر حالتها وضعیت مناسبتری را فراهم می‌کند، این گزینه برای ادامه‌ی فرآیند تحلیل فراوانی منطقه‌ای سیلاپ انتخاب شد.

پس از تشکیل این سه منطقه، با استفاده از شاخص نکویی برازش Z ، توزیع منطقه‌ای بهینه برای هر یک از مناطق تعیین شد. در جدول (۴-۸) مقادیر شاخص نکویی برازش Z برای هر یک از پنج توزیع سه پارامتری مورد آزمایش در هر یک از سه منطقه ارائه شده است. همچنین در جدول (۴-۹) مقادیر برآورده شده پارامترهای هر یک از توزیع‌های منطقه‌ای درج شده است.

جدول (۴-۸) مقادیر شاخص نکویی برازش Z در هر یک از مناطق حالت ۳ منطقه‌ای ایجاد شده با استفاده از نگاشت‌های خودسازمانده

توزیع فراوانی

شماره منطقه	لjestیک	تعمیم یافته	تعمیم یافته	نرمال تعمیم	پیرسون تیپ	پارتوی تعییم یافته
۱	۰/۸۹۵۹	-۰/۲۰۷۳	*-۰/۲۹۳۹	-۰/۵۹۴۶	III	-۰/۶۰۸۹
۲	*۱/۲۴۷۷	۰/۰۹۵۴	*-۰/۰۷۳۸	-۰/۴۹۲۶		-۰/۴۸۸۵۳
۳	۴/۲۰۷	۲/۷۹۶۵	۱/۷۳۴۲	*-۰/۱۲۸۸		-۱/۰۴۰۱

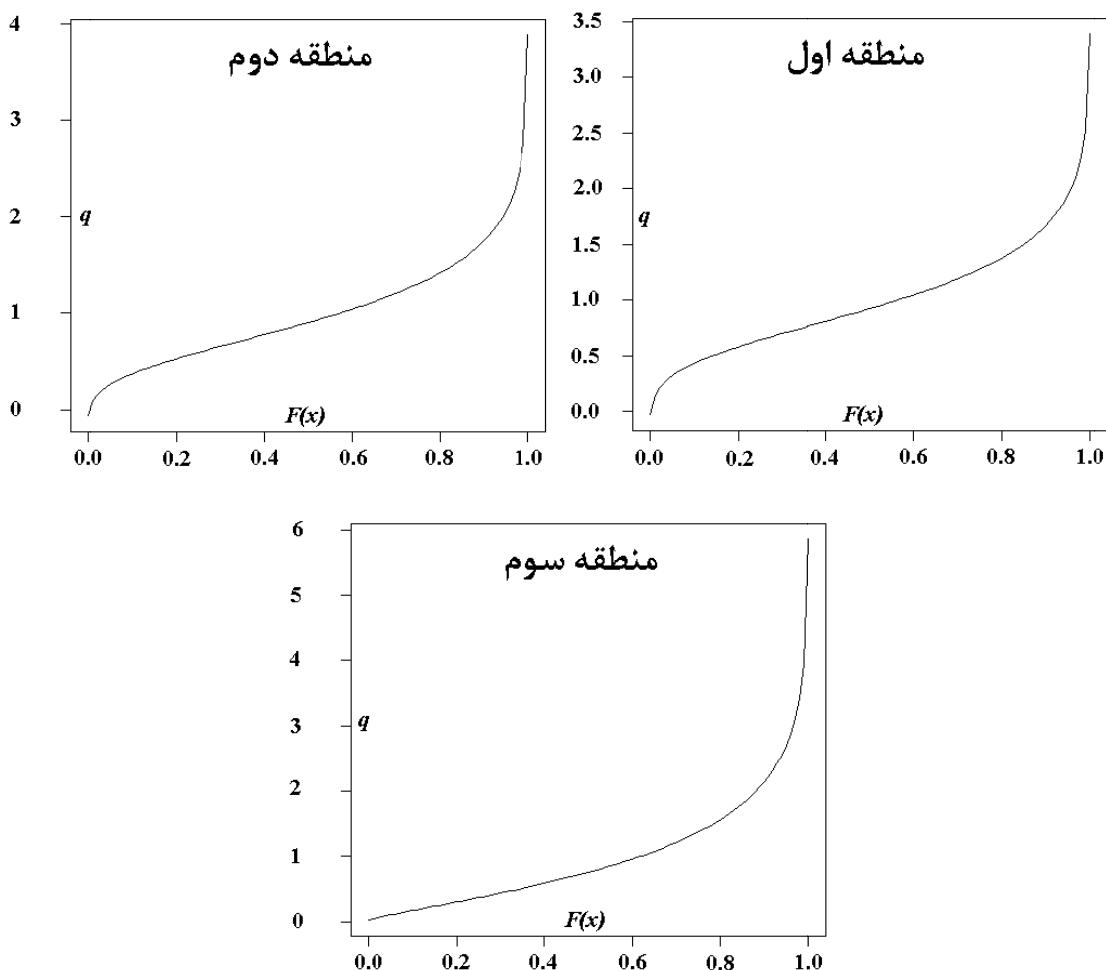
*توزیع منطقه‌ای برگزیده

جدول (۴-۹) مقادیر برآورده شده پارامترهای توزیع‌های منطقه‌ای در هر یک از مناطق حالت ۳ منطقه‌ای ایجاد شده با استفاده از نگاشت‌های خودسازمانده

شماره منطقه	توزیع منطقه‌ای	پارامترها	برآورده پارامتر
۱	مقادیر حدی تعمیم یافته	ξ	۰/۷۷۲۲
		α	۰/۴۰۷۸
		k	۰/۰۲۲۱
۲	نرمال تعمیم یافته	ξ	۰/۹۰۲۰
		α	۰/۵۲۱۹
		k	-۰/۳۶۳۲
۳	پیرسون تیپ III	μ	۱/۰۰
		σ	۰/۸۶۷۶
		γ	۱/۷۶۷۴

۴-۷-۴- برآورد چندک‌های منطقه‌ای سیلاب و دبی سیلاب ایستگاهها

پس از انتخاب توزیع بهینه برای هر منطقه و برآورد پارامترهای آن، چندک‌های منطقه‌ای سیلاب برای هر یک از مناطق محاسبه شد. در شکل (۵۱-۴) چندک‌های منطقه‌ای محاسبه شده برای سه منطقه‌ی ایجاد شده با استفاده از نگاشت‌های خودسازمانده، نشان داده شده است.



شکل (۵۱-۴) چندک‌های سیلاب منطقه‌ای (q) در برابر احتمال عدم تجاوز متناظر ($F(x)$) برای سه منطقه‌ی تشکیل شده با استفاده از نگاشت‌های خودسازمانده

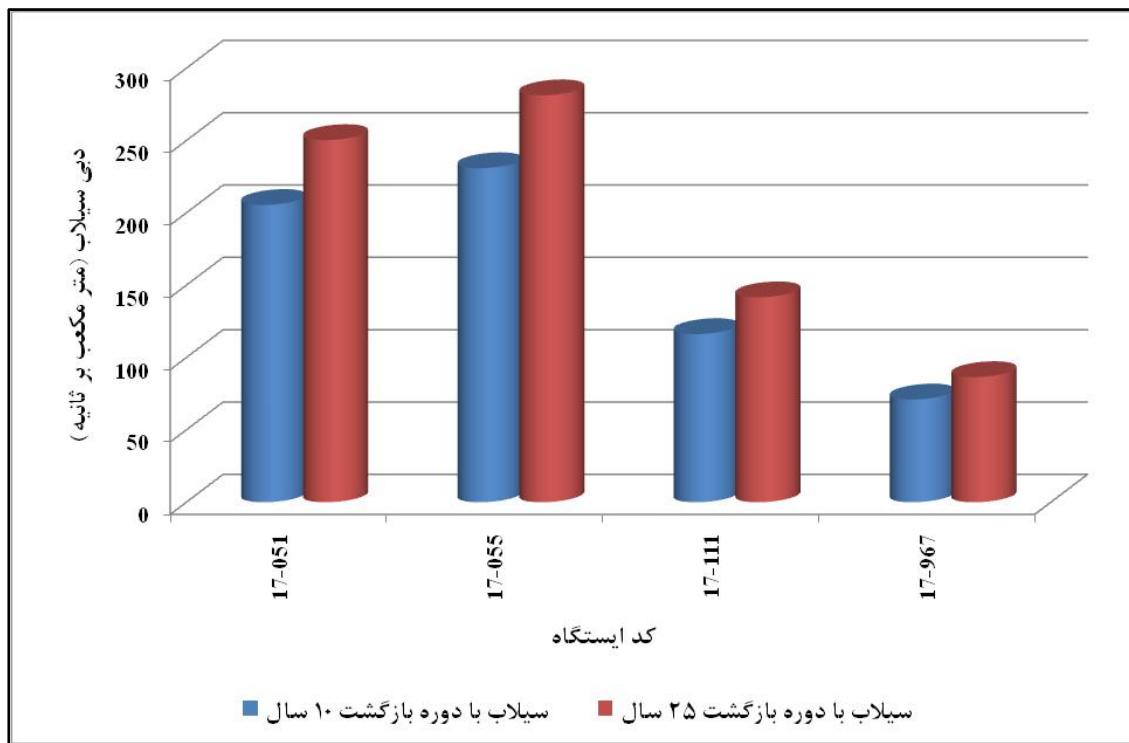
برای هر سایت یا ایستگاه، توزیع فراوانی منطقه‌ای با استفاده از عاملی که شاخص سیلاب یا سیلاب نمایه نامیده می‌شود، تبدیل به توزیع فراوانی ویژه‌ی هر سایت یا ایستگاه می‌شود. در واقع با

ضرب عامل سیلاب نمایه‌ی هر ایستگاه که در این مطالعه برابر میانگین آمار سالانه‌ی حداکثر دبی لحظه‌ای سیلاب هر سایت در چندک‌های منطقه‌ای سیلاب با دوره‌های بازگشت مختلف در نظر گرفته شده است، برآورد سیلاب با دوره‌های بازگشت مورد نظر انجام می‌شود.

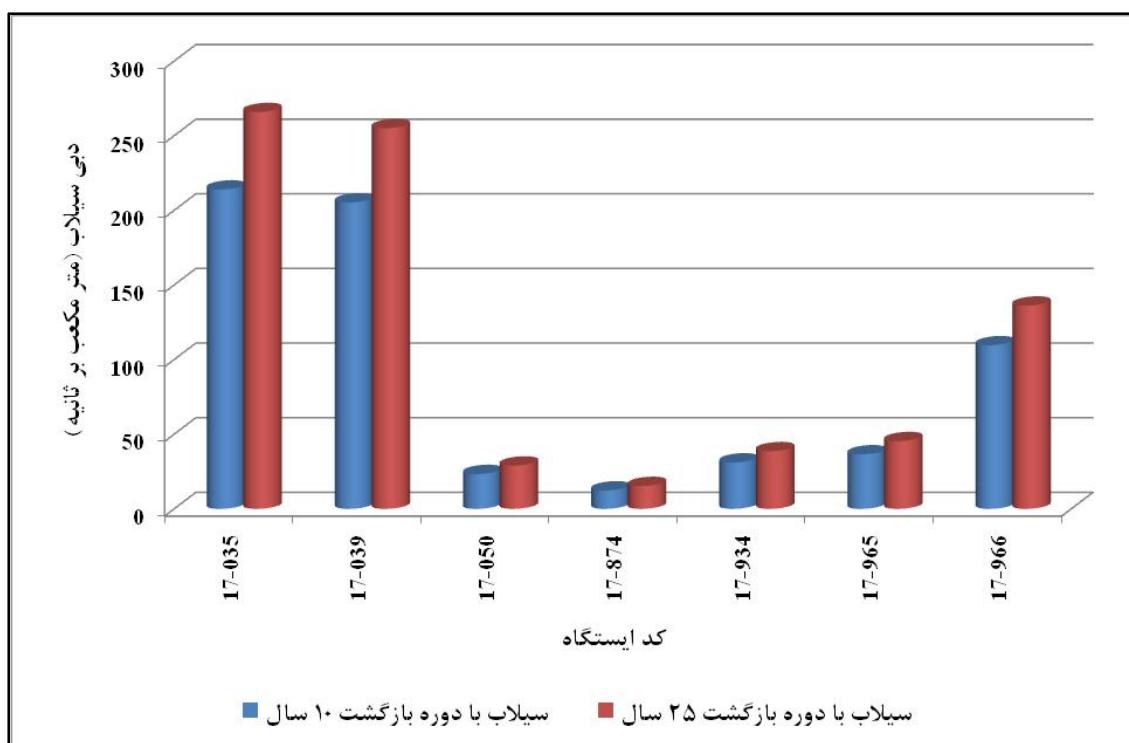
شایان ذکر است که تعداد داده‌های آماری مناطق اول تا سوم به ترتیب برابر ۱۱۰، ۱۵۰ و ۵۹۵ ایستگاه- سال است، از این رو برآورد سیلاب در مناطق اول و سوم برای دوره‌های بازگشت ۱۰ و ۲۵ سال، و در منطقه‌ی دوم برای دوره‌های بازگشت ۱۰، ۲۵، ۵۰ و ۱۰۰ سال انجام شد. در شکل‌های (۵۲-۴)، (۵۳-۴) و (۵۴-۴)، مقادیر برآورد شده‌ی حداکثر سیلاب با دوره‌های بازگشت مختلف برای ایستگاه‌های موجود در سه منطقه‌ی تشکیل شده به وسیله‌ی نگاشته‌ای خود- سازمانده به نمایش در آمده است.

در میان ایستگاه‌های متعلق به منطقه‌ی اول، ایستگاه ۱۷-۰۵۵ دارای بزرگ‌ترین مقادیر برآورد شده‌ی سیلاب است و کمترین مقادیر به ایستگاه ۱۷-۹۶۷ اختصاص دارد. در منطقه‌ی دوم، بالاترین مقادیر برآورد شده مربوط به ایستگاه ۱۷-۰۳۵ و پایین‌ترین مقادیر مربوط به ایستگاه ۱۷-۸۷۴ هستند. در منطقه‌ی سوم، بزرگ‌ترین مقادیر دبی‌های سیلاب برآورد شده به سایت ۱۷-۰۳۳ اختصاص دارد و کوچک‌ترین مقادیر متعلق به سایت ۱۷-۱۰۱ هستند.

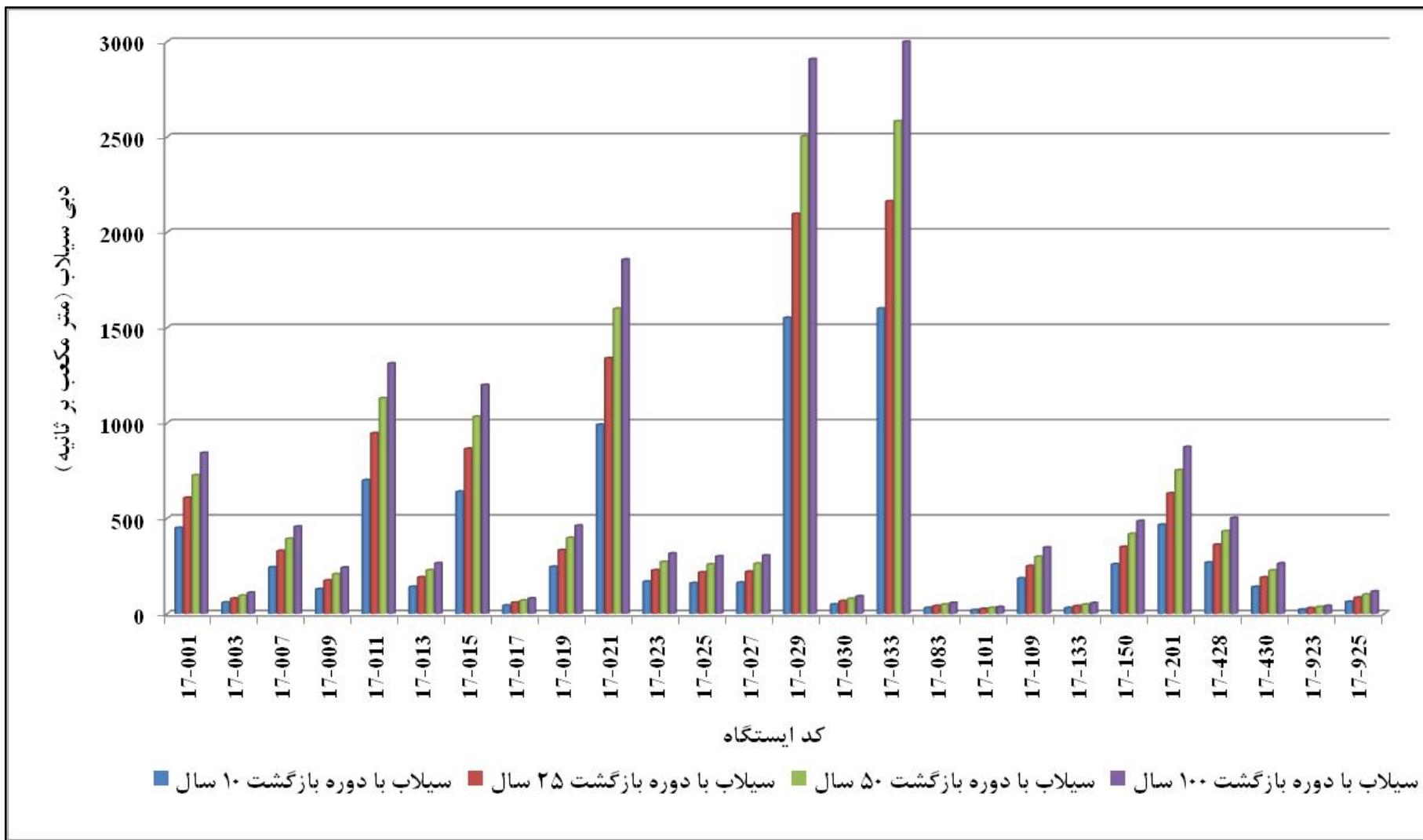
بزرگ‌ترین میزان اختلاف بین بزرگ‌ترین و کوچک‌ترین مقادیر برآورد شده‌ی سیلاب ایستگاه‌ها نسبت به بزرگ‌ترین مقدار برآورد شده در منطقه‌ی اول به طور تقریبی برابر $69/25$ درصد است. در منطقه‌ی دوم، بزرگ‌ترین مقدار اختلاف مذکور تقریباً معادل $92/24$ درصد بزرگ‌ترین مقدار برآورد شده‌ی سیلاب در این منطقه است. مقدار مورد نظر در منطقه‌ی سوم، حدوداً مساوی $98/85$ درصد بزرگ‌ترین مقدار برآورد شده‌ی سیلاب در این منطقه است. شایان ذکر است که با توجه به ثابت بودن توزیع فراوانی منطقه‌ای برای هر منطقه، برای برآورد سیلاب با دوره‌های بازگشت مختلف تنها عامل فراوانی ضرب شونده در توزیع تغییر می‌کند، که این موضوع تأثیری در میزان اختلاف نسبی محاسبه شده برای مناطق مختلف نخواهد داشت.



شکل (۵۲-۴) مقادیر برآورد شدهی سیلاب با دورههای بازگشت ۱۰ و ۲۵ سال برای ایستگاههای هیدرومتری منطقهی اول تشکیل شده به وسیلهی نگاشتهای خودسازمانده



شکل (۵۳-۴) مقادیر برآورد شدهی سیلاب با دورههای بازگشت ۱۰، ۲۵ و ۵۰ سال برای ایستگاههای هیدرومتری منطقهی دوم تشکیل شده به وسیلهی نگاشتهای خودسازمانده

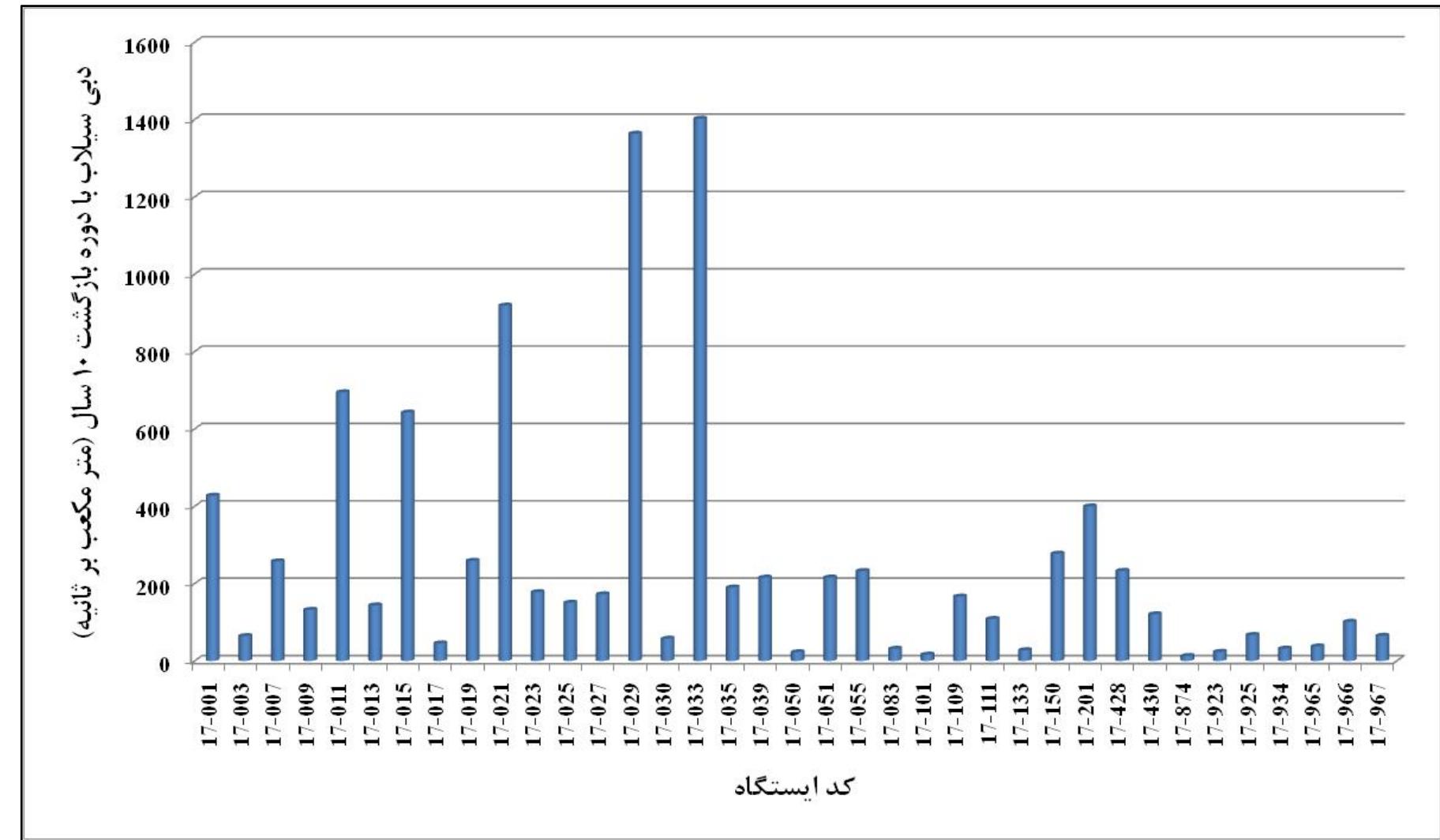


شکل (۵۴-۴) مقادیر برآورد شده سیلاب با دورهای بازگشت ۱۰، ۲۵، ۵۰ و ۱۰۰ سال برای ایستگاه‌های هیدرومتری منطقه‌ی سوم تشکیل شده به وسیله‌ی نگاشت‌های خودسازمانده

۴-۸-۴- تحلیل فراوانی نقطه‌ای سیلاپ

پس از منطقه‌بندی حوضه‌ی آبریز سفیدرود بزرگ به وسیله‌ی روش‌های مختلف تحلیل خوش‌های و برآورد سیلاپ با دوره‌های بازگشت مختلف با استفاده از تحلیل فراوانی منطقه‌ای سیلاپ، برآورد سیلاپ با استفاده از تحلیل فراوانی نقطه‌ای نیز برای هر یک از ۳۷ ایستگاه نهایی مورد مطالعه نیز انجام گرفت. بدین منظور مقادیر گشتاورهای خطی و نسبت‌های گشتاور خطی برای داده‌های آماری حداکثر دبی لحظه‌ای سیلاپ سالانه‌ی هر ایستگاه محاسبه شد. از طریق بررسی انطباق نسبت‌های چولگی خطی و کشیدگی خطی داده‌های آماری هر ایستگاه با نسبت‌های نظیر برای پنج توزیع سه پارامتری مورد آزمایش، توزیع فراوانی بهینه برای هر سایت انتخاب شد و سپس مقادیر پارامترهای مربوط به توزیع برآورد شدند. در ادامه مقادیر حداکثر دبی سیلاپ با دوره بازگشت ۱۰ سال برای هر یک از سایت‌ها مورد محاسبه قرار گرفت. انتخاب دوره بازگشت ۱۰ سال با توجه به تعداد داده‌های آماری ایستگاهها که در محدوده ۱۰ تا ۴۰ قرار دارد، صورت گرفت. در شکل (۴-۵۵) مقادیر برآورد شده‌ی سیلاپ با دوره بازگشت ۱۰ سال برای هر یک از ۳۷ ایستگاه مورد مطالعه مشاهده می‌شود. همان‌طور که در شکل دیده می‌شود، بزرگ‌ترین مقادیر برآورد شده‌ی دبی سیلاپ با دوره بازگشت ۱۰ سال، به ترتیب مربوط به ایستگاه‌های ۳۳-۰۳، ۲۹-۰۷، ۲۱-۰۷ و ۲۱-۰۷ هستند. در مقابل، کوچک‌ترین مقادیر برآورد شده‌ی دبی سیلاپ با دوره بازگشت ۱۰ سال، به ترتیب به ایستگاه‌های ۸۷۴-۰۷، ۱۰۱-۰۷ و ۵۰-۰۷ اختصاص دارند.

در بخش آینده، مقادیر برآوردهای سیلاپ حاصل از استفاده از روش‌های مختلف خوش‌های برای منطقه‌بندی و تحلیل فراوانی منطقه‌ای حوضه‌ی آبریز سفیدرود بزرگ با یکدیگر مورد مقایسه قرار خواهند گرفت. همچنین مقایسه‌ای میان متوسط نتایج حاصل از روش‌های منطقه‌ای و برآوردهای به دست آمده از طریق تحلیل فراوانی نقطه‌ای ارائه خواهد شد.



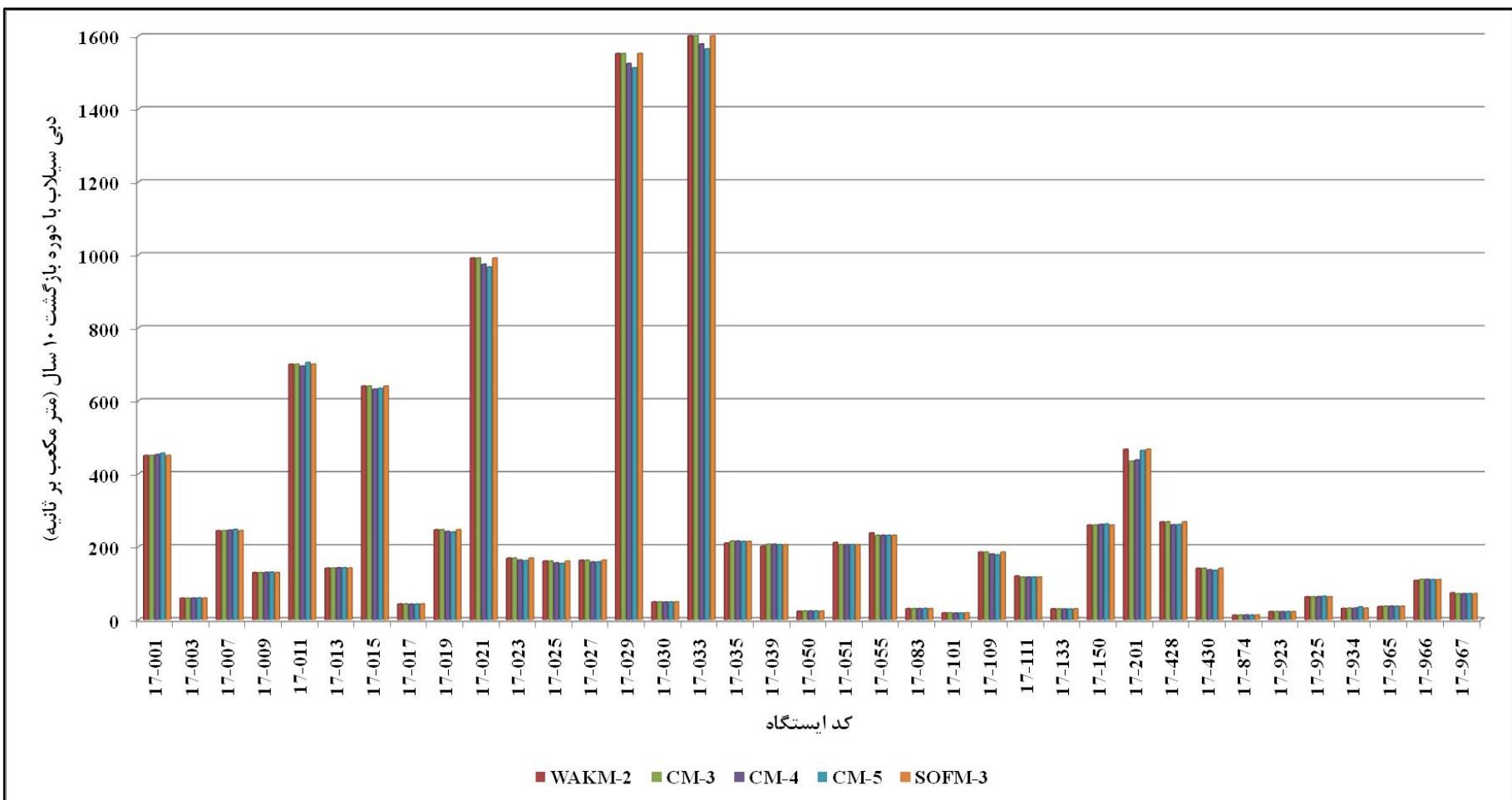
شکل (۴-۵۵) مقادیر برآوردهی سیلاب با استفاده از تحلیل فراوانی نقطه‌ای هیدرومتری حوضه‌ی آبریز سفیدرود بزرگ

۴-۹-۴- مقایسه‌ی مقادیر برآوردها

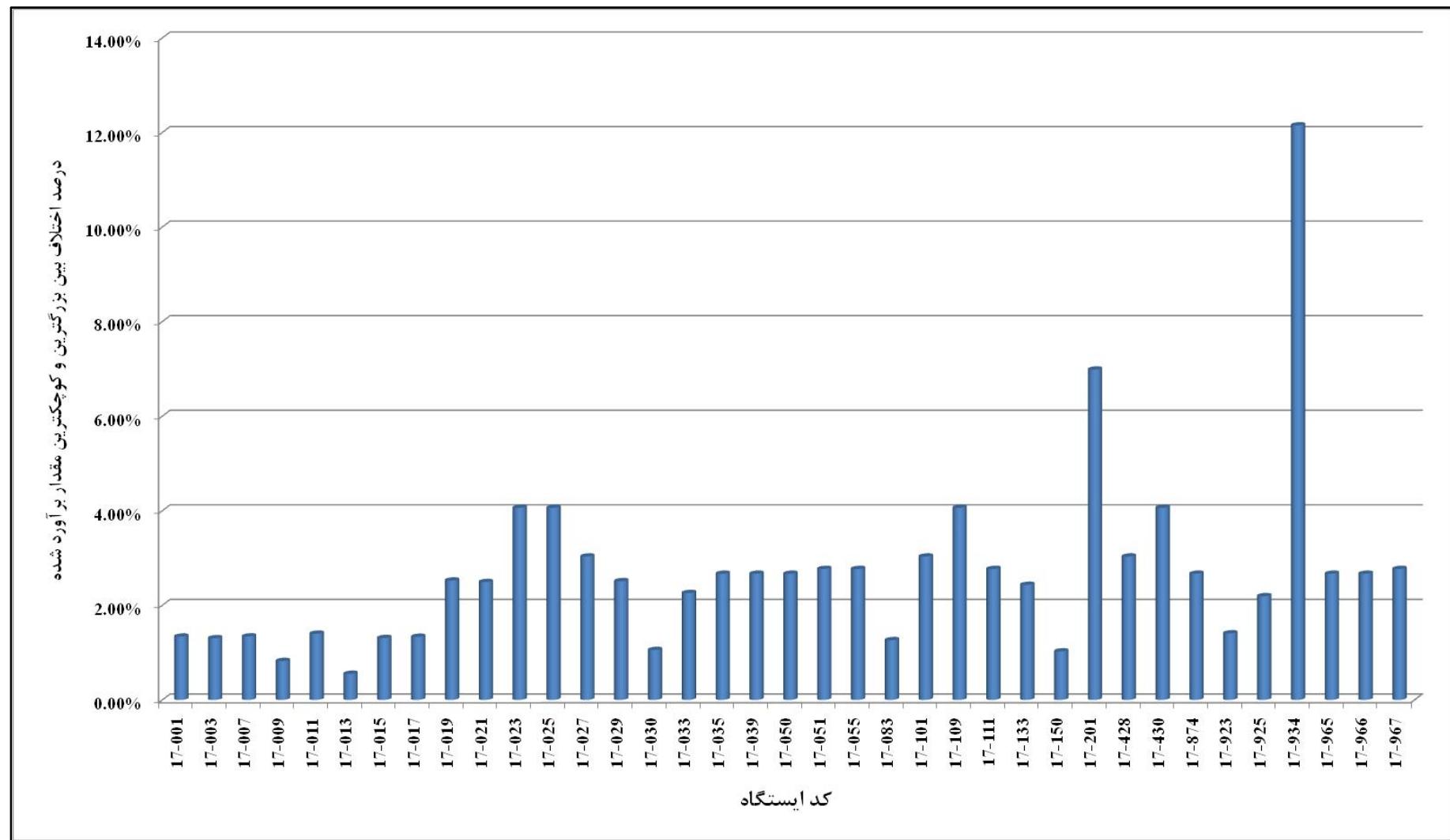
در شکل (۵۶-۴) مقادیر برآورد شده‌ی دبی حداکثر سیلاب با دوره بازگشت ۱۰ سال با استفاده از تحلیل فراوانی منطقه‌ای انجام شده بر اساس پنج گزینه‌ی نهایی منطقه‌بندی به وسیله‌ی روش‌های تحلیل خوش‌های ارائه شده است. این پنج گزینه عبارتند از: حالت دو منطقه‌ای حاصل از خوش‌هه بندی ترکیبی WAKM، حالت‌های سه، چهار و پنج منطقه‌ای به دست آمده از طریق الگوریتم خوش‌هه بندی فازی c-means، و حالت سه منطقه‌ای منطقه‌بندی با استفاده از نگاشت‌های خود سازمانده کوهونن. همچنین در شکل (۵۷-۴)، اختلاف نسبی بزرگ‌ترین و کوچک‌ترین برآورد حاصل از پنج گزینه‌ی فوق الذکر برای هر یک از ۳۷ ایستگاه مورد مطالعه به نمایش در آمده است. برای محاسبه‌ی این مقدار برای هر ایستگاه، اختلاف بزرگ‌ترین و کوچک‌ترین دبی برآورد شده برای آن ایستگاه بر مقدار بزرگ‌ترین دبی برآورد شده تقسیم شده و حاصل به صورت درصد بیان شده است.

شکل (۵۸-۴) نشان دهنده‌ی مقادیر برآورد شده‌ی دبی حداکثر سیلاب با دوره بازگشت ۱۰ سال با استفاده از تحلیل فراوانی نقطه‌ای، در کنار میانگین مقادیر برآورد شده‌ی دبی حداکثر سیلاب با همان دوره‌ی بازگشت با استفاده از پنج گزینه‌ی نهایی تحلیل فراوانی منطقه‌ای است. در شکل (۵۹-۴) اختلاف نسبی برآورد دبی سیلاب ۱۰ ساله با استفاده از تحلیل فراوانی نقطه‌ای و میانگین دبی‌های سیلاب ۱۰ ساله‌ی برآورد شده به وسیله‌ی پنج گزینه‌ی تحلیل فراوانی منطقه‌ای، برای ایستگاه‌های مورد مطالعه نشان داده شده است. برای محاسبه‌ی این مقدار برای هر ایستگاه، قدر مطلق اختلاف دبی برآورد شده‌ی نقطه‌ای و میانگین برآوردهای منطقه‌ای برای آن ایستگاه بر مقدار میانگین دبی‌های برآورد شده‌ی منطقه‌ای تقسیم شده و حاصل به صورت درصد بیان شده است.

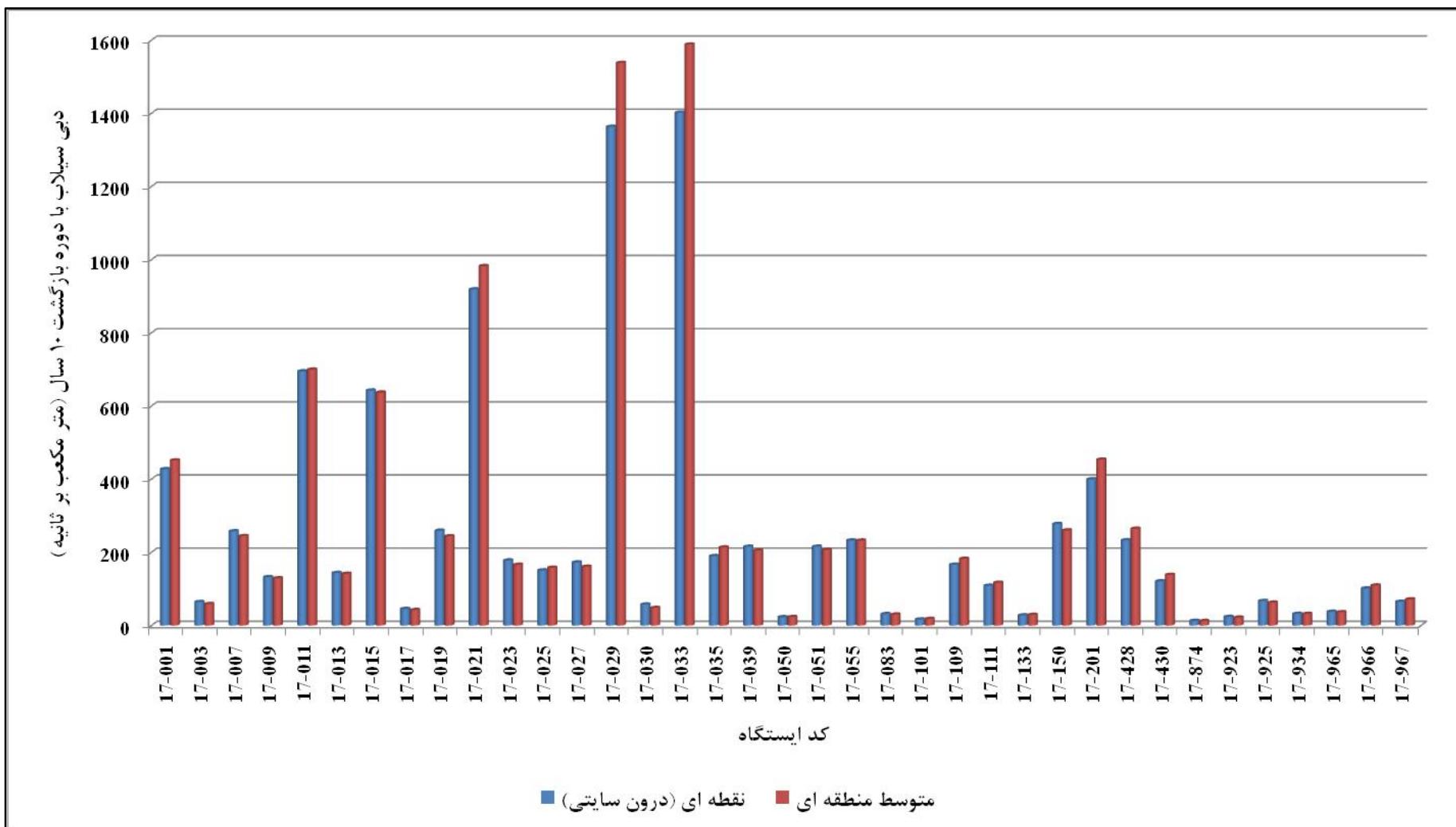
در شکل (۶۰-۴) مقادیر میانگین قدر مطلق اختلاف بین برآوردهای دبی سیلاب با دوره بازگشت ۱۰ سال به دست آمده از هر یک از گزینه‌های نهایی تحلیل منطقه‌ای با برآوردهای سیلاب ۱۰ ساله حاصل از تحلیل فراوانی نقطه‌ای نشان داده شده است.



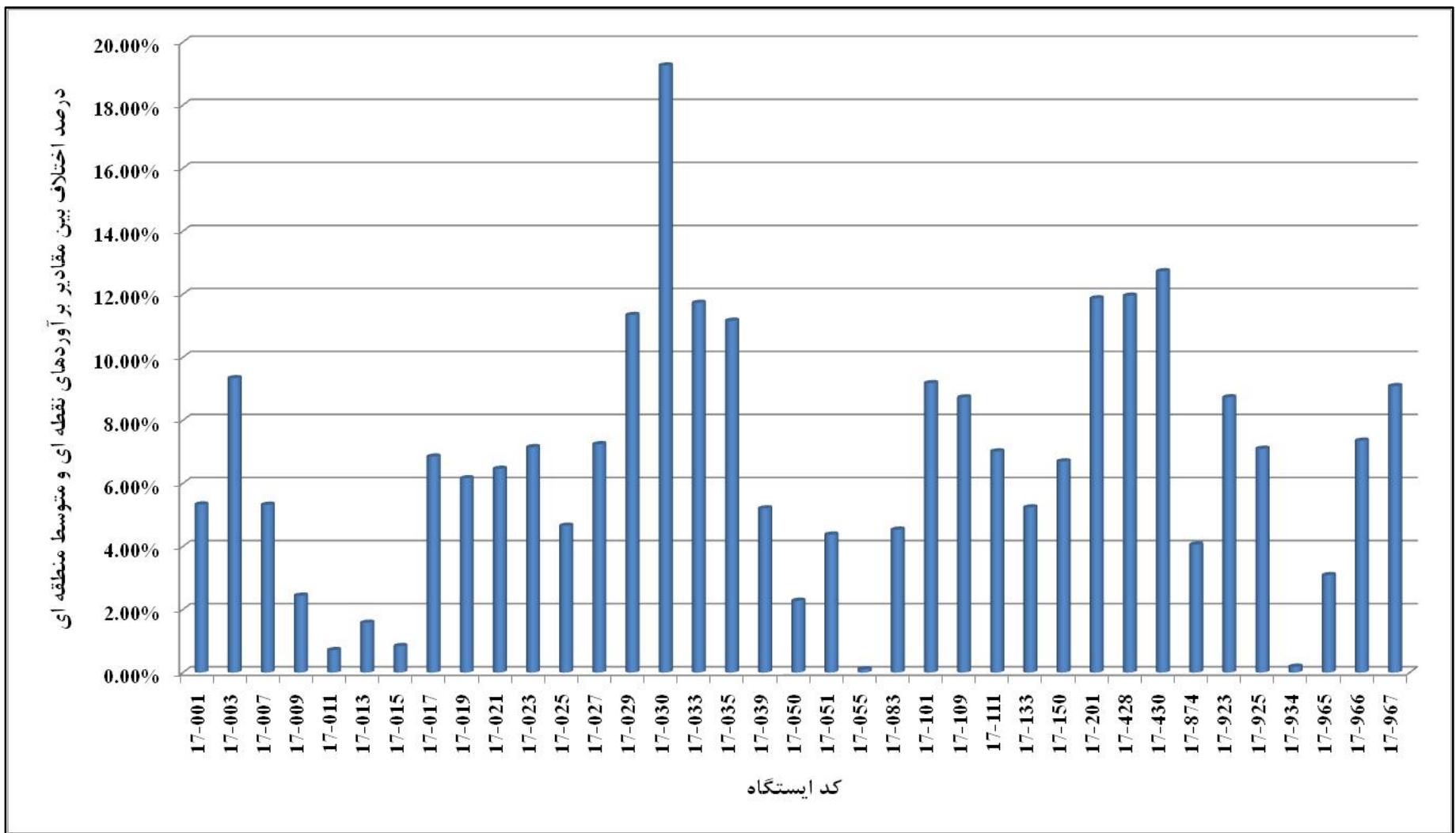
شکل (۵۶-۴) مقادیر برآورده سیلاب با استفاده از روش‌های تحلیل فراوانی منطقه‌ای با دوره بازگشت ۱۰ سال برای ایستگاه‌های هیدرومتری حوضه‌ی آبریز سفیدرود بزرگ. معرف حالت دو منطقه‌ای حاصل از خوشبندی ترکیبی WAKM، CM-3، CM-4، CM-5 و SOFM-3 به ترتیب نماد حالت‌های سه، چهار و پنج منطقه‌ای به دست آمده از طریق الگوریتم خوشبندی فازی c-means و SOFM-3 نشان دهنده‌ی حالت سه منطقه‌ای منطقه‌بندی با استفاده از نگاشت‌های خود سازمانده کوهونن هستند.



شکل (۴-۵) اختلاف نسبی بزرگترین و کوچکترین برآورد حاصل از پنج گزینه‌ی نهایی تحلیل فراوانی منطقه‌ای برای هر یک از ۳۷ ایستگاه هیدرومتری مورد مطالعه



شکل (۵۸-۴) مقادیر برآورده شده‌ی دبی حداکثر سیلاب با استفاده از تحلیل فراوانی نقطه‌ای و میانگین مقادیر برآورده شده‌ی دبی حداکثر سیلاب با استفاده از پنج گزینه‌ی نهایی تحلیل فراوانی منطقه‌ای برای دوره بازگشت ۱۰ سال



شکل (۴-۵۹) اختلاف نسبی برآورد سیلان ۱۰ ساله با استفاده از تحلیل فراوانی نقطه‌ای و میانگین دیهای سیلان ۱۰ ساله برآورد شده به وسیله پنج گزینه تحلیل فراوانی منطقه‌ای

مقدار اختلاف مذکور برای هر یک از الگوریتم‌های منطقه‌ای نهایی با تعداد n ایستگاه نهایی در مناطق، با استفاده از رابطه‌ی (۲-۴) محاسبه شده است.

$$d_{alg}^{abs} = \frac{\sum_{i=1}^n |Q_{i,10_{alg}} - Q_{i,10_{at-site}}|}{n} \quad (2-4)$$

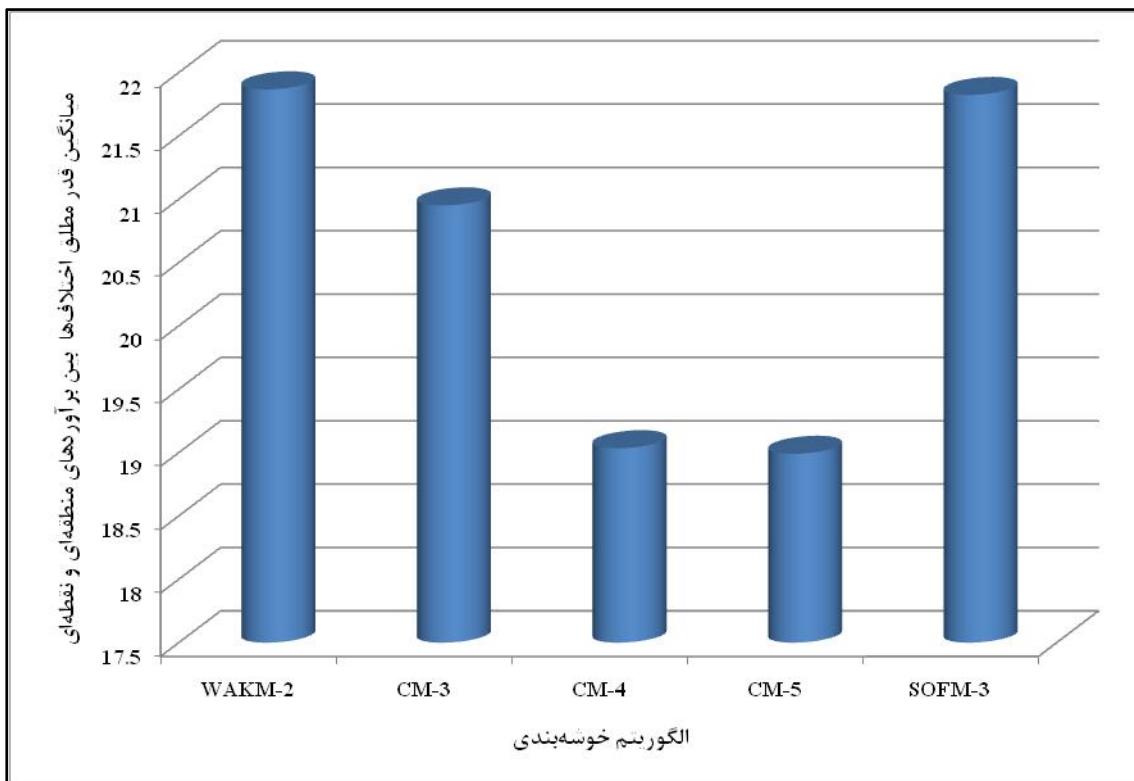
در این رابطه، d_{alg}^{abs} مقدار اختلاف مذکور بین یک الگوریتم منطقه‌ای مورد نظر و مقادیر برآوردهای نقطه‌ای است (alg با نام الگوریتم جایگزین می‌شود). $Q_{i,10_{alg}}$ مقدار دبی سیلاب ۱۰ ساله‌ی برآورده شده در ایستگاه i با استفاده از الگوریتم مورد نظر و $Q_{i,10_{ave}}$ مقدار میانگین دبی سیلاب برآورده شده‌ی تحلیل فراوانی نقطه‌ای با دوره بازگشت ۱۰ سال در ایستگاه i است.

آن چنان که در شکل (۶۰-۴) مشخص است، بزرگ‌ترین مقدار این اختلاف مربوط به گزینه‌ی مبتنی بر منطقه‌بندی با استفاده از الگوریتم ترکیبی WAKM در حالت ۲ منطقه‌ای (WAKM-2) است و پس از آن با فاصله‌ی اندکی گزینه‌ی مربوط به حالت ۳ منطقه‌ای حاصل از خوشبندی با استفاده از نگاشتهای خودسازمانده قرار دارد. اما کمترین اختلاف مربوط به حالت ۵ منطقه‌ای به دست آمده به وسیله‌ی الگوریتم c-means (CM-5) است.

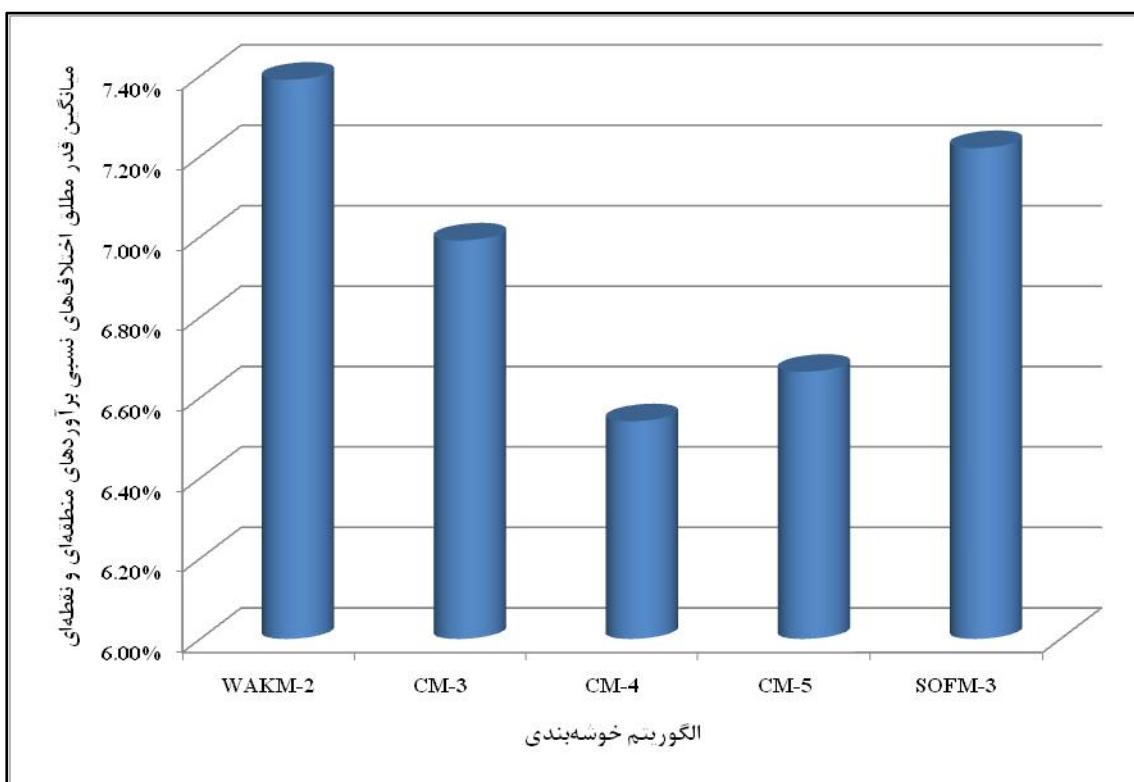
اگر مقدار اختلاف مورد نظر برای هر ایستگاه مطابق رابطه‌ی (۳-۴) به صورت نسبی محاسبه شود، آنگاه مقادیر میانگین قدر مطلق اختلاف نسبی بین برآوردهای دبی سیلاب با دوره بازگشت ۱۰ سال به دست آمده از هر یک از گزینه‌های نهایی تحلیل منطقه‌ای با برآوردهای سیلاب ۱۰ ساله‌ی حاصل از تحلیل فراوانی نقطه‌ای به صورت شکل (۶۱-۴) خواهد بود.

$$rd_{alg}^{abs} = \frac{\sum_{i=1}^n |(Q_{i,10_{alg}} - Q_{i,10_{at-site}})/Q_{i,10_{at-site}}|}{n} \quad (3-4)$$

همان طور که در شکل مشاهده می‌شود، به جز دو گزینه‌ی CM-4 و CM-5 که جایگاه آن‌ها با یکدیگر عوض می‌شود، جایگاه سایر گزینه‌ها نسبت به شکل (۶۰-۴) دستخوش تغییر نمی‌شود.



شکل (۶۰-۴) مقادیر میانگین قدر مطلق اختلافهای بین برآوردهای دبی سیلاب با دوره بازگشت ۱۰ سال به دست آمده از هر یک از گزینه‌های نهایی تحلیل منطقه‌ای و برآوردهای سیلاب ۱۰ ساله در تحلیل نقطه‌ای

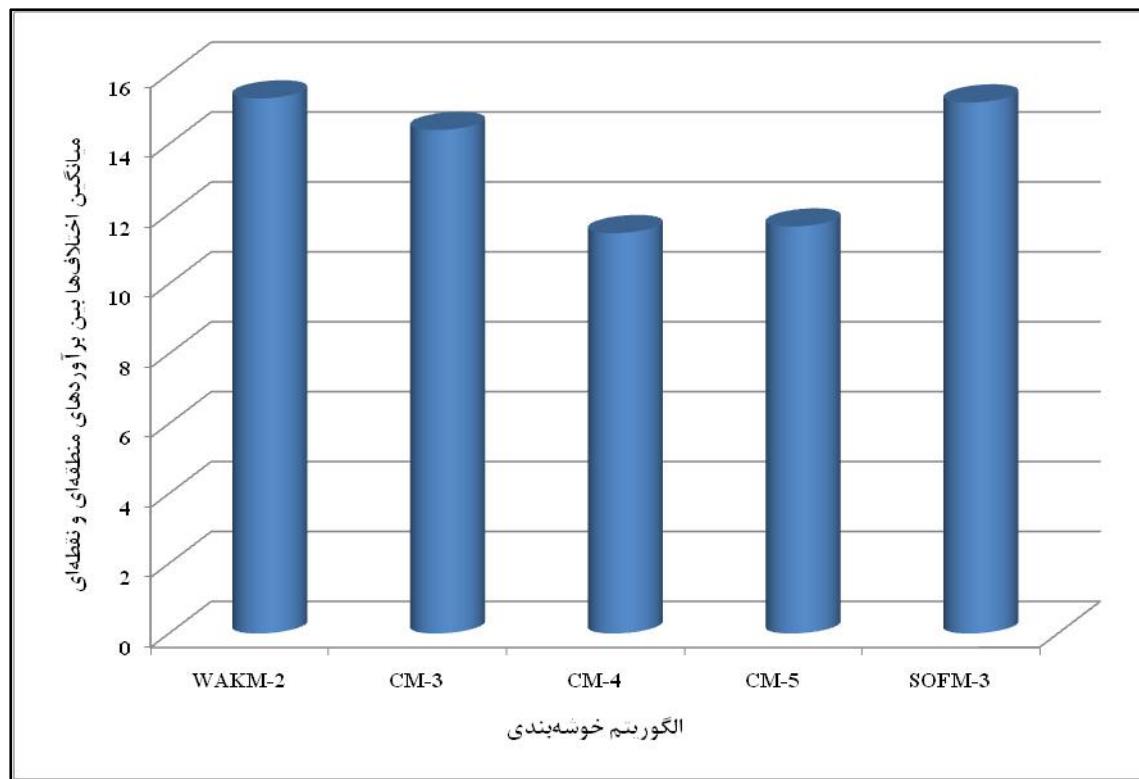


شکل (۶۱-۴) مقادیر میانگین قدر مطلق اختلافهای نسبی بین برآوردهای دبی سیلاب با دوره بازگشت ۱۰ سال به دست آمده از هر یک از گزینه‌های نهایی تحلیل منطقه‌ای و برآوردهای سیلاب ۱۰ ساله در تحلیل نقطه‌ای

با بر نتایج منعکس در شکل (۶۱-۴) بیشترین مقدار میانگین قدر مطلق اختلافهای نسبی مربوط به گزینه‌ی ۲ WAKM است، در حالی که کمترین مقدار مربوط به گزینه‌ی CM-4 است. اگر در محاسبه‌ی مقادیر اختلاف از قدر مطلق استفاده نشود و علامت مثبت یا منفی مقادیر که نشانه‌ی بیشتر یا کمتر بودن مقدار برآورد منطقه‌ای نسبت به مقدار برآورد نقطه‌ای است، در محاسبات منظور شود، آنگاه رابطه‌ی (۴-۴) به رابطه‌ی (۴-۲) تبدیل خواهد شد. نتیجه‌ی محاسبه‌ی میانگین اختلافها با استفاده از این رابطه در شکل (۶۲-۴) نشان داده شده است.

$$d_{alg} = \frac{\sum_{i=1}^n (Q_{i,10_{alg}} - Q_{i,10_{at-site}})}{n} \quad (4-4)$$

در این حالت بالاترین مقدار اختلاف مربوط به منطقه‌بندی حاصل از عملکرد الگوریتم ترکیبی WAKM-2، و پایین‌ترین مقدار متعلق به گزینه‌ی CM-4 است.

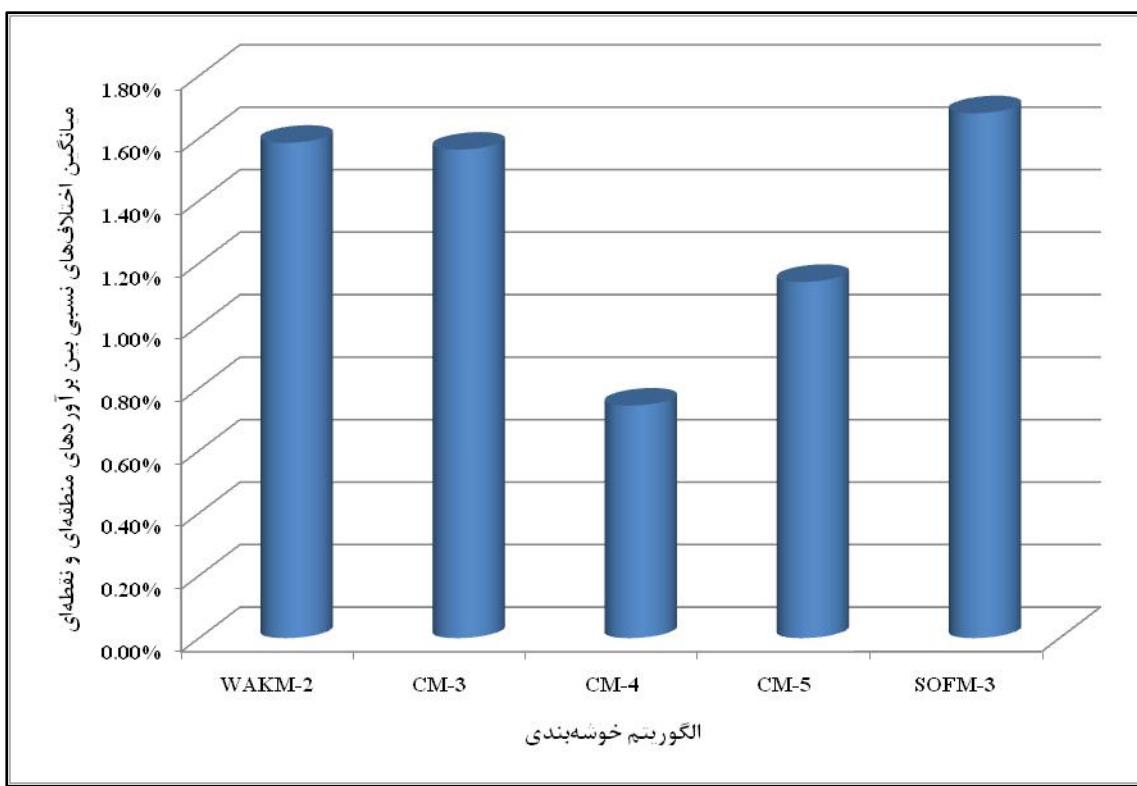


شکل (۶۲-۴) مقادیر میانگین اختلاف بین برآوردهای دبی سیلاب با دوره بازگشت ۱۰ سال به دست آمده از هر یک از گزینه‌های نهایی تحلیل منطقه‌ای و برآوردهای سیلاب ۱۰ ساله در تحلیل نقطه‌ای

اگر مقدار اختلاف نشان داده شده در شکل (۶۲-۴) برای هر ایستگاه مطابق رابطه‌ی (۵-۴) به صورت نسبی محاسبه شود، آنگاه مقادیر میانگین اختلاف نسبی بین برآوردهای دبی سیلاب با دوره بازگشت ۱۰ سال به دست آمده از هر یک از گزینه‌های نهایی تحلیل منطقه‌ای با برآوردهای سیلاب ۱۰ ساله‌ی حاصل از تحلیل فراوانی نقطه‌ای به صورت شکل (۶۳-۴) خواهد بود.

$$rd_{alg} = \frac{\sum_{i=1}^n \left[(Q_{i,10alg} - Q_{i,10at-site}) / Q_{i,10at-site} \right]}{n} \quad (5-4)$$

در این شکل، گزینه‌ی SOFM-3 بالاترین مقدار و گزینه‌ی CM-4، پایین‌ترین مقدار را به خود اختصاص داده است و به نظر می‌رسد روند کلی تغییرات نسبت به شکل (۶۱-۴) تفاوت چندانی ندارد، اما در این شکل مقادیر میانگین محاسبه شده کاهش یافته‌اند.



شکل (۶۳-۴) مقادیر میانگین اختلاف‌های نسبی بین برآوردهای دبی سیلاب با دوره بازگشت ۱۰ سال به دست آمده از هر یک از گزینه‌های نهایی تحلیل منطقه‌ای و برآوردهای سیلاب ۱۰ ساله در تحلیل نقطه‌ای

فصل پنجم:

نتیجہ گیری نهایی و پیشنهادات

۱-۵- تحلیل نتایج و نتیجه‌گیری نهایی

پس از بررسی نتایج به دست آمده از فرآیندهای منطقه‌بندی و تحلیل فراوانی منطقه‌ای سیلاب در حوضه‌ی آبریز سفیدرود بزرگ که در این مطالعه حاصل شد، این موارد را می‌توان به صورت ذیل خلاصه کرد:

- ۱) بر اساس مقادیر محاسبه شده‌ی شاخص ناجوری D برای ۴۰ ایستگاه هیدرومتری اولیه‌ی حوضه‌ی سفیدرود که برای این مطالعه انتخاب شدند، دو ایستگاه ۱۷-۰۳۷ و ۱۷-۹۲۶ از نظر آماری با سایر ایستگاه‌های مورد مطالعه ناسازگار بوده و از مجموعه‌ی ایستگاه‌های مورد بررسی کنار گذاشته شدند. همچنین از آنجا که ایستگاه ۱۷-۰۴۱ در اغلب موارد موجب افزایش قابل توجه مقادیر شاخص‌های ناهمگنی و ایجاد ناهمگنی در خوشهای که به آن تعلق می‌گرفت می‌شد، این ایستگاه نیز از محاسبات نهایی منطقه‌بندی و تحلیل فراوانی حذف شد.
- ۲) در یک الگوریتم خوشبندی ترکیبی، انتظار می‌رود که الگوریتم خوشبندی سلسه مراتبی K-means عملکرد در فراهم کردن مراکز اولیه‌ی مناسب برای به کارگیری در الگوریتم K-means مفیدی داشته باشد به طوری که الگوریتم K-means خروجی بهتری داشته باشد. بررسی نتایج به دست آمده نشان داد که در حوضه‌ی آبریز سفیدرود بزرگ، مدل‌های ترکیبی از نظر کمینه‌سازی تابع هدف و ارائه‌ی مقادیر مناسب برای شاخص صحت خوشبندی میانگین عرض silhouette، عملکرد کلی بهتری نسبت به الگوریتم‌های سلسه مراتبی و K-means از خود ارائه می‌کنند. با این حال نمی‌توان با قطعیت در مورد برتری مطلق خروجی خوشبندی ترکیبی نسبت به تمام حالات اجرای الگوریتم K-means در هر منطقه‌ی مورد مطالعه اظهار نظر کرد. در میان چهار مدل ترکیبی ارائه شده در این مطالعه، الگوریتم WAKM که در واقع ترکیب الگوریتم سلسه مراتبی Ward و الگوریتم افزایی K-means است، از نظر کمینه‌سازی تابع هدف، ارائه‌ی مقادیر مناسب شاخص صحت خوشبندی و پراکنش یکنواخت‌تر سایتها

بین مناطق از جهت تعداد، نتایج مناسب‌تری را حاصل کرد. این نتیجه با نتیجه‌ی حاصل از پژوهش رائو و سرینیواس (۲۰۰۶a) مطابقت دارد.

۳) از آنجا که در میان حالت‌های ۲ تا ۵ منطقه‌ای حاصل از به کارگیری الگوریتم WAKM،
حالت دو منطقه‌ای هم از نظر مقدار شاخص صحت خوشبندی، هم از جهت مقادیر شاخص-
های ناهمگنی و هم از نظر تشکیل مناطقی با تعداد داده‌های آماری بیش‌تر بهترین وضعیت را
دارا بود، این حالت به عنوان حالت نهایی و بهینه برای اجرای تحلیل فراوانی منطقه‌ای انتخاب
شد.

۴) در مورد خوشبندی فازی یا استفاده از الگوریتم c-means، خوشبندی‌های تشکیل شده نسبت به
تغییر مقدار μ برای یک تعداد مشخص از خوشبندی‌ها حساس هستند. از این رو مقدار بهینه‌ی μ
برای منطقه‌بندی باید با بررسی دقیق و گام به گام تعیین شود. استفاده از مقدار $2 = \mu$ که
در برخی مطالعات به صورت پیش فرض به عنوان مقدار بهینه در نظر گرفته می‌شود، نمی-
تواند چندان مناسب باشد.

۵) امکان عضویت همزمان یک سایت در دو یا چند منطقه، عمده‌ترین مزیت خوشبندی فازی
نسبت به خوشبندی سخت است که موجب افزایش تعداد داده‌های آماری موجود در مناطق
می‌شود و امکان برآورد سیلاپ با دوره‌های بازگشت طولانی‌تر را فراهم می‌کند. در این مطالعه
مشاهده شد در صورتی که آستانه‌ی I/c به عنوان آستانه‌ی عضویت در مناطق انتخاب شود،
معمولًاً با افزایش تعداد مناطق (c) که متراffد با کاهش مقدار I/c است، مجموع داده‌های
آماری موجود در کل مناطق افزایش می‌یابد، اما باید توجه داشت که افزایش تعداد مناطق
خود به طور معمول تقسیم بیش‌تر داده‌ها و کوچک شدن مناطق (از نظر تعداد داده‌های
موجود) را به دنبال دارد، لذا برای انتخاب تعداد بهینه‌ی مناطق از این حیث نیز باید هر دو
مسئله مورد توجه قرار گیرند.

۶) درباره‌ی عملکرد شاخص‌های صحت خوشبندی فازی، مشاهده شد که به ازای تعداد ثابتی از

خوشه‌ها همواره با افزایش مقدار μ ، مقدار ضریب افزایش کاهش می‌یابد. همچنین در صورت ثابت ماندن تعداد خوشه‌ها، با افزایش مقدار μ ، منحنی تغییرات انتروپی افزایش و شاخص عملکرد فازی، حالت اکیداً صعودی خواهد داشت. از این رو می‌توان گفت که انتخاب مقدار بهینه‌ی μ برای تعداد مشخصی از خوشه‌ها بر مبنای مقادیر محاسبه شده‌ی این سه شاخص امکان پذیر نیست. بررسی تغییرات مقدار شاخص فوکویاما و سوگنو به ازای تغییر مقدار μ در فاصله‌ی ۱/۱ تا ۳ برای ۲ تا ۵ خوشه نیز بیانگر آن است که مقدار این شاخص با افزایش تعداد خوشه‌ها به صورت یکنواخت کاهش می‌یابد و از این رو به نظر می‌رسد با توجه به عملکرد مشابه این شاخص در مطالعات مشابه پیشین، شاخص مذکور برای تعیین تعداد بهینه‌ی خوشه‌ها برای منطقه‌بندی حوضه‌ی آبریز مورد مطالعه کارایی چندانی ندارد. بر مبنای نتایج به دست آمده در مورد شاخص‌های صحت خوشه‌بندی فازی به نظر می‌رسد دو شاخص ژی-بنی توسعه یافته و کوان در زمینه‌ی ارزیابی کیفیت خوشه‌بندی فازی و تعیین تعداد بهینه‌ی خوشه‌ها کارایی بیشتری دارند. ناتوانی سایر شاخص‌های صحت خوشه‌بندی در تشخیص تعداد بهینه‌ی مناطق ممکن است به دلیل وجود کمبودهایی در روابط مربوط به آن‌ها باشد. تعداد قابل توجهی از این روابط تنها مبتنی بر درجه‌ی عضویت فازی سایتها هستند و ارتباطی با ساختار داده‌ها ندارند. رائو و سرینیواس (۲۰۰۶b) نیز به نتیجه‌ی مشابهی دست یافتند.

(۷) با توجه به مقادیر محاسبه شده برای شاخص‌های صحت خوشه‌بندی فازی، همچنین مقادیر شاخص‌های ناهمگنی محاسبه شده برای مناطق در حالت‌های ۲ تا ۵ منطقه‌ای و نیز تعداد ایستگاه‌ها و داده‌های آماری اختصاص یافته به مناطق در حالات مختلف منطقه‌بندی فازی، سه حالت سه خوشه‌ای با $\mu = 1.8$ ، چهار خوشه‌ای با $\mu = 2.8$ و پنج خوشه‌ای با $\mu = 2.8$ به عنوان حالت‌های بهینه و نهایی برای اجرای ادامه‌ی عملیات تحلیل فراوانی منطقه‌ای بر روی مناطق فازی تشکیل شده برگزیده شدند.

۸) در حالت فازی سه منطقه‌ای با $\mu = 1.8$ ، اگرچه هر سه منطقه همگن بودند، اما افزایش قابل توجهی در تعداد حاصل جمع ایستگاه‌های عضو در سه منطقه دیده نشد و این تعداد تنها با یک واحد افزایش از ۳۷ به ۳۸ رسید. در واقع تنها یکی از ایستگاه‌ها هم‌زمان در دو منطقه عضویت یافت و سایر ایستگاه‌ها هر کدام تنها به یک منطقه اختصاص یافتند. در حالتهای فازی برگزیده‌ی دوم و سوم، یعنی حالت‌های چهار و پنج خوش‌ای هر دو با $\mu = 2.8$ ، حاصل جمع تعداد ایستگاه‌های عضو در کل مناطق این دو حالت به ترتیب برابر ۴۹ و ۶۲ ایستگاه شمارش شد. در هر دو حالت اخیر یک منطقه‌ی ناهمگن به وجود آمد که ناهمگنی آن با حذف ایستگاه ۱۷-۹۳۴ از منطقه‌ی ناهمگن، برطرف شد.

۹) در منطقه‌بندی حوضه‌ی آبریز با استفاده از نگاشت‌های خودسازمانده کوهونن، با محاسبه‌ی شاخص صحت خوش‌بندی میانگین عرض silhouette برای حالت‌های ۲ تا ۵ خوش‌ای مشخص شد که مقدار این شاخص در حالت ۳ خوش‌ای بیش از سایر حالت‌ها است و بر این اساس، تعداد ۳ منطقه را می‌توان، تعداد بهینه از نظر وضعیت خوش‌بندی ایستگاه‌های مورد بررسی دانست. در هر چهار حالت خوش‌بندی، مقدار میانگین این شاخص، عددی بین ۰/۴ تا ۵/۰ را اختیار نمود که برای حالت ۳ خوش‌ای مقدار این شاخص برابر ۰/۴۹۲۳ ثبت شد.

۱۰) در میان حالت‌های ۲ تا ۵ خوش‌ای حاصل از منطقه‌بندی با استفاده از نگاشت‌های خودسازمانده، مناطق تشکیل شده در حالت ۳ منطقه‌ای از نظر همگنی بهترین وضعیت را فراهم کردند. لذا این حالت که هم از نظر کیفیت خوش‌بندی و هم از لحاظ وضعیت همگنی مناطق به عنوان بهترین گزینه‌ی منطقه‌بندی با استفاده از نگاشت‌های خودسازمانده شناخته شد، برای اجرای ادامه‌ی تحلیل فراوانی منطقه‌ای سیلاب مورد استفاده قرار گرفت.

۱۱) پس از برآورده سیلاب با به کارگیری تحلیل فراوانی منطقه‌ای، برآورده سیلاب با استفاده از تحلیل فراوانی نقطه‌ای نیز برای هر یک از ۳۷ ایستگاه نهایی مورد مطالعه نیز انجام گرفت. بزرگ‌ترین مقادیر برآورده شده‌ی دبی سیلاب با دوره‌ی بازگشت ۱۰ سال، به ترتیب به

ایستگاههای ۱۷-۰۳۳، ۱۷-۰۲۹ و ۱۷-۰۲۱ تعلق گرفت. در مقابل، کوچکترین مقادیر برآوردهای سیلاب با دوره‌ی بازگشت ۱۰ سال، به ترتیب به ایستگاههای ۱۷-۸۷۴، ۱۷-۱۰۱ و ۱۷-۰۵۰ اختصاص یافت.

(۱۲) مقایسه‌ی مقادیر برآوردهای سیلاب ۱۰ ساله در روش تحلیل نقطه‌ای با میانگین دبی‌های سیلاب برآورد شده با دوره‌ی بازگشت ۱۰ سال در تحلیل فراوانی منطقه‌ای با استفاده از گزینه‌های مختلف منطقه‌بندی، نشان داد در مورد بزرگ‌تر یا کوچک‌تر بودن مقادیر میانگین برآوردهای به دست آمده در تحلیل فراوانی منطقه‌ای نسبت به مقادیر دبی‌های برآوردهای ایستگاهها در تحلیل فراوانی نقطه‌ای رابطه‌ی مطلق و روشنی بر کلیه ایستگاهها حکم‌فرما نیست. بررسی نتایج نشان داد که مقدار میانگین برآوردهای حاصل از تحلیل منطقه‌ای نسبت به مقدار به دست آمده از تحلیل نقطه‌ای، در ۱۷ ایستگاه بزرگ‌تر و در ۲۰ ایستگاه کوچک‌تر است. انتخاب دوره‌ی بازگشت ۱۰ سال با توجه به تعداد داده‌های آماری ایستگاهها که در محدوده‌ی ۱۰ تا ۴۰ قرار دارد، صورت گرفت.

(۱۳) در میان ایستگاههای مورد مطالعه، بیشترین درصد اختلاف نسبی مقادیر میانگین برآوردهای به دست آمده در تحلیل فراوانی منطقه‌ای و مقادیر دبی‌های برآوردهای ایستگاهها در تحلیل فراوانی نقطه‌ای، مربوط به ایستگاه ۱۷-۰۳۰ با اختلافی حدود ۱۹/۲۶٪ است. سایر ایستگاههایی که از نظر مقدار این اختلاف نسبی در چارک بالای ایستگاههای مورد مطالعه قرار می‌گیرند، به ترتیب عبارتند از ایستگاههای ۱۷-۴۳۰، ۱۷-۴۲۸، ۱۷-۲۰۱، ۱۷-۰۳۳، ۱۷-۰۳۵، ۱۷-۰۰۳، ۱۷-۱۰۱، ۱۷-۰۰۹ و ۱۷-۹۶۷. مقادیر اختلاف نسبی مربوط به این ایستگاهها در محدوده‌ی تقریبی ۹٪ تا ۱۳٪ قرار می‌گیرد. از سوی دیگر، کمترین درصد اختلاف نسبی مربوط به ایستگاه ۱۷-۰۵۵ و در حدود ۱٪ می‌باشد. سایتها دیگری که در محدوده‌ی تقریبی چارک پایین مقدار اختلاف نسبی واقع می‌شوند به ترتیب ایستگاههای ۱۷-۹۳۴، ۱۷-۰۱۱، ۱۷-۰۱۵، ۱۷-۰۱۳، ۱۷-۰۵۰، ۱۷-۰۰۹، ۱۷-۰۵۰، ۱۷-۹۶۵، ۱۷-۸۷۴، ۱۷-۰۵۱ و ۱۷-۰۵۱

(۱۴) در مجموع به نظر می‌رسد برآوردهای حاصل از الگوریتم فازی c-means در حالت‌های ۴ و ۵ منطقه‌ای (CM-4, CM-5) نزدیک‌ترین برآوردهای سیلاب با دوره بازگشت ۱۰ سال را نسبت به برآوردهای حاصل از تحلیل فراوانی نقطه‌ای فراهم می‌کنند و در مقابل، برای حالت ۲ منطقه‌ای حاصل از منطقه‌بندی یا استفاده از الگوریتم خوش‌بندی ترکیبی WAKM و حالت ۳ منطقه‌ای مربوط به منطقه‌بندی با استفاده از نگاشتهای خودسازمانده (SOFM-3) دورترین برآوردها نسبت به برآوردهای حاصل از تحلیل فراوانی نقطه‌ای محاسبه می‌شوند. با این حال، به نظر می‌رسد که مقدار اختلاف برآوردهای این دو گزینه نیز نسبت به سایر گزینه‌های مورد بررسی در تحلیل فراوانی منطقه‌ای چندان چشمگیر نیست.

۲-۵- پیشنهادات

- با توجه به نتایج حاصل از این مطالعه، موارد زیر جهت بررسی در پژوهش‌های آینده پیشنهاد می‌گردد:
- (۱) بررسی تأثیر انتخاب ویژگی‌های مختلف جغرافیایی، فیزیوگرافیک و هواسناسی به عنوان ویژگی‌های مورد استفاده برای منطقه‌بندی حوضه‌ی آبریز بر نتایج حاصل از تحلیل فراوانی منطقه‌ای سیلاب به منظور شناسایی عوامل مؤثر بر فرآیند تولید سیلاب در نواحی مختلف؛
 - (۲) بررسی ترکیب الگوریتم‌های سلسله مراتبی با الگوریتم‌های افزایی دیگری غیر از K-means مانند الگوریتم K-medoids و ...؛
 - (۳) مطالعه در مورد چگونگی اثرگذاری نحوه‌ی تعیین مراکز اولیه‌ی خوش‌ها در الگوریتم c-means بر نتایج نهایی خوش‌بندی؛
 - (۴) بررسی نحوه‌ی تأثیر تغییرات مقدار μ در الگوریتم فازی c-means و مقدار آستانه‌ی عضویت

در منطقه‌بندی فازی بر وضعیت همگنی مناطق و تعداد آمار موجود در مناطق؛

۵) بررسی نحوه‌ی عملکرد حالت دو بعدی نگاشتهای خودسازمانده کوهون در منطقه‌بندی

حوضه‌های آبریز و همچنین ترکیب خروجی‌های آن با سایر روش‌های خوشبندی.

۶) استفاده از روش‌های تحلیل خوشهای و تحلیل فراوانی منطقه‌ای و بررسی عملکرد آن‌ها در

زمینه‌های مرتبط با پدیده‌ی خشکسالی.

پیوست الف:

زبان برنامه‌نویسی R

الف-۱- مقدمه

در این پایان‌نامه بخش عمده‌ی عملیات مربوط به برنامه‌نویسی و محاسبات در فضای نرم‌افزاری نسخه-۲.۱۵.۳ R اجرا شده است. مشخصات بسته‌های محاسباتی که در مراحل مختلف انجام کار به R اضافه شده‌اند به شرح زیر است. (نسخه‌ی هر بسته پس از علامت "_" جلوی نام آن درج شده است):

cluster_1.14.4 -۱

lmomRFA_2.5 -۲

بسته نرم‌افزاری مورد نیاز:

lmom_2.1

e1071_1.6 -۳

بسته‌های نرم‌افزاری مورد نیاز:

class_7.3-5

kohonen_2.0.10 -۴

بسته‌های نرم‌افزاری مورد نیاز:

MASS_7.3-28

همچنین فهرست توابعی از R که در بخش‌های مختلف این مطالعه از آن‌ها استفاده شده است، در

ادامه آورده شده است. توجه به دو نکته‌ی زیر در ارتباط با این فهرست ضروری است:

۱- نام بسته‌ای که هر تابع در آن قرار گرفته است درون `{}` نوشته شده است. هر جا که از

واژه‌ی base استفاده شده است، بدین معناست که نیازی به اضافه کردن بسته‌ی

نرم‌افزاری نمی‌باشد و تابع مورد نظر در هسته‌ی R موجود است.

۲- در آرگومان‌های هر تابع وجود «...» به معنای آن می‌باشد که یا در حال حاضر آرگومان

دیگری برای تابع مورد نظر وجود ندارد، یا این که آرگومان‌های دیگری وجود دارند که از

اهمیت کمتری برخوردار هستند. در این صورت برای مشاهده‌ی آن‌ها باید به راهنمای R

مراجعه نمود.

الف-۲- توابع مورد استفاده در این مطالعه

الف-۲-۱- ماتریس فاصله

dist (x, method = “euclidean”, ..., p = 2) {base}

- x : یک ماتریس یا چارچوب^۱ عددی شامل داده‌ها (که در این مطالعه ویژگی‌های سایت‌ها بوده‌اند).

• method : شاخص فاصله‌ی مورد استفاده که باید یکی از گزینه‌های “euclidean”، “minkowsky”، “binary”，“canberra”，“manhattan”，“maximum” این زیرشته‌های^۲ تعریف شده می‌تواند مورد استفاده قرار گیرد. در این مطالعه از شاخص فاصله‌ی اقلیدسی (“euclidean”) استفاده شده است.

• p : توان فاصله‌ی مینکوفسکی.

الف-۲-۲- خوشبندی سلسله‌مراتبی

a) hclust (d, method = “complete”, ...) {base}

- d : یک ساختار فاصله یا اختلاف که به وسیله‌یتابع dist تولید می‌شود.
- method : روش خوشبندی سلسله‌مراتبی متراکم کننده‌ی مورد استفاده که باید یکی از گزینه‌های اختصاری تعریف شده‌ی “ward”，“average”，“complete”，“single” و یا “centroid” و یا “median”，“mcquitty” باشد. در این مطالعه چهار گزینه‌ی نخست مورد استفاده قرار گرفته‌اند.

b) cutree (tree, k, h) {base}

¹ data frame

² substrings

- tree : یک ساختار از داده‌های مربوط به نمودار درختی یا dendrogram حاصل از خوشه‌بندی سلسله مراتبی که توسطتابع hclust تولید می‌شود.
- k : یک کمیت عددی یا برداری صحیح که تعداد خوشه‌های مورد نظر را مشخص می‌کند.
- h : یک کمیت عددی (اسکالر) یا برداری صحیح که ارتفاع محل برش نمودار درختی را معین می‌کند.

الف-۳- خوشه‌بندی با استفاده از الگوریتم K-means

- kmeans (x, centers, iter.max = 10, nstart = 1, algorithm = “Hartigan-Wong”) {base}
- x : یک ماتریس عددی داده‌ها (که در این مطالعه ویژگی‌های سایت‌ها بوده‌اند) یا شیئی که می‌تواند مانند ماتریس‌ها مورد استفاده قرار گیرد، مثل بردارها یا چارچوب‌های عددی که تمام ستون‌های آن شامل داده‌های عددی باشند.
 - centers : مقدار این آرگومان می‌تواند به صورت یک عدد k که تعداد خوشه‌های مورد نظر است باشد یا ساختاری (ماتریس یا چارچوب داده‌ها) شامل مراکز خوشه‌های ابتدایی باشد (که هر سطر آن مربوط به مقادیر ویژگی‌های یکی از مراکز است). اگر مقدار، یک عدد مانند k باشد، یک مجموعه شامل k سطر از ماتریس x به صورت تصادفی به عنوان مراکز اولیه انتخاب می‌شود.

- iter.max : بیشترین تعداد مجاز تکرارهای عددی الگوریتم که در مطالعه‌ی حاضر برابر ۱۰۰۰ انتخاب شده است.

- centers : اگر یک عدد باشد، مقدار این آرگومان مشخص می‌کند که چند مجموعه‌ی تصادفی باید انتخاب شود؟

- algorithm : الگوریتم مورد استفاده برای اجرای خوشه‌بندی k-means که باید یکی از گزینه‌های اختصاری تعریف شده‌ی “Forgy”，“Lloyd”，“Hartigan-Wong” و یا

باشد. در این مطالعه از الگوریتم "Hartigan-Wong" استفاده شده است.

الف-۴-۴- شاخص عرض silhouette

a) silhouette (x, dist, dmatrix, ...) {cluster}

x : یک شیء از نوع مناسب؛ در حالت پیش فرض یک بردار صحیح با k کد مختلف صحیح

برای معرفی خوشه‌ها یا یک لیست^۳ شامل یک جزء به صورت x\$clustering. باید توجه

داشت که آماره‌های silhouette فقط در صورتی که $1 \leq k \leq n - 2$ باشد، تعریف می‌شوند.

dist : یک ماتریس (یا شیء) اختلاف از نوعتابع dist. اگر این مورد معین نشود، آرگومان dmatrix باید حتماً مشخص شود.

یک ماتریس متقارن اختلاف (n^*n) که در صورت نبود dist معین می‌شود.

b) summary (object, FUN = mean, ...) {cluster}

.silhouette : یک شیء از نوع object

FUN : تابع مورد استفاده برای خلاصه‌سازی نتایج مربوط به عرض‌های silhouette. در این

مطالعه از میانگین (mean) استفاده شده است.

الف-۵- گشتاورهای نمونه‌ی مجموعه‌های چندگانه‌ی داده‌ها

regsamlmu (x, nmom = 5, sort.data = TRUE, lcv = TRUE) {lmomRFA}

x : یک ماتریس عددی یا لیستی از بردارهای عددی. در این مطالعه هر ستون این ماتریس

شامل داده‌های حداکثر سیلاب لحظه‌ای سالانه‌ی یک ایستگاه هیدرومتری بوده است.

nmom : تعداد گشتاورهای خطی مورد نظر برای محاسبه.

³ list

sort.data : یک آرگومان منطقی^۴ که اگر مقدار آن TRUE باشد، داده‌ها به صورت صعودی

یا نزولی مرتب می‌شوند.

lcv : یک آرگومان منطقی که اگر مقدار آن برابر TRUE انتخاب شود، گشتاور خطی دوم به

صورت نسبتی از میانگین بیان می‌شود، به بیان روش‌تر مقدار محاسبه شده برابر $L-CV$

نمونه ($t = l_2/l_1$) خواهد بود. اگر مقدار این آرگومان برابر FALSE انتخاب شود، گشتاور

خطی مرتبه‌ی دوم به صورت همان مقدار l_2 نشان داده می‌شود.

الف-۲-۶- محاسبه‌ی شاخص‌ها و آماره‌های آزمون H D و Z برای

تحلیل فراوانی منطقه‌ای

a) regtst (regdata, nsim=1000) {lmomRFA}

regdata : داده‌های ورودی به صورت یک چارچوب از داده‌ها که هر یک از سطرهای آن

محتوی داده‌های مربوط به یک سایت هستند (یک شیء از نوع regdata). هفت ستون اول به

ترتیب شامل نام سایتها، طول آمار، گشتاورهای خطی و نسبت‌های گشتاورهای خطی l_1

(میانگین)، t_3 ($L-kurtosis$)، t_4 ($L-skewness$) و t_5 ($L-CV$) هستند.

باید توجه داشت که ستون چهارم حتماً باید شامل مقادیر $L-CV$ باشد و نه مقادیر l_2 .

تابع regsmalmu با تنظیمات پیش فرض آرگومان‌های آن یک شیء از نوع regdata را ارائه

می‌کند.

nsim : تعداد شبیه‌سازی‌های مورد استفاده در محاسبه‌ی شاخص‌های ناهمگنی و نکویی

برازش. اگر مقدار این آرگومان از ۲ کمتر باشد، فقط مقدار شاخص ناجوری D محاسبه می-

شود. در این پژوهش مقدار این آرگومان برابر ۱۰۰۰ انتخاب شده است.

⁴ logical

b) summary (regsts, prob = c(0.01,0.02,0.05,0.1,0.2,0.5,0.8,0.9,0.95,0.98,0.99,0.999),
...) {lmomRFA}

• شیئی از نوع regst که معمولاً خروجیتابع regst است.

• احتمال‌های عدم تجاوز برای براورددهای چندک‌هایی که باید نمایش داده شوند.

الف-۲-۷- میانگین (متوسط) وزنی منطقه‌ای گشتاورهای خطی

regavlmom (regdata, weight) {lmomRFA}

• یک شیء از نوع regdata، شامل آماره‌های خلاصه‌ی داده‌ها برای سایت‌های یک

منطقه.

• برداری شامل وزن‌های مورد استفاده برای سایت‌ها. اگر این آرگومان حذف شود،

وزن هر یک از سایت‌ها برابر اندازه‌ی نمونه در آن سایت در نظر گرفته می‌شود که از ستون

دوم regdata گرفته می‌شود. اگر این آرگومان برابر یک مقدار واحد انتخاب شود، به تمام

سایت‌ها وزن یکسانی داده خواهد شد.

الف-۲-۸- برازش یک توزیع فراوانی منطقه‌ای

regfit (regdata, distr) {lmomRFA}

• یک شیء از نوع regdata، شامل آماره‌های خلاصه‌ی داده‌ها برای سایت‌های یک

منطقه.

• رشته‌ای که نوع توزیع مورد نظر برای برازش را مشخص می‌کند و می‌تواند گزینه‌هایی

مانند "glo", "gpa", "pe3", "gno", "gev" و "wak" به ترتیب برای توزیع‌های لجستیک

تعمیم یافته، مقادیر حدی تعمیم یافته، نرمال تعمیم یافته، پیرسون تیپ III، پارتوفی تعمیم

یافته و ویکبی باشد.

الف-۲-۹- چندک‌ها و توابع چندک یک توزیع فراوانی منطقه‌ای

a) regquant (f, rfd) {lmomRFA}

f : بردار احتمالات. •

rfd : یک شیء از نوع rfd که معمولاً شامل خروجیتابع regfit است و حاوی مشخصات یک

توزیع فراوانی منطقه‌ای است. •

b) regqfunc (rfd) {lmomRFA}

rfd : یک شیء از نوع rfd که معمولاً شامل خروجیتابع regfit است و حاوی مشخصات یک

توزیع فراوانی منطقه‌ای است. •

الف-۲-۱۰- رسم نمودار تابع

curve (expr, from = 0.001, to = 0.999, ..., xlab = xname, ylab = yname, ...) {base}

expr : نام یک تابع یا فراخوانی یا یک عبارت نوشته شده به صورت تابعی از x که در این

مطالعه خروجیتابع regqfunc است. •

from : ابتدای بازه‌ای که تابع در آن بازه ترسیم می‌شود. •

to : انتهای بازه‌ای که تابع در آن بازه ترسیم می‌شود. •

xlab : عنوان محور افقی یا محور x. •

ylab : عنوان محور عمودی یا محور y. •

الف-۲-۱۱- چندک‌ها و توابع چندک منحصر به فرد سایت‌ها در یک منطقه

sitequant (f, rfd, sitenames, index, drop = TRUE) {lmomRFA}

f : بردار احتمالات. •

- rfd : یک شیء از نوع rfd که معمولاً شامل خروجیتابع regfit است و حاوی مشخصات یک توزیع فراوانی منطقه‌ای است.
- sitenames : برداری شامل نام سایت‌ها.
- drop : اگر مقدار آن برابر TRUE باشد و تنها یک سایت یا یک مقدار احتمال وجود داشته باشد، خروجی sitequant به صورت یک بردار و در غیر این صورت یک ماتریس خواهد بود.

الف-۱۲-۲- خوشبندی فازی با استفاده از الگوریتم c-means

cmeans (x, centers, iter.max = 1000, verbose = FALSE, dist = "euclidean", method = "cmeans", m = 2, ...) {e1071}

- x : ماتریس داده‌ها که ستون‌های آن متناظر با متغیرها یا ویژگی‌ها و سطرهای آن مربوط به مشاهدات یا سایت‌ها هستند.

centers : تعداد خوش‌ها یا مقادیر اولیه‌ی ویژگی‌های مرکز خوش‌ها.

- iter.max : حداقل تعداد تکرارهای عددی که در این مطالعه برابر ۱۰۰۰ در نظر گرفته شده است.

verbose : اگر مقدار آن برابر TRUE انتخاب شود، برخی خروجی‌ها در طول آموزش در حالت یادگیرنده تولید می‌شود. در این پژوهش مقدار این آرگومان برابر FALSE انتخاب شده است.

- dist : شاخص فاصله که می‌تواند یکی از گزینه‌های "euclidean" یا "manhattan" باشد.
- method : اگر برابر "cmeans" باشد، آنگاه روش خوشبندی فازی c-means مورد استفاده قرار می‌گیرد و اگر برابر "ufcl" باشد، به هنگام سازی فعال اجرا می‌شود.

m : یک مقدار بزرگ‌تر از ۱ که درجه‌ی فازی بودن را مشخص می‌کند. مقدار این آرگومان در واقع همان μ است که در این مطالعه در بازه‌ی ۰/۱ تا ۳ با تغییراتی در فواصل ۰/۱ مورد

بررسی قرار گرفته است.

الف-۲-۱۳- شاخص‌های صحت خوشبندی فازی

fclustIndex (y, x, index = "all") {e1071}

y : یک شیء حاصل از خوشبندی فازی از نوع fclust که معمولاً خروجی تابع

است.

x : ماتریس داده‌های ورودی مربوط به ویژگی‌های سایت‌ها.

index : شاخص‌های مورد نظر برای محاسبه، شامل گزینه‌های "xie.beni", "gath.geva"

"partition.entropy", "partition.coefficient", "fukuyama.sugeno"

و "separation.index", "proportion.exponent" برای هر یک از شاخص‌های مورد نظر

گزینه‌ی "all" برای تمام شاخص‌ها.

الف-۲-۱۴- نگاشت‌های خودسازمانده کوهونن

a) somgrid (xdim = 3, ydim = 1, topo = c("rectangular", "hexagonal")) {kohonen}

xdim و ydim : ابعاد شبکه یا نگاشت.

topo : توپولوژی یا شکل شبکه که می‌تواند یکی از گزینه‌های "hexagonal" (شش ضلعی) و

یا "rectangular" (مستطیلی) باشد.

b) som (data, grid = somgrid(), ...) {class}

data : یک ماتریس که هر سطر آن معرف یک شیء (مربوط به ویژگی‌های نرمال شده‌ی یک

سایت) است.

grid : یک شبکه از معرف‌ها از نوع خروجی تابع somgrid

c) plot (x, type = c("codes", "changes", "counts", "dist.neighbours", "mapping", "property", "quality"), ...) {kohonen}

• x : یک شیء از نوع kohonen که می‌تواند خروجی تابع scale باشد.

• "counts", "changes", "codes" : نوع نمودار که می‌تواند یکی از گزینه‌های "counts", "changes", "codes" باشد.

"mapping", "property", "quality" و "dist.neighbours" : نحوه اختصاص سایت‌ها به گره‌ها را نشان می‌دهد و گزینه‌های "mapping", "property", "quality" و "dist.neighbours"

نحوه اختصاص سایت‌ها به گره‌ها را نشان می‌دهد و گزینه‌های "codes" و "property" هر

یک به گونه‌ای مقدار متوسط ویژگی‌های تجدید مقیاس شده‌ی سایت‌های متعلق به یک گره

را مشخص می‌کنند. گزینه‌ی "count" تعداد سایت‌های تعلق گرفته به هر یک از گره‌ها را

معین می‌کند.

منابع و مراجع

شرکت سهامی آب منطقه‌ای گیلان، ۱۳۸۸، مطالعات یکپارچه‌ی برنامه‌ریزی منابع آب رودخانه قزل اوزن-سفیدرود، گزارش نیازهای آب شرب و صنعت محدوده‌ی مطالعاتی، گیلان.

موسوی ندوشنی س، ۱۳۹۱، آشنایی با زبان محاسبات آماری *R*، ویراست ۱/۸/۰، دانشگاه صنعت آب و برق (شهید عباسپور)، تهران.

مهدوی م، ۱۳۸۶، هیدرولوژی کاربردی، جلد دوم، چاپ پنجم، انتشارات دانشگاه تهران، تهران، ص ۹۸.

Acreman M.C. and Sinclair C.D. (1986) "Classification of drainage basins according to their physical characteristics: An application for flood frequency analysis in Scotland". *Journal of Hydrology* 84(3–4): 365–380.

ASCE Task Committee on Application of Artificial Neural Networks in Hydrology (2000a) "Artificial neural networks in hydrology, I: Preliminary concepts". *Journal of Hydrologic Engineering*, ASCE 5(2): 115–123.

ASCE Task Committee on Application of Artificial Neural Networks in Hydrology (2000b) "Artificial neural networks in hydrology, II: Hydrologic applications". *Journal of Hydrologic Engineering*, ASCE 5(2): 124–137.

Backer E. and Jain A.K. (1981) "A clustering performance measure based on fuzzy set decomposition". *IEEE Transactions on Pattern Analysis and Machine Intelligence* 3(1): 66–75.

Ball G. and Hall D. (1967) "A clustering technique for summarizing multivariate data". *Behavioral Science* 12: 153–155.

Bargaoui Z-K., Fortin V., Bobee B. and Duckstein L. (1998) "A fuzzy approach to the delineation of region of influence for hydrometric stations". *Revue des sciences de l'eau* 11(2): 255–282 (In French).

Bezdek J.C. (1974a) "Numerical taxonomy with fuzzy sets". *Journal of Mathematical Biology* 1: 57–71.

Bezdek J.C. (1974b) "Cluster validity with fuzzy sets". *Journal of Cybernetics* 3(3): 58–72.

Bezdek J.C. (1975) "Mathematical models for systematics and taxonomy". 8th International Conference on Numerical Taxonomy. Estabrook G. (Ed.). San Francisco, CA.

Bezdek J.C. (1981) “*Pattern recognition with fuzzy objective function algorithms*”. Plenum Press, New York.

Bezdek J.C. (1987). Partition structures: A tutorial. In: “*The Analysis of Fuzzy Information*”. Bezdek J. C. (Ed.). CRC Press, Boca Raton, FL.

Bhaskar N.R. and O'Connor C.A. (1989) “Comparison of method of residuals and cluster analysis for flood regionalization”. *Journal of Water Resources Planning and Management* 115(6): 793–808.

Black A.R. and Werritty A. (1997) “Seasonality of flooding: A case study of north Britain”. *Journal of Hydrology* 195(1–4): 1–25.

Burn D.H. (1989) “Cluster analysis as applied to regional flood frequency”. *Journal of Water Resources Planning and Management* 115(5): 567–582.

Burn D.H. (1990a) “An appraisal of the “region of influence” approach to flood frequency analysis”. *Hydrological Sciences Journal* 35(2): 149–165.

Burn D.H. (1990b) “Evaluation of regional flood frequency analysis with a region of influence approach”. *Water Resources Research* 26(10): 2257–2265.

Burn D.H. (1997) “Catchment similarity for regional flood frequency analysis using seasonality measures”. *Journal of Hydrology* 202: 212–230.

Burn D.H. and Goel N.K. (2000) “The formation of groups for regional flood frequency analysis”. *Hydrological Sciences Journal* 45(1): 97–112.

Burn D.H., Zrinji Z. and Kowalchuk M. (1997) “Regionalization of catchments for regional flood frequency analysis”. *Journal of Hydrologic Engineering* 2(2): 76–82.

Chow V.T., Maidment D.R. and Mays L.W. (1988) “*Applied hydrology*”. McGraw-Hill Inc., New York.

Cunderlik J.M. and Burn D.H. (2002a) “The use of flood regime information in regional flood frequency analysis”. *Hydrological Sciences Journal* 41(1): 77–92.

Cunderlik J.M. and Burn D.H. (2002b) “Analysis of the linkage between rain and flood regime and its application to regional flood frequency estimation”. *Journal of Hydrology* 261: 115–131.

Cunderlik J.M., Ouarda T.B.M.J. and Bobee B. (2004a) “Determination of flood Seasonality from hydrologic records”. *Hydrological Sciences Journal* 49(3): 511–526.

Cunderlik J.M., Ouarda T.B.M.J. and Bobee B. (2004b) “On the objective identification of flood seasons”. *Water Resources Research* 40: W01520, doi:10.1029/2003WR002295.

Dalrymple T. (1960) “*Flood frequency analysis*”. U.S. Geological Survey, Water Supply Paper 1543-A., U.S. Department of the Interior, Washington, DC.

Dave R.N. and Krishnapuram R. (1997) “Robust clustering methods: A unified view”. *IEEE Transactions on Fuzzy Systems* 5(2): 270–293.

Dixon W.J. (Ed.) (1975) “*BMDP biomedical computer programs*”. University of California Press, Berkeley, CA.

Dunn J.C. (1974) “Well separated clusters and optimal fuzzy partitions”. *Journal of Cybernetics* 4: 95–104.

Fausett L.V. (1994) “*Fundamentals of neural networks: Architectures, algorithms, and applications*”. Englewood Cliffs, Prentice Hall, NJ.

Fisher N.I. (1993) “*Statistical analysis of circular data*”. Cambridge University Press, New York.

Fukuyama Y, Sugeno M (1989) “A new method of choosing the number of clusters for the fuzzy c-means method”. Fifth Fuzzy Systems Symposium, pp. 247–250 (In Japanese).

Gath I. and Geva A.B. (1989) “Unsupervised optimal fuzzy clustering”. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 11(7): 773–781.

Govindaraju R.S. and Rao A.R. (Eds.) (2000) “*Artificial neural networks in hydrology*”. Kluwer Academic Publishers, Holland, p. 329.

Greenwood J.A., Landwehr J.M., Matalas N.C. and Wallis J.R. (1979) “Probability weighted moments: Definition and relation to parameters of distribution expressible in inverse form”. *Water Resources Research* 15(5): 1049–1054.

Guenoche A., Hansen P. and Jaumard B. (1991) “Efficient algorithms for divisive hierarchical-clustering with the diameter criterion”. *Journal of Classification* 8(1): 5–30.

Güler C. and Thine G.D. (2004) “Delineation of hydrochemical facies distribution in a regional groundwater system by means of fuzzy c-means clustering”. *Water Resources Research* 40: W12503, doi:10.1029/2004WR003299.

Halkidi M., Batistakis Y. and Vazirgiannis M. (2001) “On clustering validation techniques”. *Journal of Intelligent Information systems* 17(2/3): 107–145.

Hall M.J. and Minns A.W. (1998) “Regional flood frequency analysis using artificial neural network”. Third International Conference on Hydroinformatics, Copenhagen, Denmark, vol. 2. Babovic V. and Larsen LC (Eds.), Balkema, Rotterdam, pp. 759–763.

Hall M.J. and Minns A.W. (1999) “The classification of hydrologically homogeneous regions”. *Hydrological Sciences Journal* 44(5): 693–704.

Hall M.J., Minns A.W. and Ashrafuzzaman A.K.M. (2002) “The application of data mining techniques for the regionalization of hydrological variables”. *Hydrology and Earth System Sciences* 6(4): 685–694.

Hartigan J.A. and Wong M.A. (1979) “Algorithm AS 136: A K-means clustering algorithm”. *Applied Statistics* 28: 100–108.

Haykin S. (2003) “*Neural networks: A comprehensive foundation*”. Fourth Indian Reprint, Pearson Education, Singapore, p. 842.

Hosking, J.R.M. (1990). “L-moments: Analysis and estimation of distributions using linear combinations of order statistics”. *Journal of the Royal Statistical Society, Series B*, 52, 105-24.

Hosking, J.R.M. and Wallis, J.R. (1987). “Parameter and quantile estimation for the generalized Pareto distribution”. *Technometrics*, 29, 339-49.

Hosking J.R.M. and Wallis J.R. (1997) “*Regional frequency analysis: An approach based on L-moments*”. Cambridge University Press, New York.

Huang Z. (1997) “A fast clustering algorithm to cluster very large categorical data sets in data mining”. SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery, Tucson, Arizona.

Huang Z. (1998) “Extensions to the K-means algorithm for clustering large data sets with categorical values”. *Data Mining and Knowledge Discovery* 2: 283–304.

Huang Z. and Ng M.K. (2003) “A note on K-modes clustering”. *Journal of Classification* 20(2): 257–261.

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.

Kaufman L. and Rousseeuw P. (1990) “*Finding groups in data: An introduction to cluster analysis*”. Wiley, New York.

Kohonen T. (1982) “Self-organized formation of topologically correct feature maps”. *Biological Cybernetics* 43: 59–69.

Kohonen T. (1997) “*Self-organizing maps*”, second edition. Springer Verlag, Berlin.

Kohonen T., Oja E., Simula O., Visa A. and Kangas J. (1996) “Engineering applications of the selforganizing map”. *Proceedings of the IEEE* 84(10): 1358–1384.

Kwon S.H. (1998) “Cluster validity index for fuzzy clustering”. *Electronics Letters* 34(22): 2176–2177.

Landwehr J.M., Matalas, N.C, and Wallis, J.R. (1979) “Probability weighted moments compared with some traditional techniques in estimating Gumbel parameters and quantiles”. *Water Resources Research*, 15, 1055-64.

Lecce S.A. (2000) “Seasonality of flooding in North Carolina”. *Southeastern Geographer* 41(2): 168–175.

MacQueen J. (1967) “Some methods for classification and analysis of multivariate observations”. In: Fifth Berkeley Symposium on Mathematical Statistics and Probability, vol. 1. Le Cam L. M. and Neyman J. (Eds.), University of California Press, Berkeley, CA, pp. 281–297.

Mardia K.V. (1972) “*Statistics of directional data*”. Academic Press, San Diego, California.

Mosley M.P. (1981) “Delimitation of New Zealand hydrological regions”. *Journal of Hydrology* 49: 173–192.

Murtagh F. (1983) “A survey of recent advances in hierarchical-clustering algorithms”. *Computer Journal* 26(4): 354–359.

Nathan R.J. and McMahon T.A. (1990) “Identification of homogeneous regions for the purposes of regionalization”. *Journal of Hydrology* 121: 217–238.

Natural Environment Research Council (NERC) (1975) “*Flood studies report*”, vol. I, Hydrological Studies, NERC, London.

Ng R. and Han J. (1994) “Efficient and effective clustering methods for spatial data mining”. 20th VLDB conference, Santiago, Chile.

Obermayer K. and Sejnowski T.J. (Eds.) (2001) “*Self-organizing map formation foundations of neural computation*”. MIT Press, Cambridge, MA.

Ouarda T.B.M.J., Cunderlik J.M., St-Hilaire A., Barbet M., Bruneau P. and Bobee B. (2006) “Data-based comparison of seasonality-based regional flood frequency methods”. *Journal of Hydrology* 330(1–2): 329–339.

Pal N.R. and Bezdek J.C. (1995) “On cluster validity for the fuzzy c-means model”. *IEEE Transactions on Fuzzy systems* 3(3): 370–379.

Potter K.W. and Faulkner E.B. (1987) “Catchment response time as a predictor of flood quantiles”. *Water Resources Bulletin* 23(5): 857–861.

Ralambondrainy H. (1995) “A conceptual version of the K-means algorithm”. *Pattern Recognition Letters* 16: 1147–1157.

Rao A.R. and Srinivas V.V. (2006a) “Regionalization of watersheds by hybrid cluster analysis”. *Journal of Hydrology* 318(1–4): 37–56.

Rao A.R. and Srinivas V.V. (2006b) “Regionalization of watersheds by fuzzy cluster analysis”. *Journal of Hydrology* 318(1–4): 57–79.

Rao A.R. and Srinivas V.V. (2008). “*Regionalization of Watersheds - An Approach Based on Cluster Analysis*”, Series: Water Science and Technology Library, Vol. 58, Springer Publishers, pp.248.

Reed D.W., Jakob D. and Robson A.J. (1999) Selecting a pooling group. In: “*Statistical procedures for flood frequency estimation, Flood estimation handbook*”, vol. 3, Robson A. J. and Reed D.W. (Eds.), Institute of Hydrology, Wallingford, UK (chapter 6, pp. 28–39).

Ross T.J. (1995) “*Fuzzy logic with engineering applications*”. McGraw-Hill, New York.

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

Rousseeuw P.J. (1987) “Silhouettes: A graphical aid to the interpretation and validation of cluster analysis”. *Journal of Computational and Applied Mathematics* 20: 53–65.

Ruspini E.H. (1969) “A new approach to clustering”. *Information and Control* 15(1): 22–32.

Ruspini E.H. (1970) “Numerical methods for fuzzy clustering”. *Information Sciences* 2: 319–350.

Sadri S., Burn D.H. (2011) “A Fuzzy C-Means approach for regionalization using a bivariate homogeneity and discordancy approach”. *Journal of Hydrology* 401(2011): 231–239.

Savarese S.M., Boley D.L., Bittanti S. and Gazzaniga G. (2002) “Cluster selection in divisive clustering algorithms”. 2nd SIAM ICDM, Arlington, VA, pp. 299–314.

Sokal R.R. and Sneath P.H.A. (1963) “*Principles of Numerical Taxonomy*”. W.H. Freeman and Co., San Francisco, CA.

SPSS (1988) “*Statistical package for the social sciences-X, user’s guide*”. SPSS Inc., McGraw-Hill, New York.

Su M.C. and Chang H.T. (2000) “Fast self-organization feature map algorithm”. *IEEE Transactions on Neural Networks* 11(3): 721–733.

Tasker G.D. (1980) “Hydrologic regression with weighted least squares”. *Water Resources Research* 16(6): 1107–1113.

- Tasker G.D. (1982) “Comparing methods of hydrologic regionalization”. *Water Resources Bulletin* 18(6): 965–970.
- Theodoridis S. and Koutroubas K. (1999) “*Pattern recognition*”. Academic Press, New York.
- Wallis J.R. (1981). Risk and uncertainties in the evaluation of flood events for the design of hydraulic structures. In “*Piene e Siccità, Fondazione Politecnica del Mediterraneo*”, edited by Guggino E., Rossi G., and Todini E., pp. 3-36. Catania, Italy.
- Wallis J.R. (1982). Hydrologic problems associated with oilshale development. In “*Environmental Systems and Management*”, edited by Rinaldi S., pp. 85-102. North-Holland, Amsterdam.
- Wallis J.R., Matalas, N.C. and Slack, J.R. (1997). “Just a moment!”. *Water Resources Research*, 10, 211-9.
- Ward J.H. Jr. (1963) “Hierarchical grouping to optimize an objective function”. *Journal of the American Statistical Association* 58: 236–244.
- Webster R., Burrough P.A. (1972) “Computer-based soil mapping of small areas from sample data, II: Classification smoothing”. *Journal of Soil Science* 23(2): 222–234.
- Willshaw D.J., Malsburg C.V. (1976) “How patterned neural connections can be set up by selforganization”. Royal Statistical Society London B 194: 431–445.
- Wiltshire S.E. (1986) “Regional flood frequency analysis II. Multivariate classification of drainage basins in Britain”. *Hydrological Sciences Journal* 31(3): 335–346.
- Windham M.P. (1981) “Cluster validity for fuzzy clustering algorithms”. *Fuzzy Sets and Systems* 5(2): 177–185.
- Windham M.P. (1982) “Cluster validity for the fuzzy c-means clustering algorithm”. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 4(4): 357–363.
- 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.
- Zadeh L.A. (1965) “Fuzzy sets”. *Information and Control* 8(3): 338–353.

Abstract

One of the first steps in the regional flood frequency analysis is assigning of gauging stations to regions that their stations are expected to have similar flood producing mechanisms. This process is known as regionalization. Regionalization methods based on cluster analysis can be useful in identifying groups of stations that their flood producing mechanisms are similar. In this study, performances of various procedures of cluster analysis such as hybrid clustering, fuzzy clustering using fuzzy c-means algorithm and clustering using self-organizing feature maps for regionalization of Sefidrood watershed in order to perform regional flood frequency analysis have been assessed. Also assessment of homogeneity of regions and regional flood frequency analysis has been performed by using L-moments. Results showed that combination of Ward algorithm and K-means algorithm would be best option among hybrid clustering algorithms. Also results showed that fuzzy c-means algorithm has an acceptable performance in forming homogenous regions and obtaining appropriate estimates in regional flood frequency analysis using L-moments based algorithm in interested watershed. In addition, this point has been observed that using of fuzzy clustering can provide reliable flood estimates for longer return periods. In the case of self-organizing feature maps this has been observed regions obtained by this technique in 3 regions state, provide suitable values for cluster validity measures and heterogeneity measures.

Keywords: *regional flood frequency analysis, regionalization, cluster analysis, hybrid clustering, fuzzy clustering, self-organizing feature maps, L-moments*



Shahrood University of Technology

Faculty of Agricultural Engineering

**Assessment of Performances of Cluster Analysis Methods for
Regionalization of Watersheds and Their Effects on Regional Flood
Frequency Analysis**

Ali Ahani

Supervisors:

Samad Emamgholizadeh

S. Saeid Mousavi Nadoushani

Co-Supervisor:

Khalil Azhdary

September 2013