

کاربرد خوشه‌بندی در تهیه نقشه‌های شبه زمین‌شناسی با استفاده از داده‌های ژئوفیزیک هوابردی

محسن شادمان¹، بهزاد تخم‌چی^{2*}، حسن خیراللهی³

1- دانشجوی کارشناسی ارشد، دانشکده مهندسی معدن، دانشگاه تهران، mohseneshademan@yahoo.com

2- استادیار، دانشکده مهندسی معدن، نفت و ژئوفیزیک، دانشگاه صنعتی شاهرود، tokhmechi@ut.ac.ir

3- کارشناس ارشد اکتشاف، گروه ژئوفیزیک هوایی، سازمان زمین‌شناسی، h.khayrollahi@gmail.com

(دریافت 30 بهمن 1389، 26 تیر 1391)

چکیده

از جمله کاربردهای داده‌های ژئوفیزیک هوابردی، تهیه نقشه شبه زمین‌شناسی است. در این نوع نقشه‌ها، محدوده هر واحد زمین‌شناسی و عوارض ساختاری چون خطواره‌ها را از روی تغییرات جانبی و تلفیق انواع داده‌های ژئوفیزیکی مشخص می‌کنند، اما جنس واحدهای زمین‌شناسی نامشخص خواهد بود. البته در ادامه امکان تطابق‌دهی بین این نقشه و عوارض زمینی و مارک‌دار کردن نقشه مربوطه وجود خواهد داشت. در این مقاله، با استفاده از داده‌های ژئوفیزیک هلیکوپتری دیگر بخش‌های از برگه 1:100,000 آلوت توسط روش‌های خوشه‌بندی میانگین‌های -k و میانگین‌های -k فازی، نقشه توزیع واحدهای زمین‌شناسی تهیه شده و اعتبار آن ضمن مقایسه با نقشه 1:100,000 زمین‌شناسی بررسی شده است. علاوه بر این از مجموعه داده‌های مختلف ژئوفیزیکی به عنوان ورودی خوشه‌بندی کننده استفاده شده و به این ترتیب بهترین مجموعه از نظر دقت تطابق با مرز واحدهای زمین‌شناسی معرفی شده است. نتایج خوشه‌بندی داده‌های ژئوفیزیک هوابردی نشان داد که در بین مجموعه داده‌های مختلف، با ترکیب نتایج دو مجموعه داده (راديوعناصر و نسبت آن‌ها) و (مقاومت الکتریکی فرکانس‌های مختلف و سیگنال تحلیلی مغناطیس کل) می‌توان نقشه شبه زمین‌شناسی معتبرتری ارائه کرد.

کلمات کلیدی

ژئوفیزیک هوابردی، نقشه شبه زمین‌شناسی، خوشه‌بندی، میانگین‌های -k، میانگین‌های -k فازی.

1- مقدمه

داده ژئوفیزیک هوابرد در مقیاس ناحیه‌ای برداشت می‌شود و اغلب به عنوان ابزاری برای شناسایی تغییرات جانبی لیتولوژی و مناطق با پتانسیل بالای کانی‌سازی استفاده می‌شود [1]. از آنجایی که استفاده از یک پارامتر ژئوفیزیکی برای رسیدن به هدف مشخص می‌تواند دارای عدم قطعیت بالایی باشد، می‌توان با استفاده از چند پارامتر، این عدم قطعیت را تا حدود زیادی کاهش داد. برای تعیین محدوده‌های با پتانسیل بالای کانی‌سازی اغلب از روش‌های وزن‌دهی و تلفیق مانند وزن‌های نشانگر، منطق فازی و رگرسیون منطقی استفاده می‌شود که در حوزه GIS به دفعات استفاده شده اند [2]، [3].

در این راستا می‌توان از روش‌های تشخیص الگو برای تهیه نقشه‌های شبه زمین‌شناسی، به خصوص در مقیاس کوچک و تعیین هدف برای بازدیدهای صحرایی استفاده کرد. به نظر می‌رسد، در صورت استفاده از چند نوع داده مختلف و ترکیب آن‌ها با یکدیگر، بتوان به نتایج مطلوب‌تر و معتبرتری رسید. از جمله تکنیک‌های تشخیص الگو که در علوم زمین بسیار مورد توجه قرار گرفته است، روش‌های خوشه‌بندی و یا کلاسه‌بندی غیر نظارتی¹ است [4]، [5]، [6]، [7]، [8].

در این مقاله همزمان از داده‌های رادیومتری، الکترومغناطیسی و مغناطیسی به منظور خوشه‌بندی واحدهای زمین‌شناسی به روش‌های میانگین‌های k و میانگین‌های k فازی استفاده شده است. در ضمن ایده جدیدی برای تفسیر نتایج خوشه‌بندی داده‌ها معرفی شده است.

2. الگوریتم خوشه‌بندی میانگین‌های K و میانگین‌های K فازی²

در خوشه‌بندی با روش‌های میانگین‌های k و میانگین‌های K فازی، تعداد خوشه‌ها (K) مفروض در نظر گرفته می‌شود و توسط الگوریتمی بهینه تعلق داده‌ها به خوشه‌های مفروض، صورت می‌گیرد. سپس با تغییر تعداد خوشه‌های مفروض (تغییر K)، عملیات فوق تکرار شده و نسبت به بهینه‌سازی تعداد خوشه‌ها (K) اقدام می‌شود. نقطه قوت این روش‌ها، راندمان محاسباتی بالای آن‌ها است. این دو روش، توسط نرم‌افزار Matlab کد نویسی شده و برای داده‌های این مقاله استفاده شده‌اند، که در ادامه تشریح خواهند شد.

2-1- روش خوشه‌بندی میانگین‌های K

الگوریتم خوشه‌بندی این روش به صورت زیر است [9]:

الف. یکسان‌سازی اهمیت ویژگی‌های مختلف (نرمالایز کردن). برای این منظور از رابطه 1 استفاده شده است.

$$X_i = \frac{x_i - m_i}{d_i} \quad (1)$$

که در آن، X_i : مقدار استاندارد شده یک داده، x_i : مقدار واقعی همان داده، m_i : میانگین داده‌ها و d_i : انحراف معیار داده‌ها است.

ب. در ابتدا داده‌ها به صورت تصادفی در K خوشه تقسیم می‌شوند. به این ترتیب یک خوشه‌بندی اولیه صورت گرفته است.

ج. محاسبه بردار میانگین ویژگی‌ها مختلف (\underline{m}_i):

$$m_i = \frac{1}{N_i} \sum_{x_j \in w_i} x_j \quad (2)$$

که در آن: N_i : تعداد داده‌های واقع در خوشه نام است.

د. محاسبه فاصله هر داده تا میانگین داده‌های هر خوشه.

ه. تعلق داده به خوشه‌ای با فاصله کمتر از خوشه موجود.

و. داده‌های دیگر به ترتیب و به صورت تصادفی انتخاب شده و نسبت به تعیین تکلیف تعلق آن‌ها به خوشه‌های مختلف اقدام می‌شود.

ی. محاسبه تابع هزینه (J):

$$J = \sum_{i=1}^k \sum_{x \in w_i} \|x - \underline{m}_i\|^2 \quad (3)$$

که در آن:

J : تابع هزینه، k : تعداد خوشه‌ها، w_i : خوشه نام، x : بردار ویژگی‌های داده‌ها، و \underline{m}_i : بردار میانگین ویژگی‌های داده‌ها در خوشه نام است.

عملیات بالا تا زمانی که شیب کوچک شدن J کم شود تکرار می‌شود. توجه شود که در اینجا شرط توقف عملیات خوشه‌بندی، کوچک شدن نرخ کاهش J از یک حد آستانه است. در صورتی که داده‌ها به درستی در خوشه خودشان طبقه‌بندی شده باشند، J کمینه خواهد شد.

برای گزینش تعداد خوشه بهینه، از اعتبارسنجی خوشه‌بندی استفاده شده که در رابطه 4 آورده شده است.

$$u_{ij} = \frac{1}{\sum_{l=1}^K \left(\frac{d(x^j, w_l)}{d(x^j, w_i)} \right)^{\frac{1}{m-1}}} \quad (7)$$

$$\sum_{i=1}^K u_{ij} = 1 \quad \text{and} \quad u_{ij} \geq 0$$

در روابط فوق می‌بایست نسبت به بهینه‌سازی m اقدام کرد که به طور معمول به شیوه سعی و خطا انجام می‌شود.

د. بهینه‌سازی تعداد خوشه‌ها با الگوریتم ضرایب افزاز PC صورت گرفته است. در این روش میزان فازی بودن نتایج نهایی خوشه‌بندی مورد بررسی قرار می‌گیرد رابطه 8:

$$PC(\underline{u}) = \frac{\sum_{i=1}^K \sum_{j=1}^N u_{ij}^2}{N} \quad (8)$$

که در آن \underline{u} ماتریس درجه عضویت داده‌ها در خوشه‌های مختلف، N تعداد داده‌ها و u_{ij} درجه عضویت داده j در خوشه i است. استفاده از توان دو برای u_{ij} ضروری است، چرا که در غیر اینصورت نتیجه در هر حالت یک می‌شود. حاصل رابطه 8 در محدوده $[1/K, 1]$ خواهد بود، که پاسخ یک به معنی غیرفازی - ترین حالت است [11]. در این روش، فاصله گرفتن از حالات فازی مزیت محسوب شده و تعداد بهینه خوشه‌ها آنی است که PC را به یک نزدیک‌تر می‌کند.

3. زمین‌شناسی منطقه مورد مطالعه

داده‌های مورد استفاده در این مقاله، در بخشی از برگه 1:100,000 آلودت، واقع در غرب - شمال غرب ایران و به موازات زون سنندج-سیرجان قرار دارد. در سیستم مختصات $UTM/Zone 38N$ عرض محدوده مورد مطالعه بین $36^{\circ}04'$ و $36^{\circ}14'$ و طول آن بین $45^{\circ}56'$ و $45^{\circ}42'$ است. مجموعه واحدهای دگرگونی پرکامبرین که به دگرگونی‌های مهاباد معروفند، کهن‌ترین واحد زمین‌شناسی منطقه است. این مجموعه در برگیرنده گنیس، شیست، اسلیت و متاریولیت بوده و محدوده گسترده‌ای از ورقه آلودت را می‌پوشاند. شکل 1 نقشه زمین‌شناسی محدوده برداشت داده‌های هواپرد ژئوفیزیکی را

$$D(\underline{U}) = \frac{\min_{i=1}^c \min_{j=1}^c d_{\min}(w_i, w_j)}{\max_{l=1}^c d_{\max}(w_l, w_l)} \quad \text{for}(i \neq j) \quad (4)$$

که در آن، صورت کسر، ملاکی از فاصله بیرونی و مخرج کسر، ملاکی از فاصله درونی خوشه‌ها است. آن تعداد خوشه بهینه خواهد بود، که رابطه 4 را بیشینه کند.

2-2- روش خوشه‌بندی میانگین‌های k فازی

تنها تفاوت منطق خوشه‌بندی میانگین‌های K فازی، با منطق خوشه‌بندی میانگین‌های K در این است که در این روش، تعلق داده‌ها به خوشه‌های مختلف به صورت مطلق نبوده و فازی است. بنابراین برای هر داده یک احتمال تعلق به خوشه‌های مختلف تعریف می‌شود [9]، [11]. در ادامه تنها نحوه تغییر روابط مورد اشاره توضیح داده می‌شود:

الف. در اینجا \underline{m}_i (بردار میانگین ویژگی‌های داده‌ها در خوشه i ام) به صورت رابطه 5 تعریف می‌شود:

$$\underline{m}_i = \frac{\sum_j u_{ij} \times x^j}{\sum_j u_{ij}} \quad (5)$$

که در آن u_{ij} میزان تعلق داده j ام به خوشه i ام و x^j معرف بردار ویژگی‌های داده j ام است.

ب. تابع عضویت فازی به صورت رابطه 6 تعریف می‌شود:

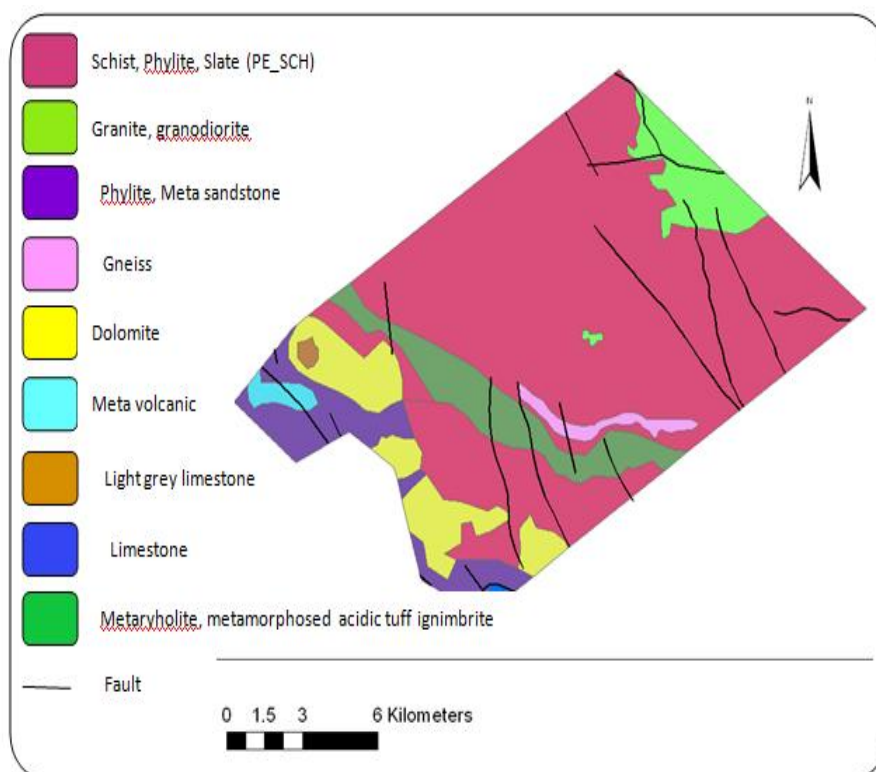
$$J = \sum_{i=1}^k \sum_{x \in w_i} m_{w_i}^m(x) \left\| x - \underline{m}_i \right\|^2 \quad (6)$$

که در آن، $m_{w_i}^m(x)$ تابع عضویت فازی داده x در خوشه i ام (w_i یعنی خوشه i ام) است. m پارامتر شدت فازی³ است [11].

ج. در خوشه‌بندی فازی u_{ij} ها و m می‌بایست به نحوی یافت شوند که تابع هزینه J را کمینه کنند [11]. u_{ij} های بهینه از رابطه 7 به دست می‌آیند:

آلبیتی به صورت ضعیف تا متوسط دیده شده است [19].

نشان می‌دهد. همان‌طور که در نقشه نیز مشخص است، در این منطقه بیشتر سنگ‌های دگرگونی مانند ماسه سنگ‌های دگرگون شده، فیلیت، ولکانیک‌های دگرسان شده و گنایس و کمتر سنگ‌های با منشاء رسوبی مانند آهک و دولومیت یافت می‌شوند [12]. معدن فعالی در این منطقه وجود ندارد، اما بر اساس مطالعات حسنی پاک به نظر می‌رسد که این منطقه پتانسیل بالای کانی‌سازی طلا داشته باشد. در این منطقه، آلتراسیون سیلیسی، پروپلیتی، آرژلیتی و



شکل 1: نقشه زمین‌شناسی منطقه مورد مطالعه [12]

رادئومتر بوده است [13]. از داده‌های الکترومغناطیسی که با استفاده از 3 فرکانس 56000، 7200 و 900 هرتز برداشت شده است، برای سه عمق مختلف نقشه‌های توزیع مقاومت الکتریکی با استفاده از الگوریتم وارون‌سازی سیمون⁵ تولید شده‌اند [14].

داده‌های بکار رفته در این منطقه، به وسیله خطوط پرواز موازی در جهت شمال شرقی- جنوب غربی و با فاصله 200 متر برداشت شده‌اند. در مجموع از داده‌های 465 خط پرواز با 67707 نقطه اندازه‌گیری برای هر پارامتر برای انجام این مطالعه استفاده شده است. به منظور مشاهده توزیع پارامتر-های ژئوفیزیکی در منطقه مورد مطالعه، نقشه‌های

4- لایه‌های ژئوفیزیکی منطقه

داده‌های ژئوفیزیکی مورد استفاده، با سیستم هلیکوپتری دیگم برداشت شده و شامل داده‌های رادئومتری، الکترومغناطیسی و مغناطیسی هستند. سیستم دیگم عبارتست از مجموعه‌ای از سیم‌پیچ‌ها و سنسورهای بسیار دقیق در یک لوله استوانه‌ای به طول 9 متر که میدان‌های الکترومغناطیس ثانویه زمین را به ازای هر فرکانس مشخص برای هر سیم پیچ مشخص اندازه‌گیری می‌کند. برای نمونه- برداری در این سیستم هواپرد از قابلیت تفکیک بالایی برخوردار بوده و با توجه به سرعت پرواز در حدود 4 متر برای مغناطیس‌سنجی و الکترومغناطیس و 10 متر برای داده‌های

شکل 2: نقشه‌های توزیع پارامترهای ژئوفیزیک مورد استفاده در مقاله؛ الف: مقاومت الکتریکی فرکانس 56 کیلوهرتز، ب: سیگنال تحلیل میدان مغناطیسی، ج: نقشه سه گانه رادیو عناصر بر اساس مدل رنگی RGB، قرمز: پتاسیم، سبز: توریم و آبی: اورانیوم

U/Th تطابق بهتری با نقشه زمین‌شناسی دارند، که در این مقاله نیز از این مجموعه داده‌ها استفاده شده است [16]. گدهای خوشه‌بندی در *Matlab* نوشته شده و داده‌های منطقه مورد مطالعه از فایل ژئودیتابیس اصلی جدا شده‌اند. از آنجا که زمان نمونه‌برداری داده‌های رادیومتری با داده‌های مغناطیس و الکترومغناطیسی متفاوت است، تنها نمونه‌هایی استفاده شده که هر سه پارامتر مغناطیس، رادیومتری و مقاومت الکتریکی اندازه‌گیری شده است. در ضمن برای مقایسه نقشه خوشه‌بندی با نقشه زمین‌شناسی 1:100,000 آلود، فواصل نمونه‌برداری 200 متر انتخاب شده‌اند. در مجموع 13520 داده برای ورود به کد آماده شد. در این مقاله عملیات خوشه‌بندی برای دو تا 10 خوشه انجام شده و سپس با توجه به نقشه خروجی خوشه‌بندی، اهمیت هر واحد سنگی در خوشه‌بندی با استفاده از داده‌های بکار رفته مشخص شده است. مجموعه داده‌های مختلفی برای این منظور استفاده شده‌اند که تعدادی از نتایج در اینجا ارائه خواهند شد:

الف. مجموعه داده توریم، پتاسیم، اورانیوم و U/Th (شکل 3)
ب. مجموعه داده توریم، پتاسیم، اورانیوم، U/Th ، Th/K و U/K (شکل 4)

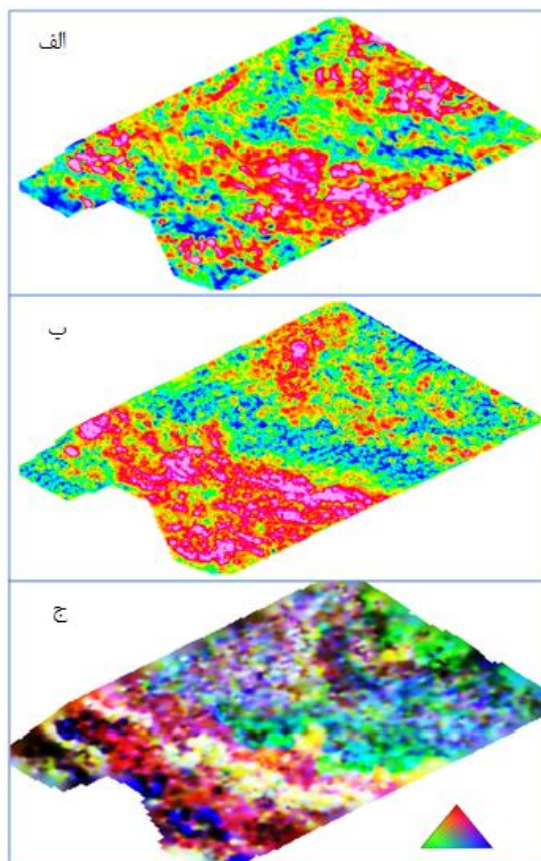
ج. مجموعه داده (ب) به همراه داده‌های مقدار سیگنال تحلیلی (AS) و مقاومت الکتریکی فرکانس 56 کیلوهرتز (شکل 5)

د. مجموعه داده مقدار سیگنال تحلیلی (AS) و مقاومت الکتریکی فرکانس‌های 7200 و 56000 هرتز (شکل 6)
اعداد موجود در اشکال 3 تا 6، مختصات نقشه در سیستم مختصات UTM/Zone 38N است که محور عرض در توان 5 عدد 10 و محور طول در توان 6 عدد 10 ضرب می‌شوند. در بین مقالات موجود، کمتر از داده مغناطیسی و مقاومت الکتریکی استفاده شده ولی در این مقاله، از این داده‌ها نیز برای خوشه‌بندی استفاده شده است. به دو دلیل از داده‌های سیگنال تحلیلی به جای خود میدان مغناطیسی استفاده شده است.

الکترومغناطیس برای فرکانس‌های 900 و 56000 هرتز و نقشه‌های رادیومتر عناصر پتاسیم، اورانیوم و توریم در شکل 2 آورده شده است. لازم به ذکر است که در مورد تمام داده‌های استفاده شده، تصحیحات بطور کامل صورت گرفته است. تخمین داده‌های ژئوفیزیکی شکل 2 با استفاده از روش حداقل انحنای⁶ صورت گرفته که روش معمول در داده‌های هواپردی است، چرا که در داده‌های هواپردی، چگالی نمونه‌برداری در جهت پرواز بسیار بیشتر از چگالی نمونه‌برداری در جهت عمود بر آن است و توصیه نشده از روش‌هایی همچون کریجینگ استفاده شود [15].

5- خوشه‌بندی داده‌های ژئوفیزیکی

هریس از روش میانگین‌های- k برای تفسیر داده‌های طیف سنجی اشعه گاما استفاده کرده است. وی در مطالعات خود از نسبت‌های رادیو عناصر به علاوه خود رادیو عناصر استفاده کرده و به این نتیجه رسیده است که خوشه‌های به دست آمده از مجموعه داده‌های 6 بُعدی (پتاسیم، اورانیوم، توریم، U/Th ، Th/K و U/K) و 4 بُعدی (پتاسیم، توریم، اورانیوم و



نیست. به عنوان مثال اپرل تعداد کلاس‌ها را برای تهیه نقشه شبه زمین‌شناسی یک منطقه با استفاده از یک مجموعه داده ژئوفیزیکی (پتاسیم، توریم، مشفق قائم میدان مغناطیسی و خودپذیری ظاهری) برابر شش انتخاب کرد در حالی که در نقشه زمین‌شناسی منطقه، بیش از هشت رخنمون سنگی وجود داشت [7]. در این مقاله برای هشت مجموعه داده، معیار دیویس-بولدین برای تعداد کلاس 2 تا 10 محاسبه شد و نمودار مربوطه ترسیم شد [10]. با توجه به اینکه در این روش‌ها پاسخ هر مرحله با پاسخ مرحله قبل متفاوت است این کار برای هر مجموعه داده حداقل شش مرتبه تکرار شد. نتایج این عملیات به صورت نمودارهایی در شکل (7) آورده شده است. همان‌طور که مشاهده می‌شود در مجموعه داده‌هایی که از پارامترهای سیگنال تحلیلی و مقاومت الکتریکی استفاده شده است نتایج هر بار اجرای برنامه با یکدیگر بسیار متفاوت است و نمودارها از روند مورد انتظار تبعیت نمی‌کنند. با مشاهده هیستوگرام توزیع مقادیر نرمالیزه شده می‌توان فهمید که تمام پارامترهای رادیومتری، رادیو عناصر و نسبت آن‌ها، دارای توزیع نرمال بوده، در حالی که هیستوگرام توزیع مقاومت الکتریکی و سیگنال تحلیلی توزیع نرمال ندارند (شکل 8). وقتی از لگاریتم مقادیر مقاومت الکتریکی و سیگنال تحلیلی استفاده شد، نمودارهایی به دست آمد که به نمودار-های مورد انتظار شبیه‌ترند (شکل 9). با توجه به نتایج به‌دست آمده به نظر می‌رسد تعداد هشت کلاستر برای خوشه‌بندی به روش میانگین‌های k و میانگین‌های K فازی مناسب باشد.

الف. از آنجا که تغییرات مقدار این پارامتر با نزدیک شدن به کنتاکت بسیار سریع‌تر است، مقدار این پارامتر نسبت به خود میدان کل مغناطیسی و میدان برگردان به قطب، تطابق بهتری با تغییرات جانبی زمین‌شناسی دارد.

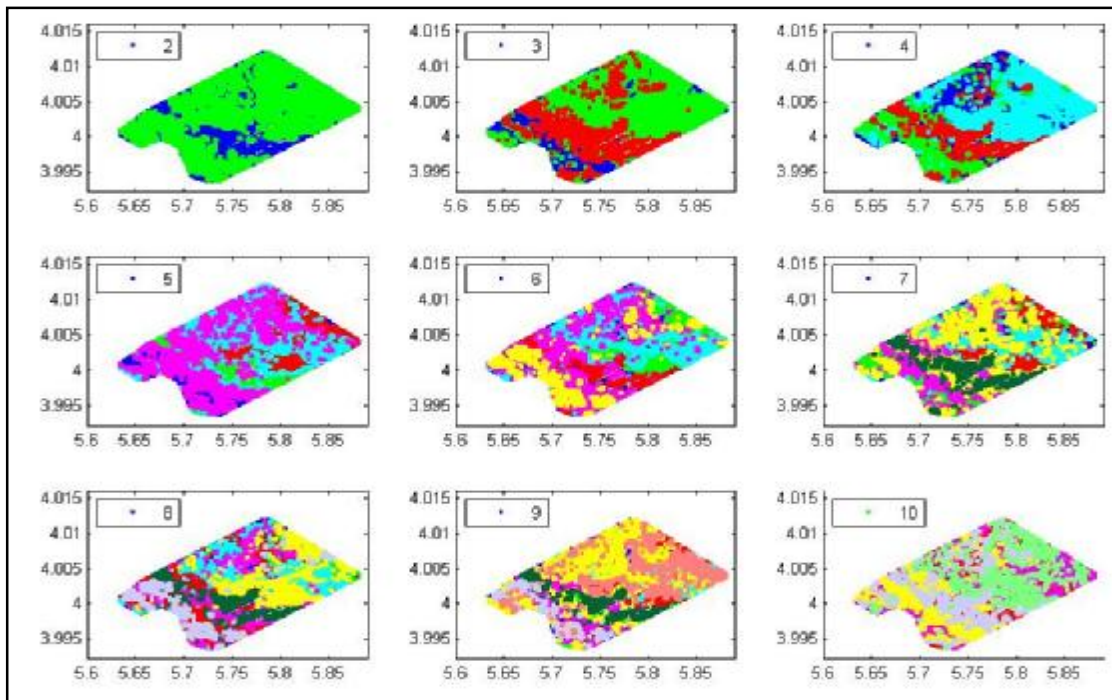
ب. قدر مطلق سیگنال تحلیلی، میزان تغییرات خودپذیری مغناطیسی ظاهری را در یک منطقه کوچک (محدود، مثل یک برگه 1:100,000 زمین‌شناسی) و سطح عمق ثابت نشان می‌دهد و مانند پارامترهای مقاومت الکتریکی و غلظت رادیو عناصر، یک ویژگی از سنگ و نه میدان را بیان می‌کند (رابطه 9):

$$\Delta K = |A|G \quad (9)$$

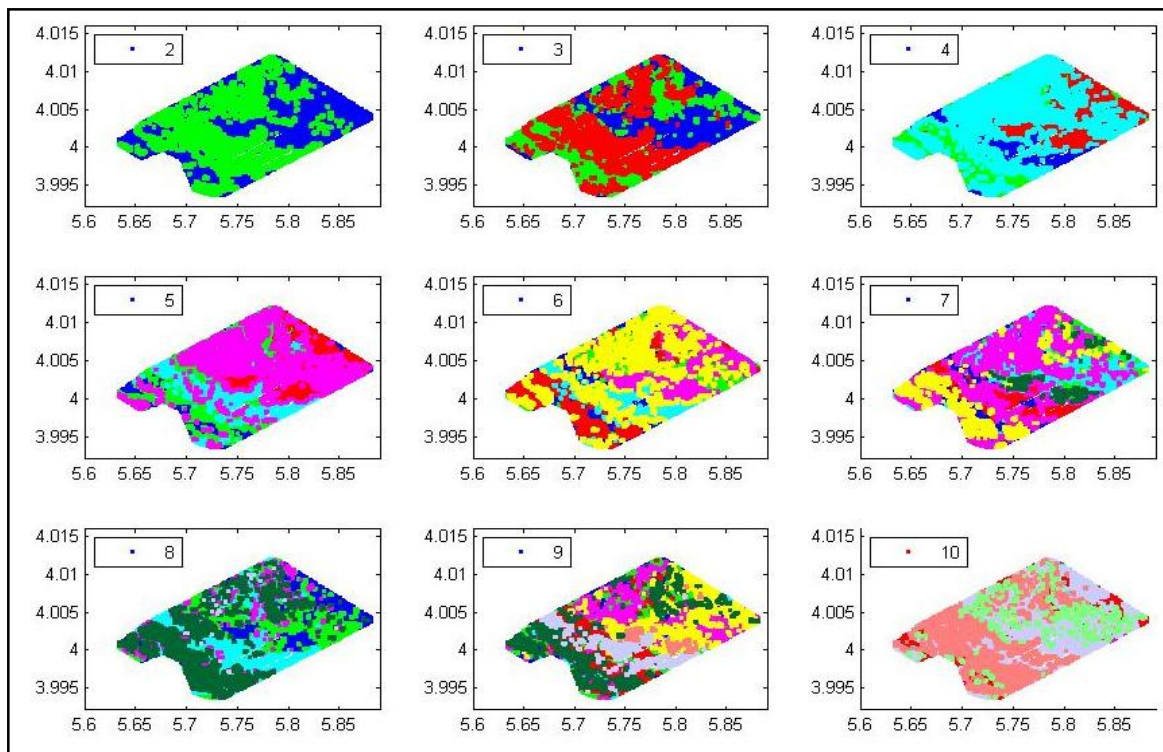
که در آن، $|A|$ قدر مطلق مقدار سیگنال تحلیلی، ΔK خودپذیری ظاهری و $G=f(z, IGRF, d, i)$ است. برای محاسبه پارامتر G در یک منطقه کوچک و یک عمق و جهت پرواز ثابت، مانند داده‌های منطقه مورد مطالعه در این مقاله، مقدار این پارامتر برای کل منطقه ثابت خواهد بود و خودپذیری برای کل منطقه تنها به مقدار سیگنال تحلیلی بستگی خواهد داشت [17].

6- بهینه‌سازی تعداد خوشه‌ها

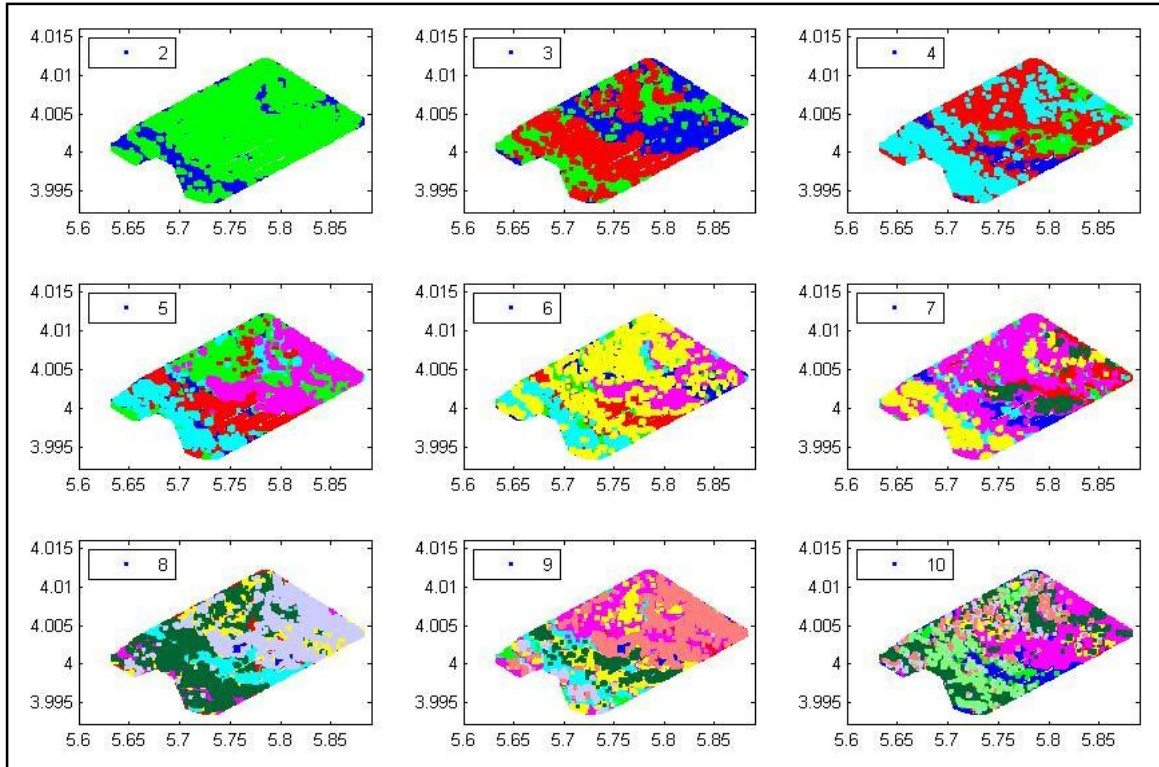
در خوشه‌بندی داده‌های ژئوفیزیک به منظور تهیه نقشه شبه زمین‌شناسی، می‌توان از تعداد واحدهای سنگی مشخص شده بر روی نقشه زمین‌شناسی به عنوان یک ایده استفاده کرد. اما در اکثر موارد این مقدار با تعداد صحیح خوشه‌ها یکسان



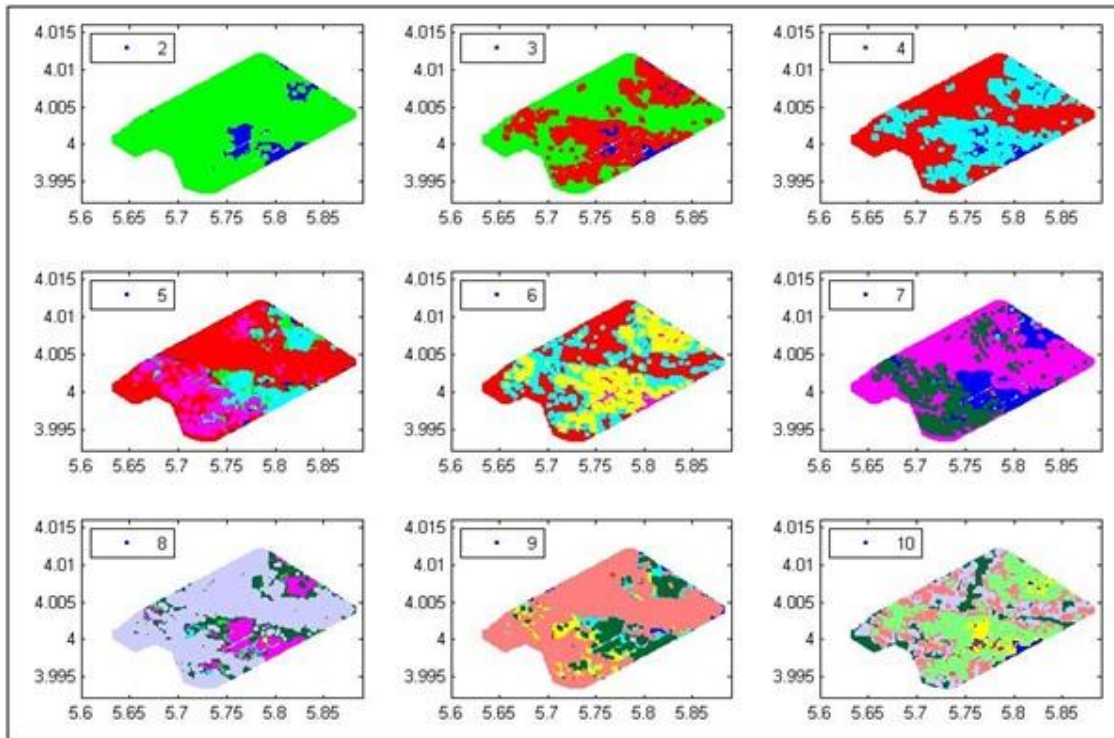
شکل 3: نتایج خوشه‌بندی مجموعه داده $\{K, Th, U, Th/U\}$ از تعداد خوشه 2 تا 10



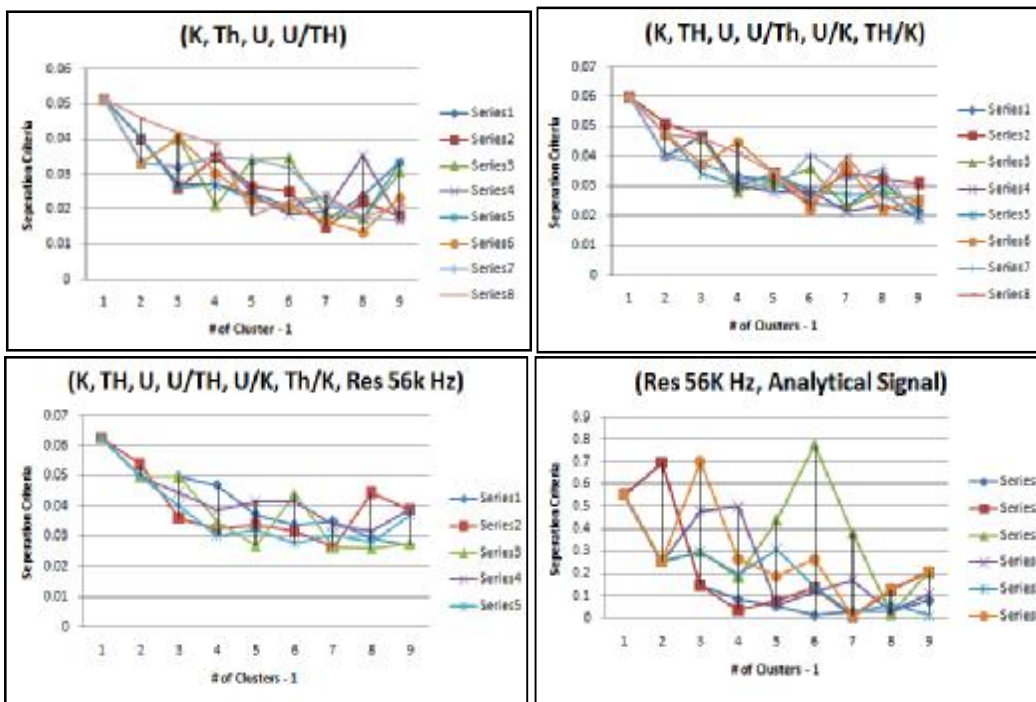
شکل 4: نتایج خوشه‌بندی مجموعه داده $\{K, Th, U, Th/U, U/K, Th/K\}$ از تعداد خوشه 2 تا 10



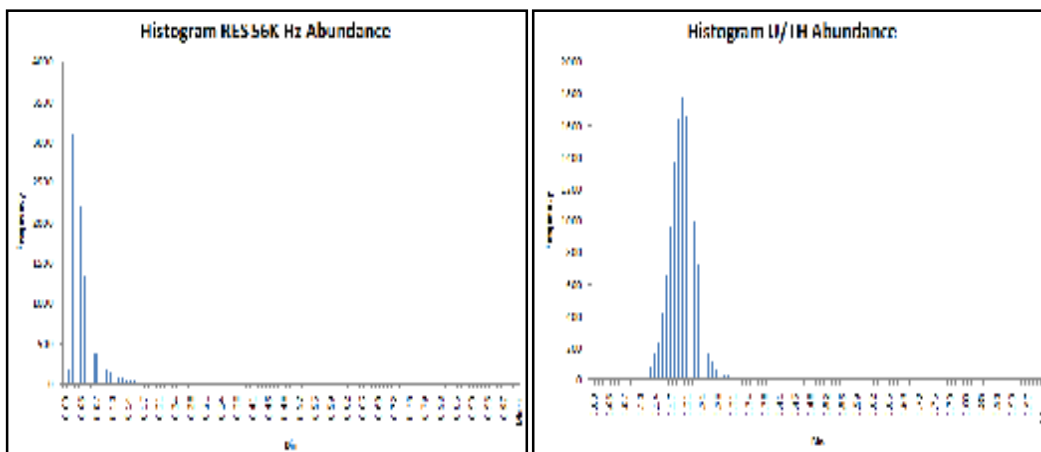
شکل 5: نتایج خوشه‌بندی مجموعه داده { $K, Th, U, Th/U, U/K, Th/K, AS, RES56K$ } از تعداد خوشه 2 تا 10 (مقدار سیگنال تحلیلی)



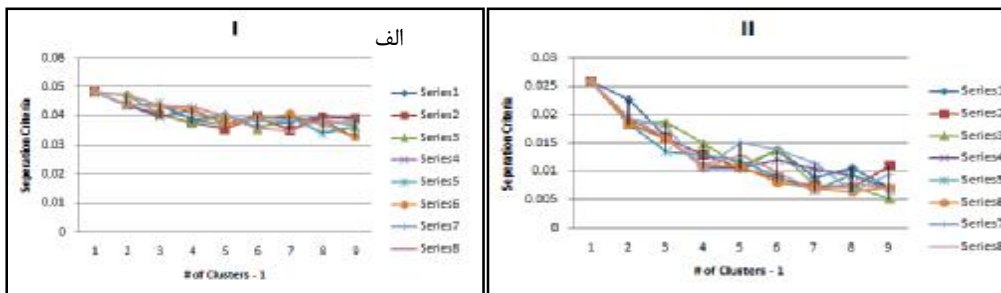
شکل 6: نتایج خوشه‌بندی مجموعه داده { $AS, RES56K, RES7200$ } از تعداد خوشه 2 تا 10 (مقدار سیگنال تحلیلی)



شکل 7: پیدا کردن تعداد بهینه خوشه‌ها برای چهار مجموعه داده در حداقل 6 مرتبه تکرار. محور قائم معیار جدایش و محور افقی تعداد خوشه‌ها است.



شکل 8: هیستوگرام توزیع فراوانی مقاومت الکتریکی و نسبت رادیو عناصر U/TH



ب

شکل 9: نمودار تعیین تعداد خوشه‌های بهینه برای لگاریتم پارامترهای مقاومت الکتریکی و سیگنال تحلیلی در مجموعه داده الف: (راديو عناصر، نسبت راديو عناصر، مقاومت الکتریکی و سیگنال تحلیلی) و ب: (مقاومت الکتریکی و سیگنال تحلیلی). محور قائم معیار جدایش و محور افقی تعداد خوشه‌ها است.

7- تفسیر نتایج خوشه‌بندی

نکات قابل توجه در نتایج مختلف خوشه‌بندی در ادامه مورد اشاره قرار می‌گیرند:

الف. همان‌طور که در نقشه زمین‌شناسی این منطقه (شکل 1) نیز دیده می‌شود، در آن نه واحد سنگی وجود دارد که واحدهای آهکی به رنگ‌های آبی تیره و قهوه‌ای از مساحت بسیار اندکی برخوردارند و نمی‌توان آن‌ها را شناسایی کرد. بنابراین واقع‌بینانه خواهد بود اگر تعداد واحدهای زمین‌شناسی منطقه مورد مطالعه، هشت واحد در نظر گرفته شود. از این رو، همان‌طور که در شکل‌های 3 تا 6 نیز دیده می‌شود، در حالت‌گزینه‌ش بیش از هشت خوشه مفروض، عملیات خوشه‌بندی تفسیر مشخصی نداشته و نتایج قابل‌تعمیم نیستند.

ب. پارامتر خطا در بهینه‌سازی الگوریتم خوشه‌بندی یک پارامتر غیر خطی است و می‌تواند اکسترم‌های محلی زیادی داشته باشد. در الگوریتم‌های تهیه شده سعی شد تا حداقل خطای محاسبه شده به مینیمم مطلق نزدیک باشد ولی با این حال در نقشه‌های با تعداد خوشه بیش از هشت، پاسخ خوشه‌بندی هر اجرا نسبت به اجرای قبل بسیار متفاوت است که این امر را می‌توان ناشی از افزایش کمینه‌های محلی و ضعف روش در پیدا کردن مینیمم خطای نزدیک به مقدار مینیمم مطلق دانست.

ج. در مواردی که تعداد خوشه‌ها، بیش از نه در نظر گرفته می‌شود، الگوریتم‌های خوشه‌بندی، منطقه را به طور عمده به هفت یا هشت خوشه تقسیم می‌کنند. از این رو احتمالاً تعداد خوشه‌های زمین‌شناسی که در این منطقه، می‌توان با داده‌های ژئوفیزیک هوابرد شناسایی کرد، همین تعداد است.

د. انتخاب بازه 2 تا 10 واحد زمین‌شناسی به منظور بررسی تغییرات خوشه‌بندی داده‌ها و تفسیر بهتر زمین‌شناسی است و به عنوان مثال می‌توان به‌طور تقریبی این گونه استنباط کرد که به‌طور مثال در تعداد واحد زمین‌شناسی برابر 2، واحدهای دگرگونه به‌طور مشخصی از سایر واحدهای غیر دگرگونه جدا شده است. این نقشه

می‌تواند در تهیه نقشه زمین‌شناسی با جزئیات بیشتر راهنمای خوبی باشد.

د. نکته دیگری که باید به آن توجه کرد، ترتیب تفکیک واحدهای زمین‌شناسی با افزایش تعداد خوشه‌ها است. این ترتیب عبارت است از:

واحد متاریولیتی، واحد دولومیتی، واحد گنایسی، در بعضی مواقع واحد فیلیت-ماسه سنگ دگرگونی و در صورتی که از داده‌های مقاومت الکتریکی استفاده می‌شود (شکل 6) واحد گرانیتی جزء واحدهایی است که خیلی سریع، به عنوان یک خوشه مشخص از سایر داده‌ها تفکیک می‌شود.

ه. در تمام نقشه‌های تهیه شده از روش‌های خوشه‌بندی، یک تا دو خوشه به‌نسبت وسیع در واحد زمین‌شناسی شیت-فیلیت-اسلیت به وجود آمده است که در نقشه شکل (4) محدوده آن‌ها مشخص شده است. این خوشه‌ها حتی در نقشه‌هایی که از داده‌های رادیومتری استفاده نشده نیز دیده می‌شوند. این محدوده‌ها، با منطقه آنومالی طلای مرتبط با گسل‌های کم شیب و زون‌های برشی⁷ انطباق خوبی دارند که در سال 1378 در مطالعات ژئوشیمیایی منطقه به ثبت رسیده‌اند.

و. از آنجا که داده‌های مقاومت الکتریکی و مغناطیس بر خلاف داده‌های رادیومتری، اطلاعات اعماق بیشتری را ارائه می‌کنند، ممکن است بعضی از واحدهای زمین‌شناسی که در سطح دیده می‌شوند، در عمق وجود نداشته باشند. از این رو همان‌طور که در شکل (6) نیز می‌توان مشاهده کرد، نقشه‌های به دست آمده از این داده‌ها از پیچیدگی کمتری برخوردارند و تمام واحدهایی که در نقشه زمین‌شناسی سطحی دیده می‌شوند، در شکل (6) قابل تشخیص نیستند.

ی. به نظر می‌رسد برای مجموعه داده‌های غیر رادیومتری، تمایل به داشتن حداکثر 4 خوشه وجود دارد. یکی از دلایل به وجود آمدن این مساله تعداد کم پارامترهای ورودی در خوشه‌بندی داده‌های منطقه است. بنابراین انتخاب بهینه تعداد پارامترهای ورودی نیز موضوعی مهم است که پیش از خوشه‌بندی می‌بایست به آن توجه کرد.

منابع:

- [1] Dickson, B.L., Scott, K.M., 1997, *Interpretation of aerial gamma ray surveys-adding the geochemical factors*. AGCO Journal of Australian Geology & Geophysics, Vol. 17, No. 2, pp. 187-200.
- [2] Bonham-Carter, G.F., Agterberg, F.P., Wright, D.F., 1989. *Weights of evidence modeling: A new approach to mapping mineral potential*. In Statistical Application in the Earth Sciences, Geology Survey of Canada, pp. 171-183.
- [3] Caranza, E.J.M, Hale, M., Faassen, C., 2008. *Selection of coherent deposit-type locations and their application in data-driven mineral prospectively mapping*. Ore Geology Reviews, Vol. 33, pp. 536-558.
- [4] Paasche, H., Eberle, D.G., 2010. *Regional geochemical mapping constrained by airborne radiometric data and satellite imagery*. EMG 2010 International Workshop.
- [5] Krishna, P.T.K., Phoha, V.V., Iyengar, S.S., 2008. *Classification of radio elements using mutual information: A tool for Geological mapping*. International Journal of Applied Earth Observation and Geoinformation.
- [6] Martlet, G., Traffert, C., Tourlier, B., Ledre, P., Perrin, J., 2006. *Classifying airborne radiometry data with agglomerative hierarchical clustering: A tool for geological mapping in context of rainforest*. International Journal of Applied Earth Observation and Geoinformation, No. 8, pp. 208-223.
- [7] Eberle, D., Tessema, Abera., Muiwane, E.A., Daudi, E.X., 2009. *Integration of Geophysical and satellite imagery data from the Alto de Liogoha pegmatite fields, northern Mozambique: implications for the control of mineralization*. 11th SAGA Biennial Technical Meeting and Exhibition Swaziland, pp. 21-26.
- [8] Paasche, H., Eberle, D., 2009. *Automated integration of large geophysical data using three partitioning cluster algorithms: A Comparison*. 11th SAGA Biennial Technical Meeting and Exhibition Swaziland, pp. 286-291.
- [9] Velmurugan, T., 2010. *Performance evaluation of K-Means and Fuzzy K-Means clustering algorithms*

8- نتیجه‌گیری

با توجه به نتایج مشاهده می‌شود که استفاده همزمان از چندین مجموعه داده نتایج دقیق‌تری را در تفسیر ارائه می‌دهد. به عنوان مثال مجموعه داده‌های رادیومتری تفکیک خوبی از تمام واحدها ارائه می‌دهند ولی قادر به شناسایی واحد گرانیتهی نیستند. این در حالی است که در مورد مجموعه داده مقاومت الکتریکی و سیگنال تحلیلی، تفکیک بهتری بین واحدهای ولکانیکی (گرانیت و متاریولیت) از سایر واحدها صورت گرفته است.

با مقایسه دو روش بکار رفته می‌توان دریافت که پاسخ نهایی هر دو الگوریتم برای یک حالت خاص، تا حدود زیادی از دقت یکسانی برخوردار است ولی در مورد روش میانگین‌های K فازی تعداد اجرای برنامه برای رسیدن به یک پاسخ قابل قبول کمتر از روش میانگین‌های K است. علاوه بر این، در این مقاله نشان داده شد که نرمال بودن توزیع مقادیر پارامتر-های ورودی امری ضروری است و تعداد اجرای برنامه را برای رسیدن به یک پاسخ قابل قبول کمتر می‌کند. همچنین استفاده از پارامترهایی که بیان‌گر ویژگی زمین (خودپذیری یا سیگنال تحلیلی به جای میدان و مشتقات میدان مغناطیسی یا استفاده از مقاومت الکتریکی به جای بخش حقیقی و موهومی داده‌های الکترومغناطیس) هستند و نه ویژگی‌های میدان، توصیه می‌شود.

علاوه بر این توصیه می‌شود در مطالعات دیگری از روش-های جدیدی مانند الگوریتم ژنتیک برای پیدا کردن اکستریم-های مطلق به منظور پیدا کردن کمینه خطا استفاده شود.

قدردانی و تشکر

بر خود لازم می‌دانیم از زحمات آقایان دکتر علیرضا عرب امیری و مهندس داود رجبی که در کدنویسی و تهیه نقشه-های مقاومت الکتریکی داده‌های الکترومغناطیس هلیکوپتری ما را بسیار یاری نمودند کمال قدردانی و تشکر را بجا بیاوریم. همچنین تشکر ویژه‌ای نیز از مسئولین محترم سازمان زمین شناسی داریم که اجازه استفاده از داده‌های ژئوفیزیک هواپردی را برای انجام این تحقیق دادند.

spectrometry data, International Atomic Energy Agency, TECDOC 1363, IAEA Publications.

[16] Harris, J.R., 1989. *Clustering of gamma ray spectrometer data using a computer image analysis system*. In Statistical Applications in the Earth Sciences, edited by Agterberg and Bonham-Carter, Geological Survey of Canada, Paper 89-9, pp. 9-31.

[17] Thurston, J.B., Smith, R.S., 1997. *Automatic conversion of magnetic data to depth, dip and susceptibility contrast using the SPI (TM) method*. Geophysics, Vol. 62, No. 3, pp. 807-813.

[18] Graham, D.F., Bonham-Carter, G.F., 1993. *Airborne radiometric data: A tool for reconnaissance geological mapping using a GIS*. Photogrammetric Engineering and Remote Sensing, Vol. 58, No. 8, pp. 1243-1249.

Tajoldin, H., 2006. *Introducing Favorable Mineral Zones in 1:100,000 Alut Sheet*, Report of Geology Survey of Iran.

for statistical distributions of input data points. European Journal of Scientific Research, Vol. 46, No. 3, pp. 320-330.

[10] Davis, D.L., Bouldin, D.W., 1979. *A cluster separation measure*, IEEE Trans. Pattern Anal. Machine Intell., Vol. 1, pp. 224-227.

[11] Lucieer, V., Lucieer, A., 2009. *Fuzzy clustering for seafloor classification*. Marine Geology, Vol. 264, pp. 230-241.

[12] Omrani, A., Khabaziyan, R., 2003. *Geology Report for Alut 1:100,000 scale Sheets*, Geology Survey of Iran.

[13] Hodges, G., Fugro Airborne Systems, 2003. *Airborne electromagnetic surveys*. HEM Manual, Fugro Publications.

[14] Siemon, B., 2001. *Improved and new resistivity - depth profiles for helicopter electromagnetic data*. Journal of Applied Geophysics, No. 46, pp. 65- 76.

[15] Nicolet, J.P., Erdi-Hrausz, G. 2003. *Guidelines for radioelement mapping using gamma ray*

¹ . *clustering, unsupervised classification*

² . *K- means and fuzzy K- means clustering methods*

³ . *fuzzy fire*

⁴ . *partition coefficient*

⁵ . *Siemon's inversion algorithm*

⁶ . *Minimum Curvature*

⁷ . *shear zones*