



## Identity of Hydrologic Homogenous Regions using Fuzzy C-Means Method

M. Rayati<sup>1\*</sup>, M. Gharavi<sup>2</sup> and B. Saghafian<sup>3</sup>

### Abstract

Earth science methods for mapping, discretization, and separation of homogenous regions are recently used extensively to estimate the natural landscape response to hydrologic processes. For more homogenous study areas, the accuracy of the precision increases and the operation of water resources management system and/or natural wizard controls are less complicated. There are several methods to natural landscape discretization to identify homogenous regions. Because of continuity in landscape specifications, the use of Fuzzy classification/ clustering methods such as Fuzzy C-Mean (FCM) can be a desirable choice. In this study the FCM was used for discretization of homogenous regions in regards to runoff generation for Walnut Gulch Experimental Watershed (WGEW) located in southeast Arizona. This watershed is selected to benefit from the quality and adequacy of the hydrologic data. The process of choosing and preparation of data was addressed earlier in texts. The Principal Components Analysis (PCA) was performed and then the three essential steps for making decisions based on the FCM method was taken; determination of appropriate fuzzifier; controlling of Confusion Index (CI), and optimization of number of clusters using validation indices. It was shown that use of these validation indices would surely not lead to determination of optimal number of clusters and therefore a rational method was proposed.

**Keywords:** Hydrologic homogenous regions, Principal Components Analysis, Fuzzy C-Mean, Clustering validation index.

Received: October 12, 2009  
Accepted: March 15, 2011

## شناسایی نواحی همگن هیدرولوژیک با استفاده از روش میان مرکز فازی

مرتضی رعیتی دماوندی<sup>۱\*</sup>، مجتبی غروی<sup>۲</sup> و بهرام تقفیان<sup>۳</sup>

### چکیده

امروزه استفاده از روش‌های پهنه‌بندی، گسسته سازی و تفکیک نواحی همگن در علوم زمینی جهت ارزیابی پاسخ عرصه‌های طبیعی به فرآیندهای هیدرولوژیک مورد مطالعه از موضوعات جاری می‌باشد. زیرا هر چه محدوده مورد مطالعه همگن‌تر باشد موجب بالا رفتن دقت و سهولت در عملیات مدیریت منابع آب و یا کنترل خسارات طبیعی خواهد شد. برای شناسایی نواحی همگن، می‌توان از روش‌های مختلف گسسته‌سازی استفاده نمود. به علت پیوستگی در ویژگی‌های زمینی، استفاده از روش‌های دسته‌بندی/ خوشه‌بندی فازی نظیر روش خوشه‌بندی میان مرکز فازی (FCM)<sup>۱</sup> می‌تواند یک انتخاب مطلوب محسوب گردد. در این مطالعه و بعنوان نمونه، تفکیک نواحی همگن از نظر تولید رواناب با بکارگیری این روش فازی، با انتخاب حوضه معرف والنات گولچ (WGEW)<sup>۲</sup> واقع در جنوب شرقی ایالت آریزونا آمریکا به جهت کفایت و کیفیت داده‌های هیدرولوژیک، مدنظر می‌باشد. برای اینکار ابتدا به نحوه انتخاب و آماده‌سازی داده‌ها اشاره شده است، سپس تحلیل مولفه‌های اصلی انجام گردید و نهایتاً به هر سه مرحله اساسی و عبارتی تصمیم‌گیری در استفاده از روش (FCM) یعنی: تعیین مقدار فازیت مناسب، کنترل شاخص ابهام و تعیین تعداد بهینه خوشه‌ها با استفاده از شاخص اعتباریابی خوشه‌بندی پرداخته شده است. نشان داده شده است استفاده از روش‌های مختلف و موجود اعتباریابی خوشه‌بندی، لزوماً منجر به تعیین تعداد بهینه خوشه‌ها بصورت واضح و قطعی نخواهد شد و برای این منظور می‌توان از یک روش منطقی پیشنهادی استفاده نمود.

**کلمات کلیدی:** نواحی همگن هیدرولوژیک، تحلیل مولف‌های اصلی، میان مرکز فازی، سنجش اعتبار خوشه‌بندی.

تاریخ دریافت مقاله: ۲۰ مهر ۱۳۸۸  
تاریخ پذیرش مقاله: ۲۴ اسفند ۱۳۸۹

1- Ph.D. candidate in Water Engineering, Iran University of Science and Technology, Tehran, Iran, Email: [m\\_rayati@iust.ac.ir](mailto:m_rayati@iust.ac.ir)  
2- Assistant Professor, Civil Engineering School, Iran University of Science and Technology, Tehran, Iran, Email: [Gharavi@iust.ac.ir](mailto:Gharavi@iust.ac.ir)  
3- Professor -Science and Research Branch, Islamic Azad University, Tehran, Iran, Email: [Saghafian@scwmri.ac.ir](mailto:Saghafian@scwmri.ac.ir)  
\*- Corresponding Author

۱- دانشجوی دکتری مهندسی آب دانشگاه علم و صنعت ایران- تهران- ایران.  
۲- استادیار دانشکده عمران دانشگاه علم و صنعت ایران- تهران- ایران  
۳- استاد دانشگاه آزاد واحد علوم و تحقیقات- تهران- ایران  
\*- نویسنده مسئول

## ۱- مقدمه

در دهه‌های اخیر، وجود اطلاعات نسبتاً دقیق و توسعه نرم افزارهای پردازشگر و تخصصی برای علوم زمینی و طبیعی، کاربران را به استفاده و توسعه مدل‌های هیدرولوژیکی توزیعی (زمانی- مکانی) به منظور برآوردهای دقیق‌تر و یا پیش‌بینی خسارات با اطمینان بیشتر تشویق نموده است. علاوه بر آن، ایده بکارگیری گسسته‌سازی اولیه حوضه‌های متوسط و به‌ویژه حوضه‌های بزرگ به تعدادی زیر حوضه با بهره‌مندی از روش‌های نظیر مرتبه (Strahler or Shreve) و یا تفکیک آنها به تعدادی نواحی محدودتر قبل از استفاده از این مدل‌های توزیعی قوت گرفت. علاوه بر آن، با وجود سخت‌افزارهای جدید و توسعه نرم‌افزاری در علوم طبیعی، ضرورت گسسته‌سازی ثانوی به‌منظور شناسایی نواحی همگن هیدرولوژیکی در این زیرحوضه‌ها نیز از چالش‌های جدی محسوب گردید. یک ناحیه همگن هیدرولوژیکی، طبق تعریف محدوده‌ای است که تغییرات پاسخ نقاط (عرصه‌های کوچک) در آن به فرآیند مورد مطالعه نسبتاً کم بوده و به عبارتی رفتار مشابهی داشته باشند. بدین خاطر توسعه تکنیک‌های گسسته‌سازی به کمک تئوری‌های زمین آمار و همچنین روش‌های مختلف دسته‌بندی و یا خوشه‌بندی نظیر فازی، سلسله مراتبی، شبکه عصبی و غیره توسعه روزافزونی یافته‌اند.

## ۲- پیشینه مطالعات

در دهه‌های اخیر، توسعه و به‌کارگیری مدل‌های هیدرولوژیکی توزیعی (زمانی- مکانی) نظیر: MODCLARK (Kull and Feldman, 1998) و KINEROS (Smith et al. 1995) و غیره در جهت تخمین و برآوردهای دقیق‌تر در علوم زمینی نیز رشد روزافزونی داشته است. از آنجا که در استفاده بعضی از این مدل‌ها، کالیبراسیون و اعتباریابی همه پارامترهای دخیل و در حد قابل قبول امکان‌پذیر نمی‌باشد، لذا بکارگیری این مدل‌های توزیعی بعد از تفکیک نواحی همگن، ضمن صرفه‌جویی در مراحل کالیبراسیون و اعتباریابی آنها، منجر به درک بهتری از پاسخ این سطوح به فرآیند مورد نظر و سهولت در تجزیه و تحلیل نتایج خواهد شد. از سویی دیگر، در حوضه‌های فاقد آمار استفاده از این مدل‌های توزیعی عملاً امکان‌پذیر نمی‌باشد. بدین خاطر روش‌های نظیر نواحی مقدماتی نماینده (REA) توسط Wood et al. (1988)، روش حوضه‌های مقدماتی معرف (REWs) توسط Reggiani et al. (1998)، روش (Reggiani et al. 2000), Reggiani et al. (1999) و روش نواحی و عرصه‌های زمینی هیدرولوژیکی (HLRs) توسط Wolock et al. (2004) پیشنهاد گردید.

(Dehotin and Braud 2008) نیز با بهره‌گیری از روش (Lagacherie et al. 2001) یک متدولوژی انعطاف‌پذیر را برای شناسایی واحدهای همگن هیدرولوژیکی ارائه نمودند که دست‌مایه تحقیق حاضر نیز می‌باشد. آنان پیشنهاد نمودند که بطور کلی گسسته‌سازی حوضه‌های متوسط و بزرگ حداقل در دو سطح (مرحله) انجام پذیرد، بدین صورت که در مرحله اول تقسیم یک حوضه متوسط یا بزرگ به زیر حوضه‌های کوچک‌تر و فرضاً به کمک مرتبه‌بندی (Strahler or Shreve) انجام و در مرحله بعد در این زیرحوضه‌ها، شناسایی واحدهای هیدرولوژیکی همگن با استفاده از روش‌های مناسب گسسته‌سازی انجام گیرد. آنان برای این شناسایی (تفکیک)، موضوع برآورد بیلان آب در حوضه مورد نظر را انتخاب نمودند و کار ایشان را می‌توان در چهار بخش خلاصه نمود: اول، انتخاب تعدادی لایه موثرتر (لایه‌های شیب، کاربری و زمین شناسی) از بین تعدادی لایه‌های مرتبط با بیلان آب. دوم تجزیه این لایه‌های انتخابی به تعدادی دسته و سپس انتخاب تعدادی نواحی مرجع از ترکیب و همپوشانی آنها. سوم، با در نظر گرفتن یک پنجره و توصیف‌گر همسایگی و با تحلیل ترکیب همسایگی در این پنجره و همچنین به کمک معیار فاصله، شناسایی واحدهای همگن (متناظر با نواحی مرجع انتخابی) انجام گردید، چهارم، کفایت تعداد نواحی مرجع انتخابی با استفاده از نقشه فاصله بررسی گردید. آنان جهت تکمیل و تدقیق مراحل پیشنهادی خود، ضرورت بکارگیری تکنیک و روش‌های مناسب که اولاً: منجر به شناسایی لایه‌های موثرتر گردد ثانیاً: انتخاب منطقی تعداد نواحی مرجع (واحدهای همگن مورد نظر در این تحقیق) و ثالثاً تعیین حد مشخصی جهت توقف عملیات گسسته‌سازی با معیاری مناسب را توصیه نمودند که در تحقیق حاضر به هر سه آنها پرداخته شده است. بایستی اضافه نمود، از آنجا که هدف اصلی شناسایی (تفکیک) واحدهای همگن می‌باشد، لذا استفاده از روش فازی و بویژه روش خوشه‌بندی میان مرکز فازی (FCM) بنا به توصیه محققین مختلف و بعنوان نمونه (Bruin and Stein 1998) و (Rao and Srinivas 2007) که آنرا روشی مناسب برای تفکیک و پهنه‌بندی عرصه‌های طبیعی دانسته‌اند، مبنای این تفکیک قرار گرفته است.

## ۳- مواد و روش‌ها

### ۳-۱- محدوده تحقیق

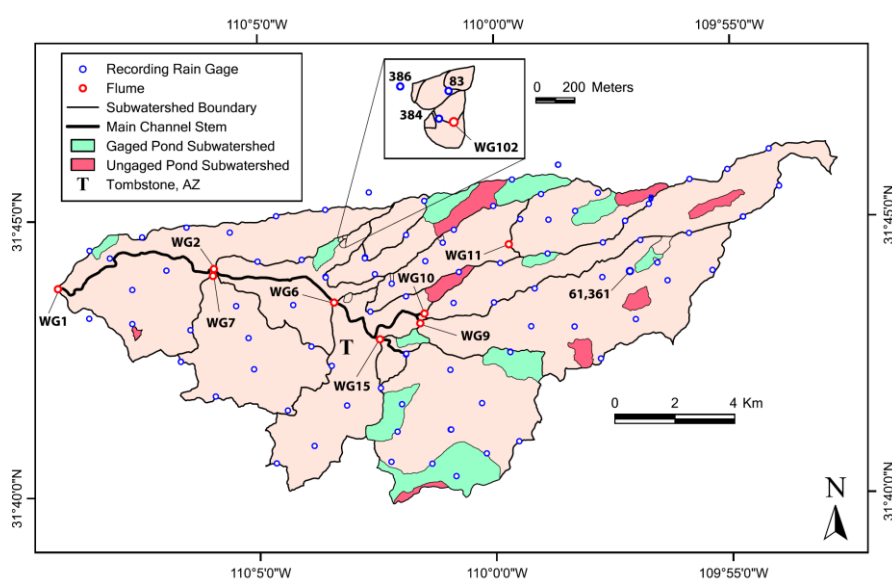
در این تحقیق، از اطلاعات و داده‌های حوضه معرف والنات گولچ (WGWE) که یکی از زیر حوضه‌های رود Upper SanPedro واقع در جنوب شرقی ایالت آریزونای آمریکا، با موقعیت

### ۳-۲- انتخاب و آماده‌سازی لایه‌ها

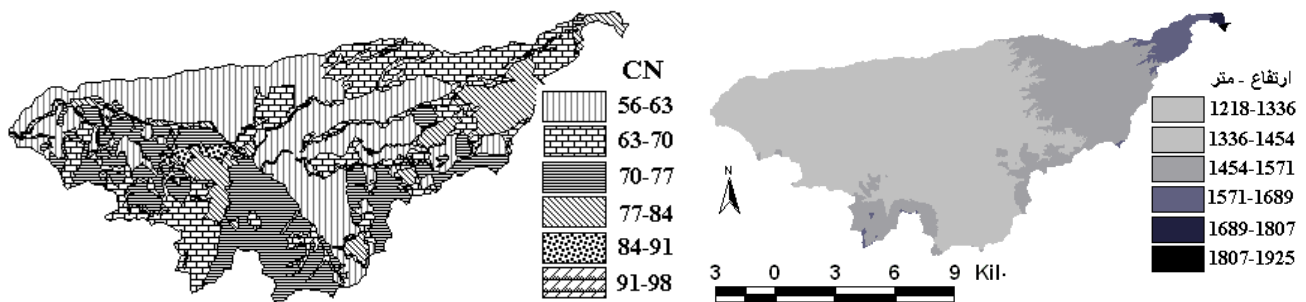
از آنجا که در تحقیق حاضر، شناسایی واحدهای همگن از نظر پاسخ به تولید رواناب (سیل‌خیزی) مدنظر می‌باشد، لذا سعی گردیده تا حدامکان لایه‌های موثر بر این فرآیند انتخاب و یا تولید شوند. این لایه‌ها می‌توانند شامل مدل‌رقومی ارتفاع (DEM)، شیب (Slope)، جهت شیب (Aspect) از نظر ارتباط با جهت حرکت ابرها، برفگیر بودن و یا تاثیر بر نرخ ذوب برف و رواناب، انحناء (Curvature) در پروفیل و یا در پلان که تاثیرگذار از نظر واگرایی یا همگرایی رواناب و نیز شتاب جریان می‌باشد، عدد منحنی (CN)، شاخص گیاهی نرمال شده (NDVI)<sup>۶</sup> معرف تراکم پوشش گیاهی، شاخص موقعیت توپوگرافیک (TPI)<sup>۷</sup> معرف عوارض ارتفاعی (Weiss, 2001)، شکل زمین (Landforms) متمایز کننده مورفولوژی سطوح هیدرولوژیک و لایه متوسط بارش سالیانه (MAR)<sup>۸</sup> می‌باشند. تمامی لایه‌های رستری مورد نیاز با ابعاد سلول ۳۰ متر تهیه و یا تولید گردیده‌اند که به عنوان نمونه لایه رستری DEM و لایه برداری CN در (شکل ۲) نشان داده شده‌اند.

در این مرحله ذکر دو نکته ضروری است: اول اینکه شاید بتوان پارامترها، متغیرها و یا شاخص‌های بیشتری انتخاب نموده و در این کار مشارکت داد، ولی ترجیح داده شد که معروفترین و در دسترس‌ترین آنها گزینش شوند، چرا که در بعضی از کارهای مشابه این تحقیق، حتی از تعداد کمتری از پارامترهای دخیل در فرآیند مورد مطالعه مورد نظرشان استفاده گردیده و بیشتر بر روی تدقیق متدولوژی پیشنهادی خود متمرکز شده‌اند.

(31° 45' N, 110° 03' W) می‌باشد، استفاده شده است (شکل ۱). این حوضه معرف، با وسعت حدود ۱۵۰ km<sup>2</sup> دارای انواع تجهیزات اندازه‌گیری و بصورت متراکم برای مطالعات هیدرولوژیک، در واقع یک آزمایشگاه صحرایی در مطالعات هیدرولوژیک برای نواحی نیمه خشک در دنیا محسوب می‌شود. به‌عنوان نمونه در این وسعت محدود، ۹۰ ایستگاه ثبات و معمولی باران سنجی وجود دارد. بعلاوه مقالات مختلفی نیز در تایید اصالت و دقت نقشه‌ها، اطلاعات مشاهداتی و ویژگی‌های زمینی و اقلیمی آن در مجلات معتبر به چاپ رسیده است (Heilman et al., 2008). قابل ذکر است که بخش زیادی از اطلاعات هیدرولوژیک این حوضه شامل فیزیوگرافی، زمین شناسی و هواشناسی و همچنین برخی گزارش‌ها و تحقیقات انجام شده در آن از سایت <http://www.tucson.ars.ag.gov/dap> قابل دریافت می‌باشد. علت عمده استفاده از این حوضه به دلیل این بوده که تفکیک (تعیین مرز مشخص) عرصه‌های طبیعی پیوسته با استفاده از یک روش مناسب فازی با وجود همخوانی به نسبت زیاد این دو (پیوستگی در تغییرات زمینی و غیر قطعی بودن نتایج)، بطور متعارف به معنی پذیرش عدم قطعیت نسبتاً قابل توجه خواهد بود. اگر در این میان کمیت و کیفیت داده‌ها نیز از کفایت و سلامت کمی برخوردار باشند، بی‌تردید اتخاذ سه تصمیم مناسب و مورد نیاز در این روش که قابل‌توصیه نیز باشد، چندان آسان و قابل اعتماد نخواهد بود. نکته دیگر اینکه، به جهت بالابردن دقت و کاهش عدم قطعیت خوشه‌بندی، ابعاد سلول لایه‌های رستری ۳۰ متر (به‌عبارتی حدود ۱۶۰۰۰۰ سلول برای این حوضه کوچک)، در نظر گرفته شده است.



شکل ۱- محدوده و موقعیت حوضه WGEW واقع در جنوب شرقی ایالت آریزونا، آمریکا



شکل ۲- لایه‌رستری مدل رقومی ارتفاع (DEM) با اندازه سلول ۳۰ متر و لایه عدد منحنی (CN)

اندازه‌گیری شده داده‌ها (لایه‌ها) مختلف باشد، لایه‌های با اعداد بزرگتر، در خوشه‌بندی (با معیار فاصله) حاکم شده و منجر به نتایج اوریب خواهد شد (Chiu, 2005). بدین‌خاطر بایستی قبل از خوشه‌بندی این اختلاف با استفاده از یک روش مناسب استاندارد سازی، نظیر رابطه (۱) بر طرف گردد.

$$x_{jn} = \frac{y_{jn} - y_{j(\min)}}{y_{j(\max)} - y_{j(\min)}}, \quad j = 1, 2, \dots, t \quad (1)$$

که در آن  $x_{jn}$  و  $y_{jn}$  به ترتیب داده‌های استاندارد شده و واقعی نام از لایه‌زام مشارکت کننده در خوشه‌بندی،  $y_{j(\max)}$  و  $y_{j(\min)}$  به ترتیب مقادیر واقعی حداکثر و حداقل لایه زام و  $n$  تعداد سلول لایه‌ها می‌باشد. لازم به توضیح است که در خوشه‌بندی موردنظر  $t=3$  می‌باشد، زیرا لایه‌های CN, DEM, TPI در خوشه‌بندی مشارکت خواهند کرد.

#### ۳-۴-۱- تعیین مقدار فازیت (Fuzzifier)

در خوشه‌بندی FCM میزان فازیت (غیر قطعی بودن) تعلق هر سلول با خوشه مربوطه، با پارامتر  $m$  تعیین می‌گردد. از آنجا که در این روش، مقدار فازیت ( $m$ ) و تعداد خوشه‌ها ( $c$ ) در خوشه‌بندی به لحاظ عملکردی متضاد هم عمل می‌نمایند، لذا بایستی هر دو بهینه گردند. زیرا افزایش در مقدار فازیت، موجب کاهش متوسط درجه عضویت سلول‌ها به خوشه‌ها و در عین حال موجب همپوشانی بیشتر خوشه‌های مجاور در نواحی مرزی خواهد شد. (نمایش تدریجی بیشتر تغییرات در طبیعت). در حالی که با افزایش تعداد خوشه‌ها، شباهت نقاط درون یک خوشه به مرکز خود بیشتر خواهد شد، هر چند در تفکیک نواحی این افزایش مطلوب و به‌عبارتی کاربردی نمی‌باشد. لذا جهت تعیین محدوده فازیت و مقدار مطلوب آن در یک خوشه‌بندی فازی، به پیشنهاد (Rao and Srinivas, 2007) می‌توان از روابط اعتبارسنجی (Bezdek, 1981)، یعنی ضریب جداسازی خوشه‌ها

دوم، در جهت اجتناب از حجیم شدن مطالب، فرض بر این است که خوانندگان با شاخص‌های فوق آشنا بوده و برای آشنایی بیشتر و یا نحوه بدست آمدن آنها نیز می‌توانند از الحاقیات نرم‌افزارهای Arc GIS و ArcView, ILWIS استفاده نمایند.

#### ۳-۳- تحلیل مولفه‌های اصلی (PCA)

به علت ارتباط معنی‌داری که بین پارامترهای مشارکت‌کننده در یک فرآیند هیدرولوژیک وجود دارد، استفاده از این تکنیک، علاوه بر کمک به شناسایی لایه‌های اصلی (مستقل)، می‌تواند در جهت کاهش ابعاد داده‌های ورودی به عملیات خوشه‌بندی FCM نیز مفید باشد.

#### ۳-۴- خوشه بندی میان مرکز فازی (FCM)

بطور کلی، استفاده از روش‌های خوشه‌بندی فازی جهت شناسایی نواحی همگن ضمن نمایش تغییرات و یا ناهمگنی‌ها در عرصه‌های طبیعی، به علت غیر قطعی (فازی) بودن خروجی‌ها، می‌تواند تا حدود زیادی همانند طبیعت، پیوستگی در این تغییرات را در مرز خوشه‌ها حفظ نماید، زیرا میزان تعلق هر سلول به یک خوشه با درجه عضویت تعیین می‌گردد، که مقداری بین (۰-۱) دارد. بدین‌خاطر الگوریتم FCM که توسط (Bezdek, 1981) پیشنهاد گردیده، برای این خوشه‌بندی (گسسته‌سازی) استفاده شده است. در این روش، مبنای اتلاق یک سلول به یک خوشه کمترین فاصله اقلیدسی بین مقدار سلول و مقدار مرکز خوشه می‌باشد. لازم بذکر است، به لحاظ بدیهی بودن این روش، از ارائه الگوریتم و مراحل انجام آن صرفنظر شده و خوانندگان می‌توانند به مراجع مختلف نظیر، غضنفری و رضائی (۱۳۸۵) مراجعه نمایند. همچنین این روش برای خوشه‌بندی داده‌های (لایه‌های) چندگانه مناسب است که ارتباط خطی بین آنها وجود نداشته باشد. علاوه بر آن اگر بزرگی مقدار

### ۳-۴-۳- عدم قطعیت خوشه‌بندی

برای تحلیل عدم قطعیت خوشه بندی، می‌توان از شاخص پیشنهادی Burrough et al. (1997) معروف به شاخص ابهام (CI) و با رابطه زیر استفاده نمود.

$$CI_k = 1 - \left( u_{k, \max} - u_{k, 2^{nd} \max} \right) \quad (4)$$

که در آن  $u_{k, \max}$  بیشترین مقدار درجه عضویت سلول  $k$  و  $u_{k, 2^{nd} \max}$  ماقبل بیشترین (دومین) درجه عضویت سلول  $k$

می‌باشند. در واقع این شاخص تعیین‌کننده میزان قطعیت یک سلول به خوشه مربوطه بوده و همچنین مبین میزان همپوشانی خوشه‌های مجاور در نواحی مرزی خوشه‌ها می‌باشد. مقدار کمی این شاخص بین [۰،۱] می‌باشد و بدین صورت که سلول با  $CI$  بزرگتر (نزدیک به ۱) دارای عدم قطعیت بیشتر در تعلق به خوشه اختصاص یافته خواهد بود و بر عکس. جهت بررسی کیفی این پارامتر (میزان ابهام) هر سلول اختصاص یافته به خوشه موردنظر، می‌توان محدوده کلی این شاخص ابهام [۰،۱] را به پنج بازه مساوی تقسیم و نام‌گذاری نمود (Chiu, 2005). به عبارتی چنانچه مقدار این شاخص برای هر سلول بین ۰ تا ۰/۲ باشد ابهام آن سلول در خوشه اختصاص یافته ناچیز می‌باشد و به همین شکل، مقادیر بین ۰/۲ تا ۰/۴ و ۰/۴ تا ۰/۶ و ۰/۶ تا ۰/۸ و ۰/۸ تا ۱ بترتیب معرف ابهام خیلی کم، ابهام کم، ابهام زیاد و ابهام خیلی زیاد می‌باشد. لذا تهیه نقشه این شاخص با تعیین آن (به صورت کمی و یا کیفی) برای تمامی سلول‌ها می‌تواند جهت صحت‌یابی عملیات خوشه‌بندی مفید باشد.

### ۳-۴-۴- اعتبارسنجی خوشه‌بندی (Clustering Validation)

اعتبارسنجی در خوشه‌بندی فازی بمعنی سنجش همزمان تراکم داده‌های درون خوشه‌ایی و فاصله بین خوشه‌ایی می‌باشد که با تأمین مقدار حداکثر برای آنها، تعداد بهینه خوشه‌ها حاصل خواهد شد. برای سنجش اعتبار خوشه بندی فازی روابط مختلفی توسط محققین ارائه گردیده که در جدول ۱ لیست گردیده و در تحقیق حاضر نیز جهت مقایسه و بررسی از آنها استفاده شده است.

علاوه بر روابط مذکور برای اعتبارسنجی خوشه‌ها، از رابطه پیشنهادی Liu and George (2005) که توسعه رابطه Xie and Beni (1991) بکمک ضریب همبستگی پیرسون (رابطه ۹) می‌باشد، نیز استفاده شده است (رابطه ۱۰).

( $V_{PC}$ ) که معرف متوسط مجذور درجه عضویت ( $u_{ik}$ ) تمامی سلول‌ها می‌باشد (رابطه ۲) و آنتروپی خوشه‌ها ( $V_{CE}$ ) معرف بی‌نظمی در خوشه‌ها (رابطه ۳) می‌باشند، بصورت غیر مستقیم استفاده نمود. زیرا از آنجا که با افزایش مقدار فازیت از متوسط درجه عضویت سلول‌ها کاسته شده و در عین حال به مقدار بی‌نظمی سلول‌ها در تعلق به خوشه‌های مجاور خوشه مربوطه افزوده می‌گردد، لذا با ترسیم همزمان تغییرات مقدار این دو رابطه بازم مقادیر مختلف ( $m$ ) و ( $c$ )، بایستی بتوان محدوده فازیت و مقدار مناسب  $m$  را یافت که در ادامه نیز نشان داده شده است.

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

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

در رابطه (۳)، مبنای  $a$  می‌تواند مقداری بین ۱ تا بی‌نهایت باشد که پیش فرض آن ۱۰ می‌باشد و در این مطالعه نیز در نظر گرفته شده است.

### ۳-۴-۲- غیرفازی سازی (Defuzzifying)

به عبارتی، از میان یک دسته نتایج خروجی فازی، می‌بایست یک معیار منطقی برای درجه عضویت مناسب جهت اتلاق یک سلول به یک خوشه اتخاذ گردد. اگر چه استفاده از روش ماکزیمم درجه عضویت، نسبت به روش‌های دیگر غیرفازی سازی عمومیت بیشتری دارد، ولی در این روش، ارتباط قوی دیگر خوشه‌ها با یک سلول مدنظر قرار نمی‌گیرد و از طرف دیگر، پذیرش تعلق یک سلول به یک خوشه زمانی که ماکزیمم درجه عضویت آن (سلول) مقدار کمی باشد زیر سوال خواهد رفت. در چنین مواقعی استفاده از یک حد قبولی برای درجه عضویت بنام برش آلفا ( $\alpha$ -cut) می‌تواند راهگشا باشد. هر چند تعیین این حد نیز کار مشکلی است. بدین منظور، توسط Chiu (2005) دو برش آلفا ( $\alpha_{High}$ ,  $\alpha_{Low}$ ) برای مقدار درجه عضویت برحسب تعداد خوشه ( $c$ ) پیشنهاد گردیده است. در حدود ایشان،  $\alpha_{High} \geq 1-1/c$  بوده و چنانچه مقدار تابع عضویت هر سلول برابر این حد باشد اتلاق آن سلول به خوشه موردنظر قطعی می‌باشد و همچنین  $\alpha_{Low} < 1/c$  می‌باشد که این مقدار درجه عضویت برای هر سلول، بطور منطقی جهت رد قاطع تعلق آن سلول به خوشه اختصاص یافته لازم می‌باشد. همچنین سلول‌های با درجه عضویت بین دو حد مذکور در محدوده انتقالی قرار داشته و در صورت قابل توجه بودن این سلول‌ها، اضافه کردن تعدادی خوشه و فقط در این محدوده برای تأمین درجه عضویت قابل قبول توصیه می‌گردد.

لازم است اشاره گردد در این تحقیق سعی گردیده از ضریب تعیین ( $R^2$ ) که در واقع نسبت واریانس داده‌های تخمین زده شده (خوشه‌بندی‌شده) به واریانس کل داده‌های واقعی می‌باشد، به دو منظور استفاده گردد. اول جهت جایگزینی آن با ضریب پیرسون (رابطه ۹) در رابطه (۱۰) برای اعتبارسنجی خوشه‌ها و دوم، به عنوان یکی از معیارهای اساسی در تعیین تعداد بهینه خوشه‌ها در روش پیشنهادی نگارندگان که در ادامه به آن اشاره شده است. لذا با به کارگیری این ضریب تعیین (رابطه ۱۲) رابطه اعتبارسنجی Liu and George (2005) را می‌توان بصورت رابطه (۱۳) نشان داد.

$$R^2 = \frac{\sum (Y - \bar{X})^2}{\sum (X - \bar{X})^2} \quad \text{or} \quad (12)$$

$$R = \left[ \frac{\sum (Y - \bar{X})^2}{\sum (X - \bar{X})^2} \right]^{1/2}$$

$$S(R) = \frac{\sum_{i=1}^c \sum_{k=1}^n \sum_{j=1}^t (u_{ik})^2 (1 - R_{ki})}{n \min(1 - R_o)} \quad (13)$$

$$P = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum (X - \bar{X})^2 \sum (Y - \bar{Y})^2}} \quad (9)$$

$$S(p) = \frac{\sum_{i=1}^c \sum_{k=1}^n \sum_{j=1}^t (u_{ik})^2 (1 - P_{ki})}{n \min(1 - P_o)} ; \quad (10)$$

$$P_{ki} = \frac{\sum_{i=1}^c \sum_{j=1}^t (x_{kj} - \bar{x}_j)(v_{ij} - \bar{v}_j)}{\sqrt{\sum_{j=1}^t (x_{kj} - \bar{x}_j)^2 \sum_{i=1}^c \sum_{j=1}^t (v_{ij} - \bar{v}_j)^2}}$$

$$P_o = \frac{\sum_{i=1}^c \sum_{j=1}^t (v_{oj} - \bar{v}_j)(v_{ij} - \bar{v}_j)}{\sqrt{\sum_{j=1}^t (v_{oj} - \bar{v}_j)^2 \sum_{i=1}^c \sum_{j=1}^t (v_{ij} - \bar{v}_j)^2}} ; \quad (11)$$

$$\bar{v}_j = \frac{\sum_{i=1}^c v_{ij}}{c} ; \quad i, o = 1, 2, \dots, c$$

جدول ۱- روابط پیشنهادی محققین برای اعتبار سنجی خوشه بندی (بهینه یابی تعداد خوشه‌ها)

پیشنهاد شده	رابطه اعتبار سنجی
Bezdek, (1981)	$V_{FPI}(U) = 1 - \frac{c \times V_{pc}(U) - 1}{c - 1}$ (5)
Fukuyama and Sugeno, (1989)	$V_{FS}(U, V : X) = \sum_{i=1}^c \sum_{k=1}^n \sum_{j=1}^t (u_{ik})^m \ v_{ij} - x_{kj}\ ^2 - \sum_{i=1}^c \sum_{k=1}^n \sum_{j=1}^t (u_{ik})^m \ v_{ij} - \bar{x}_j\ ^2$ $\bar{x}_j = \sum_{k=1}^n x_{kj} / n \quad j = 1, 2, \dots, t$ (6)
Xie and Beni index, (1991)	$V_{XB,m}(U, V : X) = \frac{\sum_{i=1}^c \sum_{k=1}^n \sum_{j=1}^t (u_{ik})^2 \ v_{ij} - x_{kj}\ ^2}{n \min \ v_i - v_o\ ^2} ; \quad i \neq o \quad i, o = 1, 2, \dots, c$ (7)
Kwon, (1998)	$V_K(U, V : X) = \frac{\sum_{i=1}^c \sum_{k=1}^n \sum_{j=1}^t (u_{ik})^m \ v_{ij} - x_{kj}\ ^2 + \frac{1}{c} \sum_{i=1}^c \sum_{j=1}^t \ v_{ij} - \bar{x}_j\ ^2}{\min \ v_i - v_o\ ^2}$ (8)
اغلب پارامترها قبلاً تعریف شده‌اند و فقط $V_{ij}$ مقدار مرکز خوشه $i$ ام لایه ورودی $j$ ام، $V_i$ مقدار مرکز خوشه $i$ ام و $V_o$ مقدار مرکز خوشه غیر از خوشه $i$ ام می‌باشد.	

قابل توجه اینکه، در رابطه (۱۳) پارامترهای  $R_{ki}$  و  $R_o$  به مانند روابط (۱۰) و (۱۱) قابل استخراج و استفاده می‌باشند که جهت رعایت اختصار از نوشتن آنها صرفنظر شده است.

می‌باشد. زیرا فرم رابطه این فاصله (رابطه ۱۲) با فرم فاصله اقلیدسی بکار رفته در تابع هدف روش FCM هم‌خوانی بیشتری دارد.

$$d_k = d(X - v) = \sum_{j=1}^t |x_{kj} - v_{kj}| \quad (12)$$

در این رابطه  $d_k$  فاصله مقدار هر سلول ( $X$ ) با مقدار مرکز خوشه خود ( $v$ ) می‌باشد و عبارتی فاصله مقدار سلول  $k$ ام لایه  $j$ ام ( $x_{kj}$ ) با مقدار مرکز خوشه سلول  $k$ ام لایه  $j$ ام ( $v_{kj}$ ) بوده و همچنین  $t$  تعداد لایه عوارض زمینی مشارکت کننده در خوشه‌بندی می‌باشد.

#### ۴- نتایج و بحث

**تعیین لایه‌های اصلی:** در تکنیک PCA، با در نظر گرفتن این معیار که یک مولفه اصلی بایستی حداقل ۱۰٪ از واریانس کل ( $\geq 10\%$ ) مقدار ویژه را نمایندگی نماید (Wolock et al., 2004). مطابق جدول ۲ تعداد مولفه‌های اصلی به سه مولفه کاهش یافت. البته بایستی اضافه نمود که معیار فوق یک معیار کیفی است و بسته به موضوع و یا تجربه کاربر دارای انعطاف می‌باشد و فرضاً با کمی ارفاق می‌توان مولفه چهارم را نیز در نظر گرفت. در این تحقیق روال و گام‌های اساسی این گسسته‌سازی بیشتر مدنظر می‌باشد. متعاقباً، در هر یک از این سه مولفه، لایه‌ای که دارای بیشینه واریانس مشترک در ماتریس همبستگی (در حالت چرخش یافته یا بدون چرخش) باشد، به عنوان لایه اصلی معین گردید. مطابق جدول ۳ که در حالت چرخش یافته محاسبه گردیده، لایه‌های اصلی عبارتند از: TPI, DEM, CN می‌توان فقط از آنها بعنوان داده‌های ورودی استفاده نمود که با پیشنهاد و البته بدون استناد (Dehotin and Braud 2008) به عنوان لایه‌های موثرتر (حاکم) در تولید رواناب نیز هم‌خوانی دارد. این جدول نیز معمول است که در هر مولفه اصلی، لایه‌ایی که دارای بیشترین مقدار باشد، انتخاب گردد.

ولی به مانند جدول ۲ این معیار نیز کیفی است. به‌عنوان مثال در مولفه اصلی اول، هر چند اعداد مربوط به TPI و Land بسیار به هم نزدیک می‌باشند ولی از آنجا که در تهیه لایه Landform (شاخص شکل زمین)، نیاز به لایه‌های TPI و Slope می‌باشد، لذا ضمن اینکه نزدیکی این دو مقدار کاملاً معنی‌دار و درست می‌باشد، اولویت انتخاب در این جدول و در مولفه اصلی اول با TPI می‌باشد. نکته دیگر بررسی نمایندگی کیفی لایه‌های اصلی استخراجی از PCA می‌باشد. به عبارتی هر کدام از آنها، بایستی نماینده تعدادی لایه حذف شده (وابسته) بوده و در عین حال از نظر فیزیکی تاثیر قابل توجهی در فرآیند مورد مطالعه داشته باشند.

جهت بهینه‌یابی تعداد خوشه‌ها با استفاده از روابط مذکور بایستی اذعان نمود، تعداد بهینه در رابطه (۵) متناظر با حداکثر مقدار  $V_{FPI}$  بوده و در سایر روابط (۶، ۷، ۸، ۱۰ و ۱۳) نیز متناظر با مقادیر اپتیمم (حداقل) تابع آنان ( $V_{FS}, V_{XB, mb}, V_K, S(P), S(R)$ ) می‌باشد. از آنجا که این مطالعه نشان داده است که استفاده از این روش‌های اعتبارسنجی لزوماً و یا در هر شرایطی منتج به تعیین تعداد بهینه خوشه‌ها بصورت واضح و قطعی نخواهد شد (شکل ۴) که البته مطالعه (Rao and Srinivas 2007) نیز موید آن می‌باشد، لذا نگارندگان جهت تعیین عدد بهینه خوشه‌ها استفاده از یک روش استدلالی را پیشنهاد می‌نمایند. در حقیقت این روش مبتنی بر در نظر گرفتن تابع هدف و پارامترهای مورد استفاده در روش FCM و همچنین به کارگیری و ترسیم دو اندرکنش پیشنهادی زیر میسر می‌باشد.

**الف- اندرکنش  $\alpha_{High}$  و  $R^2$ :** به عبارتی بازاء مقادیر مختلف خوشه ( $c$ ) می‌توان تعداد سلول‌های با درجه عضویت برابر یا بزرگتر از  $\alpha_{High}$  و همچنین مقدار ضریب تعیین ( $R^2$  - رابطه ۱۱) نظیر را بدست آورده و در یک دستگاه ترسیم گردند. زیرا به‌طور معمول، با افزایش تعداد خوشه‌ها از تعداد  $\alpha_{High}$  کاسته شده ولی به مقدار  $R^2$  افزوده می‌گردد.

**ب- اندرکنش مقدار فاصله و شاخص ابهام (CI):** از آنجا که با افزایش تعداد خوشه‌ها، فاصله بین مقادیر واقعی سلول‌ها با مقادیر مراکز خوشه‌های متناظر کاهش می‌یابد، ولی به مقدار شاخص ابهام یا CI سلول‌ها افزوده می‌گردد، بنابراین با ترسیم همزمان این دو بازاء مقادیر مختلف  $c$  می‌توان منجر به تعیین تعداد بهینه خوشه‌ها گردد.

لازم به توضیح می‌باشد، در هر خوشه‌بندی جهت سنجش فاصله (شباهت) بین داده‌های واقعی در خوشه با مراکز خوشه‌های متناظر (به دست آمده) حتی بعد از اعتبار سنجی، می‌توان از یک معیار فاصله دیگر در جهت کنترل مضاعف نیز استفاده نمود. روش‌های مختلفی نظیر فاصله مودال<sup>۱۱</sup>، فاصله کلموگراف<sup>۱۲</sup> یا فاصله مانهاتان<sup>۱۳</sup> وجود دارد که از میان آنها، روش آخر به پیشنهاد (Robbez-Masson 1994) برای خوشه بندی مورد نظر مناسب‌تر

جدول ۲- واریانس تبیین شده به وسیله مولفه‌های اصلی

	شماره مولفه								
	۱	۲	۳	۴	۵	۶	۷	۸	۹
مقادیر ویژه <sup>a</sup>	۲/۴۴	۱/۸۲	۱/۲۵	۰/۹۵	۰/۷۸	۰/۷۲	۰/۴۶	۰/۴۵	۰/۱۲
نسبت واریانس تبیین شده <sup>b</sup>	۰/۲۷	۰/۲۰	۰/۱۴	۰/۱۱	۰/۰۹	۰/۰۸	۰/۰۵	۰/۰۵	۰/۰۱
واریانس تبیین شده تجمعی <sup>c</sup>	۲۷	۴۷	۶۱	۷۲	۸۱	۸۹	۹۴	۹۹	۱۰۰

a- Eigen value b- Proportion of variance explained  
c- Cumulative variance explained

سازی ۱/۷ - ۱/۱ و بهترین مقدار برای ضریب فازیت ( $m$ ) برابر ۱/۵ بدست آمد است. لازم بذکر است، بازاء ( $m \geq 1/8$ ) اعتبارسنجی‌ها دچار ناپایداری شده، بگونه‌ای که نمودار آنان قابل ترسیم نبوده است.

**اعتبارسنجی خوشه‌ها:** مطابق اشکال ۴ (الف و ب) و ۵ (الف) که با استفاده از روابط (۵)، (۶) و (۸) ترسیم شده‌اند، به علت تعدد نقاط حداقل و همچنین وجود روند پیوسته افزایشی یا کاهشی، تعیین عدد بهینه خوشه‌ها در آنها بوضوح میسر نمی‌باشد که البته (Rao and Srinivas (2007) نیز به این موضوع اشاره نموده‌اند، ولی در تحقیق خود پیشنهادی برای تعیین عدد بهینه ارائه نکرده‌اند. با این وجود به نظر می‌رسد، نمودارهای شکل ۵ (ب) که بازاء  $m=1/5$  ترسیم گردیده‌اند، بیشتر قابل بررسی باشند. در این شکل مشاهده می‌گردد، نمودارها از نظر رفتار تقریباً مشابه هم بوده و بر خلاف سایرین دارای روند کاهشی یا افزایشی نمی‌باشند. ولی پذیرش تعداد ۱۱ خوشه به عنوان عدد بهینه نیز به دو علت کمی مشکل خواهد بود. اول بدلیل رفتار سیکی نسبتاً شدید نمودارها و دوم اینکه تفاوت معنی‌داری در مقدار  $S$  متناظر با خوشه‌های ۷ و ۱۱ (بعنوان عدد بهینه) مشاهده نمی‌گردد.

لذا با محاسبه و تعیین دو اندرکنش پیشنهادی که در شکل ۶ به صورت گرافیکی نشان داده شده، به وضوح می‌توان تعداد بهینه خوشه و به عبارتی تعداد نواحی همگن را برابر ۸ انتخاب نمود.

بعد از انتخاب تعداد نواحی همگن بایستی مقادیر مهم و تعیین‌کننده در یک خوشه‌بندی مناسب کنترل گردند که در جداول ۴ و ۵ نیز لیست گردیده‌اند. مطابق جدول ۴ مقدار متوسط درجه عضویت در تمامی خوشه‌ها بیش از ۰/۸ بوده و علاوه بر آن مطابق جدول ۵، متوسط درجه عضویت کل حوضه نیز برابر ۰/۸۸ می‌باشد.

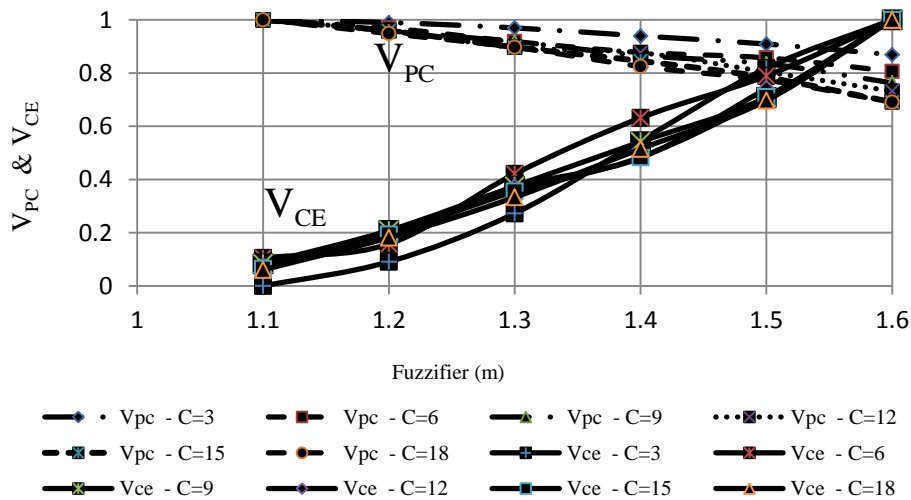
جدول ۳- تعیین بیشترین واریانس در ماتریس همبستگی

	ماتریس مولفه‌های چرخش یافته		
	۱	۲	۳
TPI	۰/۹۴۹	۰/۰۱۲	-۰/۰۲۸
Landform	۰/۹۱۶	۰/۰۲۸	۰/۰۲۶
Curvature	۰/۸۰۸	-۰/۰۲۲	۰/۰۱۳
DEM	۰/۰۶۳	۰/۷۶۸	۰/۳۵۰
MAR	-۰/۰۱۹	۰/۷۴۰	-۰/۰۲۹
NDVI	-۰/۰۱۹	۰/۷۱۴	۰/۱۱۱
CN	-۰/۰۰۳	۰/۱۶۱	۰/۷۱۵
Slope	۰/۱۲۹	۰/۰۶۸	۰/۷۱۲
Aspect	۰/۱۲۰	۰/۲۱۱	-۰/۴۶۱

با بررسی کیفی لایه‌های اصلی مشاهده می‌گردد، لایه اصلی عدد منحنی (CN) می‌تواند به‌عنوان نماینده جنس‌زمین، کاربری و تراکم پوشش گیاهی باشد. لایه ارتفاع رقومی (DEM) به عنوان تعیین‌کننده شیب، جهت شیب، انحنا سطوح (در پلان یا پروفیل) می‌باشد. از سویی با وجود ارتباط معنی‌دار (عموماً خطی) که معمولاً در حوضه‌ها بین ارتفاع نقاط و متوسط بارندگی سالیانه آنان وجود دارد، تقریباً نمایندگی ارتفاع از متوسط بارندگی سالیانه دور از انتظار نمی‌باشد. لایه شاخص موقعیت توپولوژیک (TPI) نیز می‌تواند نماینده عوارض ارتفاعی نظیر چاله، تپه، دره و نواحی مسطح یک سلول باشد که این عوارض سهم قابل توجهی در تولید رواناب خواهند داشت.

**تعیین فازیت مناسب:** با ترسیم همزمان روابط (۲) و (۳) و مطابق شکل ۳ محدوده تغییرات و قابل استفاده فازیت در این مدل





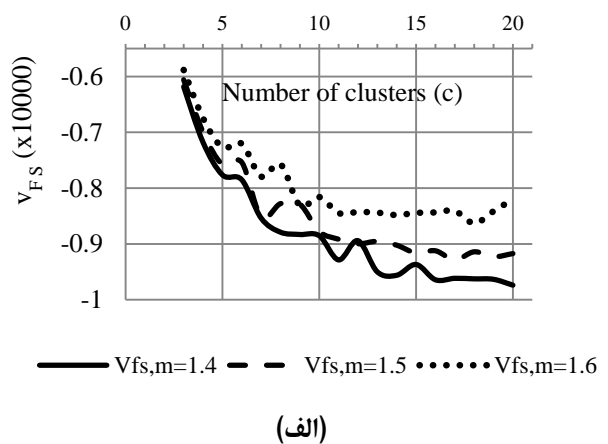
شکل ۳- اندرکنش ضریب جدایی بین خوشه‌ها (VPC) و آنتروپی درون خوشه‌های (VCE) بازاء مقادیر مختلف مقدار فازیت (m) و تعداد خوشه‌ها (C)

به عبارتی این مقدار حتی از  $\alpha_{High}$  که با در نظر گرفتن  $C=8$  برابر  $0.875$  خواهد شد که حد قبولی نسبتاً بالایی می‌باشد نیز بیشتر است. همچنین مطابق جدول ۵، حداکثر شاخص ابهام برابر  $0.33$  می‌باشد که در محدوده خیلی کم ( $0.2-0.4$ ) قرار دارد. قابل ذکر اینکه با انتخاب ۸ خوشه، بیشترین مقدار متوسط قدرمطلق خطای نسبی (MARE) برابر  $17/7$  درصد می‌باشد (جدول ۴). در حالی که با انتخاب ۱۱ خوشه به عنوان عدد بهینه که در شکل ۵ (ب) خودنمایی می‌کند، با محاسبه مقادیر کنترلی جداول ۴ و ۵ مشاهده گردید، علاوه بر بهبود نیافتن متوسط درجه عضویت و متوسط شاخص ابهام کل حوضه (تمامی سلول‌ها)، بیشترین مقدار متوسط قدرمطلق خطای نسبی (MARE) در یکی از خوشه‌ها (در این حالت) برابر  $50/7$  درصد گردید.

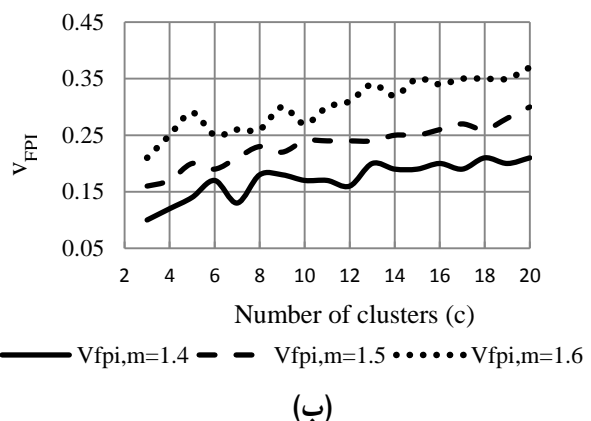
در این مرحله با توجه به مطلوبیت مقادیر کنترل کننده جداول ۴ و ۵، اقدام به تهیه و نمایش نقشه‌های خوشه‌بندی (۸ ناحیه همگن)، درجه عضویت سلول‌ها به خوشه‌های مربوطه، شاخص ابهام سلول‌ها و نهایتاً نقشه فاصله مانهاتان گردید (شکل ۷).

#### ۵- خلاصه و جمع‌بندی

با توجه جمع‌بندی نتایج بدست آمده، بحث و موارد یاد شده در طول تحقیق، نکات زیر را می‌توان به صورت خلاصه برشمرد:

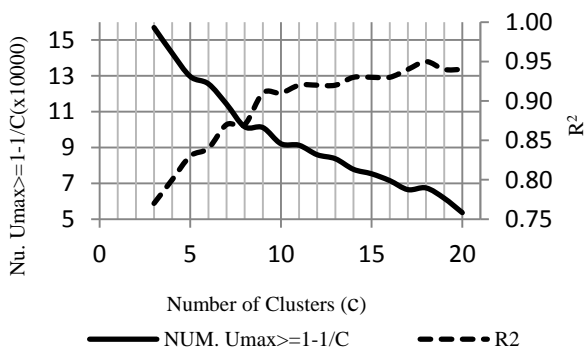


(الف)

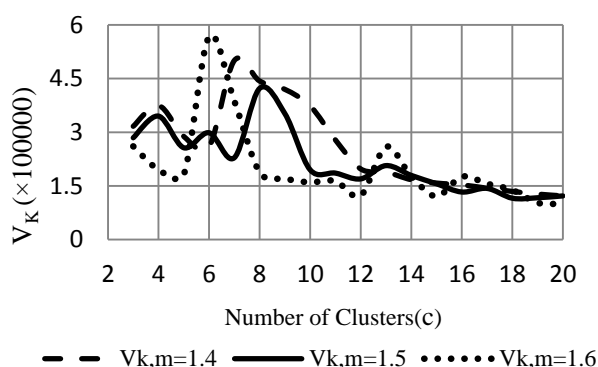


(ب)

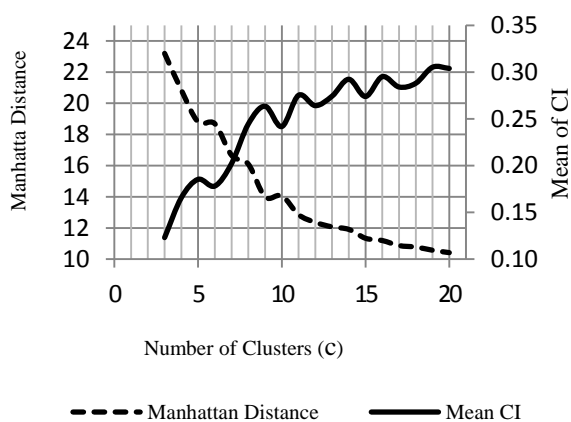
شکل ۴- نمایش گرافیکی، (الف) رابطه اعتبار سنجی Bezdek (VFPI) (ب) رابطه Fukuyama and Sugeno (VFS)



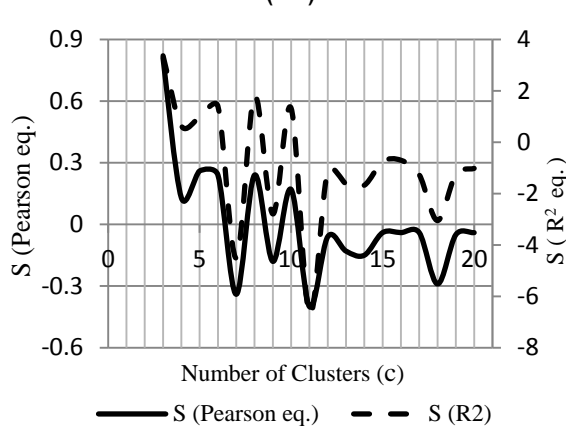
(الف)



(الف)



(ب)



(ب)

شکل ۶- (الف) اندرکنش تعداد سلول‌های با تابع عضویت مساوی یا بیشتر از  $1-1/c$  با  $R^2$  (ب) اندرکنش متوسط فاصله مانهاتان و متوسط شاخص CI

شکل ۵- (الف) نمایش گرافیکی رابطه اعتبارسنجی Kwon (ب) رابطه Liu and George (S) (VK)

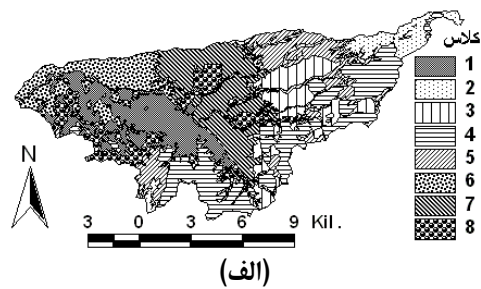
جدول ۴- مقایسه متوسط مقادیر واقعی هر خوشه با مراکز خوشه‌های نظیر ( $m=1/5$ )

C	U <sup>a</sup>	CI <sup>b</sup>	متوسط مقادیر واقعی			مقادیر مراکز خوشه‌ها			MARE(%)
			CN	DEM(m)	TPI	CN <sub>c</sub>	DEM <sub>c</sub> (m)	TPI <sub>c</sub>	
۱	۰/۸۴۴	۰/۲۷	۷۰/۰۲	۱۵۶۷	۰/۱۸۹	۶۹/۷۱	۱۵۶۶/۱	۰/۲۱۶	۴/۳
۲	۰/۸۰۱	۰/۳۳	۷۸/۴۳	۱۳۳۰	-۰/۰۹۵	۷۹/۲۰	۱۳۳۳/۷	-۰/۱۰۰	۲/۰
۳	۰/۹۲۸	۰/۱۴	۵۶/۰۳	۱۳۱۱	-۰/۰۱۲	۵۶/۰۰	۱۳۱۲/۳	-۰/۰۱۸	۱۰/۵
۴	۰/۸۱۹	۰/۳۱	۶۹/۷۵	۱۴۶۸	-۰/۰۰۸	۶۹/۶۰	۱۴۶۸/۹	-۰/۰۱۲	۱۱/۴
۵	۰/۸۴۳	۰/۲۷	۷۷/۵۵	۱۴۲۹	۰/۰۱۰	۷۷/۵۶	۱۴۳۱/۵	۰/۰۰۶	۱۷/۷
۶	۰/۸۵۵	۰/۲۵	۶۹/۴۶	۱۳۷۹	-۰/۰۷۶	۶۹/۱۸	۱۳۷۶/۹	-۰/۰۹۲	۵/۹
۷	۰/۹۴۶	۰/۱۰	۵۶/۰۱	۱۴۲۳	۰/۰۰۵	۵۶/۰۰	۱۴۲۲/۶	۰/۰۰۴	۷/۵
۸	۰/۸۲۱	۰/۳۰	۷۸/۶۷	۱۵۲۲	۰/۰۹۷	۷۸/۷۲	۱۵۳۵/۵	۰/۱۰۶	۳/۳

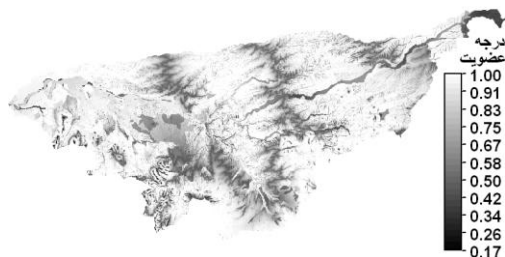
U<sub>a</sub> و CI<sub>b</sub> بترتیب متوسط درجه عضویت و شاخص ابهام هر خوشه و (MARE) درصد متوسط قدرمطلق خطای نسبی c معرف مقدار هر پارامتر در مرکز هر خوشه

## جدول (۵) خلاصه نتایج خوشه‌بندی در کل حوضه

مقدار	شاخص‌های تعیین کننده
۰/۸۷۵	ضریب همبستگی ( $R^2$ )
۰/۸۸۰	متوسط درجه عضویت کل حوضه (تمام)
۰/۶۹۰	متوسط درجه عضویت سلول‌ها در نواحی
۰/۲۴۰	متوسط شاخص ابهام (CI) برای تمام
۰/۵۴۰	متوسط شاخص ابهام (CI) در نواحی انتقالی



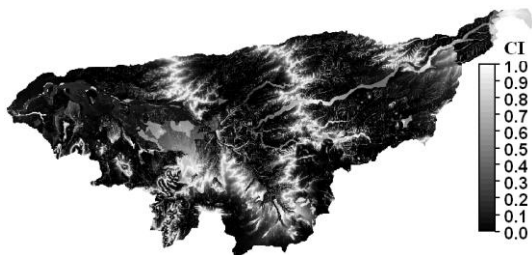
(الف)



(ب)



(ج)



(د)

شکل ۷- نقشه‌های (الف) - خوشه‌بندی و یا نواحی همگن  
(ب) - درجه عضویت سلول‌ها به خوشه‌های مربوطه (ج) -  
فاصله مانهاتان (د) - شاخص ابهام (CI) سلول‌ها

د- نشان داده شد که استفاده از روش‌های اعتبار سنجی موجود، هر چند بر اساس تئوری ریاضی و منطق فازی ارائه گردیده‌اند، ولی لزوماً در هر مدل فیزیکی و به ویژه طبیعی منجر به یک جواب بهینه واضح نخواهند شد. در حالی که روش به کار گرفته شده در مطالعه حاضر که بر مبنای اندرکنش معیارهای تصمیم‌گیری و کنترل بوده است، صراحت بیشتری در تعیین تعداد بهینه خوشه‌ها داشته است.

الف- استفاده از تکنیک PCA در موارد مشابه قابل توصیه می‌باشد. در این تکنیک، نتایج استخراجی آن کاملاً انعطاف‌پذیر بوده و تعیین معیار و یا مرز مشخص جهت انتخاب مولفه‌های اصلی و یا لایه‌های اصلی بستگی به تجربه کاربر و موضوع مورد مطالعه دارد. با استفاده از این تکنیک مشاهده گردید که انتخاب ۳ لایه اصلی (از بین ۹ لایه منتخب) یعنی: لایه عدد منحنی (CN)، لایه ارتفاع (DEM) و لایه شاخص موقعیت توپولوژیک (TPI) به عنوان لایه‌های موثرتر بر روی تولید رواناب دور از واقعیت نبوده و با پیشنهاد Dehotin and Braud (2008) مبنی بر حاکم بودن این لایه‌ها بر تولید رواناب همخوانی دارد.

ب- برای تعیین مقدار فازیت که یکی از پارامترهای تاثیرگذار در همپوشانی خوشه‌ها در نواحی مرزی می‌باشد، رابطه مشخصی وجود ندارد و به بازه‌ها و مقادیر پیشنهادی برای این ضریب نیز نمی‌توان اکتفا نمود. در این تحقیق برای تفکیک مورد نظر، مقدار فازیت ( $m$ ) در محدوده ۱/۷ - ۱/۸ قابل استفاده و بهترین مقدار آن ۱/۵ تعیین گردید.

ج- استفاده از درجه عضویت ماکزیمم برای تعیین خوشه برنده منطقی نبوده و به یک حد قابل قبول ( $\alpha$ -cut)، جهت کنترل نیاز می‌باشد که در این تحقیق استفاده از  $\alpha_{High} = 1 - 1/c$  مد نظر قرار گرفت. هر چند تأمین این حد برای تمامی سلول‌ها محقق نگردید، ولی تعداد بهینه خوشه‌ها به گونه‌ای تعیین شد که از طرفی متوسط درجه عضویت کل سلول‌ها برابر ۰/۸۸ (بیشتر از ۰/۸۷۵  $\alpha_{High}$ ) بدست آمد و از طرف دیگر متوسط شاخص ابهام (CI) به عنوان پارامتر کنترل‌کننده عدم قطعیت برابر ۰/۲۴ حاصل گردید که در بازه خیلی کم (۰/۲ - ۰/۴) قرار دارد. حتی در نواحی انتقالی نیز این مقادیر در حد کم (۰/۴ - ۰/۶) قرار گرفت.

## پی‌نوشت‌ها

1. Fuzzy C-Means
2. Walnut Gulch Experimental Watershed (WGEW)
3. Representative Elementary Area (REA)
4. Representative Elementary Watersheds (REWs)
5. Hydrologic Landscape Regions (HLRs)
6. Normalized Difference Vegetation Index (NDVI)
7. Topographic Position Index (TPI)
8. Mean Annual Rainfall (MAR)
9. Principal Component Analysis (PCA)
10. Confusion Index (CI)
11. Modal Distance
12. Kolmogorov Distance
13. Manhattan Distance

## ۷- مراجع

غضنفری، م. و رضائی، م. (۱۳۸۵) "مقدمه‌ای بر نظریه مجموعه‌های فازی" انتشارات دانشگاه علم و صنعت ایران، تهران

Bezdek, J. C. (1981), "Pattern recognition with fuzzy objective function algorithms", Plenum Press, New York, 220p.

Bruin, S. D. and Stein, A. (1998), "Soil-landscape modeling using fuzzy c-means clustering of attribute data derived from a Digital Elevation Model (DEM)", *Geoderma*, Vol. 183, No. 3-4, pp.17-33.

Burrough, P.A., Van Gaans, P.F.M and Hoostmans, R. (1997), "Continuous classification in soil survey: spatial correlation, confusion and boundaries", *Geoderma*, Vol. 77, No. 1, pp. 115- 135.

Chiu, W.Y. (2005), "Wetland mapping through semivariogram guided Fuzzy segmentation of multispectral imagery", M.Sc. Dissertation, 107p .

Dehotin, J., and Braud, I. (2008), "Which spatial discretization for which distributed hydrological models? Proposition of a methodology and illustration for medium to large-scale catchments", *Hydrology Earth System Science*, Vol. 112, No. 3, pp. 769-796.

Flügel, W.A. (1995), "Delineating Hydrological Response Units by Geographical Information System analyses for regional hydrological modeling using PRMS/MMS in the drainage basin of the river Bröl in Germany", *Hydrology Processes*, Vol. 19, No. 3-4, pp. 423-436.

Fukuyama, Y., and Sugeno, M. (1989), "A new method of choosing the number of clusters for the Fuzzy C-

ه- با مقایسه همزمان نقشه‌های خروجی (شکل ۷) به سهولت می‌توان به تفسیر نتایج و کنترل پیوستگی پهنه‌بندی (شناسایی نواحی همگن) پرداخت. در این شکل، همخوانی بسیار بالای نقشه‌های درجه عضویت و شاخص ابهام قابل توجه می‌باشد. به عبارتی، مناطقی که دارای درجه عضویت بالا می‌باشند، مقدار شاخص ابهام در آن نقاط بسیار پایین می‌باشد. علاوه بر آن، مقادیر زیاد شاخص ابهام غالباً در نواحی مرزی خوشه‌ها (نواحی انتقالی) و با عرض باریک مشاهده می‌گردند که موید خوشه‌بندی پیوسته در یک عرصه طبیعی است.

و- در نقشه فاصله (۷ ج)، تعداد سلول‌های با مقدار فاصله نسبتاً زیاد و به عبارتی مقدار آنها دو انحراف معیار بیش از میانگین مقدار فاصله می‌باشد، حدود ۵/۵٪ بوده و بیشتر در مناطقی با تغییرات زیاد نظیر خط‌الرس‌ها و یا محدوده رودخانه‌ها مشاهده می‌گردد. لذا اضافه کردن تعداد بیشتر خوشه در این مناطق کمک زیادی به کاهش عدم قطعیت خوشه‌بندی نخواهد کرد.

## ۶- نتیجه‌گیری

هرچند در یک فرآیند هیدرولوژیک، پارامترها و یا متغیرهای زیادی می‌توانند دخیل باشند، ولی وابستگی آنها به یکدیگر در طبیعت، امکان ارائه و یا توسعه مدل‌های هیدرولوژیک با پارامترهای کمتر و کاربردی‌تر را برای محققان فراهم می‌سازد و بعضاً موجب سهولت در تفکیک نواحی همگن هیدرولوژیک نیز خواهد شد. از سویی دیگر با وجود پیوستگی در طبیعت، استفاده از روش‌های فازی جهت هرگونه تفکیک و یا پهنه‌بندی عرصه‌های طبیعی همخوانی بین تفکیک شونده و تفکیک کننده را برجسته می‌کند. زیرا در روش‌های فازی، درجه تعلق (عضویت) هر سلول به خوشه اختصاص یافته با مقداری بین ۰ و ۱ تعیین می‌گردد که این مقادیر بخوبی می‌توانند پیوستگی در مقدار درجه عضویت سلول‌های مجاور را به نمایش گذارد. در چنین مواردی چنانچه سودمندی روش‌های فازی را باور داشته باشیم، بایستی به این نکته نیز توجه نمائیم که در این روش‌ها جهت تعیین تعداد بهینه خوشه‌ها، روابط اعتبارسنجی گوناگونی توسط محققان ارائه و یا توسعه یافته‌اند که ممکن است هر یک از آنها در شرایط خاص و یا نوع استفاده موردنظر آنان پیشنهاد گردیده و در هر شرایطی منتج به نتایج مطلوب کاربر نگردند. لذا در این مواقع، بایستی کاربر حتی‌المقدور تلاش نموده تا تعداد بهینه خوشه‌ها را به نحو منطقی بدست آورده و از انتخاب روش دیگری برای خوشه‌بندی (تفکیک و یا پهنه‌بندی) اجتناب نماید.

- Reggiani, P., Sivapalan, M., Hassanizadeh, S. M., and Gray, W. G. (1999) "A unifying framework for watershed thermodynamics: constitutive relationships", *Advances in Water Resources*, Vol. 23, No. 1, pp. 15–39.
- Reggiani, P., Sivapalan, M., and Hassanizadeh, S. M. (2000), "Conservation equation governing hillslope response", exploring the physical basis of water balance", *Water Resources Research*, Vol. 36, No. 7, pp. 1845–1863.
- Robbez-Masson, J. M. (1994), "*Reconnaissance et de limitation de motifs d' organization spatiale. Application à la cartographies des p'edopaysages*", PhD. Dissertation, Ecole Nationale Sup'erieure Agronomique de Montpellier, 191p.
- Smith, R. E., Goodrich, D. C., Woolhiser, D. A., and Unkrich C. L. (1995), "KINEROS-A kinematic runoff and erosion model", *Water Resources Publications*, Vol. 20, pp. 697–732.
- Weiss, A., (2001), "Topographic Position and Landforms Analysis", *Poster presentation, ESRI User Conference*, San Diego, CA.
- Wolock, D. M., Winter, T. C., and McMahon, G. (2004), "Delineation and Evaluation of Hydrologic-Landscape Regions in the United States Using Geographic Information System Tools and Multivariate Statistical Analyses", *Environmental Management* Vol. 34, No. 1, pp. S71–S88.
- Wood, E. F., Sivapalan, M., Beven K., and Band L. (1988), "Effects of spatial variability and scale with implication to hydrologic modeling", *Journal of Hydrology*, Vol. 102, No. 1-4, pp. 29–47.
- Xie, X., and Beni, G. A. (1991), "Validity measure for fuzzy clustering", *IEEE Transactions on Pattern Analysis and machine Intelligence*, Vol. 13, No. 8, pp. 841–847.
- Means method", *Proceedings of Fifth Fuzzy Systems Symposium*, pp. 247–250, In Japanese.
- Heilman, P., Nichols M. H., Goodrich, D. C., Miller, S. N., and Guertin, D. P. (2008), "Geographic information systems database, Walnut Gulch Experimental Watershed, Arizona, United States", *Water Resources Research*, Vol. 44, W05S11, pp. 1-6.
- Kull, D. W., and Feldman A. D. (1998), "Evolution of Clark's unit hydrograph method to spatially distributed runoff", *Journal of Hydrology Engineering*, Vol. 13, No. 131, pp. 9–19.
- Kwon, S. H. (1998), "Cluster validity index for fuzzy clustering Analysis", *Electronics Letters*, Vol. 34, No. 22, pp. 2176–2177.
- Lagacherie, P., Robbez-Masson, J. M., Nguyen-The, N. and Barthes, J. P. (2001), "Mapping of reference area representativity using a mathematical soilscape distance", *Geoderma*, Vol. 101, No. 3–4, pp. 105–118.
- Liu, Z., and George R. (2005), "Mining weather data using fuzzy cluster", *Fuzzy modeling with spatial information for geographic problems*, New York, pp. 105-115.
- Pal N. R., and Bezdek, J. C. (1995), "On cluster validity for the fuzzy c-means model", *IEEE Transactions on Fuzzy systems*, Vol. 13, No. 3, pp. 370–379.
- Rao, R. A., Srinivas, V. V. (2007), "*Regionalization of Watersheds*", Springer, USA, 248p.
- Reggiani, P., Sivapalan M., and Hassanizadeh, S. M. (1998) "A unifying framework for watershed thermodynamics: balance equations for mass, momentum, energy and entropy, and the second law of thermodynamics", *Advances in Water Resources*, Vol. 22, No. 4, pp. 367–598.