اکوهیدرولوژی

دورهٔ ۵، شمارهٔ ۱، بهار ۱۳۹۷، ص ۳۰۵-۲۹۳

تحلیل چندپارامترهٔ آلودگی آبخوان قزوین بر مبنای نقشهٔ کاربری اراضی و با استفاده از تکنیک خوشه‌بندی K-means

سامان جوادی[1*]، مهدی هاشمی[2]، مهدی سوخته‌زاری[3]

۱. استادیار گروه مهندسی آب، پردیس ابوریحان، دانشگاه تهران

۲. استادیار گروه مهندسی آب، پردیس ابوریحان، دانشگاه تهران

۳. کارشناسی ارشد عمران آب، گرایش منابع آب

(تاریخ دریافت ۱۳۹۶/۰۱/۳۱؛ تاریخ تصویب ۱۳۹۶/۰۵/۳۱)

چکیده

از راهکارهای اساسی برای پیشگیری افزایش آلودگی آب زیرزمینی، شناسایی مناطق بحرانی آلوده‌شده با توجه به نوع کاربری در آن آبخوان است. در بیشتر پژوهش‌های پیشین منابع آلاینده و آلودگی آبخوان به‌صورت تک‌پارامتری یا منحصر به یک نوع کاربری در نظر گرفته شده و تحلیل کیفی آبخوان فقط مربوط به یک آلاینده است. از این‌رو، در تحقیق حاضر با به‌کارگیری تکنیک خوشه‌بندی و با توجه به کاربری اراضی وضع موجود، نقشهٔ پهنه‌بندی کیفی بر اساس ترکیب چند پارامتر استخراج شد. به بیان دیگر، با درنظرگرفتن سه پارامتر کیفی نیترات (NO_3)، آرسنیک (As) و کل مواد جامد محلول (TDS)، به‌ترتیب برای سه نوع کاربری کشاورزی، صنعتی و شرب با به‌کارگیری تکنیک خوشه‌بندی کلاسیک، پهنه‌بندی آلودگی آبخوان قزوین تهیه شد. تعداد خوشه‌های بهینه بر اساس شاخص دیویس- بولدین پنج خوشه تعیین شد. نتایج نشان می‌دهد با توجه به استاندارد جهانی آب شرب و درنظرگرفتن سه پارامتر منتخب، مناسب‌ترین خوشه (C1) با مساحت ۲۲ درصد در نواحی شمالی آبخوان بوده و به سمت نواحی مرکزی، که تمرکز فعالیت‌های کشاورزی و صنایع است، آلودگی در آبخوان (خوشه‌های C4 و C5) با مساحت ۳۵ درصد به عنوان نامناسب‌ترین منطقه شناخته شده است.

کلیدواژگان: آبخوان قزوین، آلودگی آب زیرزمینی، پهنه‌بندی آلودگی، تحلیل کیفی چندپارامتره، خوشه‌بندی.

* نویسندهٔ مسئول

Email: javadis@ut.ac.ir

مقدمه

آب‌های زیرزمینی از مهم‌ترین منابع آبی کشور به‌ویژه در مناطق خشک و نیمه‌خشک هستند و در شرایط کنونی بخش درخور توجهی از مصارف آب کشور ایران، به‌خصوص در بخش شرب، توسط منابع آب زیرزمینی تأمین می شود. بنابراین، حفاظت کیفی آب زیرزمینی اهمیت دارد، به‌طوری که در بسیاری از موارد در کشور آلودگی آب‌های زیرزمینی بعد از آلوده‌شدن چاه‌های آب شرب شناسایی می‌شوند. از طرف دیگر، رفع آلودگی آب‌های زیرزمینی بسیار پرهزینه بوده و فرایند آن بسیار طولانی است و در بیشتر مواقع رفع کامل آن نیز امکان‌پذیر نیست. معمولاً اغلب زمانی آلودگی تشخیص داده می‌شود که رفع آلودگی آبخوان تقریباً غیرممکن می شود. بنابراین، شناسایی، بررسی و تحلیل کیفی این منابع ارزشمند به‌منظور حفظ آنها برای نسل‌های بعدی، امری اجتناب‌ناپذیر و ضروری است [۱]. مدیریت کیفی آب‌های زیرزمینی در گام نخست مستلزم شناخت کافی و جامع از همهٔ آلاینده‌های موجود و میزان آن در آبخوان است و در گام دوم نیاز به ابزاری دارد تا بتوان واکنش تنش‌های مختلف کمی و کیفی وارد به سفره را در شرایط فعلی و آینده پیش‌بینی کرد. بدون شک، بهترین حالت شناخت رفتارهای یک سیستم سفرهٔ آب زیرزمینی، تحقیقات درازمدت برای هر منطقه است که با توجه به میزان محدود بودجه‌های تحقیقاتی، تا حدود زیادی در کشور امکان‌پذیر نیست. در این میان، با ابزارهای محاسباتی و شبیه‌سازها می‌توان با دقت قابل قبول به شاخص‌هایی دست یافت که بر اساس آن بتوان تصمیم‌های لازم را در خصوص مدیریت آبخوان به کار برد [۲].

تا کنون بیشتر روش‌های به‌کاررفته در مطالعات و تحلیل کیفی آب زیرزمینی در کشور به‌صورت روش‌های گرافیکی بوده که معمولاً نتایج تجزیهٔ کیفی نمونه‌های آب توسط نمودارهای مختلفی نظیر پایپر، استیف، شولر و ویل‌کاکس تحلیل می‌شد [۳ و ۴]. یکی از محدودیت‌های اصلی این روش‌ها تعداد نمونه و متغیر استفاده شده در این نمودارهاست، به‌طوری که در بیشتر آنها از آنیون‌ها و کاتیون‌ها استفاده شده است و پارامترهای آلودگی همچون نیترات و فلزات سنگین در این نمودارها جایی ندارند. همچنین، نتایج خروجی از این نمودارها وضعیت کامل و جامعی را از همهٔ ابعاد کیفی سفرهٔ آب زیرزمینی بیان نمی‌کند. بنابراین، ضروری است از روش‌ها و تکنیک‌هایی استفاده شود که بتوان پارامترهای مختلفی را در یک نقشه ادغام کرد و نقشهٔ استخراج‌شده همهٔ ابعاد کیفی آبخوان را با توجه به نوع کاربری اراضی روی سطح آن نشان دهد. به بیان دیگر، در یک نقشهٔ واحد می‌توان سفرهٔ آب زیرزمینی را از چند منظر مانند شرب، صنعت و کشاورزی تحلیل کرد. حال آنکه در مطالعات گذشته تحلیل‌ها بر اساس کیفیت یک پارامتر و یا یک نوع کاربری خاص در نظر گرفته می‌شد [۵]. از این‌رو، در مطالعات کیفی آب زیرزمینی ضروری است از روش‌هایی استفاده شود تا بتوان تأثیر چندین پارامتر کیفی را هم‌زمان با توجه به نوع کاربری اراضی موجود در سطح آبخوان، در یک نقشه تحلیل کرد.

یکی از روش‌های آماری برای ارائهٔ نقشهٔ پراکندگی و توزیع کیفی چندپارامتره در سفره‌های آب‌های زیرزمینی، استفاده از تحلیل خوشه‌ای است. اگر کیفیت آب‌های زیرزمینی مناطق مختلف یک دشت از نظر چند پارامتر، خواص اندازه‌گیری مشابه داشته باشند، در فضای n بعدی بسیار نزدیک به یکدیگر قرار می‌گیرند. بنابراین، مشابهت‌های این نمونه‌ها به کمک اندازه‌گیری فاصلهٔ بین آنها در این فضا بررسی می شود که نتیجهٔ آن شاخصی به نام ضریب مشابهت است و شباهت کیفیت منابع آب را تعیین می‌کند. هدف از تحلیل خوشه‌ای، گروه‌بندی انواع متغیرها است که این متغیرها می‌توانند انواع پارامترهای کیفی آب باشند. ممکن است متغیرهای داخل یک گروه تا حد امکان همگن باشند، ولی خصوصیات آنها در بین دیگر گروه‌ها متمایز خواهد بود، بنابراین در یک گروه مشخص قرار می‌گیرند [۶].

همان گونه که قبلاً نیز یاد شد اغلب روش‌های نموداری استفاده شده، محدودیت تعداد نمونه را داشته و به خلاف روش‌های نموداری، روش‌های آماری همچون خوشه‌بندی قابلیت استفاده از همهٔ داده‌ها را خواهند داشت. البته، شایان یادآوری است محدودیتی که روش‌های تحلیل خوشه‌ای (اغلب روش‌های آماری) نسبت به روش‌های نموداری دارند این است که اطلاعاتی دربارهٔ ترکیب شیمیایی نمونه‌ها را در هر گروه ارائه نمی‌کنند. به بیان دیگر، در روش تحلیل خوشه‌ای همچون مدل‌های فراکاوشی، نوع داده‌ها برای این روش‌ها فرقی نداشته و فقط ترکیب مکانی آنها در یک نقشه تعیین‌کنندهٔ تعداد خوشه‌هاست. اگرچه تحلیل خوشه‌ای روشی بسیار کارآمد در طبقه‌بندی نمونه بر پایهٔ شباهت‌های فیزیکی و شیمیایی است، نتایج آن را نمی‌توان به‌سرعت در تفسیر روند و مشکلات مربوط به فرایندهای شیمیایی آب به

کـار بـرد. ترکیـب دو روش خوشـه‌بنـدی و نمـوداری سـبب می‌شود که مزایای هر روش باقی بماند و محـدودیت‌هـای هـر روش به کمترین حد برسد [۷].

تا کنون از تکنیک خوشـه‌بنـدی در شـاخه‌هـای مختلـف مهندسی آب استفاده شده است که از آن جمله مـی‌تـوان بـه مطالعات کیفی آب سطحی، زیرزمینی، طبقه‌بندی رودخانه‌هـا و حوضه‌هـای آبریـز اشـاره کـرد. یکـی از مراحـل مهـم روش خوشه‌بندی، نوع الگوی به‌کـار گرفتـه‌شـده بـرای گـروه‌بنـدی داده‌هاست. در ابتدایی‌ترین تحقیقـات صـورت‌گرفتـه بـا روش خوشه‌بندی در خصوص کاربردهای تشخیص الگو در داده‌های علوم آب، از آنالیزهای جداکنندهٔ[1] کلاسیک اسـتفاده شـد [۸]. روش‌های خوشـه‌بنـدی بـه دو گـروه کلـی سلسـله‌مراتبـی و تجزیـه‌ای[2] تفکیـک شـده‌انـد. روش‌هـای خوشـه‌بنـدی سلسله‌مراتبی مجموعـه‌ای از اجـزای بـه‌هـم‌پیوسـته در یـک نمودار درختی[3] به وجود می‌آورند. با استفاده از نمودار درختی میزان مشابهت در سطوح مختلف تعیین شده و خوشـه‌بنـدی انجام می شود. زمانی که خوشه‌بندی دربارهٔ مجموعـه‌هـایی بـا تعداد زیاد داده اعمال مـی‌شـود، روش سلسـله‌مراتبـی عملـی نیست، زیرا ساخت نمودار درختی از نظر محاسباتی وقـت‌گیـر و پیچیده خواهد بود. در ایـن مـوارد روش‌هـای خوشـه‌بنـدی تجزیه‌ای پیشنهاد می شود. پراستفاده‌ترین تابع استاندارد به‌کار گرفته‌شده در روش خوشه‌بندی تجزیه‌ای، معیار خطای مربع[4] اسـت [۸]. در ایـن میـان روش K-Means آسـان‌تـرین و پرکاربردترین الگوریتم استاندارد خطای مربع است. ایـن روش با یک تقسیم‌بندی تصادفی اشیاء شروع به کـار مـی‌کنـد و در ادامه براساس میزان مشابهت اشیاء و اصلاح تقسیم‌بندی قبلی ادامه پیدا می‌کند. روش K-Means شـهرت زیـادی دارد زیـرا به‌راحتی قابل اجرا است و پیچیدگی اجرا ندارد [۸]. در ادامـه، به چند مورد از کاربردهای روش‌های خوشه‌بندی در مطالعـات منابع آب پرداخته خواهد شد.

در تحقیقـی Usman و همکـارانش در سـال ۲۰۱۴ بـا اسـتفاده از تکنیـک‌هـای آمـاری چنـدمتغیره ماننـد تحلیـل خوشه‌ای (CA)، تحلیل مشخص (DA) و تحلیـل مؤلفـه‌هـای اصـلی (PCA) بـا اهـداف تعیـین تغییـرات مکـانی آب‌هـای زیرزمینی و شناسایی منابع آلودگی که در حال حاضر آب‌های زیرزمینـی را تحـت تـأثیر قـرار داده را در منطقـهٔ ترنگگـانوی مالزی بررسی کردند. فاکتورهای اصـلی بـه‌دسـت‌آمـده نشـان می‌دهد پارامترهای مؤثر در خوشه‌بندی کیفیت آب زیرزمینی بیشتر مرتبط به آلودگی‌های طبیعی (ازهم‌پاشـیدگی خـاک و سنگ)، آلودگی‌های متمرکز (فاضلاب‌های شهری و صنعتی) و غیرمتمرکز (کشـاورزی) در منطقـه اسـت [۹]. همچنـین، در تحقیقی دیگر Zou و همکارانش در سال ۲۰۱۵ بـا اسـتفاده از الگـوریتم k-means کیفیـت آب رودخانـهٔ Haihe را در چـین بررسی کردند که نتایج شبیه‌سازی نشان می‌دهد این الگوریتم و الگو برای تجزیه و تحلیل کیفیت آب رودخانهٔ Haihe کارآمد و قابل اعتماد است. عـلاوه بـر ایـن، الگـوریتم را مـی‌تـوان بـه ماتریس داده‌های پیچیده‌تـر بـا ابعـاد بـالا بـرای بـه‌کـارگیری پارامترهای بیشتر استفاده کرد [۱۰]. در مطالعه‌ای دیگـر نیـز Azhar و همکـارانش در سـال ۲۰۱۵ اقـدام بـه طبقـه‌بنـدی کیفیت آب نُه ایستگاه حوضهٔ رودخانهٔ مـودا در شـمال غربـی مالزی با استفاده از تکنیک آماری تحلیل خوشه‌ای کردند کـه نتیجهٔ به‌دست‌آمده از این تحقیق نشان می‌دهد تحلیـل نتـایج متغیرهای اندازه‌گیری‌شده از وضعیت کیفیت آب رودخانـه بـا طبقه‌بندی آب توسط سازمان حفاظت محیط زیست آن کشور همبسـتگی زیـادی دارد [۱۱]. در تحقیقـی دیگـر، نیـز Oorkavalan و همکارانش با اسـتفاده از تحلیـل خوشـه‌ای در سال ۲۰۱۶ ارزیابی کیفیت آب زیرزمینی منطقه‌ای در جنـوب شبه‌قارهٔ هنـد بـرای شـرب، مصـارف خـانگی و کشـاورزی بـا استانداردهای WHO را بررسی کردند. نتیجهٔ این تحقیق نشان می‌دهد تحلیل خوشه‌ای، گروه‌بنـدی بـر اسـاس ویژگـی‌هـای آلودگی کیفیت آب زیرزمینی را بر اساس استاندارد پشـتیبانی می‌کند. این مطالعه همچنین نشان می‌دهد تحلیل‌های آماری چندمتغیره به‌منظور درک بهتر آب‌هـای زیرزمینـی و ارزیـابی کیفیت آب زیرزمینی استفاده می‌شود [۱۲].

همـان گونـه کـه مشـاهده مـی‌شـود، بیشـتر تحقیقـات صورت‌گرفته بـه‌منظـور مطالعـات منـابع آب سـطحی بـوده و تحقیقات انجام‌شده در خصوص آب‌های زیرزمینی نیز معطوف به آنیون‌ها و کاتیون‌هاست. از طرف دیگر، استفادهٔ روزافزون از تکنیک خوشه‌بندی در تحلیل کیفی منـابع آب مـی‌توانـد بـه عنوان مزیت تحقیقات یادشده لحاظ شود، اما تک‌بعـدی‌بـودن مطالعات یادشده که فقـط منحصـر بـه یـک نـوع کـاربری یـا آلودگی است می‌تواند جزء معایـب تحقیقـات در نظـر گرفتـه شود. به‌طور نمونه، در تحقیـق Oorkavalan تنهـا پارامترهـای

1. Discriminant Analysis
2. Partitional
3. Dendrogram
4. Squared Error Algorithms

آنیون‌ها و کاتیون‌ها (که معمولاً یک رفتار کیفی دارد) به عنوان کیفیت آبخوان در نظر گرفته‌شده و به پارامتر آلودگی مربوط به کاربری صنعتی در منطقه اشاره‌ای نشده است. حال آنکه در تحقیق حاضر ترکیبی از پارامترهای مختلف آلودگی با توجه به نوع کاربری مد نظر قرار گرفته است و نقشهٔ خروجی و تحلیل کیفی آن، منحصر به یک ویژگی کاربری نظیر شرب یا کشاورزی نیست. همچنین، در خصوص تأثیر کاربری اراضی بر کیفیت آب زیرزمینی نیز تحقیقاتی صورت گرفته است که می‌توان به پژوهش Ouyang و همکارانش در سال ۲۰۱۴ اشاره کرد که در آن تحقیق اثر سه نوع کاربری کشاورزی، مخازن کودهای شیمیایی و جنگل بر میزان نیترات، فسفات موجود در آب زیرزمینی در یک دورهٔ شش‌ساله با استفاده از گراف‌های مثلثی بررسی شد. نتایج نشان می‌دهد نیترات و فسفات در نواحی مخازن کود بیشترین میزان را در آب زیرزمینی داراست [۵]. در پژوهشی دیگر Fianko و همکارانش در سال ۲۰۰۹ تأثیر کاربری اراضی را بر کیفیت آب زیرزمینی از لحاظ آنیون‌ها و کاتیون‌ها بررسی کردند. در این تحقیق مشخص شد غلظت کلر و TDS در مناطق شهری و میزان سدیم، کلسیم و سولفات در مناطق کشاورزی بیشتر از حد مجاز است [۱۳]. همچنین، تحقیقات دیگری نیز در این زمینه صورت گرفته که در آنها نیز نتایج نسبتاً مشابهی ارائه شده است [۱۴ و ۱۵]. همان گونه که مشاهده می شود در کلیهٔ تحقیقات یادشده، پارامترهای منتخب با توجه به نوع کاربری اراضی به‌صورت مجزا بررسی شده‌اند. به طور نمونه، در یک مطالعه فقط اثر کاربری بر میزان نیترات بررسی شده در حالی که آثار هم‌زمان پارامترها با هم دیده نشده است. مهم‌ترین اثر کاربری اراضی در تعیین نوع آلودگی و شناسایی اولیهٔ آلودگی است. از آنجا که معمولاً هر کاربری آلایندهٔ خاصی را به آبخوان وارد می کند، به طور مثال، در جایی با کاربری کشاورزی فلزات سنگین افزایش نمی‌یابد، پس، با توجه به نوع کاربری اراضی، انتخاب پارامتر آلاینده برای جلوگیری از هزینه‌های بیشتر اندازه‌گیری، می‌تواند مفید باشد.

همان طور که مشاهده می‌شود، تأثیر هم‌زمان آلودگی‌های مختلف ناشی از کاربری‌های متفاوت به هیچ عنوان در تحقیقات گذشته دیده نمی‌شود. به بیان دیگر، در همهٔ تحقیقات رابطهٔ یک پارامتر آلودگی با یک نوع کاربری بررسی می شود، در صورتی که ضروری است اثر پارامترهای آلودگی با وجود کاربری‌های متفاوت هم زمان تجزیه و تحلیل شود.

از این‌رو، برای نخستین‌بار در تحقیق حاضر سعی شده است بر اساس نوع کاربری موجود، آلاینده‌های مد نظر انتخاب و سپس در گام بعدی با تکنیک خوشه‌بندی، نقشهٔ پهنه‌بندی آلودگی چندپارامتره در آبخوان قزوین استخراج شود که از نوآوری مهم این تحقیق به حساب می‌آید. به بیان دیگر، هدف از این مطالعه شناسایی مناطق همگن از نظر آلودگی آب زیرزمینی با استفاده از داده‌های سه پارامتر آلودگی مهم آب و تفکیک مناطق همگن در آبخوان قزوین با استفاده از تکنیک خوشه‌بندی K-Means است. در این پژوهش سه پارامتر کیفی کل مواد جامد محلول (TDS)، نیترات و آرسنیک به‌ترتیب برای کاربری‌های مختلف شرب، کشاورزی و صنعتی استفاده شده است. روش و الگوریتم استفاده‌شده در این تحقیق نیز خوشه‌بندی K-Means بوده که از پرکاربردترین روش‌های خوشه‌بندی کلاسیک است و در تحقیقات زیادی استفاده شده است. این روش مانند سایر روش‌های خوشه‌بندی کلاسیک قادر است مرزی مشخص بین خوشه‌ها ترسیم کرده و هر داده را فقط به یک خوشه اختصاص دهد، اما این قابلیت در مواردی که هدف به‌کارگیری تصمیم‌گیری قطعی از نتایج به دست آمده باشد، مطلوب است و در مواردی که داده‌ها وابستگی شدیدی به یک خوشه ندارند، تخصیص یک داده فقط به یک خوشه نتایج را غیرواقعی می‌سازد.

در نتیجه، با توجه به نقشهٔ کاربری و توزیع و ترکیب چند پارامتر آلودگی در منطقه، می‌توان برنامهٔ مدیریتی و تصمیم‌گیری برای پیشگیری آبخوان از آلودگی بیشتر را ارائه کرد.

منطقهٔ مطالعه‌شده

منطقهٔ مطالعه‌شده در این تحقیق دشت قزوین است که در تقسیمات کشوری، بخشی از استان قزوین را شامل می‌شود و در فاصلهٔ حدود ۱۵۰ کیلومتری شمال غربی تهران قرار گرفته است (شکل ۱). محدودهٔ مطالعاتی قزوین بین طول‌های شرقی ۱۰ دقیقه و ۴۹ درجه و ۴۰ دقیقه و ۵۰ درجه و عرض‌های شمالی ۲۰ دقیقه و ۳۵ درجه و ۲۰ دقیقه و ۳۶ درجهٔ جغرافیایی قرار گرفته و ارتفاع آن

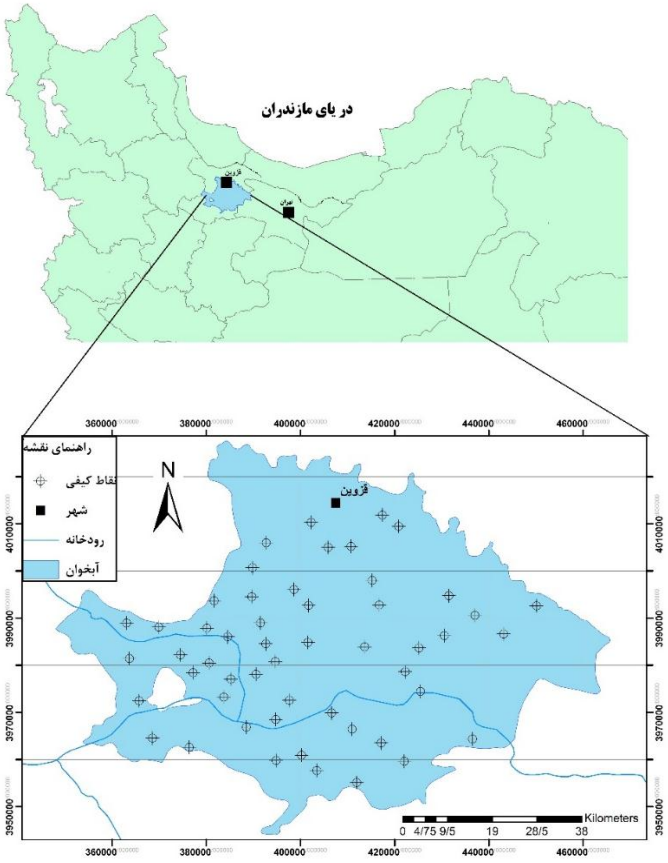
حداکثر برابر ۲۹۷۱ متر و حداقل ۱۱۰۰ متر و متوسط ارتفاع منطقه حدود ۱۲۵۰ متر از سطح دریاست.

تغییرات عمق سنگ کف آبخوان کمتر از ۷۰ متر تا بیش از ۳۷۰ متر در نوسان است. کمترین مقدار عمق سنگ کف مربوط به محدودهٔ شرقی دشت، که بخش باتلاقی دشت قزوین است، مشاهده می‌شود. در این محل سنگ کف بالا آمده و ضخامت کمی از لایه‌های ریزدانهٔ رسی روی آن قرار گرفته است. همین عامل سبب تمرکز آب در این منطقه و به‌وجودآمدن حالت باتلاقی در این قسمت از دشت شده است. همچنین، ضخامت منطقه اشباع در محدودهٔ دشت بین ۵۰ تا بیش از ۳۵۰ متر متغیر است. به طور کلی، در محدوده‌های ورودی و خروجی آبخوان قزوین ضخامت منطقهٔ اشباع آبخوان کم بوده که به دلیل عمق کم سنگ کف در این نواحی است. بررسی‌های انجام‌شده نشان می‌دهد زمین‌های آبرفتی منطقهٔ قزوین به غیر از منطقه شوره‌زار، نفوذپذیری مناسبی داشته و هر نوع مواد زائدی که در سطح زمین تولید شود، بر اثر بارندگی و یا مجاورت با منابع آبی در داخل زمین نفوذ می‌کند و مخازن آب‌های زیرزمینی، که به صورت پیوسته و وسیع در منطقه گسترش دارند، آلوده می‌شوند [۱۶]. به این دلیل آبخوان منطقهٔ یادشده از نظر استعداد آلودگی آب‌های زیرزمینی آن باید بررسی شود. شکل ۱ محدودهٔ آبخوان قزوین را به همراه نقاط نمونه‌برداری کیفی آن نشان می‌دهد.

مواد و روش‌ها

خوشه‌بندی

امروزه، میزان داده‌های تولیدشده و ذخیره‌شده به دلیل پیشرفت شگفت‌انگیز ابزارهای تولید، اندازه‌گیری و جمع‌آوری اطلاعات به طور چشمگیری افزایش یافته است. به علت وجود ظرفیت شناسایی دانش از داده‌های اندازه‌گیری‌شده، رشته‌های جداگانهٔ علمی ایجاد شد که به



شکل ۱. نقشهٔ موقعیت آبخوان قزوین به همراه نقاط نمونه‌برداری کیفی

علم «کشف دانش از بانک داده‌ها»[1] (KDD) شناخته می‌شود. داده‌کاوی در واقع کاربرد یک الگوریتم خاص برای استخراج الگوها از داده‌ها است که امکان استخراج دانش ناشناخته در زمینهٔ یک فرایند علمی و یا یک سیستم را فراهم می‌سازد. انواع مختلفی از روش‌های داده‌کاوی ارائه و توسعه داده شده‌اند که پرکاربردترین آنها خوشه‌بندی[2] است [۱۷]. بر این اساس در تحقیق حاضر، نتایج به‌دست‌آمده از خوشه‌بندی در هر مرحله از نظر توجیه‌پذیری در تحلیل آلودگی آبخوان بررسی شد.

هدف از آنالیز خوشه‌بندی، دسته‌بندی اشیاء براساس میزان مشابهت بین آنها و سازماندهی گروهی داده‌هاست. اصلی‌ترین پتانسیل خوشه‌بندی، شناسایی ساختار اصلی نهفته در داده‌هاست. روش‌های متفاوت طبقه‌بندی داده‌ها براساس الگوریتم‌های مختلف خوشه‌بندی ایجاد می‌شود [۱۸]. روش خوشه‌بندی برای داده‌های کمی (عددی)، کیفی و یا ترکیب این دو می‌تواند به کار بسته شود. داده‌های کمی یا عددی معمولاً از مشاهدات فرایندهای فیزیکی به‌دست می‌آیند. هر مشاهده شامل n متغیر اندازه‌گیری‌شده است که در یک بردار n بعدی $X_k=[x_{k1},x_{k2},...,x_{kn}]^T, X_k\in R^n,\ X_k$ قرار می‌گیرد. مجموعه‌ای از N مشاهدات به صورت $X=\{x_k|k=1,2,...,N\}$ نوشته می‌شود که بیان کنندهٔ یک ماتریس $N\times n$ است. روش means-K آسان‌ترین و پرکاربردترین الگوریتم استاندارد خطای مربع است. تابع هدف الگوریتم خطای مربع مطابق رابطهٔ ۱ است[۱۷]:

$$J(X;V)=\sum_{i=1}^{c}\sum_{k\in i}\left\|x_k^{(i)}-v_i\right\|^2 \qquad (۱)$$

که X عبارت از ماتریس داده‌ها یا اشیاء، V بردار حاوی مراکز خوشه‌ها، $x_k^{(i)}$، فاصلهٔ شیء kام از خوشهٔ i از مبدأ مختصات است و V_i فاصلهٔ مرکز جرم خوشهٔ i از مبدأ مختصات به شمار می‌رود. مرکز خوشه (v_i) از رابطهٔ ۲ به دست می‌آید [۱۹]:

$$v_i=\frac{\sum_{k=1}^{N_i}x_k}{N_i},\ x_k\in A_i \qquad (۲)$$

که در آن N_i تعداد اعضای مجموعه A_i و A_i مجموعه اعضای خوشهٔ iام و مانند رابطهٔ ۱ $x_k^{(i)}$ فاصلهٔ شیء kام از خوشهٔ i از مبدأ مختصات است.

1. (KDD) Knowledge Discovery in Database
2. Clustering

پیش‌پردازش[3] خوشه‌بندی

فرایند پیش‌پردازش داده‌ها، حذف داده‌های غیرواقعی و داده‌های پرت، نرمال‌سازی داده‌ها و نیز کاهش و انتخاب تعداد خصوصیت‌ها، ستون‌های ماتریس داده‌هاست. از این‌رو، در مطالعهٔ حاضر به علت صحت‌سنجی مقدماتی داده‌ها و نیز انتخاب آگاهانهٔ خصوصیات بررسی شدهٔ هر سلول مکانی، از روش نرمال‌کردن کمینه- بیشینه استفاده شده است. این روش بر فاصلهٔ میان نقاط کمینه $Min(X)$ و بیشینه $Max(X)$ تأکید دارد و مقیاس‌سازی را براساس تفاوت این دو مقدار طبق رابطهٔ ۳ انجام می‌دهد [۲۰]:

$$X^*=\frac{X_i-Min(X)}{Max(X)-Min(X)} \qquad (۳)$$

که در آن X^* متغیر نرمال‌سازی‌شده و X_i متغیر اصلی است. در این روش همهٔ مقادیر داده‌ها در محدودهٔ صفر تا یک قرار می‌گیرند [۲۰].

شاخص صحت‌سنجی دیویس- بولدین

به‌منظور ارزیابی تعداد بهینهٔ خوشه‌ها برای مجموعه‌ای از داده‌ها در اینجا از شاخص دیویس- بولدین استفاده شد [۲۱]. از نگاه این شاخص خوشه‌بندی خوب زمانی رخ می‌دهد که خوشه‌های ایجادشده کروی‌شکل، متراکم و قابل تفکیک از هم باشند. با فرض تفکیک n داده به g خوشه، مقدار کلی شاخص DB برای هر بار خوشه‌بندی از روابط ۴ و ۵ به دست می‌آید. مقدار این شاخص برای دو خوشهٔ i و k از رابطهٔ ۶ به دست می‌آید. با توجه به توضیحات ارائه‌شده، روابط ۴- ۷ به این شرح است [۲۰]:

$$I_{DB}=\frac{1}{g}\sum_{i=1}^{g}R_j. \qquad (۴)$$

$$R_j=\max_{k=1,2,...,g;k\neq j}R_{jk}. \qquad (۵)$$

$$R_{jk}=\frac{\sigma_j+\sigma_k}{\left\|\mu_j-\mu_k\right\|},\quad j,k=1,2,...,g;\quad k\neq j \qquad (۶)$$

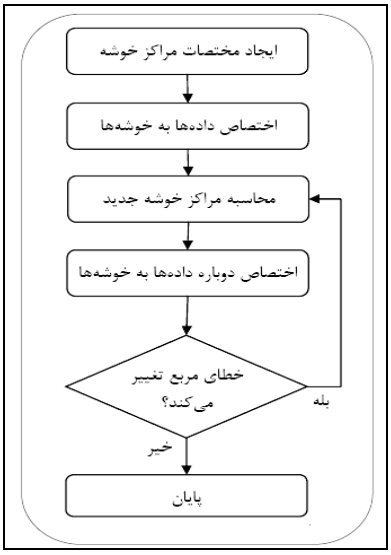
در این روابط I_{DB} مقدار شاخص دیویس- بولدین، n تعداد داده، g تعداد خوشه‌های انتخاب‌شدهٔ دو خوشه تا تعداد حداکثر، R_j شاخص مقایسهٔ دوبه‌دوی خوشه‌ها،

3. Pre-processing

R_{jk} معیار مقایسهٔ پراکندگی درونی و فاصلهٔ هر دو خوشهٔ j,k از هم است. همچنین، μ_j برابر میانگین فواصل همهٔ اشیای واقع‌شده در خوشهٔ j و σ_j میزان پراکندگی درونی خوشهٔ j است که از رابطهٔ ۷ به دست می‌آید. مقدار R_{jk} زمانی به کمترین مقدار خود می‌رسد که مطابق رابطهٔ ۷، فاصلهٔ بین دو خوشه از یکدیگر زیاد و پراکندگی درونی خوشه‌ها به کمترین مقدار خود کاهش یابد [۸]:

$$\sigma_j = \sqrt{\frac{1}{n_j}\sum_{x_i \in C_j} \|x_i - \mu_j\|^2} \qquad (۷)$$

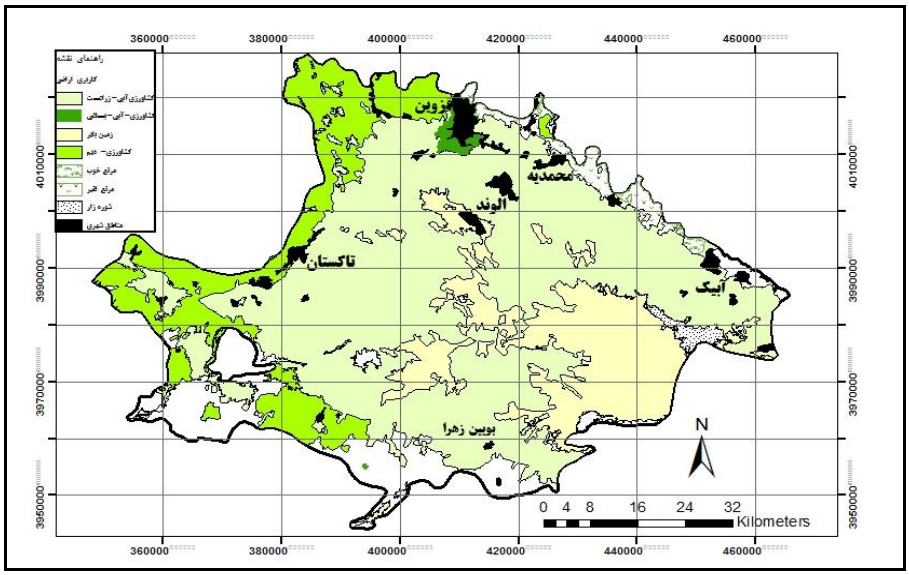
در این رابطه C_j مجموعه داده‌های متعلق به خوشهٔ j و n_j تعداد این داده‌هاست. شاخص DB (دیویس و بولدین) برابر با میانگین مقادیر R_j تعریف می‌شود که R_j برابر بیشترین فاصلهٔ R_{jk} به‌دست‌آمده از مقایسهٔ خوشهٔ j با سایر خوشه‌هاست. برای تعیین محدودهٔ اجرای مدل در روش خوشه‌بندی توصیه شده است مدل خوشه‌بندی را به تعداد m مرتبه که $2 \le m \le \sqrt{N}$، که N برابر تعداد داده‌ها یا همان ردیف‌های ماتریس داده‌ها است، اجرا کرده و در هر بار ماتریس مقادیر عضویت و معیارهای صحت‌سنجی محاسبه شود. تعداد بهینهٔ خوشه‌ها در نقطه‌ای که شاخص DB به کمترین حد برسد یا اختلاف شیب بین خط قبل و بعد از آن نقطه تفاوت معناداری داشته باشد، به دست می‌آید. در نهایت، متدولوژی روش خوشه‌بندی استفاده شده در این تحقیق مطابق شکل ۲ است:



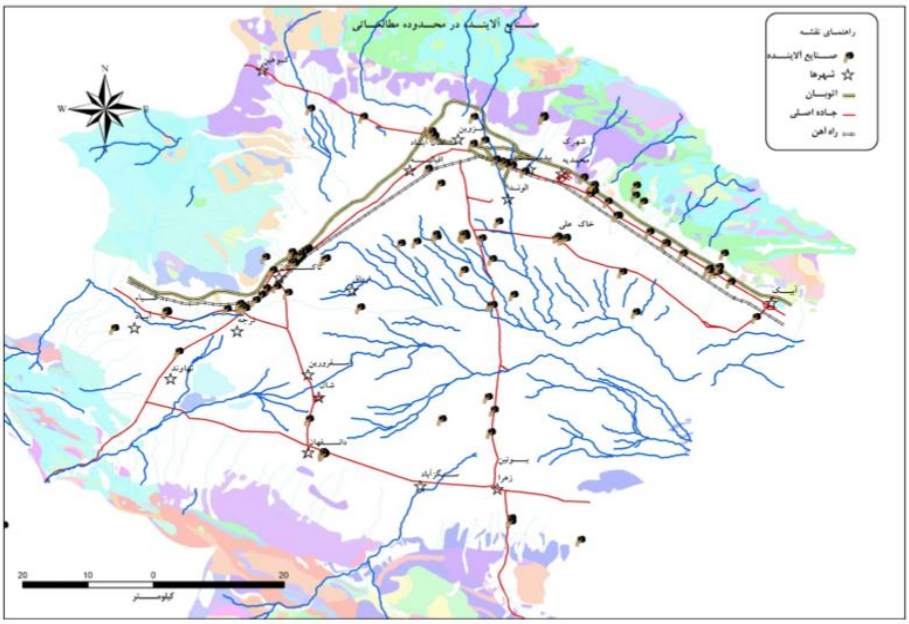
شکل ۲. مراحل انجام الگوریتم خوشه‌بندی K-Means [۲۱]

نمونه‌برداری کیفی

همان گونه که پیش‌تر نیز گفته شد، در بیشتر مطالعات گذشته فقط آنیون‌ها و کاتیون‌ها تجزیه و تحلیل شده‌اند و به پارامترهای آلودگی به‌صورت مجزا پرداخته شده است. در این مطالعه برای نخستین‌بار سه پارامتر کیفی براساس کاربری منطقه هم‌زمان در نظر گرفته شده است. پارامتر مجموع مواد محلول (TDS) که مجموعهٔ آنیون‌ها و کاتیون‌ها را در بر گرفته برای کاربری شهری، پارامتر آرسنیک (AS) که از فلزات سنگین است برای کاربری صنعتی منطقه و پارامتر آلودگی نیترات (NO3) برای کاربری کشاورزی انتخاب شد. به این‌منظور، از ۵۲ چاه منطقه نمونه‌برداری کیفی صورت گرفت (شکل ۱) و نتایج آن پس از تجزیه و تحلیل‌های اولیه وارد خوشه‌بندی شد. شکل ۳ کاربری اراضی آبخوان قزوین را نشان می‌دهد. همان طور که مشاهده می‌شود، بیشتر نواحی تحت تأثیر کشاورزی آبی بوده که بیشتر زراعی است. با توجه به کشت زراعی، معمولاً کشاورزان منطقه کود و سموم زیادی نیز مصرف می‌کنند. براساس اطلاعات سالنامهٔ آماری کشاورزی حدود ۸۲ هزار تن کود در سطح اراضی کشاورزی استان مصرف می‌شود که بیشتر کود مصرفی براساس سالنامهٔ آماری اوره، فسفات آمونیم، نیترات آمونیوم و سولفات آمونیوم است [۱۶]. با توجه به نوع کودهای مصرفی پارامتر نیترات برای تهیهٔ نقشهٔ آلودگی چندپارامتره آبخوان انتخاب شد. همچنین، در شکل ۳ مناطق شهری نیز مشاهده می شود که بیشتر شهرهای پرجمعیت و پرمخاطره (از نظر ورود آلودگی) آبخوان قزوین، شهرهای قزوین، بویین‌زهرا، آبیک، الوند و محمدیه است. با توجه به مصرف دترجنت‌ها و دیگر مواد فاضلاب شهری پارامتر TDS انتخاب شد. شکل ۴ نیز موقعیت واحدهای صنعتی را در سطح آبخوان قزوین نشان می‌دهد [۱۶]. با توجه به گزارش‌های موجود چهار صنعت عمدهٔ غذایی، نساجی، فلزی و شیمیایی بیشترین سهم را در صنایع استان دارند که از این بین سهم صنعت شیمیایی در آلوده‌کردن آبخوان بیشتر است. از این‌رو، با تحقیقات صورت‌گرفته مشخص شد در برخی نقاط آبخوان قزوین میزان فلزات سنگین حتی بیشتر از حد مجاز است و از بین آنها پارامتر آرسنیک بیشترین تأثیر را دارد. در نتیجه، برای کاربری صنعتی نیز پارامتر آرسنیک در نظر گرفته شد.



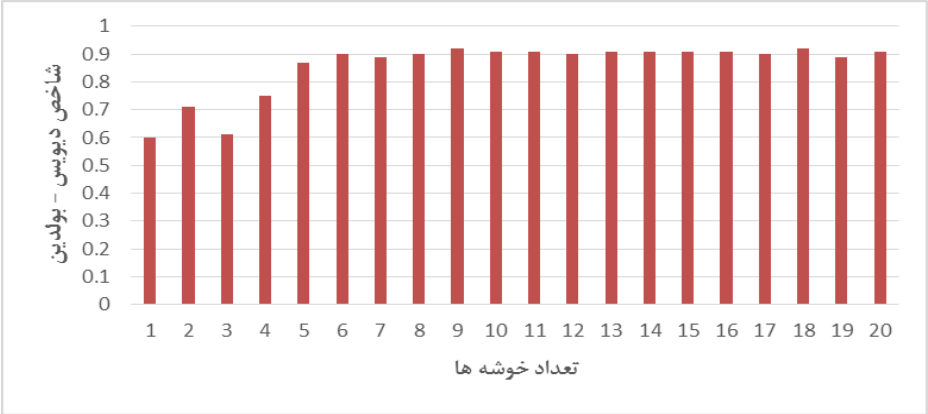
شکل ۳. نقشهٔ کاربری اراضی آبخوان دشت قزوین



شکل ۴. موقعیت آلاینده‌های صنعتی آبخوان دشت قزوین [۱۶]

نتایج و بحث

برای انجام خوشه‌بندی کلاسیک به روش K-Means، از جعبه‌ابزار تشخیص الگو در نرم‌افزار Matlab استفاده شد. پس از اعمال روش پیش‌پردازش نرمال‌سازی مطابق رابطهٔ ۳، خوشه‌بندی کلاسیک K-Means انجام و برای هر بار اجرای این الگوریتم، مقدار شاخص صحت‌سنجی DB محاسبه شد. براساس توضیح رابطهٔ ۷ برای تعیین تعداد بهینهٔ خوشه‌ها از روش دیویس- بولدین، مدل خوشه‌بندی باید از تعداد دو تا ۷۱ خوشه اجرا شود، که در این تحقیق برای اطمینان از تعیین دقیق‌تر تعداد بهینهٔ خوشه‌ها، تعداد ۲۰ خوشه به عنوان حد بالایی خوشه‌بندی انتخاب شد. تغییرات شاخص صحت‌سنجی DB در مقابل تعداد خوشه‌ها در شکل ۵ نشان داده شده است. همان طور که در شکل ۵ مشاهده می شود، در عمل از تعداد هشت خوشه به بعد تغییرات به‌وجودآمده در مقادیر شاخص صحت‌سنجی درخور توجه نیست. بنابراین، تعداد بهینهٔ خوشه‌ها برای داده‌های این تحقیق براساس شاخص DB برابر پنج خوشه است.



شکل ۵. مقادیر شاخص صحت‌سنجی دیویس - بولدین در برابر تعداد خوشه‌ها

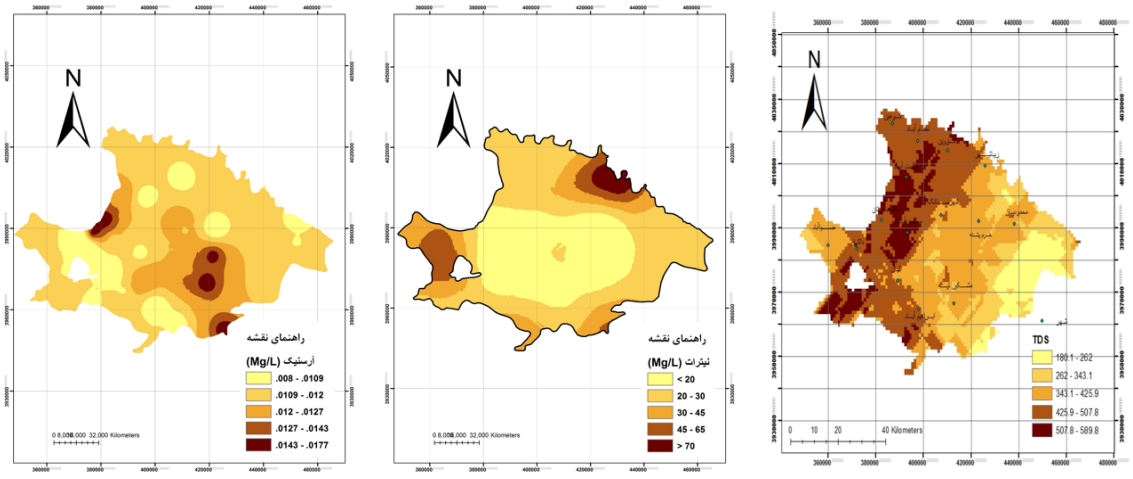
پس از انجام مراحل خوشه‌بندی در نهایت هر یک از خوشه‌ها با پارامترهای کیفی با دامنهٔ تغییرات آن (بیشینه، کمینه و متوسط) استخراج می شود که نتایج آن در جدول ۱ ارائه شده است. همان گونه که مشاهده می‌شود، هر خوشه برای هر پارامتر رنج حداکثر و حداقلی داشته و نیاز به معیاری دارد که خوشه‌ها باز تعریف شوند. به بیان دیگر، پس از مشخص‌شدن تعداد خوشه‌ها و همچنین مقادیر بیشینه و کمینهٔ پارامترها در هر خوشه، ضروری است خوشه‌ها براساس معیاری مشخص تعریف شوند چرا که تکنیک خوشه‌بندی بدون توجه به ماهیت نوع داده‌ها فقط بر اساس مشابهت‌های موجود، داده‌ها را دسته‌بندی می کند. از آنجا که مقادیر پارامترها و یا نوع پارامترها در تئوری خوشه‌بندی تعریف‌شده نیست و فقط این تکنیک براساس شباهت خوشه‌ها را تعیین و مشخص می‌کند، در این مرحله لازم است خوشه‌ها بر اساس طبیعت و ذات آنها بازتعریف شوند. به این‌منظور، در تحقیق حاضر از استاندارد جهانی آب شرب، که توسط سازمان بهداشت جهانی (WHO) ارائه شده است، برای تعریف خوشه‌ها استفاده می‌شود.

طبق استاندارد سازمان بهداشت جهانی، مقدار استاندارد نیترات در آب‌های آشامیدنی حداکثر ۵۰ میلی‌گرم در لیتر، برای آرسنیک حداکثر ۰/۰۱ میلی‌گرم در لیتر و مقدار مجاز TDS در آب آشامیدنی طبق این استاندارد حداکثر ۱۵۰۰ میلی‌گرم در لیتر تعیین شده است. با توجه به جدول ۱ ملاحظه می‌شود که بیشترین مقدار TDS در کل منطقه به‌ترتیب ۵۸۹/۷۹ میلی‌گرم در لیتر است که با توجه به استاندارد بهداشت جهانی کاملاً در حد قابل قبولی است و نشان می دهد که منطقه از نظر پارامتر TDS از نظر شرب دچار مشکل نیست. همچنین، دیده می‌شود که پارامتر نیترات و آرسنیک از گروه اول به سمت گروه پنجم افزایش می‌یابد به‌طوری که در خوشه‌های ۴ و ۵ (C4 و C5) به طور میانگین بیشترین مقدار نیترات و آرسنیک در همهٔ خوشه‌ها مشاهده می‌شود که نشان‌دهندهٔ آلودگی محدوده در بخش کشاورزی و صنعت است. در نتیجه، براساس نتایج جدول ۱ و استاندارد آب شرب برای پارامترهای یادشده، خوشه‌بندی ارائه‌شده توسط این تکنیک براساس میزان آرسنیک و نیترات براساس C1 تا C5 به‌ترتیب از ناچیز تا خیلی‌زیاد طبقه‌بندی می‌شود.

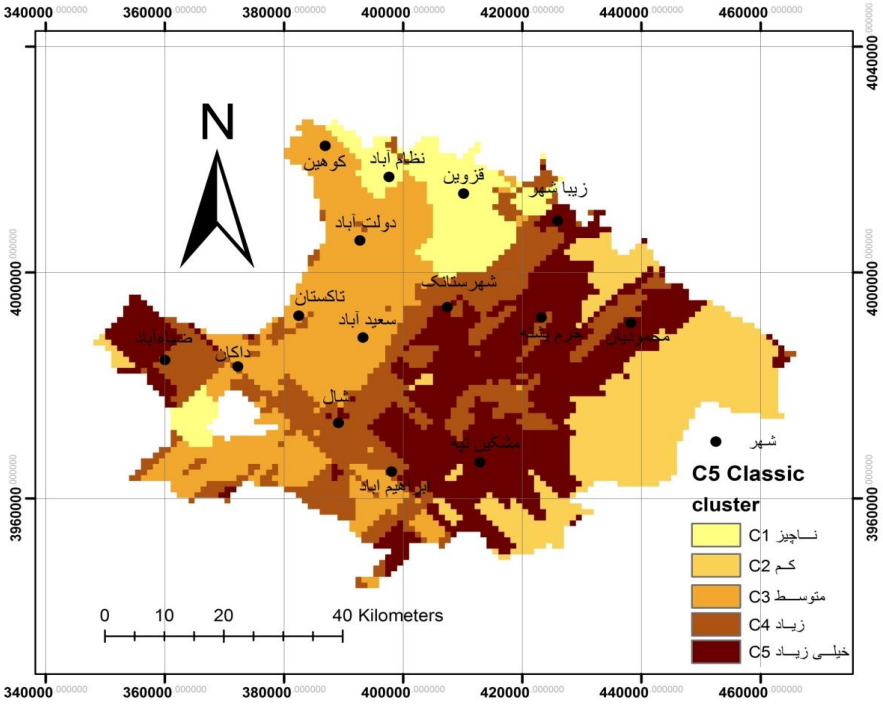
جدول ۱. دامنهٔ تغییرات پارامترهای آلودگی (برحسب میلی‌گرم در لیتر) به تفکیک خوشه‌های به‌دست‌آمده از روش کلاسیک برای کل خوشه‌ها

ردیف	نام خوشه	تعداد	نیترات (No_3)			آرسنیک (As)			کل مواد محلول (TDS)		
			حداقل	میانگین	حداکثر	حداقل	میانگین	حداکثر	حداقل	میانگین	حداکثر
۱	C1	۱۱۰۴	۵/۳۶	۲۲/۷۱	۷۸/۹۶	۰/۰۰۴۵۴	۰/۰۰۸۸۲	۰/۰۱۷۶۶	۳۷۷/۹۳	۴۰۷/۹۱	۴۴۷/۸۸
۲	C2	۹۵۴	۷/۳۵	۳۱/۰۴	۱۲۱/۷۰	۰/۰۰۳۹۶	۰/۰۰۹۴۰	۰/۰۱۴۴۲	۴۷۷/۶۶	۵۱۴/۶۲	۵۸۹/۷۹
۳	C3	۱۱۵۱	۵/۲۷	۲۲/۹۱	۷۲/۱۰	۰/۰۰۳۸۲	۰/۰۱۰۰۸	۰/۰۱۶۲۲	۲۹۸/۲۶۴	۳۴۷/۶۸	۳۷۷/۶۳
۴	C4	۹۳۸	۷/۲۸	۵۲/۶۸	۱۹۴/۳۷	۰/۰۰۳۶۷	۰/۰۰۷۹۹	۰/۰۱۵۶۹	۴۲۷/۵۱	۴۶۸/۷۵	۵۰۹/۹۴
5	C5	۷۹۹	۱۳/۶۶	۲۹/۶۹	۵۹/۵۲	۰/۰۰۵۷۹	۰/۰۱۲۰۸	۰/۰۱۴۸۰	۱۸۰/۰۱	۲۵۰/۱۸	۲۹۸/۳۴

پس از تعریف هر یک از خوشه‌ها براساس استاندارد جهانی آب شرب در نهایت نقشهٔ پهنهٔ آلودگی هر سلول مطابق شکل ۶، که خروجی مدل خوشه‌بندی روی آبخوان دشت قزوین است، استخراج می‌شود. در این شکل چگونگی توزیع مکانی خوشه‌ها به تصویر کشیده شده است. همچنین، برای مقایسهٔ بهتر نقشهٔ ترکیبی (شکل ۷) با هر یک از پارامترهای آلودگی، نقشهٔ توزیع مکانی آلاینده‌ها به تفکیک با تکنیک کریجینگ در شکل ۵ نمایش داده شده است. همان طور که مشاهده می شود، توزیع آلودگی آلاینده‌های مجزا با ترکیب آنها توسط مدل خوشه‌بندی کاملاً متفاوت است. بنابراین، ارزش نقشهٔ کیفی چندپارامتره ترکیبی بر اساس استاندارد آب شرب مشخص می‌شود. در نقشهٔ ترکیبی هر خوشه با یک رنگ مشخص شده است و با توجه به نام و رنگ هر بازه می‌توان عضویت آنها را به هر خوشه مشخص کرد. به بیان دیگر، هر بخش از آبخوان که کلاس C1 داشته باشد از نظر کیفی (آنیون‌ها، کاتیون‌ها، نیترات و فلزات سنگین) وضعیت مناسب داشته و هر چه به C5 نزدیک‌تر شود، وضعیت کیفی آبخوان بحرانی‌تر می شود. به بیان دیگر، نقشهٔ به‌دست‌آمده از ترکیب آلودگی‌های مختلف است و منحصر به یک آلودگی خاص نظیر نیترات نیست که از نوآوری‌های مهم این تحقیق است. با توجه به شکل ۷، خوشهٔ ۱ (C1) با مساحت ۲۲/۳ درصد در ناحیهٔ شمالی قرار گرفته و از لحاظ وسعت مکانی در میان خوشه‌ها رتبهٔ دوم را دارد و دو شهر قزوین و نظام‌آباد در این محدوده قرار دارند. همچنین، خوشهٔ ۲ (C2) با مساحت ۱۹/۳ درصد به‌صورت یک محدودهٔ مجزا در گوشهٔ شرقی استان در محدودهٔ شهر آبیک قرار دارد. خوشهٔ ۳ (C3) نیز با مساحت ۲۳/۳ درصد به لحاظ مکانی گسترده‌ترین خوشه در منطقهٔ مطالعه‌شده است که در قسمت‌های غرب و شمال غربی گسترده شده است. در خوشهٔ سوم مناطق کوهین، دولت‌آباد، سعیدآباد، تاکستان جای می‌گیرند. همچنین، در خوشهٔ چهارم (C4) با مساحت ۱۹ درصد که بیشتر به‌صورت گسترده و لکه‌هایی در مناطق مرکزی و تا حدودی غربی مشاهده می‌شود. مناطق روستایی خرم‌پشته، شهرستانک، شال، ضیاآباد در این خوشه قرار می‌گیرند که قطب‌های کشاورزی منطقه‌اند. در نهایت، خوشهٔ ۵ (C5) با مساحت ۱۶/۱ درصد در قمست مرکزی به سمت شرق قرار دارد. مناطق مشکین‌تپه، محمودیان، زیباشهر در محدودهٔ خوشهٔ ۵ قرار دارند. در این منطقه نیز فعالیت‌های صنعتی و کشاورزی زیادی وجود دارد. در واقع، نقشهٔ به‌دست‌آمده نشان می‌دهد کدام مناطق هم‌زمان از نظر شرب، کشاورزی و صنعتی مشکل نیترات، فلزات سنگین و آنیون‌ها و کاتیون‌ها (TDS) را دارند. نتایج این تحقیق در مقایسه با تحقیقات پیشین همچون Usman و همکارانش در سال ۲۰۱۴ هم‌راستا است به‌طوری که در این تحقیق نیز بیشتر پارامترهای مؤثر در خوشه‌بندی کیفیت آب زیرزمینی به آلودگی‌های متمرکز (فاضلاب های شهری و صنعتی) و غیرمتمرکز (کشاورزی) در منطقه مرتبط است [۹]. همچنین، در مقایسه با تحقیق Oorkavalan و همکارانش در سال ۲۰۱۶ [۱۲]، نتایج این تحقیق بیان می‌کند که درنظرگرفتن پارامترهایی همچون آنیون یا کاتیون یا یک آلودگی همچون نیترات به صورت نقشهٔ پهنه‌بندی آلودگی به‌تنهایی کافی نیست و ضروری است در تهیهٔ نقشهٔ آلودگی، کاربری‌های دیگری نظیر صنعت نیز در نظر گرفته شود. حال در این مطالعه چنانچه منطقه‌ای خوشهٔ C4 یا C5 داشته باشد، نشان می‌دهد آن مناطق مشکل آلودگی زیادی از لحاظ چندین پارامتر دارند که می‌تواند ناشی از ترکیب هم‌زمان کاربری صنعتی و کشاورزی باشد. در واقع، روش خوشه‌بندی تکنیکی است که در این مطالعه کمک می‌کند با یک روش آماری سه نقشهٔ مهم کیفی (نیترات، آنیون‌ها و کاتیون‌ها و فلزات سنگین) هم‌زمان در یک چارچوب علمی و منطقی با هم ترکیب شوند. از این‌رو، یکی از کاربردهای مهم نقشهٔ به‌دست‌آمده این است که در نواحی مرکزی آبخوان از چندین نوع آلودگی در آسیب است و باید شبکهٔ پایش در آن نواحی پوشش بیشتری داشته باشد. از طرف دیگر، از آنجا که نقشهٔ به‌دست‌آمده براساس استاندارد آب شرب طبقه‌بندی شده است در نتیجه باید تمهیدات بیشتری در نواحی مرکزی همچون نظارت بر کوددهی در کشاورزی و یا تصفیهٔ فاضلاب صنعتی صورت گیرد.



شکل ۶. پراکندگی مقادیر آلودگی آرسنیک، نیترات و TDS در آبخوان قزوین



شکل ۷. پراکندگی خوشه‌های مربوط به خوشه‌بندی کلاسیک برای پنج خوشهٔ آبخوان قزوین

نتیجه‌گیری

در این مطالعه از تکنیک خوشه‌بندی برای ارزیابی تغییرات مکانی چندپارامترهٔ کیفی آب زیرزمینی ۵۲ حلقه چاه در آبخوان دشت قزوین استفاده شد. بر اساس تحلیل خوشه‌ای چاه‌های مد نظر، نقشهٔ آلودگی چندپارامترهٔ آبخوان قزوین به پنج گروه یا خوشه تقسیم شد که از گروه ۱ تا ۵ به‌ترتیب از ناچیز به خیلی‌زیاد دسته‌بندی شده است. بهترین کیفیت آب از لحاظ سه پارامتر اندازه‌گیری‌شدهٔ نیترات (NO_3)، آرسنیک (As) و کل مواد جامد محلول (TDS) در نمونه‌های واقع در گروه اول قرار گرفته است و بدترین ردهٔ آب متعلق به خوشهٔ پنجم است که از نظر کشاورزی و صنایع در شرایط بحرانی است. با توجه به آنالیز خوشه‌ای انجام‌شده روی نمونه‌ها و چگونگی توزیع آنها در دشت قزوین، می‌توان این‌گونه دریافت کرد که نمونه‌های خوشهٔ ۱ (وضعیت مطلوب) بیشتر در نواحی شمال تمرکز یافته‌اند. این مناطق بیشتر آب با کیفیت مناسب دارند، اما این نکته درخور یادآوری است که در برخی

مناطق نمونه های با کیفیت متفاوت نیز مشاهده می‌شود. به‌طور مثال، در ناحیهٔ جنوبی دشت در برخی مناطق نمونه های متعلق به خوشه های ۳ـ ۵ مشخص است و یا در ناحیهٔ شمالی در مناطقی نمونه‌های متعلق به خوشهٔ ۴ دیده می‌شود. این امر بیشتر به‌دلیل جهت جریان آب زیرزمینی از بخش شمالی آبخوان به نواحی مرکزی بوده که همهٔ آلودگی را در این نواحی متمرکز کرده است. از طرف دیگر، چنانچه فقط براساس آنیون‌ها و کاتیون‌ها طبقه‌بندی صورت می‌گرفت، آبخوان مشکل حادی نداشت، حال آنکه مشاهده می‌شود با دخالت پارامترهای دیگر آلودگی نظیر نیترات و آرسنیک، بخش‌هایی از آبخوان مخاطره‌آمیز است. در نهایت، می‌توان نتیجه گرفت خوشه‌بندی همهٔ خصوصیات را به‌طور ترکیبی و هم‌زمان در جداسازی خوشه ها از یکدیگر در نظر می‌گیرد و فقط به چند خصوصیت محدود برای خوشه‌بندی اکتفا نکرده است. بنابراین، با ارائهٔ یک شبکهٔ بهینهٔ پایش و همچنین تمهیدات و نظارت بیشتر در نواحی بحرانی کیفی، می‌توان یک مدیریت بهینهٔ کیفی روی آبخوان اعمال کرد.

منابع

[1]. Neshat A, Pradhan B, Javadi S. Risk assessment of groundwater pollution using monte carlo approach in an agriculture region: an example from Kerman plain, Iran. Computers, Environment and urban system. 2015; 50(1): 66-73.

[2]. Al-adamat R.A.N, Foster I.D.L, Baban S.M.J. Groundwater vulnerability and risk mapping for the Basaltic aquifer of the Azraq basin of Jordan using GIS and Remote sensing and DRASTIC, Applied Geography. 2003; 23(4): 303-324.

[3]. Samani S, Kalantari N, Rahimi, M.H. Evaluation of groundwater quality by Cluster analysis technique in Avan aquifer, Journal of Water resources engineering. 2011; 4: 75-85. [Persian].

[4]. Aghdar H, Mohammadyari F. Assessment of groundwater quality using Cluster analysis method in Mehran and Dehloran aquifer. The first international conference on new achievements in Agriculture, natural resources and environmental sciences. 2014. [Persian].

[5]. Ouyang Y, Jia Zh, Cui L. Estimating impacts of land use on groundwater quality using trilinear analysis. Environmental monitoring and assessment. 2014; 186(9): 5353-5362.

[6]. Ghiasi N, Arabkhedri M, Ghafari A, Hatami H. Survey on the effect of some morphometric characteristics of basins on peak discharge with different return periods (Case study north Albors basins). Research and development journal. 2004; 62: 2-10. [Persian].

[7]. Goulet V, Rocourt J, Jacquet C. Cluster of listeriosis cases in France. Euro surveillance weekly. 2002; 27(6).

[8]. Kim K J, Ahn H. A recommender system using GA K-means clustering in an online shopping market. Expert Syst. Appl. 2008; 34 (2): 1200–1209.

[9]. Usman N, Toriman M.E, Juahir H. Assessment of Groundwater Quality Using Multivariate Statistical Techniques in Terengganu. Science and Technology, 2014; 4(3): 42-49.

[10]. Zou H, Zou Z, Wang X. An Enhanced K-Means Algorithm for Water Quality Analysis of the Haihe River in China. Int. J. Environ. Res. Public Health. 2015; 12: 1400-1413.

[11]. Azhar S.C, Aris A.Z, Yussof M.K, Ramli M.F, Juahir H. Classification of river water quality using multivariate analysis. Procedia Environmental Sciences. 2015; 30: 79-84.

[12]. Oorkavalan G, Chidambaram S.M, Mariappan V, Kandaswamy G, Natarajan S. Cluster Analysis to Assess Groundwater Quality in Erode District, Tamil Nadu, India. Circuits and Systems, 2016; 7: 877-890.

[13]. Fianko J.R, Osae S, Adomako D, Achel D.J. Relationship between land use and groundwater quality in six districts in the eastern region of Ghana. Environmental Monitoring and Assessment. 2009; 153(4): 139-146.

[14]. Yongjun J, Daoxian Y, Shiyou X, Linli L, Gui Zh, Raosheng H. Groundwater quality and land use change in a typical karst agricultural region: a case study of Xiaojiang watershed, Yunnan. Journal of geographical Sciences. 2006; 16(4): 405-414.

[15]. Lerner D, Harris B. The relationship between land use and groundwater resources and quality. Land Use Policy. 2009; 26(1): 265-273.

[16]. Announcement. Hydrogeology section- the report of Quality and Quantity Modelling study in Qazvin aquifer. Qazvin Regional Water Company. 2012: 23-26. [Persian].

[17]. Han J, Kamber M. Data mining concepts and techniques. San Francisco, U.S.A, Morgan Kaufman Publisher. 2006: 110.

[18]. Hoppner F, Klawonn F, Kruse R, Runkler T. Fuzzy cluster analysis. Sussex, England: Wiley and Sons. 1999: 146.

[19]. Feil B. Fuzzy Clustering in Process of Data Mining. Ph.D. thesis, Department of Process Engineering, University of Veszprem Hungary. 2006.

[20]. Kim D.W, Lee K.H, Lee D. On cluster validity index for estimation of the optimal number of fuzzy clusters. Journal of Pattern Recognition Society. 2004; 37: 209-225.

[21]. Hashemy S.M. Spatial and Temporal Clustering in Irrigation network using classic and fuzzy technique. M.Sc. thesis, Tarbiat Modares University. 2008. [Persian].

[22]. Davies D.L, Bouldin D.W. A cluster separation measure. IEEE Trans. Pattern Anal. Mach. Intell. 1979; 1(4): 224–227.