

## شناسایی ایستگاه‌های همگن شبکه پایش کیفی آبخوان مشهد بر مبنای آلودگی نترات

مسلم اکبرزاده<sup>۱\*</sup> - بیژن قهرمان<sup>۲</sup> - کامران داوری<sup>۳</sup>

تاریخ دریافت: ۱۳۹۴/۰۲/۰۲

تاریخ پذیرش: ۱۳۹۵/۰۵/۱۱

### چکیده

در راستای پایش پایدار منابع آب، ارزیابی شبکه پایش و ارزیابی کیفیت آبخوان با تجزیه و تحلیل دقیق اطلاعات به دست آمده از مقادیر آلودگی حاصل می‌شود. از اساسی‌ترین این تحلیل‌ها سنجش دقیق پهنه‌های مستعد خطر و شناسایی ایستگاه‌های همگن از نظر شاخص آلودگی است. با شناسایی چاه‌های همگن از نظر آلودگی به نترات می‌توان سبب بهبود شرایط پایش شد. بر این اساس در این پژوهش، آبخوان شهر مشهد با دو معیار فاصله اقلیدسی و آنتروپی، خوشه‌بندی شد. نتایج نشان داد بهترین ساختار خوشه بندی با چهار خوشه همگن حاصل می‌شود. با توجه به پراکندگی چاه‌ها و میانگین گشتاورهای خطی، روش مبتنی بر معیار آنتروپی بر روش فاصله اقلیدسی برتری دارد. همچنین تغییرات نترات نقش بسزایی را در شناسایی ایستگاه‌های همگن مبتنی بر آنتروپی ایفا می‌کند. بنابراین با خوشه‌بندی مبتنی بر آنتروپی می‌توان چاه‌های همگن از نظر تغییرپذیری شاخص آلودگی نترات را شناسایی کرد که این امر می‌تواند گامی مهم و موثر در جهت پایش آبخوان مشهد و ارزیابی وضعیت کیفی آن باشد. همچنین در راستای ارزیابی و بهینه‌سازی شبکه پایش، هر چه شبکه موجود خوشه‌های کمتری به خود اختصاص دهد، همگنی بیشتری دارد. بنابراین رویکرد بهینه‌سازی از افزایش به کاهشی (تبدیل شبکه پایش) توجیه خواهد داشت. در این صورت هزینه‌های پایش اعم از حفر، تجهیز، نمونه‌برداری، نگهداری و آنالیز آزمایشگاهی نیز کاهش می‌یابد.

**واژه‌های کلیدی:** آب زیرزمینی، آنتروپی، خوشه‌بندی، ضریب نیم‌رخ، فاصله اقلیدسی، کریجینگ فضایی زمانی، گشتاورهای خطی

### مقدمه

جغرافیایی، هواشناسی، آبی و غیره ایستگاه‌ها بنا نهاده شده است (۴). از این روش‌ها می‌توان به تحلیل همبستگی متعارف (۶) و تحلیل خوشه‌ای (۱۴، ۱۹ و ۲۰) اشاره کرد. تحلیل خوشه‌ای احتمالاً مهم‌ترین روش برای شناسایی ایستگاه‌های همگن به شمار می‌رود (۲۰). به طور کلی یک گروه (شبکه پایش) وقتی همگن است که اعضای آن (ایستگاه‌ها) بر مبنای مشخصه‌هایشان با یکدیگر متشابه و با اعضای گروه‌های دیگر نامتشابه باشند (۳). این تشابه در فرایند خوشه‌بندی با معیاری کمی معروف به فاصله (تابع اندازه) اندازه‌گیری و تعیین می‌شود. بر این اساس خوشه همگن را می‌توان مجموعه‌ای نامید که بر مبنای معیاری از سنجش فاصله، به یکدیگر نزدیک و از سایر اعضای دیگر دور باشند (۸ و ۱۶). معیارهای مختلفی برای محاسبه فاصله تشابه، معرفی و بررسی شده است که در مقیاس کمی می‌توان به فاصله اقلیدسی، فاصله منتهان، فاصله حداکثری، فاصله مینکوفسکی، فاصله متوسط، فاصله ژئودیسک، فاصله بر مبنای آنتروپی، ضریب همبستگی و کانبرا اشاره نمود. استفاده از هر کدام از این روش‌ها به ماهیت داده‌ها بستگی دارد مثلاً اگر در مجموعه داده‌ها، خوشه‌هایی بصورت فشرده یا جزیره‌ای وجود داشته باشد، فاصله مینکوفسکی، اقلیدسی و منتهان خوب عمل می‌کنند (۱۵).

ارزیابی، سنجش نحوه عملکرد سیستم و برآورد کارایی آن در رسیدن به اهداف از پیش تعیین شده است. در ارزیابی وضعیت کیفی، میزان اهمیت هر ایستگاه پایش در کل شبکه پایش مهم و موثر است. به عبارت دیگر ارزش گذاری ایستگاه‌های پایش با توجه به محل قرارگیری، میزان آلودگی و سایر ویژگی‌های آن‌ها از درجه اهمیت بالایی برخوردار است (۲۶). بر این اساس و بویژه در هنگام مواجهه با حجم بالای داده‌های کیفی، ارائه روشی برای خوشه‌بندی شامل شناسایی ایستگاه‌های همگن شبکه پایش و دسته‌بندی آن‌ها در گستره‌هایی از میزان اهمیت (بر مبنای آلودگی)، از ضرورت‌های پایش آبخوان و ارزیابی وضعیت کیفی آن است.

پژوهشگران در زمینه شناسایی ایستگاه‌های پایش همگن روش‌های زیادی را بر اساس شاخص‌های مختلف ارائه کرده‌اند. این روش‌ها بر پایه شناسایی تشابه و عدم تشابه بین خصوصیات

۱، ۲ و ۳- به ترتیب دکتری آبیاری و زهکشی، (کارشناس آزمایشگاه) و استادان گروه علوم و مهندسی آب، دانشکده کشاورزی، دانشگاه فردوسی مشهد  
\* - نویسنده مسئول: (Email: akbarzade@staff.um.ac.ir)

رضوی را به مناطق همگن تقسیم نمودند و به تحلیل فراوانی منطقه‌ای بارش‌های حداکثر سالانه در استان مذکور پرداختند.

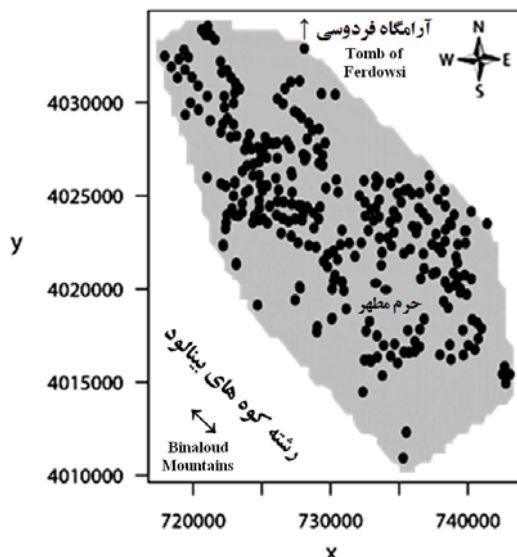
در این مقالات هیچ اشاره‌ای به استفاده از معیار تشابه (همانند فاصله اقلیدسی یا ...) نشده و صرفاً از یک مجموعه نرم افزاری همانند SPSS یا MATLAB برای خوشه‌بندی استفاده شده است. از این رو می‌توان یکی از معایب پژوهش‌های اخیر را عدم توجه به انتخاب معیار تشابه مناسب برای خوشه‌بندی بیان نمود.

بر اساس آنچه از سلسله پژوهش‌های پیشین بررسی گردید و در راستای ارزیابی کیفیت منابع آب زیرزمینی، در این پژوهش با هدف خوشه‌بندی ایستگاه‌های پایش کیفی آبخوان شهر مشهد، شناسایی ایستگاه‌های همگن از نظر تغییرپذیری شاخص آلودگی نیترات و بررسی کارایی معیار فاصله آنتروپی در مقایسه با معیار فاصله اقلیدسی در فرایند تحلیل خوشه‌ای صورت پذیرفت.

## مواد و روش‌ها

### منطقه و داده‌ها

دشت مشهد با ابعادی به طول بیش از ۱۰۰ کیلومتر، عرض متوسط ۲۵ کیلومتر و وسعتی حدود ۲۵۰۰ کیلومتر مربع، در حوضه آبریز رودخانه کشف رود واقع است که در آن سالانه حدود ۲۵۰ میلیون متر مکعب کسری مخزن وجود دارد و در نتیجه از سال ۱۳۴۸ در فهرست دشت‌های بحرانی و ممنوعه قرار گرفته است. شهر مقدس مشهد بزرگ‌ترین منطقه فشرده مسکونی این دشت است.



شکل ۱- پراکندگی ۲۸۷ حلقه چاه در محدوده آبخوان شهر مشهد بر اساس مختصات UTM

Figure 1- Dispersion of 287 water wells in Mashhad aquifer (UTM)

این شهر از نظر وسعت و جمعیتی بیش از ۳ میلیون نفر، دومین

آنتروپی که یکی از معیارهای اندازه‌گیری عدم قطعیت است، میزان آشفتگی، بی‌نظمی و میزان تصادفی بودن را تشریح می‌کند. چندین نوع اندازه برای سنجش آنتروپی وجود دارد. سینگ (۲۴ و ۲۵) در پژوهش‌های خود قابلیت‌های آنتروپی را در تحلیل و ارزیابی در مفهوم کلی آن مورد بررسی قرار داده است. در پژوهش‌های بعدی همچون ریانا و همکاران (۲۰) معیارهای دیگری مانند SNI معرفی شد. معیار SNI به عنوان یک معیار ترکیبی از آنتروپی برای خوشه‌بندی، اطلاع متقابل دو مقدار (مقادیر شاخص آلودگی دو چاه) را نسبت به آنتروپی توأم آن‌ها بیان می‌کند. این چنین معیارهایی می‌توانند به عنوان معیار فاصله تشابه برای خوشه‌بندی ایستگاه‌های پایش بکار روند.

راجسختار و همکاران (۱۸) از معیار ترکیبی DIT (مبتنی بر آنتروپی) برای خوشه‌بندی ایستگاه‌های باران سنجی استفاده کردند. آن‌ها علاوه بر تعیین تشابه ایستگاه‌ها با استفاده از معیار مذکور، برای خوشه‌بندی از روشی همانند روش سلسله مراتبی با الگوریتمی خاص استفاده کردند. برای این که بتوان از یک معیار برای بررسی فاصله تشابه اعضای یک خوشه همگن استفاده کرد، باید ویژگی‌هایی همچون غیرمنفی بودن، همانی بودن، تقارن و نامساوی مثلث را دارا باشد (۲). معیار DIT متقارن نیست و بنابراین ویژگی‌های مذکور را برای استفاده در تحلیل خوشه‌ای دارا نمی‌باشد. براین اساس ریانا و همکاران (۲۰) با اصلاح معیار DIT، معیار جدیدی به نام SNI معرفی نمودند که در آن آنتروپی توأم جایگزین آنتروپی شده و ویژگی‌های مذکور را نیز دارا است. آنان همگنی مناطق ایجاد شده با تعداد ایستگاه‌های پایش برابر را با استفاده از معیار SNI و فاصله اقلیدسی مقایسه کردند. آن‌ها با مقایسه ایستگاه‌های هیدرومتری در خوشه‌هایی با حجم متفاوت (تکرار خوشه بندی با تعداد ایستگاه‌های متفاوت) به این نتیجه رسیدند که هر چه حجم خوشه‌ها بیشتر باشد معیار SNI عملکرد بهتری خواهد داشت که این موضوع را نیز طی فرایند آزمون همگنی با شاخص‌های اعتباریابی همگنی نشان دادند. آنان منطقه مورد مطالعه را خوشه‌بندی نکردند، فقط اظهار داشتند که هرچه تعداد ایستگاه‌های موجود در هر خوشه بیشتر باشد، معیار SNI بهتر عمل می‌کند.

پژوهش‌های زیادی بر مبنای استفاده از اندازه آنتروپی برای خوشه‌بندی ایستگاه‌های پایش صورت پذیرفته و بیشتر مطالعات برای خوشه‌بندی پارامترهای کیفی آب بوده است. از جمله پژوهش‌هایی که در زمینه خوشه‌بندی ایستگاه‌های پایش صورت پذیرفته می‌توان به پژوهش یانگ و همکاران (۲۶) اشاره نمود که با تقسیم حوضه‌ای به خوشه‌های همگن، تحلیل فراوانی بارش حداکثر سالانه را انجام دادند. نگوگونندو و همکاران (۱۷) نیز در تحلیلی مشابه به تحلیل فراوانی حداکثر بارش سالانه پرداختند. شاهدی و همکاران (۲۲) پس از بررسی فرض‌های اولیه با استفاده از تحلیل خوشه‌ای، استان خراسان

دیگر معیارهای سنجش فاصله تشابه است که رابطه غیرخطی را به خوبی تشریح می‌کند. زمانی که تعداد مشاهدات زیاد باشد توزیع داده‌ها نرمال بوده و این معیار برای داده‌هایی با رابطه خطی نیز مناسب خواهد بود. برای سنجش آزمایشات مختلف و ارتباط آن‌ها با یکدیگر در آنتروپی شانون اندازه‌هایی تعریف شده است. یکی از این معیارها که شامل هر دو مورد خطی و غیر خطی شود، اطلاع متقابل بین دو متغیر است (۲۵). بطور کلی اطلاع متقابل کسب شده از مجموعه‌ای از متغیرهای تصادفی  $(X_1, \dots, X_m)$  را می‌توان توسط مجموعه‌ای دیگر از متغیرهای تصادفی  $(Y_1, \dots, Y_n)$  تعریف کرد. اگر  $X_1, \dots, X_m$  و  $Y_1, \dots, Y_n$  متغیرهای تصادفی باشند، اطلاع متقابل کسب شده راجع به آنها به صورت رابطه ۳ تعریف می‌شود (۷).

$$I(X_1, \dots, X_m; Y_1, \dots, Y_n) = H(X_1, \dots, X_m) - H(X_1, \dots, X_m | Y_1, \dots, Y_n) \quad (۳)$$

که در آن  $H(X_1, \dots, X_m)$  آنتروپی توام متغیرهای تصادفی  $X_1, \dots, X_m$  و  $H(X_1, \dots, X_m | Y_1, \dots, Y_n)$  آنتروپی شرطی  $X_1, \dots, X_m$  به شرط  $Y_1, \dots, Y_n$  است که بصورت روابط ۴ و ۵ تعریف می‌شوند.

$$H(X_1, \dots, X_m) = - \sum_{x_1, \dots, x_m} p(x_1, \dots, x_m) \log p(x_1, \dots, x_m) \quad (۴)$$

$$H(X_1, \dots, X_m | Y_1, \dots, Y_n) = - \sum_{\substack{x_1, \dots, x_m \\ y_1, \dots, y_n}} p(x_1, \dots, x_m, y_1, \dots, y_n) \log p(x_1, \dots, x_m | y_1, \dots, y_n) \quad (۵)$$

از این رو اگر  $X=(X_1, \dots, X_n)$  شاخص‌های ایستگاه‌های  $X$  شامل  $n$  شاخص (همانند مقدار نیترا و موقعیت جغرافیایی) و  $Y=(Y_1, \dots, Y_n)$  شاخص‌های ایستگاه‌های  $Y$  شامل همان  $n$  شاخص در  $Y$  باشد می‌توان رابطه ۳ را بصورت رابطه ۶ بازنویسی کرد (۷).

$$I(X; Y) = I(X_1, \dots, X_n; Y_1, \dots, Y_n) \quad (۶)$$

که در آن  $X=(X_1, \dots, X_n)$  شاخص‌های ایستگاه  $X$  و  $Y=(Y_1, \dots, Y_n)$  شاخص‌های ایستگاه  $Y$  و  $I(X; Y)$  اطلاع متقابل آن‌ها است.

برای اندازه‌گیری فاصله تشابه دو متغیر  $X$  و  $Y$  که دو نقطه (شاخص مستقل) در فضای  $n$  بعدی می‌باشند، معیار فاصله انتقال اطلاعات جهت‌دار یا همان DIT این دو نقطه بصورت رابطه ۷ تعریف می‌شود (۲۰).

$$DIT(X; Y) = \frac{I(X; Y)}{H(X)} \quad (۷)$$

که در آن  $I(X; Y)$  اطلاع متقابل  $X=(X_1, \dots, X_n)$  و  $Y=(Y_1, \dots, Y_n)$  و  $H(X)$  آنتروپی متغیر  $X=(X_1, \dots, X_n)$  است.

همان‌طور که در رابطه ۷ مشاهده می‌شود، معیار DIT مقارن نیست یعنی  $DIT(Y, X) \neq DIT(X, Y)$  است زیرا اگر جای  $X$  و  $Y$  عوض شود مقدار کسر نیز عوض خواهد شد. از این رو شرط سوم

کلانشهر ایران شناخته می‌شود. متوسط جمعیت زوار این شهر نیز در حدود ۳ میلیون نفر است.

مساحت منطقه براساس مختصات چاه‌های مورد مطالعه در حدود ۳۲۰ کیلومتر مربع برآورد می‌شود. شکل ۱ نحوه پراکندگی ایستگاه‌های پایش (چاه‌ها) در منطقه مورد مطالعه را نشان می‌دهد. در پژوهش حاضر از بین پارامترهای کیفی، مقادیر نیترا مربوط به ۲۸۷ حلقه چاه که دارای اطلاعات ثبت شده در گستره آماری ۱۳۸۱ تا ۱۳۸۹ بود به ترتیب به عنوان شاخص آلودگی استفاده شد. داده‌های هر چاه که با میانگین ۳ نوبت پایش در هر سال ثبت شده است شامل داده‌های دورافتاده یا نقاط فاقد مشاهده بود که داده‌های دورافتاده شناسایی (۱۱) و از مجموعه داده‌ها کنار گذاشته شد. سپس این مقادیر همانند نقاط فاقد مشاهده در نظر گرفته شد و کل آن‌ها (نقاط فاقد مشاهده اصلی و نقاط فاقد مشاهده حاصل از کنار گذاشتن داده‌های دور افتاده) طی فرآیند کریجینگ فضایی زمانی (۱) برآورد شد. همچنین نظر به این که فاصله تشابه به واحد داده‌ها بستگی ندارد و استاندارد نبودن داده‌ها خوشه‌بندی را تحت تاثیر قرار می‌دهد، تمامی داده‌ها با استفاده از رابطه ۱ استاندارد شد (۹).

$$z_i = \frac{x_i - \bar{x}}{s} \quad (۱)$$

که در آن  $x_i$  مشاهده نام و  $\bar{x}$  میانگین و  $s$  انحراف معیار مشاهدات شاخص (همانند مقدار نیترا و موقعیت جغرافیایی) می‌باشد.

در ادامه، فرآیند خوشه‌بندی توسط داده‌های تکمیل و استاندارد شده انجام شد.

### معیارهای فاصله تشابه

در فرآیند خوشه‌بندی شبکه پایش کیفی، فاصله تشابه (تفاوت) ایستگاه‌های پایش (چاه‌ها) محاسبه می‌شود. این تفاوت با معیارهای براساس ویژگی‌هایی از قبیل موقعیت مکانی ایستگاه‌ها و شاخص آلودگی بیان می‌شود (۱۸).

فاصله اقلیدسی به عنوان یکی از معیارهای تعیین فاصله تشابه که ویژگی‌های عمومی (۲) را دارد استفاده می‌شود. فاصله اقلیدسی برای دو نقطه  $X$  و  $Y$  در فضایی با  $n$  بعد به صورت رابطه ۲ تعریف می‌شود (۲۰).

$$D(X, Y) = \sqrt{\sum_{j=1}^n (X_j - Y_j)^2} \quad (۲)$$

که در آن  $X_j$  و  $Y_j$  شاخص‌های  $j$ ام ( $j=1, \dots, n$ ) برای نقاط  $X$  و  $Y$  هستند.

در پژوهش حاضر، از شاخص‌های نیترا و موقعیت مکانی (طول و عرض جغرافیایی) استفاده شده که در این صورت  $n=3$  است.

اغلب پژوهش‌ها اطلاعات یک آزمایش یا یک پدیده را توسط آنتروپی شانون (۲۳) اندازه‌گیری می‌کنند. به عبارت دیگر آنتروپی از

جدول ۱- تفسیر ضریب نیمرخ (۲۱)

Table 1- Interpretation of silhouette score (21)		
گستره ضریب نیمرخ	ساختار کیفی خوشه بندی	
SC range	Quality structure of clustering	
0.71 – 1	Excellent	عالی
0.51 – 0.7	Good	خوب
0.26 – 0.5	fair	متوسط
-1 – 0.25	Poor	ضعیف

برای مقایسه خوشه‌بندی‌های صورت گرفته بوسیله دو معیار اقلیدسی و آنتروپی (SNI) باید همگنی خوشه‌بندی‌های انجام شده مورد سنجش قرار گیرد. از این رو سنجش همگنی خوشه‌ها توسط روش گشتاورهای خطی انجام می‌شود (۱۲ و ۱۳). در این راستا، ناحیه‌ای شامل n ایستگاه در نظر گرفته می‌شود که شامل  $n_i$  داده اخذ شده باشد و  $t^{(i)}$  و  $t_3^{(i)}$  نسبت گشتاورهای خطی باشند. میانگین وزنی  $t^{(R)}$  و  $t_3^{(R)}$  که ضرایب تاثیر منطقه‌ای می‌باشند بر مبنای گشتاورهای خطی محاسبه می‌شوند. برای محاسبه اندازه همگنی ابتدا مقادیر روابط ۱۱ و ۱۲ محاسبه می‌شوند (۱۲ و ۱۳).

$$V_1 = \left\{ \frac{\sum_{i=1}^n n_i (t^{(i)} - t^{(R)})^2}{\sum_{j=1}^n n_j} \right\}^{\frac{1}{2}} \quad (11)$$

$$V_2 = \frac{\sum_{i=1}^n n_i \{ (t^{(i)} - t^{(R)})^2 + (t_3^{(i)} - t_3^{(R)})^2 \}^{\frac{1}{2}}}{\sum_{j=1}^n n_j} \quad (12)$$

سپس با روش‌های مونت کارلو، میانگین ( $\mu_{V_1}$  و  $\mu_{V_2}$ ) و انحراف معیار ( $\sigma_{V_1}$  و  $\sigma_{V_2}$ ) آن‌ها محاسبه می‌شود. در انتها معیارهای همگنی به وسیله روابط ۱۳ و ۱۴ محاسبه می‌شود (۱۲ و ۱۳).

$$H_1 = \frac{V_1 - \mu_{V_1}}{\sigma_{V_1}} \quad (13)$$

$$H_2 = \frac{V_2 - \mu_{V_2}}{\sigma_{V_2}} \quad (14)$$

برای بررسی میزان همگنی یک خوشه اگر  $H_1 < 1$  باشد، خوشه همگن، اگر  $1 < H_1 < 2$  باشد خوشه نسبتاً همگن و اگر  $H_1 > 2$  باشد خوشه ناهمگن است. همچنین اگر  $H_2 < 1$  باشد خوشه همگن، اگر  $1 < H_2 < 2$  باشد، نسبتاً همگن و اگر  $H_2 > 2$  باشد خوشه ناهمگن است با این تفاوت که شرط ضعیف‌تری برای پذیرش همگنی ایستگاه‌ها است (۱۲ و ۱۳).

شمای کار مراحل انجام پژوهش در شکل ۲ ارائه شده است.

ویژگی‌های عمومی آندربریگ را دارا نمی‌باشد. بنابراین با تصحیح رابطه ۷ معیار مقارنی بصورت رابطه ۸ بدست می‌آید (۲۰).

$$SNI(X; Y) = \frac{I(X; Y)}{H(X; Y)} \quad (8)$$

که در آن  $I(X; Y)$  و  $H(X; Y)$  به ترتیب اطلاع متقابل و آنتروپی توام بردارهای  $X=(X_1, \dots, X_n)$  و  $Y=(Y_1, \dots, Y_n)$  است.

پس از به دست آوردن فاصله تشابه هر چاه با چاه‌های دیگر، باید از روشی برای تفکیک آن‌ها از یکدیگر و تشکیل خوشه‌های همگن بهره برد که از آن جمله می‌توان به روش سلسله مراتبی اشاره نمود. برای تحلیل‌های هیدرولوژیکی، روش سلسله مراتبی مناسب است زیرا می‌توان فرض نمود که در چنین داده‌هایی ساختار سلسله مراتبی وجود دارد (۹).

### تحلیل خوشه‌ای

در اولین گام از فرایند خوشه‌بندی سلسله مراتبی، هر کدام از ۲۷۳ ایستگاه پایش (حلقه چاه) یک خوشه در نظر گرفته شد. سپس فاصله تشابه هر چاه با دیگر چاه‌ها محاسبه شد. بر این اساس ۲ خوشه (دو چاه) که کمترین فاصله را دارند یک خوشه می‌شوند. این فرایند تا رسیدن به k خوشه که توسط متوسط ضریب نیم رخ تعیین می‌شود ادامه پیدا می‌کند. بر این اساس اگر  $d_{ij}$  فاصله بین دو ایستگاه پایش (چاه)  $Z_j$  و  $Z_i$  باشد، ماتریس  $D=(d_{ij})$  مربوط به فاصله‌های بین تمامی چاه‌های  $Z_1, \dots, Z_n$  است.

تعداد خوشه‌ها را بر اساس کیفیت ساختاری می‌توان با آزمون ضریب نیمرخ بهینه نمود (۲۱). در این روش فاصله چاه i با هر خوشه دیگری چون C که i بدان تعلق ندارد را با  $d(i, C)$  نمایش داده و بصورت میانگین فاصله چاه i با همه چاه‌های آن خوشه تعریف می‌شود. پس از محاسبه  $d(i, C)$  برای همه خوشه‌ها مقدار  $b(i)$  بصورت کمترین مقدار  $d(i, C)$  در بین تمام C‌ها تعریف می‌شود. متوسط فاصله‌ای که چاه با اعضای خوشه خودش دارد نیز با  $a(i)$  نمایش داده می‌شود. بر این اساس شاخص  $s(i)$  بصورت رابطه ۹ تعریف می‌شود.

$$s(i) = \frac{b(i) - a(i)}{\text{Max}\{b(i), a(i)\}} \quad (9)$$

اگر میانگین  $s(i)$ ‌ها در خوشه kام با  $\bar{s}(k)$  نمایش داده شود، ضریب نیمرخ (SC)، بیشترین  $\bar{s}(k)$  در بین تمام خوشه‌ها خواهد بود (رابطه ۱۰).

$$SC = \text{Max}_k \bar{s}(k) \quad (10)$$

دامنه SC در گستره  $[-1, 1]$  تعریف می‌شود (۲۱). هر چقدر SC به یک نزدیک باشد نشان دهنده اعتبار بیشتر و ساختار بهتر خوشه‌بندی است. تفسیر مقادیر مختلف SC در جدول ۱ آورده شده است.

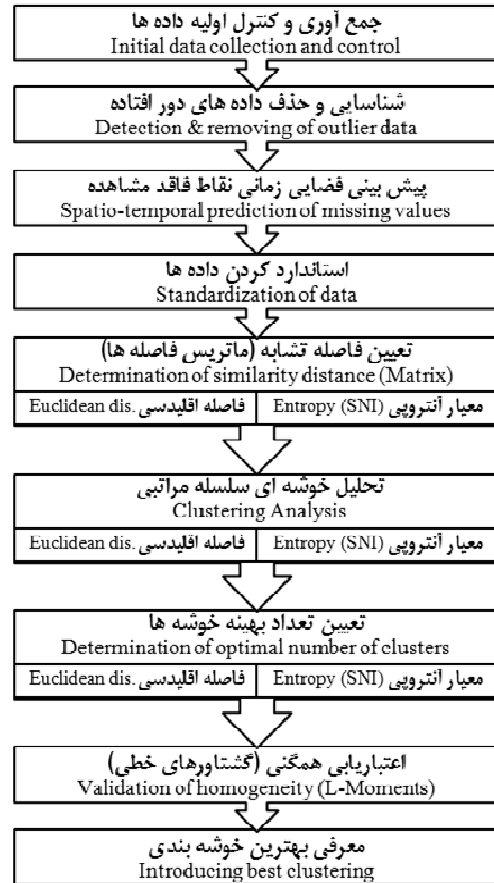
### نرم افزار

در این پژوهش برای تجزیه و تحلیل فضایی، انجام مراحل خوشه‌بندی، آزمون نیمرخ و آزمون همگنی از نسخه 3.1.2 نرم افزار R استفاده شده است. به منظور انجام فرایند نرم افزاری (کدنویسی) در تحلیل خوشه‌ای، برای هر چاه یکی از کدهای ۱ تا ۲۸۷ اختصاص داده شد به قسمی که هر کد مبین چاهی با موقعیت جغرافیایی خاص است.

### نتایج و بحث

در پژوهش حاضر و در راستای ارزیابی وضعیت کیفی و بهبود شرایط پایش، آبخوان شهر مشهد با دو معیار فاصله اقلیدسی و آنتروپی، خوشه‌بندی و چاه‌های همگن از نظر آلودگی شناسایی شد. براین اساس از داده‌های نیترا ت مربوط به موقعیت جغرافیایی ۲۸۷ حلقه چاه در گستره آماری سال‌های ۱۳۸۱ تا ۱۳۸۹ به عنوان شاخص آلودگی استفاده شد. از این رو خوشه‌بندی براساس تفاوت‌های مقادیر نیترا ت و موقعیت جغرافیایی چاه‌ها انجام شد. یعنی به هر ایستگاه پایش (چاه) سه برچسب مقدار نیترا ت، طول جغرافیایی و عرض جغرافیایی اختصاص داده شد.

ماتریس فاصله‌ها که ماتریسی با ۲۸۷ سطر و ستون است، با استفاده از فاصله اقلیدسی (رابطه ۲) و معیار SNI (رابطه ۸) بدست آمد. برخی از مقادیر این دو ماتریس در جداول ۲ و ۳ ارائه شده است. به دلیل استاندارد شدن هر یک از شاخص‌ها مقادیر ماتریس واحد ندارد.



شکل ۲- فلوجارت مراحل انجام پژوهش  
Figure 2- Flowchart of research steps

جدول ۲- نمونه نتایج ماتریس فاصله‌ها براساس معیار فاصله اقلیدسی  
Table 2- Results sample of distance matrix based on Euclidean distance

کد چاه Code	1	2	3	285	286	287
1	0	0.106	0.087	5.815	0.832	5.928
2	0.106	0	0.127	5.902	0.893	6.007
3	0.087	0.127	0	5.826	0.815	5.932
285	5.815	5.902	5.826	0	5.314	2.780
286	0.832	0.893	0.815	5.314	0	5.308
287	5.928	6.007	5.932	2.780	5.308	0

جدول ۳- نمونه نتایج ماتریس فاصله‌ها براساس معیار آنتروپی (SNI)  
Table 3- Results sample of distance matrix based on Entropy (SNI)

کد چاه Code	1	2	3	285	286	287
1	0	$1.618 \times 10^{-8}$	$1.107 \times 10^{-8}$	$1.615 \times 10^{-5}$	$3.625 \times 10^{-7}$	$1.892 \times 10^{-5}$
2	$1.618 \times 10^{-8}$	0	$2.350 \times 10^{-8}$	$1.694 \times 10^{-5}$	$4.628 \times 10^{-7}$	$1.970 \times 10^{-5}$
3	$1.107 \times 10^{-8}$	$2.350 \times 10^{-8}$	0	$1.626 \times 10^{-5}$	$3.209 \times 10^{-7}$	$1.892 \times 10^{-5}$
285	$1.615 \times 10^{-5}$	$1.694 \times 10^{-5}$	$1.626 \times 10^{-5}$	0	$1.301 \times 10^{-5}$	$2.852 \times 10^{-6}$
286	$3.625 \times 10^{-7}$	$4.628 \times 10^{-7}$	$3.209 \times 10^{-7}$	$1.301 \times 10^{-5}$	0	$1.483 \times 10^{-5}$
287	$1.892 \times 10^{-5}$	$1.970 \times 10^{-5}$	$1.892 \times 10^{-5}$	$2.852 \times 10^{-6}$	$1.483 \times 10^{-5}$	0

ارائه شده است. همان گونه که ملاحظه می‌شود، با در نظر گرفتن ۴ خوشه، ضریب نیمرخ در خوشه‌بندی با معیار فاصله اقلیدسی، برابر ۰/۹۸۸۷ و با معیار آنتروپی SNI برابر ۰/۷۴۶۰ بدست آمد. از این رو انتخاب تعداد خوشه‌ها برابر ۴ در هر دو روش موجب بروز ساختاری عالی برای خوشه‌بندی شد (جدول ۴).

با استفاده از هر کدام از ماتریس‌های بدست آمده، تحلیل خوشه‌ای بر مبنای روش سلسله مراتبی انجام شد. به منظور تعیین تعداد بهینه خوشه‌ها بطوری که با استفاده از هر دو معیار، بهترین ساختار خوشه‌بندی تولید شود، بر مبنای فرض‌های مختلف برای تعداد خوشه‌ها، ضرایب نیمرخ بدست آمد که خلاصه نتایج آن در جدول ۴

جدول ۴- خلاصه نتایج تعیین تعداد بهینه خوشه‌ها بر اساس ضریب نیمرخ

Table 4- Results summary for optimal number of clusters determination based on silhouette score

تعداد خوشه‌ها Number of clusters	ضریب نیمرخ Silhouette score		ساختار خوشه بندی Structure of Clustering	
	آنتروپی SNI	اقلیدسی Euclidean		
	2	0.502	0.533	Good
3	0.532	0.625	Good	خوب
4	0.746	0.989	Excellent	عالی
5	0.411	0.423	fair	متوسط
6	0.400	0.402	fair	متوسط
7	0.380	0.386	fair	متوسط
8	0.341	0.345	fair	متوسط
9	0.318	0.321	fair	متوسط
10	0.298	0.300	fair	متوسط

نمی‌تواند بیانگر واقعیت فیزیکی منطقه (تاثیر تغییرات نیترات در خوشه‌بندی) باشد چرا که در راستای اهداف پژوهش، باید خوشه‌بندی ارائه گردد که وابسته به مقادیر نیترات (شاخص آلودگی) باشد. از طرف دیگر در شکل ۴ ملاحظه می‌شود که تاثیر مقادیر نیترات در خوشه‌بندی توسط معیار آنتروپی SNI سبب شده خوشه‌ها شکل منظمی نداشته و در برخی موارد چاه‌های موجود در هر خوشه از نظر جغرافیایی فاصله نسبتاً زیادی نیز هم داشته‌اند. این فاصله جغرافیایی زیاد ناشی از تاثیر بیش تر مقدار نیترات در مقایسه با دو عامل طول و عرض جغرافیایی می‌باشد.

نظر به این که در خوشه‌بندی از شاخص‌های موقعیت مکانی (طول و عرض جغرافیایی) و نیترات استفاده گردید، به منظور ارائه دلایل بیشتر بر برتری روش آنتروپی در خوشه‌بندی متاثر از شاخص آلودگی (مقادیر نیترات)، با تغییر شاخص‌ها، مجدداً خوشه بندی در دو روش انجام شد. شکل ۶ نتایج بدست آمده را نشان می‌دهد. بر این اساس و همان گونه که در شکل ۶ و در مقایسه آن با شکل ۵ ملاحظه می‌شود، در خوشه‌بندی بر مبنای فاصله اقلیدسی با حذف هریک از مختصات جغرافیایی تاثیر مقادیر نیترات بیشتر شده ولی این تاثیرات باز هم بسیار کم بوده است. پس از حذف هر یک از شاخص‌های مکانی، فقط تاثیر شاخص حاضر در خوشه‌بندی به طور مشخص ملاحظه گردید و تغییرات کلی در خوشه بندی ملاحظه

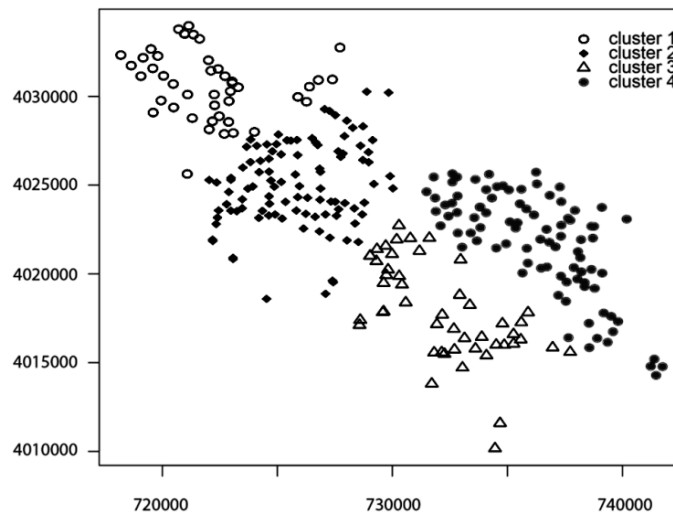
شکل‌های ۳ و ۴ نتایج حاصل از خوشه‌بندی را با در نظر گرفتن تعداد بهینه ۴ خوشه بر مبنای هر یک از دو معیار، نشان می‌دهد. در شکل‌های مذکور، پراکنش چاه‌ها در هر خوشه ارائه شده است. همان طور که در شکل ۳ مشاهده می‌شود پراکنندگی چاه‌ها از نظر جغرافیایی شکل منظمی دارد. بدین معنی که خوشه‌های تشکیل شده توسط معیار فاصله اقلیدسی بطور کامل تحت تاثیر فاصله جغرافیایی بین چاه‌ها می‌باشد.

گرچه خوشه‌هایی که به خوبی تفکیک شده باشند از نظر کاربردی ممکن است مناسب باشد ولی شرایط آب زیرزمینی منطقه و منابع آلاینده ناهمگن، ممکن است با این خوشه‌بندی هماهنگ نباشد. نظر به اینکه هدف این پژوهش خوشه‌بندی به منظور شناسایی ایستگاه‌های همگن شبکه پایش آلودگی بوده، می‌باید به گونه‌ای تاثیر نیترات در این خوشه بندی ملموس تر باشد. نتایج پژوهش‌های پیشین نشان می‌دهد که توزیع نیترات در آبخوان مشهد متقارن نیست. اکبرزاده و قهرمان (۱) به منظور ارزیابی شبکه پایش کیفی آبخوان شهر مشهد به برآورد احتمال خطر آلودگی در این شهر پرداختند که این موضوع منجر به شناخت پهنه‌های هم احتمال آلودگی شد. همان طور که در شکل ۵، مستخرج از پژوهش آنان ملاحظه می‌شود، مقادیر نیترات در برخی مناطق بیش از حد مجاز است که این مناطق لزوماً در نزدیکی یکدیگر نیستند بنابراین خوشه‌بندی شکل ۳

است که با ملاحظات فیزیکی منطبق است در صورتی که در خوشه‌بندی بر مبنای فاصله اقلیدسی این تاثیر به چشم نمی‌آید. از این رو چاه‌هایی که مقادیر نیترات آن‌ها نزدیک به یکدیگر باشند در یک خوشه قرار خواهند گرفت. بنابراین با توجه به اینکه هدف این پژوهش ارائه خوشه‌بندی چاه‌ها بر مبنای آلودگی می‌باشد باید بر روشی تکیه کرد که بتواند شرکت آب و فاضلاب را در راستای تدوین یک برنامه پایش کیفی موثر و قابل اتکا برای منابع آب زیرزمینی دشت مشهد رهنمود سازد. در این راستا به نظر می‌رسد با توجه به اینکه خوشه‌بندی بر مبنای آنتروپی تاثیرپذیری بیشتری نسبت به مقادیر و تغییرات شاخص آلودگی نیترات دارد، ارزیابی وضعیت کیفی آبخوان بر این اساس قابل توجیه و کارآمد است.

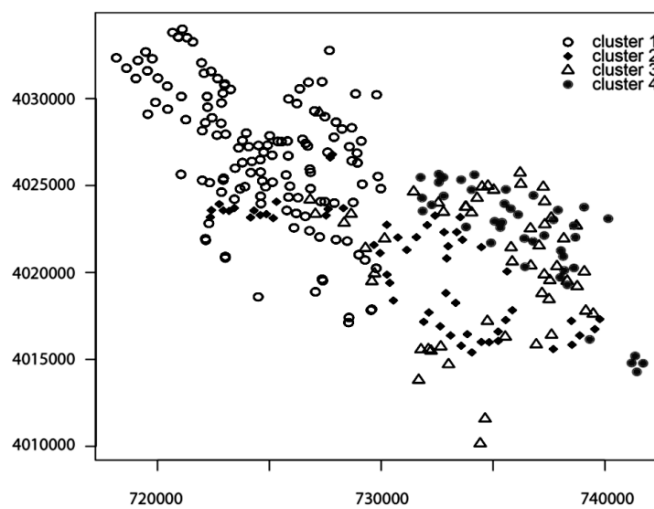
نمی‌شود. یعنی بین دو شاخص مکانی، با حذف طول جغرافیایی، چاه‌ها بصورت عرضی مرتب شده است. از طرفی پس از حذف شاخص نیترات، خوشه‌بندی بر مبنای معیار آنتروپی (SNI) به وسیله شاخص‌های مکانی انجام شد که در شکل ۶ ارائه شده است. نتایج مبین تاثیر مقادیر نیترات در تعیین خوشه‌بندی بر این مبنای می‌باشد و همان طور که ملاحظه می‌شود پس از حذف شاخص نیترات، هر خوشه شکل منظمی پیدا کرده است که در مقایسه، خوشه‌بندی ارائه شده در شکل ۴ منظم نمی‌باشد و این تغییرات با حذف نیترات صورت گرفته است.

از این رو با توجه به آنچه بیان شد نتایج حاکی از آن است که سنجش وابستگی بین نیترات چاه‌های مختلف توسط معیار آنتروپی سبب ایجاد خوشه‌هایی تحت تاثیر وابستگی مقادیر نیترات چاه‌ها شده



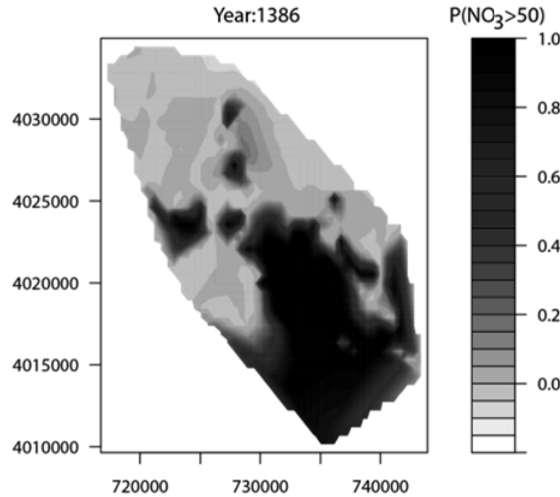
شکل ۳- پراکنش خوشه‌ها در روش فاصله اقلیدسی

Figure 3- Dispersion of clusters based on Euclidean distance



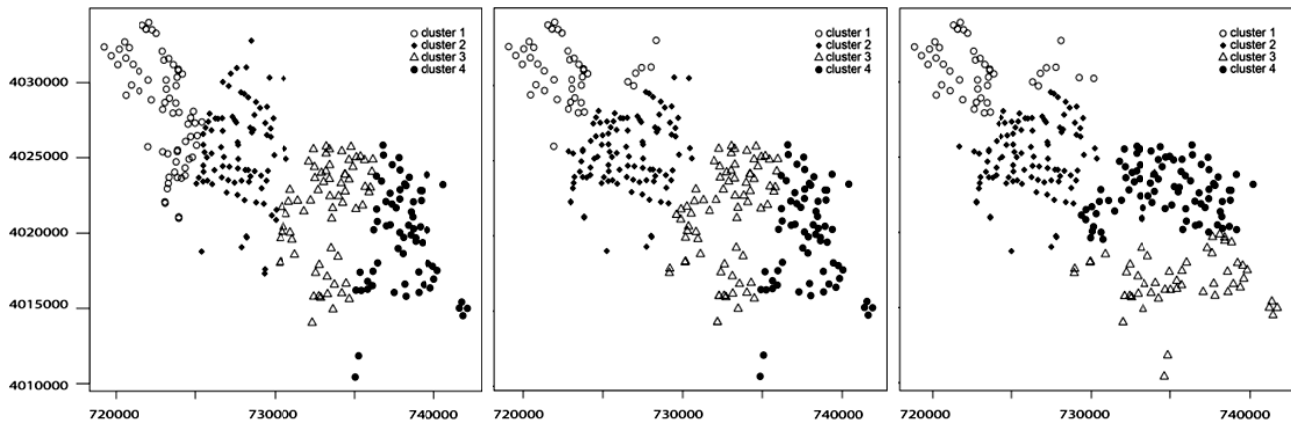
شکل ۴- پراکنش خوشه‌ها در روش آنتروپی (SNI)

Figure 4- Dispersion of clusters based on Entropy measure (SNI)



شکل ۵- منحنی‌های هم احتمال گذر نیترات آبخوان مشهد از حد مجاز آلودگی (۱)

Figure 5- Iso-probability contours of nitrate passed the pollution allowed limit (regulatory threshold) (1)



شکل ۶- خوشه بندی بر مبنای فاصله اقلیدسی با شاخص‌های عرض جغرافیایی و نیترات (راست)، بر مبنای فاصله اقلیدسی با شاخص‌های طول جغرافیایی و نیترات (وسط) و بر مبنای آنتروپی با شاخص‌های طول و عرض جغرافیایی (چپ)

Figure 6- Clustering based on Euclidean distance with latitude and nitrate index (right) Clustering based on Euclidean distance with longitude and nitrate index (middle) clustering based on entropy with latitude and longitude index (left)

که ملاحظه می‌شود میانگین معیارهای اعتباریابی همگنی ( $H_1$  و  $H_2$ ) بر مبنای آنتروپی (SNI) نسبت به فاصله اقلیدسی کمتر است. اما این تفاوت به صورت یکنواخت برای همه خوشه‌ها برقرار نمی‌باشد. به طور مثال در خوشه ۳، همگنی بر مبنای فاصله اقلیدسی بسیار بهتر عمل کرده است. اما در مجموع با توجه به مقایسه موارد دیگر و میانگین، معیار آنتروپی (SNI) عملکرد بهتری دارد. نتایج این پژوهش با پژوهش ریانا و همکاران (۲۰) همخوانی داشته و بر برتری روش آنتروپی بر فاصله اقلیدسی در حجم بالای ایستگاه‌ها، صحت می‌گذارد.

در انتها، برای هر روش خوشه‌بندی آزمون همگنی انجام شد که مقادیر آماره‌های  $H_1$  و  $H_2$  در جداول ۵ و ۶ ارائه شده است. برای تولید این جداول ابتدا در هر یک از خوشه‌ها معیارهای همگنی محاسبه گردید. سپس میانگین هر معیار همگنی بدست آمد. از این رو می‌توان با توجه به میانگین هر یک از آماره‌ها در جداول ۵ و ۶ نسبت همگنی هر یک از روش‌ها را نتیجه‌گیری نمود. با توجه به نتایج آزمون همگنی، هر دو روش، خوشه‌بندی همگنی را تشکیل داده‌اند اما خوشه‌بندی بر مبنای معیار SNI نتایج بهتری را در بر دارد. همان طور

جدول ۵- نتایج آزمون اعتباریابی همگنی برای خوشه بندی بر مبنای معیار فاصله اقلیدسی

Table 5- Results of homogeneity validation test for clustering based on Euclidean distance

معیار همگنی	خوشه ۱	خوشه ۲	خوشه ۳	خوشه ۴	میانگین
Homogeneity measure	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Average
$H_1$	1.990	-0.039	0.532	-0.471	0.503
$H_2$	1.210	-0.199	-0.095	-0.471	0.109



جدول ۶- نتایج آزمون اعتباریابی همگنی برای خوشه بندی بر مبنای معیار آنترپوی (SNI)

Table 6- Results of homogeneity validation test for clustering based on Entropy measure (SNI)

معیار همگنی	خوشه ۱	خوشه ۲	خوشه ۳	خوشه ۴	میانگین
Homogeneity measure	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Average
H <sub>1</sub>	0.343	-0.195	1.129	-0.472	0.201
H <sub>2</sub>	-0.050	-0.397	1.420	-0.576	0.099

## نتیجه گیری کلی

نمود، در معیار آنترپوی بر پایه احتمال رخداد مقادیر نیترات، تفاوت بین چاه‌ها و همبستگی بین آن‌ها تعیین می‌شود اما در معیار اقلیدسی فقط با توجه به تفاوت مقادیر نیترات، فاصله تشابه چاه‌ها تعیین می‌شود. بنابراین با توجه به این که معیار آنترپوی بر پایه احتمال رخ داد مقادیر نیترات، تفاوت بین چاه‌ها را بیان می‌کند نسبت به معیار اقلیدسی عملکرد بهتری دارد. بنابراین با خوشه‌بندی مبتنی بر آنترپوی می‌توان چاه‌های همگن از نظر تغییرپذیری شاخص آلودگی نیترات را شناسایی نمود که این امر می‌تواند گامی مهم و موثر در راستای ارزیابی وضعیت کیفی آبخوان مشهد باشد. به عبارت دیگر دو عامل شاخص آلودگی و معیار فاصله تشابه، نتیجه خوشه‌بندی و به تبع آن نتیجه ارزیابی کیفی را تحت تاثیر قرار می‌دهند.

در آینده پژوهش و در راستای ارزیابی و بهینه سازی شبکه پایش، می‌توان بر لزوم بهینه سازی شبکه موجود و انتخاب رویکرد آن تاکید کرد. بر این اساس انتظار می‌رود هر چه شبکه پایش خوشه‌های کمتری به خود اختصاص دهد، همگنی بیشتری داشته باشد. بنابراین رویکرد بهینه‌سازی از افزایشی به کاهش (تعدیل شبکه پایش) توجیه خواهد داشت. در این صورت هزینه‌های پایش اعم از حفر، تجهیز، نمونه‌برداری، نگهداری و آنالیز آزمایشگاهی نیز کاهش می‌یابد.

در پژوهش حاضر، بر پایه تحلیل خوشه‌ای سلسله مراتبی و با استفاده از داده‌های شاخص آلودگی نیترات در موقعیت ۲۸۷ حلقه چاه و در دوره آماری سال‌های ۱۳۸۱ تا ۱۳۸۹، خوشه‌هایی همگن تولید شد که ایستگاه‌هایی با تغییرپذیری یکسان را در خود جای می‌دهند. فرایند مذکور با توجه به پژوهش‌های پیشین بر مبنای دو معیار فاصله اقلیدسی و آنترپوی (SNI) انجام شد. نتایج حاکی از برتری خوشه‌بندی بر مبنای معیار فاصله آنترپوی است. هر چند با توجه به نتایج، چنین به نظر می‌رسد که از نظر جغرافیایی خوشه‌بندی بوسیله فاصله اقلیدسی همگنی بیشتری دارد اما از نظر تغییرپذیری شاخص آلودگی نیترات، خوشه‌بندی با معیار آنترپوی عملکرد بهتری دارد. همچنین به منظور ادله بیشتر برای اثبات فرضیه تاثیر شاخص آلودگی نیترات در خوشه‌بندی بر مبنای آنترپوی، با حذف این شاخص، فرایند خوشه‌بندی به طور مجدد انجام شد. بر این اساس با حذف شاخص نیترات نتایج بسیار تحت تاثیر شاخص‌های مکانی (مختصات مکانی) قرار گرفت. همچنین با حذف شاخص‌های مکانی از فرایند خوشه‌بندی مشخص شد که در خوشه‌بندی با معیار فاصله اقلیدسی، تاثیرپذیری از مقادیر نیترات بسیار کم تر است. این موضوع را می‌توان با توجه به ماهیت این دو معیار برای بیان تفاوت بین چاه‌ها تفسیر

## منابع

1. Akbarzadeh M., Ghahraman B., and Davary, K. 2016. Evaluation of groundwater quality in mashhad aquifer using the indicator kriging based on nitrate pollution. *Iranian Journal of Irrigation and Drainage*, 1(10):48-62.
2. Anderberg, M. R. 1973. *Cluster Analysis for Applications*. Academic Press, New York, USA.
3. Bailey, K.D. 1994. *Typologies and Taxonomies: An Introduction to Classification Techniques*. SAGE Publications, Inc., USA.
4. Burn, D.H. 1990. An appraisal of the "region of influence" approach to flood frequency analysis. *Journal of Hydrological Sciences*, 35(2):149-165.
5. Burn, D.H. 1990. Evaluation of regional flood frequency analysis with a region of influence approach. *Journal of Water Resources Research*, 26(10): 2257-2265.
6. Cavadias, G.S., Ouarda, T. B. M. J., Bobee, B. and Girard, C. 2001. A canonical correlation approach to the determination of homogeneous regions for regional flood estimation of ungauged basins. *Journal of Hydrological Sciences*, 46(4): 499-512.
7. Cover, T.M., and Thomas, J.A. 2006. *Elements of Information Theory*. 2nd ed., John Wiley & Sons, Inc., New Jersey, USA.
8. Estivill-Castro, V. 2002. Why so many clustering algorithms: A position paper. *ACM SIGKDD Explorations Newsletter*, 4(1): 65-75.
9. Gan, G., Ma, C., Wu, J. 2007. *Data Clustering: Theory, Algorithms and Applications*. ASA-SIAM, Philadelphia.

USA.

10. Glatfelter, D.R. 1984. Techniques for estimating magnitude and frequency of floods on streams in Indiana. US Geological Survey. Water Resources Investigations Report, 84-4134.
11. Haining, R. J. 1993. Spatial Data Analysis in the Social and Environmental Sciences. Cambridge University Press, Cambridge. UK.
12. Hosking J.R.M., Wallis J.R. 2013. Regional Frequency Analysis (An Approach Based on L-Moments). Translated by: B. Ghahraman. Tanin Ghalam, Mashhad.
13. Hosking, J., Wallis, J. 1993. Some statistics useful in regional frequency analysis. *Journal of Water Resources Research*, 29(2):271-281.
14. 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.
15. Mao, J., and Jain, A. 1996. A self-organizing network for hyper ellipsoidal clustering (HEC). *IEEE Transactions on Neural Networks*, 7(1): 16-29.
16. Murphy, K.P. 2012. *Machine Learning: A Probabilistic Perspective*. MIT Press, Cambridge.
17. Ngongondo, C.S., Xu, C.Y., Tallaksen, L.M., Alemaw, B., and Chirwa, T. 2011. Regional frequency analysis of rainfall extremes in Southern Malawi using the index rainfall and L-moments approaches. *Journal of Stochastic Environmental Research & Risk Assessment*, 25:939-955.
18. Rajsekhar, D., Mishra, A., and Singh, V.P. 2011. Drought regionalization of Brazos river using an entropy approach. *Proceedings of symposium on data-driven approaches to drought*, Purdue University, West Lafayette, Indiana. June 21-22.
19. Rao, A. R. and Srinivas, V. V. 2006. Regionalization of watersheds by hybrid-cluster analysis. *Journal of Hydrology*, 318(1-4): 37-56.
20. Rianna, M., Ridolfi, E., Lorino, L., Alfonso, L., Montesarchio, V., Di Baldassarre, G., Russo, F., Napolitano, F. 2012. Definition of homogeneous regions through entropy theory. *Proceedings of 3rd STAHY International Workshop on Statistical Methods for Hydrology and Water Resources Management*. STAHY, Tunis, Tunisia, October 1-2.
21. 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.
22. Shahedi M., Sanaiinejad S.H., Ghahraman B. 2013. Regional frequency analysis of annual maximum 1-day and 2-day rainfalls using Clustering and L-moments, case study: Khorasan Razavi province. *Journal of Water and Soil*, 27(1):80-89.
23. Shannon, C.E. 1948. A mathematical theory of communication. *Journal of Bell System Technical*, 27: 623-656.
24. Singh, V. P. 1997. The use of entropy in hydrology and water resources. *Journal of Hydrological Processes*, 11: 587-626.
25. Singh, V.P. 2013. *Entropy Theory and its Application in Environmental and Water Engineering*. Wiley-Blackwell, John Wiley & Sons, Ltd., UK.
26. Yang, T., Shao, Q., Hao, Z.C., Chen, X., Zhang, Z., Xu, C.Y., and Sun L. 2010. Regional frequency analysis and spatio-temporal pattern characterization of rainfall extremes in the Pearl River Basin, China. *Journal of Hydrology*, 380:386-405.

## Identification of Homogeneous Stations for Quality Monitoring Network of Mashhad Aquifer Based on Nitrate Pollution

M. Akbarzadeh<sup>1\*</sup> - B. Ghahraman<sup>2</sup> - K. Davary<sup>3</sup>

Received: 22-04-2015

Accepted: 01-08-2015

**Introduction:** For water resources monitoring, Evaluation of groundwater quality obtained via detailed analysis of pollution data. The most fundamental analysis is to identify the exact measurement of dangerous zones and homogenous station identification in terms of pollution. In case of quality evaluation, the monitoring improvement could be achieved via identifying homogenous wells in terms of pollution. Presenting a method for clustering is essential in large amounts of quality data for aquifer monitoring and quality evaluation, including identification of homogeneous stations of monitoring network and their clustering based on pollution. In this study, with the purpose of Mashhad aquifer quality evaluation, clustering have been studied based on Euclidean distance and Entropy criteria. Cluster analysis is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense or another) to each other than to those in other groups (clusters). SNI as a combined entropy measure for clustering calculated from dividing mutual information of two values (pollution index values) to the joint entropy. These measures apply as similar distance criteria for monitoring stations clustering.

**Materials and Methods:** First, nitrate data (as pollution index) and electrical conductivity (EC) (as covariate) collected from the related locational situation of 287 wells in statistical period 2002 to 2011. Having identified the outlying data and estimating non-observed points by spatial-temporal Kriging method and then standardizes them, the clustering process was carried out. A similar distance of wells calculated through a clustering process based on Euclidean distance and Entropy (SNI) criteria. This difference explained by characteristics such as the location of wells (longitude & latitude) and the pollution index (nitrate). Having obtained a similar distance of each well to others, the hierarchical clustering was used. After calculating the distance matrix, clustering of 287 monitoring stations (wells) was conducted. The optimal number of clusters was proposed. Finally, in order to compare methods, the validation criteria of homogeneity (linear-moment) were used. The research process, including spatial-temporal Kriging, clustering, silhouette score and homogeneity test was performed using R software (version 3.1.2). R is a programming language and software environment for statistical computing and graphics supported by R foundation for statistical computing.

**Results and Discussion:** Considering 4 clusters, the silhouette score for Euclidean distance criteria was obtained 0.989 and for entropy (SNI) was 0.746. In both methods, excellent structure was obtained by 4 clusters. Since the values of  $H_1$  and  $H_2$  are less, clusters will be more homogeneous. So the results show the superiority of clustering based on entropy (SNI) criteria. However, according to the results, it seems there is more homogeneity of clustering with Euclidean distance in terms of geography, but the measure of entropy (SNI) has better performance in terms of variability of nitrate pollution index. To prove the nitrate pollution index effectiveness in clusters with entropy criteria, the removal of nitrate index, the results was influenced by location index. Also, by removing index locations from clustering process it was found that in clusters with Euclidean distance criteria, the influence of nitrate values is much less. Also, compared to Euclidean distance, better performance was obtained by Entropy based on probability occurrence of nitrate values.

**Conclusion:** Results showed that the best clustering structure will obtain by 4 homogenous clusters. Considering wells distribution and average of the linear-moment, the method based on entropy criteria is superior to the Euclidean distance method. Nitrate variability also played a significant role in identification of homogeneous stations based on entropy. Therefore, we could identify homogenous wells in terms of nitrate pollution index variability based on entropy clustering, which would be an important and effective step in Mashhad aquifer monitoring and evaluation of its quality. Also, in order to evaluate and optimize the monitoring network, it could be emphasized on network optimization necessity and approach selection. Accordingly, less monitoring network clusters lead more homogeneous. Therefore the optimization approach will be justified from increasing to decreasing. In this case the monitoring costs, including drilling, equipment, sampling, maintenance

1, 2 and 3- Ph.D., Research Lab Specialist and Professors Department of Water Engineering, Ferdowsi University of Mashhad

(\* - Corresponding Author Email: akbarzade@staff.um.ac.ir)

and laboratory analysis, also reduce.

**Keywords:** Clustering, Entropy, Euclidean distance, Groundwater, Linear-moment, Silhouette score, Spatial-temporal Kriging