

برتری روش خوشه‌بندی C- میانگین فازی در بیان توزیع رخساره‌های هیدروشیمی آب زیرزمینی دشت ورامین

محمد نخعی^(۱)، مهدی تلخایی^(۲) و میثم ودیعتی^(۳)

۱. دانشیار هیدروژئولوژی، دانشکده علوم زمین، دانشگاه خوارزمی تهران

۲. دانش آموخته هیدروژئولوژی، دانشکده علوم زمین، دانشگاه خوارزمی تهران

۳. دانشجوی دکتری هیدروژئولوژی، دانشکده علوم زمین، دانشگاه تبریز

تاریخ دریافت: ۹۳/۷/۵

تاریخ پذیرش: ۹۳/۱۲/۲۸

چکیده

در پژوهش حاضر، خوشه‌بندی مجموعه‌ای از داده‌های هیدروشیمی دشت ورامین با استفاده از روش‌های خوشه‌بندی C- میانگین فازی (FCM) و تحلیل خوشه سلسله مراتبی (HCA) انجام شده و کاربرد آن‌ها در تغییرات رخساره‌های هیدروشیمی بحث گردید. نمونه‌های آب زیرزمینی با استفاده از بهینه کردن تعداد خوشه و درجه فازی‌شدگی با استفاده از روش C- میانگین فازی به سه گروه طبقه‌بندی شدند. از داده‌های آب زیرزمینی ۹۰ نمونه چاه عمیق و نیمه عمیق و ۹ متغیر هیدروشیمی منطقه مورد مطالعه استفاده شد. نتایج این دو روش، مراکز خوشه را تولید می‌کند که در تشخیص فرایندهای فیزیکی و شیمیایی تغییرات هیدروشیمی منطقه مورد مطالعه مؤثر است. در روش FCM تعداد خوشه بهینه توسط توابع بهینه‌یابی تعیین می‌شود اما در روش HCA براساس تجربه کاربر و سعی و خطا تعیین می‌شود. روش FCM روشی مناسب در تحلیل داده اکتشافی در بیان توزیع رخساره‌های هیدروشیمی است و زمانی که خوشه‌های پیوسته یا دارای هم‌پوشانی وجود دارند، ابزار بهتری نسبت به HCA برای خوشه‌بندی است. با ترسیم خطوط تراز مقدار عضویت هر خوشه که بر روی نقشه به صورت مکانی و پیوسته نشان داده شده، خوشه‌های نمونه‌های آب زیرزمینی به خوبی مشخص شده است. نتایج نشان داد؛ روش FCM در تحلیل داده‌های مرزی، نسبت به روش HCA که تغییراتی واضح و ناگهانی دارد؛ توانا تر است.

واژه‌های کلیدی: آب زیرزمینی، رخساره هیدروشیمی، دشت ورامین، خوشه‌بندی، منطق فازی.

مقدمه

از چندین پارامتر استفاده می‌کنند تا تحلیل مناسب‌تری از روابط بین متغیرها داشته باشند. در مطالعات هیدروشیمی، گروه‌بندی به منظور مجزا کردن داده‌ها به خوشه‌های معرف (که به عنوان گروه‌های آب یا رخساره‌های هیدروشیمی نیز شناخته می‌شوند) به کار می‌رود تا فرایندهای طبیعی

روش‌های آماری چند متغیره روشی مناسب برای تجزیه و تحلیل نمونه‌های آب زیرزمینی و ویژگی‌های هیدروشیمی است (Singh et al., 2004). این روش‌ها به صورت هم‌زمان

* نویسنده مرتبط nakhaei@khu.ac.ir

متغیرهای زبانی استفاده می‌کنند. جالب اینجاست که اگرچه سیستم‌های فازی، پدیده‌های غیردقیق و نامشخص را توصیف می‌کنند؛ با این حال تئوری فازی یک تئوری دقیق است. متغیر زبانی^۴، متغیری است که مقادیرش کلمات یا جملات یک‌زبان طبیعی باشد. به‌عنوان نمونه سن یک فرد را در نظر بگیرید؛ اگر مقادیری را که سن اختیار می‌کند با کلماتی مثل نونهال، نوجوان، جوان، مسن و پیر نشان دهیم؛ متغیر سن یک متغیر زبانی است (آذر و فرجی، ۱۳۸۶).

در نظریه کلاسیک، تابع عضویت مجموعه‌ای است که عدد یک در مرز آن و صفر بیرون آن قرار می‌گیرد. برای یک مجموعه فازی، تابع عضویت به‌صورتی تعریف می‌گردد که در فاصله بین صفر تا یک است (رابطه ۱) (تشنه‌لب و همکاران، ۱۳۸۷).

رابطه (۱)
$$\mu_A: X \rightarrow [0, 1]$$
 مجموعه A برحسب تابع عضویت عبارت است از رابطه (۲):

رابطه (۲)
$$A = \{(\mu_A(X)), x \in X, \mu_A(X) \in [0, 1]\}$$
 مهم‌ترین ویژگی منطق فازی در مقایسه با منطق کلاسیک این است که دانش و تجربه بشر را می‌تواند در قالب روابط ریاضی بیان نماید. این مهم باعث شده است که مسائل دنیای واقعی را به‌خوبی بتوان با استفاده از آن مدل‌سازی کرد (تشنه‌لب و همکاران، ۱۳۸۷).

در روش‌های خوشه‌بندی، تعداد و نوع معیارها ثابت و مشخص نیست، از این‌رو تلاش خواهد شد تا با تغییر آن‌ها به بهترین حالت ممکن دست پیدا کرد. به‌طور کلی خوشه‌بندی عبارت است از این‌که یک مجموعه مرجع مانند X شامل n داده به c دسته یا خوشه تقسیم‌بندی شود به‌طوری‌که دسته‌های مزبور دارای خصوصیات مشابهی باشند ($2 \leq c \leq n$). خوشه‌بندی را می‌توان به‌صورت فازی و غیر فازی انجام داد. الگوریتم‌های مختلفی برای خوشه‌بندی فازی و غیرفازی توسط محققین مختلف پیشنهاد شده است (Bezdek, 1981).

تغییردهنده پارامترهای هیدروشیمی تشخیص داده شود. این خوشه‌های معرف که به تعیین روندهای اصلی شیمیایی کمک می‌کنند؛ می‌توانند برای بازسازی فرایندهای اصلی مورد استفاده قرار گیرند (Barbieri et al., 2001).

روش‌های خوشه‌بندی به دو شاخه اصلی تقسیم می‌شوند. روش‌های سخت^۱ یا کلاسیک که در آن یک عنصر به‌طور کامل به یک گروه تعلق دارد و روش‌های نرم^۲ یا فازی که در آن یک عنصر به تمام خوشه‌ها با درجه عضویت‌های مختلف تعلق دارد. هدف اصلی این روش‌ها چه سخت و چه فازی تفکیک یک ماتریس داده با n نمونه و p متغیر به c زیرگروه همگن به‌وسیله دسته‌بندی دقیق نمونه‌های مرتبط به خوشه‌های مشخص است. در نتیجه تقسیم‌بندی، نمونه‌های درون یک خوشه یکسان به‌وسیله درجه بالایی از تشابه مشخص می‌شوند، درحالی‌که نمونه‌های متعلق به خوشه‌های مختلف با درجه بالایی از عدم تشابه مشخص می‌شوند. سپس این خوشه‌های همگن توسط ماتریس $c \times p$ بعدی نمایش داده می‌شوند که متشکل از c نمونه شاخص نماینده هر خوشه است (Guler and Thyne, 2004).

فرآیندهای اصلی، همیشه خروجی‌های دقیقی تولید نمی‌کنند و در سیستم‌های هیدروشیمی و هیدرولوژیکی پیچیده به‌سختی تشخیص داده می‌شوند. خصوصیات فیزیکی و شیمیایی در سیستم‌های طبیعی اغلب نه ناگهانی بلکه به‌طور پیوسته تغییر می‌کنند. به‌واسطه این پیوستگی، خوشه‌های آماری نمی‌توانند جداکننده خوبی باشند و بایستی توالی خوشه‌های هم‌پوشانی شکل گیرد (Guler et al., 2002; Goyal and Gupta, 2014).

منطق فازی^۳ برای اولین بار به همین نام توسط Zadeh (1965) ارائه شد و بسیاری از مفاهیم، متغیرها و سیستم‌های مبهم را صورت‌بندی ریاضی بخشید. بدین ترتیب زمینه را برای استدلال، استنتاج، کنترل و تصمیم‌گیری در شرایط عدم اطمینان فراهم آورد (طاهری، ۱۳۷۸). یکی از پرکاربردترین روش‌های حل مسائل خوشه‌بندی فازی روش c - میانگین فازی است (Bezdek, 1981).

قوانین فازی، قوانین ساده و آشکار برای توصیف پاسخ‌دهی مطلوب سیستم به‌جای فرمول‌های ریاضی از

1. Hard
2. Soft
3. Fuzzy logic
4. Linguistic term

خوشه‌ها بیشتر و مرز رخساره‌ها به‌صورت تدریجی و نه ناگهانی است که با واقعیت هماهنگی بیشتری دارد. این روش با موفقیت در دیگر علوم از جمله زیست‌شناسی (Zhang et al., 1995)، در هواشناسی (Moore and McBratney, 1985)، در زمین‌شناسی (Rantitsch, 2004; Burrough et al., 2000). در این پژوهش داده‌های برگرفته از آنالیز نمونه آب‌های زیرزمینی آبخوان ورامین با استفاده از روش‌های خوشه‌بندی C- میانگین و آنالیز خوشه سلسله مراتبی^۲ (HCA) دسته‌بندی می‌شوند که کاربرد این روش‌ها در شرح تحولات هیدروشیمی مورد بحث و بررسی قرار می‌گیرد. با توجه به این که تغییرات در میزان عناصر شیمیایی آب‌های زیرزمینی به‌طور پیوسته می‌باشند، استفاده از روشی که این تغییرات را به‌خوبی نمایش دهد و میزان عضویت هر نمونه آب را به گروه‌های مختلف بیان کند؛ ضروری به‌نظر می‌رسد.

مواد و روش تحقیق

منطقه مورد مطالعه در استان تهران واقع شده و از شمال به ادامه سلسله البرز و لواسانات، از شرق به ایوانکی و گرمسار، از جنوب به تپه ماهورهای جنوبی دشت ورامین و همچنین دریاچه نمک و از مغرب به دشت تهران-کرج محدود می‌شود. بیشترین شیب دشت برابر ۹/۳۲ درصد است. بیشترین ارتفاع در بخش شمالی آن برابر ۱۰۲۰ متر بوده و کمترین آن در جنوب و جنوب شرقی دشت و ۸۲۰ متر از سطح دریا است. از کل منطقه مورد مطالعه، حدود ۲۱/۴ کیلومترمربع را نمک‌زار و حدود ۰/۶۸ کیلومترمربع را زمین‌های کشاورزی تشکیل می‌دهند. میانگین ۲۵ ساله درجه حرارت سالانه، ۱۶ درجه سانتی‌گراد و متوسط بارندگی سالانه، ۱۷۲ میلی‌متر در ایستگاه سینوپتیک شهرستان ورامین است. از نظر زمین‌شناسی، این منطقه به‌واسطه فروافتادگی دوران اول و دوم ایجاد شده که داخل آن را رسوبات دوران سنوزوئیک و رسوبات جدید رودخانه جاجرود پر کرده است (آقنابتی، ۱۳۸۵). واحدهای تشکیل‌دهنده منطقه مورد مطالعه به‌ترتیب سن از جدید به قدیم شامل رسوبات دوران کواترنری، کنگلومرا، شیل،

در خوشه‌بندی قطعی، هر نمونه ورودی متعلق به یک و فقط یک خوشه است؛ بنابراین نمی‌تواند عضوی از چندین خوشه باشد؛ به‌عبارت‌دیگر خوشه‌ها همپوشانی ندارند. درحالی‌که در خوشه‌بندی فازی یک نمونه می‌تواند متعلق به بیش از یک خوشه باشد. خوشه‌بندی فازی به کشف مدل فازی داده‌های موجود می‌پردازد. روش خوشه‌بندی فازی توسط Bezdek (1981) پیشنهاد شد که روشی کارآمد برای خوشه‌بندی فازی داده‌هاست و در واقع شکل گسترش یافته‌ای از خوشه‌بندی C- میانگین سخت است.

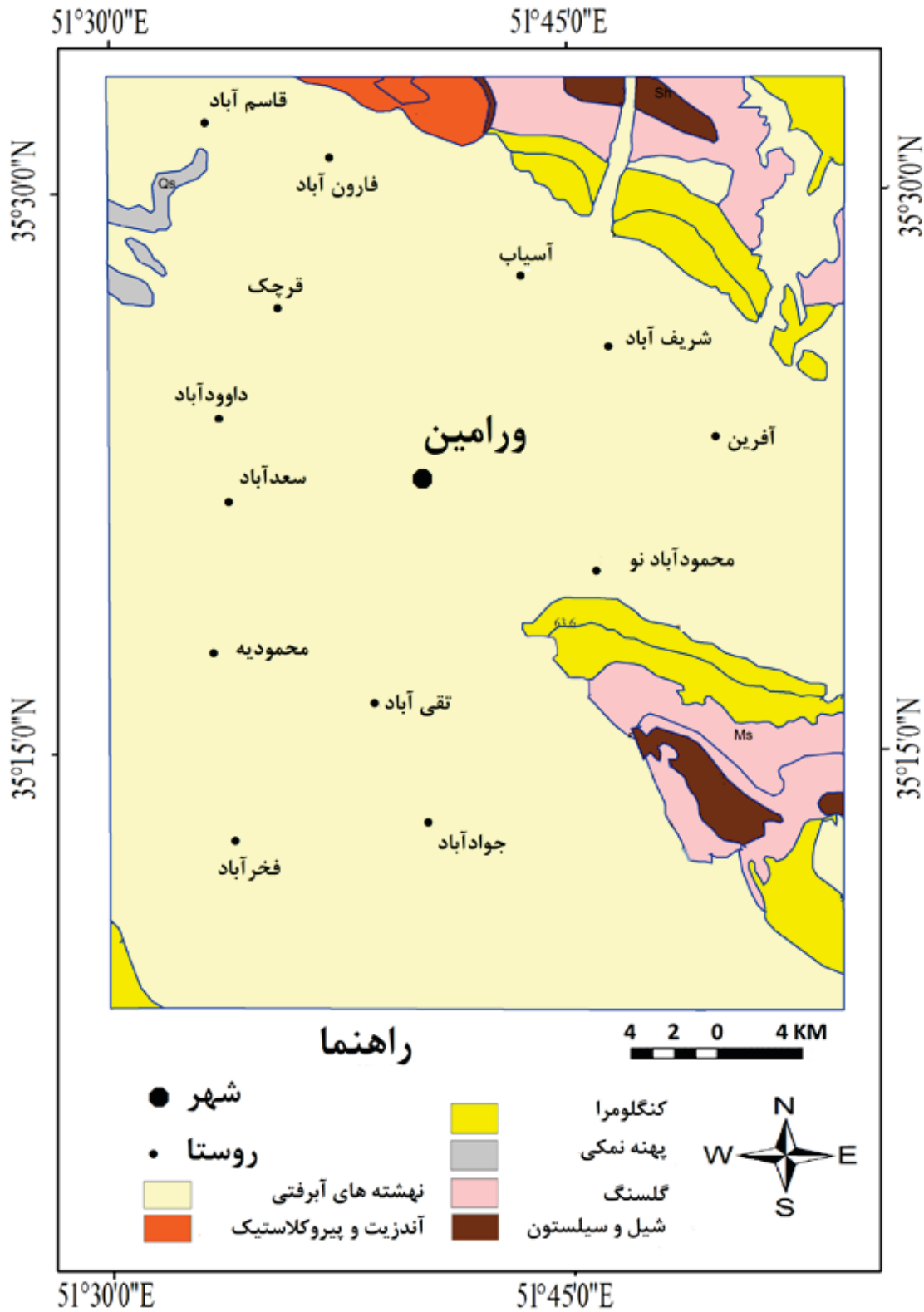
در شاخه‌های مختلف علوم زمین، مطالعات متنوعی در زمینه خوشه‌بندی فازی انجام شده است. برای مثال شکاری و باقرنژاد (۱۳۸۴) به‌منظور طبقه‌بندی خاک از روش فازی استفاده کردند. در این پژوهش، تعداد خوشه بهینه تعیین شد و نتایج آن بررسی و مقایسه شد. این پژوهش به‌صورت مطالعه موردی برای منطقه چشمه سفید کرمانشاه انجام شد. در مطالعه دیگری سلطانی محمدی و همکاران (۱۳۸۶) در خنثی‌سازی اثر سنژنتیک (مربوط به تغییرات لیتوژئوشیمیایی) داده‌های ژئوشیمیایی رسوبات آبراه‌ای در منطقه شمال تکاب از روش خوشه‌بندی فازی استفاده کردند. در زمینه مدل‌سازی نیاز اکسیژن شیمیایی توسط روش ترکیبی شبکه عصبی مصنوعی، عصبی فازی و خوشه‌بندی k میانگین نیز مطالعاتی صورت گرفته است (Ay and Kisi, 2014)

Guler و همکاران (۲۰۰۲) از روش‌های آماری چند متغیره به‌منظور طبقه‌بندی داده‌های شیمیایی کیفیت آب استفاده نمودند و به بررسی کیفیت آب زیرزمینی منطقه‌ای در ترکیه با استفاده از این روش پرداختند. در پژوهش دیگری در زمینه خوشه‌بندی C- میانگین فازی، به تخمین زمین لغزش ناشی از بارش با استفاده ترکیبی از روش شبکه عصبی و خوشه‌بندی C- میانگین فازی پرداختند (Alimohammadlou et al., 2014).

Guler and Thyne (2004) با استفاده از داده‌های هیدروشیمی آب زیرزمینی منطقه‌ای در جنوب کالیفرنیا نشان دادند که روش خوشه‌بندی C- میانگین فازی (FCM) نتایج بهتری نسبت به روش‌های دیگر دارد. در روش FCM تعداد

1. Fuzzy C-Mean clustering (FCM)
2. Hierarchical clustering analysis

سیلتستون، گل‌سنگ، آهک، مارن، گدازه‌های آندزیتی و گدازه‌های پیروکلاستی می‌شوند. در شکل ۱، نقشه و گدازه‌های پیروکلاستی آورده شده است.



شکل ۱. نقشه سنگ‌شناختی منطقه مورد مطالعه (برگرفته از سازمان زمین‌شناسی کشور، نقشه یک صد هزارم ورامین، ۱۳۸۵)

$$\sum_{i=1}^c U_{ik}^m = 1, 1 \leq k \leq n \quad \text{رابطه (۴)}$$

محدودیت دوم در ماتریس M در این است که مجموع ضرایب مرکز هر خوشه باید کمتر از تعداد اجزاء باشد (رابطه ۵):

$$\sum_{k=1}^n U_{ik} < n, 1 \leq i \leq c \quad \text{رابطه (۵)}$$

در نتایج FCM ممکن است نمونه‌ها ۱۰۰ درصد عضو خوشه نباشند؛ در عوض عضویت نمونه‌ها بین خوشه‌ها تقسیم می‌شود.

یکی از مشکلات روش‌های خوشه‌بندی، تعیین تعداد خوشه‌ها است. بعضی از روش‌های اجتناب از این مشکل، استفاده از دانش فردی یا روش‌های آماری است. در این پژوهش، تعداد خوشه‌های بهینه با استفاده از روابط ۶ و ۷ حاصل می‌شود. شاخص FPI درجه (رابطه ۶) فازی‌شدگی را برای تعداد خوشه‌های مشخص نشان می‌دهد (کوره‌پزان دزفولی، ۱۳۸۷):

$$F = \frac{1}{n} \sum_{i=1}^c \sum_{k=1}^n (U_{ik})^2 \quad \text{FPI} = 1 - \frac{cF-1}{c-1} \quad \text{رابطه (۶)}$$

شاخص مهم دیگری که درجه به هم‌ریختگی را برای تعداد گروه‌های مشخص نشان می‌دهد؛ NCE (رابطه ۷) است (کوره‌پزان دزفولی، ۱۳۸۷):

$$F = \frac{1}{n} \sum_{i=1}^c \sum_{k=1}^n U_{ik} \times \log(U_{ik}) \quad \text{NCE} = \frac{H}{\log c} \quad \text{رابطه (۷)}$$

که در آن، H تابع انتروپی^۱ است. تعداد خوشه بهینه با کمینه کردن این دو شاخص به دست می‌آید.

در خوشه‌بندی به روش فازی، تعداد خوشه، نوع فاصله، معیار توقف و پارامتر فازی‌شدگی بایستی در شروع تحلیل معین شد. به همین منظور، فاصله اقلیدسی به عنوان فاصله انتخاب شد؛ داده‌ها استاندارد شدند تا وزن یکسانی به تمام متغیرها تعلق گیرد و مقدار توقف 0.0001 تعیین شد. خوشه‌بندی برای c خوشه (بین ۳ و ۱۰) و با استفاده از مقادیر m (۳، ۱۰/۱۵، ۱/۱، ...) و برای ترکیب‌های c و m در مجموع ۳۱۲ بار اجرا شد. برای انتخاب تعداد گروه بهینه از توابع (FPI, NCE) استفاده می‌شود. برخی نتایج در شکل ۲ آمده است که اغلب نمودارها الگوی همسانی را نشان می‌دهند. تعداد خوشه بهینه در اغلب نمودارها ۳ خوشه به دست آمده است.

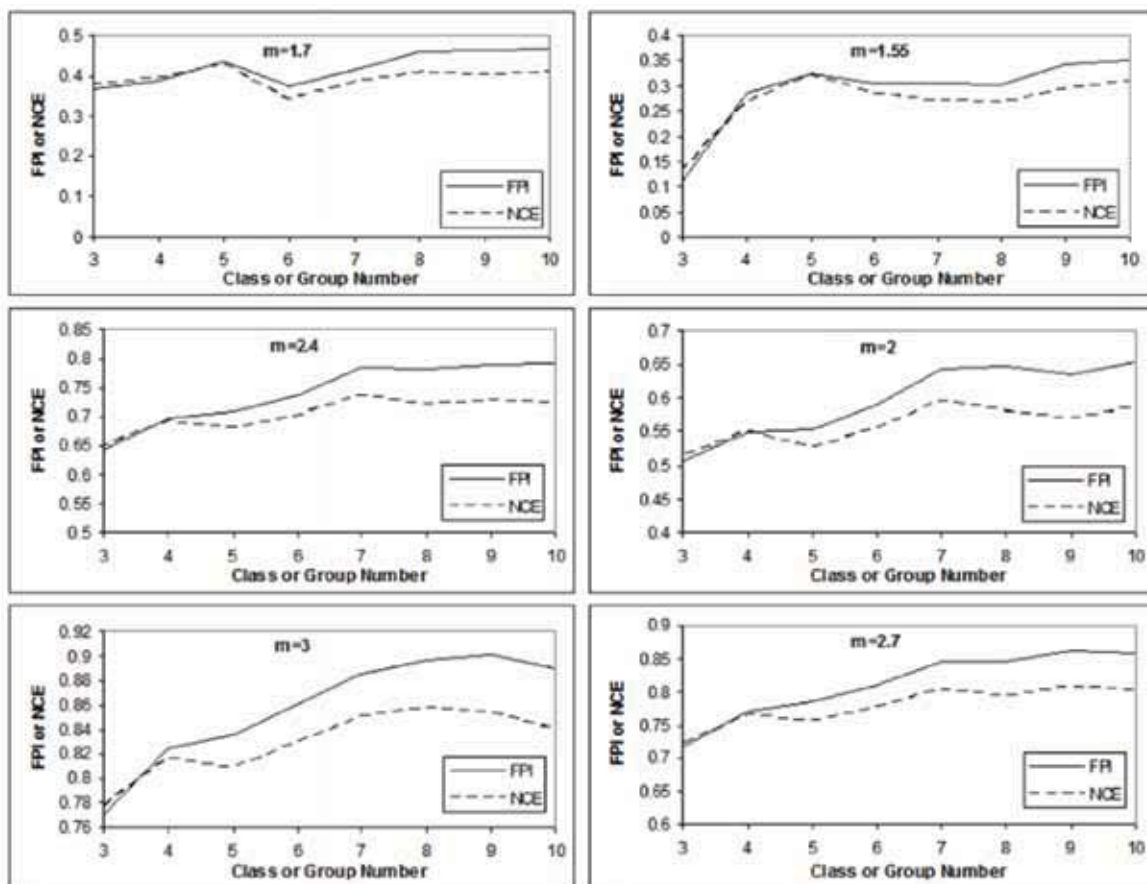
در این پژوهش، از میان متغیرهای اندازه‌گیری شده در ۹۰ نمونه چاه آب زیرزمینی دشت ورامین که توسط آب منطقه‌ای استان تهران انجام شده است، متغیرهای بی‌کربنات، سولفات، کلراید، کلسیم، منیزیم، سدیم، پتاسیم، اسیدپت و هدایت الکتریکی استفاده شدند. خطای موازنه بار یونی داده‌ها کمتر از ۰.۵٪ است که برای ارزیابی کیفیت آب زیرزمینی مناسب است. ماتریس داده‌های ورودی برای روش FCM در محدوده $+3$ تا -3 با میانگین صفر استاندارد شدند و مشکل مقایسه متغیرها با واحدها مختلف رفع شد.

الگوریتم خوشه‌بندی FCM روش تحلیل داده چند متغیره است. این الگوریتم، مجموعه‌ای از داده‌ها (فضای چندبعدی اقلیدسی (p)) را به خوشه فازی (C) تقسیم می‌کند که به وسیله مراکز خوشه‌هایشان (V_i) تعیین می‌شوند. جدا کردن داده‌ها به خوشه‌های فازی با بهینه کردن تابع هدف در رابطه ۳ به دست می‌آید:

$$J_{FCM} = (M, C) = \sum_{i=1}^c \sum_{k=1}^n U_{ik}^m \|X_k - V_i\|^2 \quad \text{رابطه (۳)}$$

در رابطه ۳، M ماتریس اعضاء، C ماتریس مراکز خوشه‌ها، c تعداد خوشه‌ها، n تعداد داده‌ها و U_{ik}^m درجه عضویت داده k در خوشه i است. اگر فاصله اقلیدسی برای هر یک از نمونه‌های آب زیرزمینی (فاصله خط مستقیم بین دو نقطه در فضای p بعدی که به وسیله p متغیر تعریف می‌شود) بین داده X_k که شامل ماتریسی از داده‌های ۹ متغیر آب زیرزمینی نمونه‌برداری شده است و مرکز خوشه V_i بالا باشد؛ J_{FCM} کم می‌شود (کوره‌پزان دزفولی، ۱۳۸۷). اگر فاصله کم باشد؛ مقدار عضویت به یک میل می‌کند. پارامتر $m \in (1, \infty)$ پارامتر وزنی است که درجه فازی بودن نتایج طبقه‌بندی را کنترل می‌کند که درجه هم‌پوشانی بین خوشه‌ها است. با حداقل مقدار معنی‌دار $m=1$ پاسخ، تقسیم‌بندی از نوع غیرفازی است. با میل m به مقدار بی‌نهایت، پاسخ به بالاترین درجه فازی‌شدگی میل می‌کند (Bezdek, 1981). انتخاب مقدار $m=2$ در بسیاری حالات به عنوان انتخاب مناسب پارامتر فازی‌شدگی پذیرفته شده است (Hathaway and Bezdek, 2001). ماتریس M به اجزاء $[0, 1]$ محدود می‌شود (رابطه ۴):

1. Entropy function



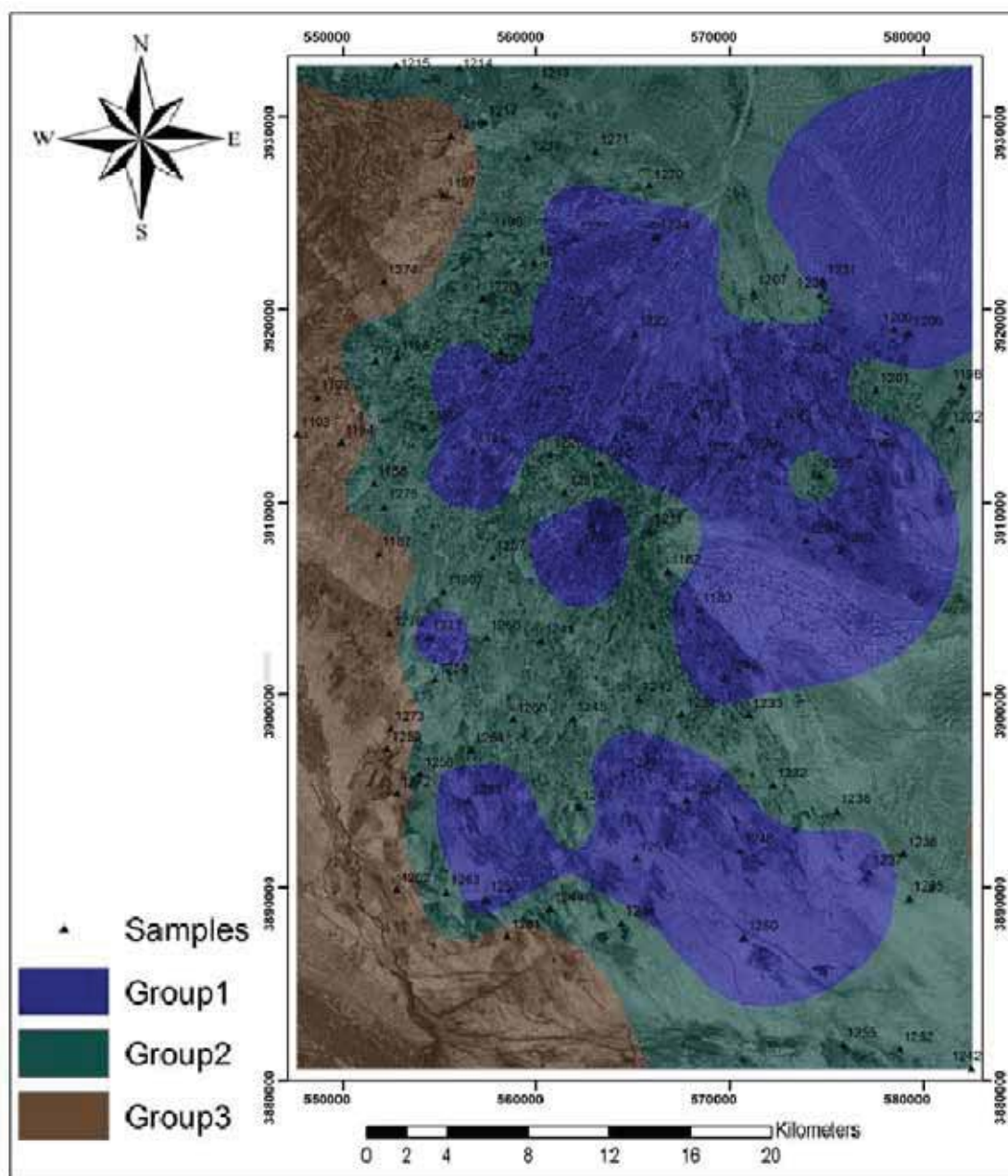
شکل ۲. نمودار مقادیر شاخص FPI و NCE در برابر تعداد خوشه

نتایج و بحث

انتخاب خوشه‌ها و فاصله اقلیدسی به‌عنوان معیار فاصله انتخاب شد. اول این‌که داده‌های مورد استفاده در این مطالعه از نوع داده‌هایی با مقیاس فاصله‌ای است (داده‌های استاندارد شده) و دلیل دوم این است که روش وارد به لحاظ ساختار خوشه‌بندی شبیه به روش c- میانگین است و این موضوع باعث می‌شود، اختلافات ساختاری دو روش تأثیری در مقایسه خروجی‌ها نداشته باشند. شکل ۳ نتایج روش HCA را نشان می‌دهد تا روند تغییرات بهتر نمایش داده شود.

فرآیندهای هیدروژنولوزیک، بسیار پیچیده هستند، از این‌رو نمی‌توان با استفاده از روش‌های مرسوم به بررسی کیفیت آب زیرزمینی در شرایط هیدروژنولوزیک پیچیده پرداخت و محققین در مطالعات هیدرولوژی و زمین‌شناسی به روش‌های جدید و دقیق‌تری روی آورده‌اند. یکی از روش‌های پرکاربرد در زمینه ارزیابی کیفیت آب و بررسی روند تغییرات آب در آبخوان، روش‌های آماری چندمتغیره و روش‌های مختلف خوشه‌بندی است. خوشه‌بندی تجمعی برای این پژوهش انتخاب شد و روش وارد^۱ به دو دلیل برای

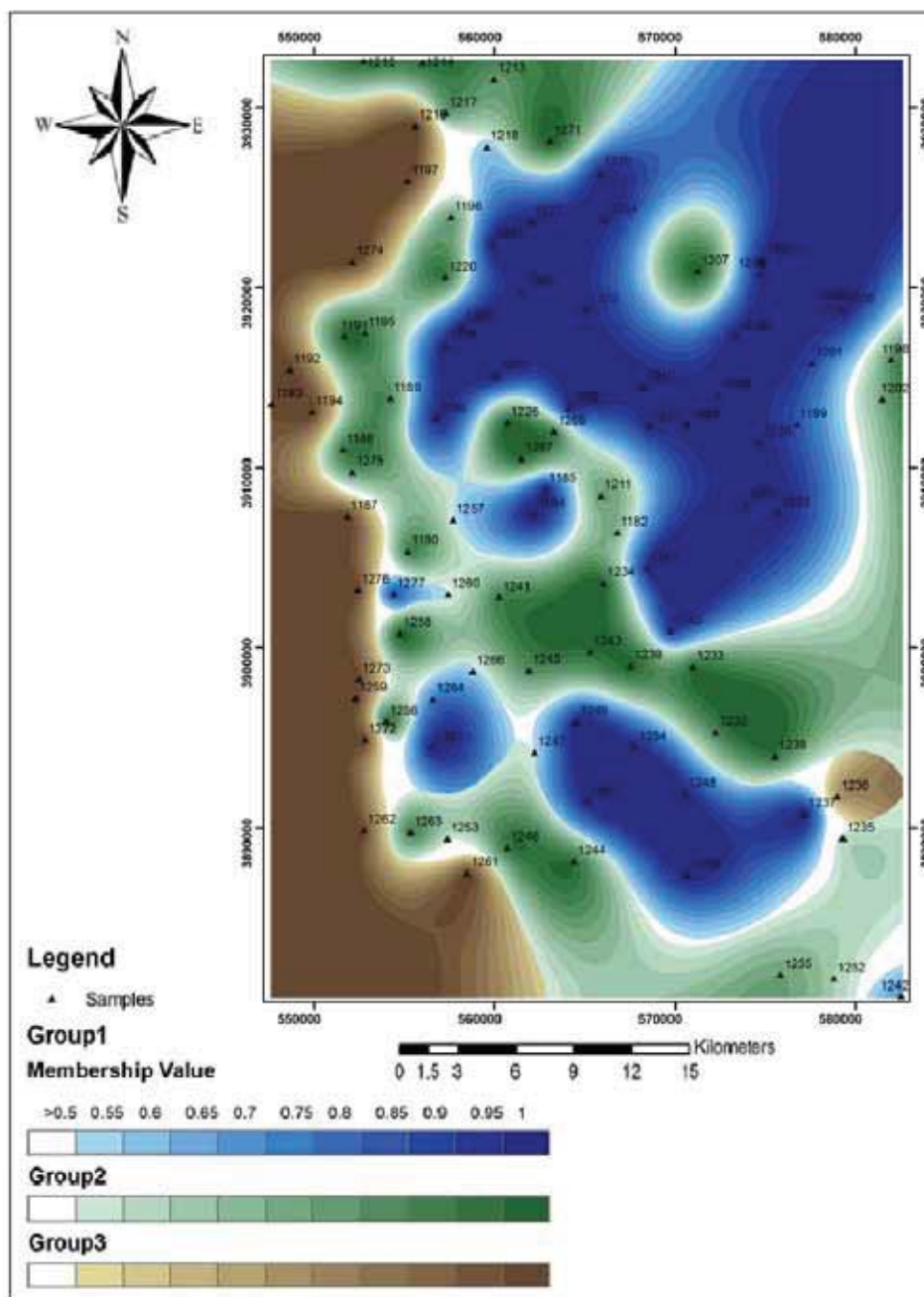
1. Ward method



شکل ۳. توزیع مکانی رخساره‌های هیدروشیمی به روش HCA

این شکل برای نمایش بهتر تغییرات عضویت در خوشه‌ها از نقشه منطقه صرف نظر شده است. مناطقی که با رنگ سفید در شکل مشخص شده است مربوط به نمونه‌هایی است که متعلق به هیچ یک از خوشه‌ها نیستند؛ یعنی مناطقی که مقدار عضویت آن‌ها به هر یک از گروه‌ها کمتر از ۰/۵ است.

همان‌طور که در شکل ۳ دیده می‌شود؛ مرز بین گروه‌ها در این روش کاملاً واضح و ناگهانی تغییر می‌کند، به طوری که یک نمونه یا به یک گروه تعلق دارد یا خیر. در شکل ۴ نتایج روش FCM نشان داده شده تا روند تغییرات بهتر نمایش داده شود. تغییرات مقادیر عضویت در بازه ۱ و ۰/۵ در خوشه‌های مختلف نشان داده شده است. در



شکل ۴- توزیع مکانی رخساره‌های هیدروشیمی به روش FCM

بین تفسیر کمی و کیفی داده‌های آب زیرزمینی در روش FCM نسبت به روش HCA وجود دارد. همچنین در روش HCA تعداد خوشه به صورت سعی و خطا تعیین می‌شود؛ در حالی که در روش FCM توسط توابع هدف (روابط ۶ و ۷) مشخص می‌گردد. همان‌طور که در شکل ۴ مشخص است، توزیع مکانی خوشه‌های FCM ارتباط نزدیکی با رخساره‌های هیدروشیمی دارد. نتایج حاصل از خوشه‌بندی داده‌ها سه

الگوریتم خوشه‌بندی FCM، روشی مؤثر برای یافتن ساختار خوشه‌بندی داده‌های پیچیده دارای هم‌پوشانی و مقادیر مبهم است (Guler and Thyne, 2004). توزیع مکانی نتایج خوشه‌بندی به روش FCM، هماهنگی معناداری با فرآیندهای فیزیکوشیمیایی رخساره‌های هیدروشیمی آب نشان می‌دهد. با مقایسه دو روش مورد استفاده در این پژوهش مشخص می‌شود که انطباق خوبی

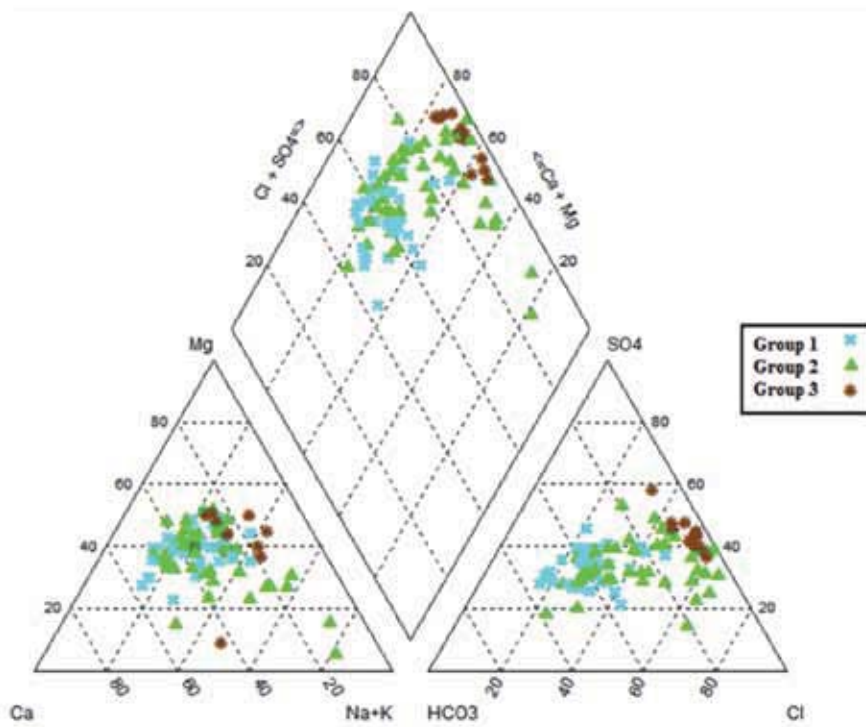
پساب صنعتی و کشاورزی در مناطقی که سطح آب زیرزمینی بالاست و برخورد آب زیرزمینی با پهنه‌های نمکی و سولفات‌ها محلی (بر روی شکل ۱، با QS نمایش داده شده) است. گروه سوم که با رنگ قهوه‌ای در شکل‌ها مشخص شده است. آب‌هایی با کیفیت پایین هستند که میانگین مواد جامد محلول در آن ۴۶۰۰ میلی‌گرم در لیتر است. آب‌های این گروه از لحاظ مقدار TDS در گروه آب‌های شور و خیلی شور قرار می‌گیرند که منشأ آن سازندهای نمکی و سولفات‌ها غرب محدوده مورد مطالعه است.

استفاده از نمودار پایپر به‌عنوان اصلی‌ترین نمودار در شناسایی رخساره‌ها یا گروه‌های هیدروشیمی در این مطالعه، امری اجتناب‌ناپذیر به‌نظر می‌رسد. نمودار پایپر می‌تواند در سنجش عملکرد روش‌های خوشه‌بندی به‌عنوان یک شاخص مورد استفاده قرار گیرد. در ادامه، رفتار دو روش خوشه‌بندی در تعلق داده‌ها به گروه‌های هیدروشیمی توسط نمودار پایپر مورد ارزیابی قرار می‌گیرد. در شکل ۵ عملکرد روش خوشه‌بندی HCA در نمودار پایپر نمایش داده می‌شود.

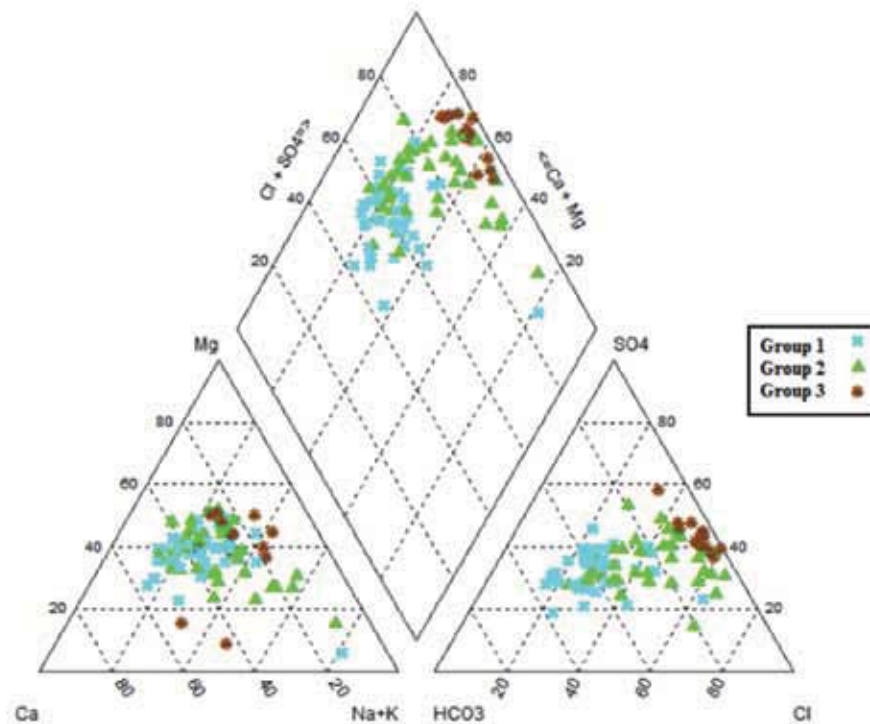
گروه معرف را نشان می‌دهد که نشان‌دهنده وجود سه تیپ آب زیرزمینی مختلف در ایجاد این گروه‌ها است.

گروه اول که با رنگ آبی در تصاویر نشان داده شده است، آب‌هایی با کیفیت بالا، املاح کم و با میانگین پایین مجموع مواد جامد محلول (۴۰۰ میلی‌گرم در لیتر) است. این گروه در طبقه‌بندی آب‌ها بر اساس (TDS (Todd and Mays, 2005 در گروه آب شیرین قرار می‌گیرد که عمدتاً مربوط به منطقه تغذیه و مناطقی می‌باشد که تحت تأثیر آلودگی‌ها قرار نگرفته‌اند.

گروه دوم که با رنگ سبز در تصاویر نشان داده شده است؛ آب‌هایی با کیفیت نسبتاً پایین که میانگین مواد جامد محلول در آن ۱۰۰۰ میلی‌گرم در لیتر است. از نظر مقدار TDS، این نمونه‌ها در گروه آب‌های لب‌شور قرار دارند که برای مصارف کشاورزی و صناعی که نیاز به آب‌های با کیفیت بالا ندارند، مناسب است. تشکیل این گروه را می‌توان نتیجه چند فرآیند دانست که به نظر می‌رسد به دلیل اختلاط آب‌های منطقه تغذیه و آب‌های شور در غرب محدوده مورد مطالعه، ارتباط



شکل ۵. نمودار پایپر برای داده‌های خوشه‌بندی شده به روش HCA



شکل ۶. نمودار پایپر برای داده‌های خوشه‌بندی شده به روش FCM

یک، دو و سه در شکل شماره ۴ با نمودار پایپر، مشخص می‌شود، روش FCM در ارزیابی تحولات هیدروشیمی منطقه مورد مطالعه نسبت به روش HCA برتری دارد. بدین ترتیب که روش FCM نسبت به روش HCA روند تحولات هیدروشیمی منطقه مورد مطالعه را بهتر نشان می‌دهد. در نمودار پایپر نیز می‌توان روند تغییرات هیدروشیمی منطقه مورد مطالعه را از گروه یک به گروه دو و در نهایت گروه سه مشاهده نمود. به صورت کلی می‌توان گروه یک را نمونه‌های ابتدایی دشت، گروه دو را نمونه‌های میان دشتی و گروه سه را نمونه‌های خروجی و انتهایی خوشه‌بندی نمود.

نتیجه‌گیری

در این پژوهش تلاش شد تا با بررسی و مقایسه دو روش پرکاربرد از روش‌های خوشه‌بندی، روشی مناسب برای تعیین رخساره‌های هیدروشیمی انتخاب شود. الگوریتم خوشه‌بندی FCM روشی مؤثر برای یافتن ساختار خوشه‌بندی داده‌های پیچیده دارای هم‌پوشانی و مقادیر

همان‌طور که مشاهده می‌شود بی‌نظمی قابل ملاحظه‌ای در گروه‌بندی داده‌ها به روش HCA دیده می‌شود. این موضوع در تفکیک گروه‌های ۱ و ۲ بهتر مشخص است. شکل ۶ نمودار پایپر را برای داده‌های خوشه‌بندی شده به روش FCM نشان می‌دهد، چنانچه در شکل دیده می‌شود؛ در این روش جدایش بهتری در نمونه‌ها دیده می‌شود.

با توجه به شکل ۶، نمونه‌های گروه یک به‌طور عمده در منطقه اختلاط و منطقه بدون تیپ غالب قرار می‌گیرند. همان‌طور که در شکل شماره ۴ مشاهده می‌شود در بخش‌های شمال شرقی منطقه مورد مطالعه اغلب نمونه‌ها در گروه یک قرار گرفته‌اند. در گروه دو نیز اغلب، نمونه‌ها در منطقه اختلاط قرار می‌گیرند اما در برخی نمونه‌ها تیپ کلراید سدیک نیز مشاهده می‌شود که به نظر می‌رسد به دلیل وجود پهنه‌های نمکی در منطقه مورد مطالعه باشد. گروه سه نیز اغلب نمونه‌ها تیپ سولفات منیزیک دارند که به دلیل انحلال سازندهای آبرفتی و پهنه‌های نمکی و ژئوسی در غرب منطقه مورد مطالعه است. با بررسی روند تغییرات گروه‌های

خوشه‌بندی فازی. امیرکبیر، ۶۷، ۲۷-۳۴.
- شکاری، پ. و باقر نژاد، م.، ۱۳۸۴. بررسی کاربرد روش فازی در طبقه‌بندی خاک، مطالعه موردی: چشمه سفید کرمانشاه. علوم و فنون کشاورزی و منابع طبیعی زمستان، ۹، ۵۵-۶۸.
- طاهری، م.، ۱۳۷۸. آشنایی با نظریه مجموعه‌های فازی. انتشارات جهاددانشگاهی مشهد.
- کوره‌پزان دزفولی، ا.، ۱۳۸۷. اصول تئوری مجموعه‌های فازی و کاربرد آن در مدل‌سازی مسائل مهندسی آب. انتشارات جهاددانشگاهی واحد صنعتی امیرکبیر.

- Alimohammadlou, Y., Najafi, A. and Gokceoglu, C., 2014. Estimation of rainfall-induced landslides using ANN and fuzzy clustering methods: A case study in Saen Slope, Azerbaijan province, Iran. *Catena*, 120, 149-162.

- Ay, M. and Kisi, O., 2014. Modelling of chemical oxygen demand by using ANNs, AN-FIS and k means clustering techniques *Journal of Hydrology*, 511, 279-289.

- Barbieri, P., Adami, G., Favretto, A., Lutman, A., Avoscan, W. and Reisenhofer, E., 2001. Robust cluster analysis for detecting physicochemical typologies of freshwater from wells of the plain of Friuli (northeastern Italy). *Analytica Chimica Acta*, 440, 161-170.

- Bezdek, J.C., 1981. *Pattern Recognition with Fuzzy Objective Function Algorithms* Plenum, 256.

- Burrough, P.A., Van Gaans, P.M. and McMillan, R.A., 2000. High resolution landform classification using fuzzy k-means. *Fuzzy Sets Systems*, 113, 37-52.

- Goyal, M.K. and Gupta, V., 2014. Identification of Homogeneous Rainfall Regimes in Northeast Region of India using Fuzzy Cluster Analysis. *Water Resources Management*, 28, 4491-4511.

- Guler, C. and Thyne, G.D., 2004. De-

توزیع مکانی نتایج خوشه‌بندی در روش FCM با فرایندهای فیزیکوشیمیایی هماهنگی معناداری با رخساره‌های هیدروشیمی آب نشان می‌دهد. با مقایسه توزیع مکانی رخساره‌های هیدروشیمی و نمودار پایپر برای هر دو روش، نتایج نشان می‌دهد؛ در روش FCM توزیع مکانی و خطای خوشه‌بندی نسبت به روش HCA کمتر است و خوشه‌بندی فازی، فرایندهای ایجادکننده تغییرات طبیعی پیوسته در منطقه را بهتر نشان می‌دهد و مقدار زیادی داده را می‌تواند بررسی و تحلیل نماید. قرار دادن نمونه‌های آب در گروه‌هایی با مرزهای مشخص باعث ایجاد خطای طبقه‌بندی داده‌ها می‌شود که از اشکالات روش HCA است. در روش خوشه‌بندی فازی با تخصیص هر نمونه به چند گروه، خطای طبقه‌بندی به حداقل می‌رسد. به‌منظور صحت‌سنجی، از نمودار پایپر استفاده شد. در نمودار پایپر ترسیم شده به روش HCA بی‌نظمی در گروه‌بندی داده‌ها دیده می‌شود. این موضوع در تفکیک گروه‌های یک و گروه دو بهتر دیده می‌شود؛ اما در نمودار پایپر ترسیم شده به روش FCM گروه‌ها نظم بهتری دارند و نمونه‌هایی با رخساره هیدروشیمی یکسان، با هم در یک گروه قرار گرفتند. به‌طور کلی می‌توان نتیجه گرفت؛ در روش HCA تغییرات در مرز بین گروه‌ها کاملاً واضح و ناگهانی است اما در روش FCM، تغییرات تدریجی است.

منابع

- آذر، ع. و فرجی، ح.، ۱۳۸۶. علم مدیریت فازی. انتشارات موسسه کتاب مهربان نشر.

- آقائباتی، ع.، ۱۳۸۵. زمین‌شناسی ایران. انتشارات سازمان زمین‌شناسی و اکتشافات معدنی کشور.

- تشنه‌لب، م.، صفاریور، ن. و افیونی، د.، ۱۳۸۷. سیستم‌های فازی و کنترل فازی. انتشارات دانشگاه صنعتی خواجه‌نصیرالدین طوسی.

- سازمان زمین‌شناسی کشور، ۱۳۸۵. نقشه زمین‌شناسی یک صد هزارم ورامین، سازمان زمین‌شناسی و اکتشافات معدنی کشور.

- سلطانی محمدی، س.، علی‌نیا، ف. و آویانی، د.، ۱۳۸۶. خنثی‌سازی اثر سنترتیک داده‌های ژئوشیمیایی رسوبات آبراهه‌ای در منطقه شمال تکاب با استفاده از

lineation of hydrochemical facies distribution in a regional groundwater system by means of fuzzy c-means clustering. *Water Resource Research*, 40, w12503. Doi: 10.1029/2004WR003299.

- Guler, C., Thyne, G.D., McCray, J.E. and Turner, A.K., 2002. Evaluation of graphical and multivariate statistical methods for classification of water chemistry data. *Hydrogeology Journal*, 10, 455- 474.

- Hathaway, R.J. and Bezdek, J.C., 2001. Fuzzy C-means clustering of incomplete data. *IEEE Transactions System Man and Cybernetics*, 31, 735- 744.

- McBratney, A.B. and Moore, A.W., 1985. Application of fuzzy sets to climatic classification. *Agriculture and Forest Meteorology*, 35, 165- 185.

- Rantitsch, G., 2000. Application of fuzzy clusters to quantify lithological background concentrations in stream-sediment geochemistry. *Journal of Geochemical Exploration*, 71, 73- 82.

- Singh, K., Malik, A., Mohan, D. and Sinha, S., 2004. Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India)-a case study. *Water Research*, 38, 3980-3992.

- Todd, K.D., and Mays, 2005. *Groundwater Hydrology*. John Wiley and Sons. 636.

- Zadeh, L.A., 1965. Fuzzy sets. *Information Control*, 8, 338- 353.

- Zhang, C.T., Chou, K.C. and Maggiora, G.M., 1995. Predicting protein structural classes from amino acid composition: Application of fuzzy clustering. *Protein Engineering*, 8, 425-435.

Archive of SID