

## مقایسه روش های یادگیری غیرنظارتی با تأکید بر تشخیص رخساره های کانالی تنگه هرمز

امیرحسین مردان<sup>۱</sup>، عبدالرحیم جواهریان<sup>۲\*</sup> و مرضیه میرزاخانیا<sup>۳</sup>

۱- دانشجوی کارشناسی ارشد، دانشکده مهندسی نفت، دانشگاه صنعتی امیرکبیر  
۲- استاد، دانشکده مهندسی نفت، دانشگاه صنعتی امیرکبیر و استاد بازنشسته مؤسسه ژئوفیزیک دانشگاه تهران  
۳- کارشناس ارشد ژئوفیزیک، مدیریت اکتشاف شرکت ملی نفت ایران

دریافت مقاله: ۱۳۹۴/۱۲/۱۲؛ پذیرش مقاله: ۱۳۹۵/۰۳/۰۱

\* نویسنده مسئول مکاتبات: javaherian@aut.ac.ir

### چکیده

### واژگان کلیدی

کانال ها یکی از انواع رخساره های زمین شناسی می باشند که به دلیل توانایی در ذخیره سیالات هیدروکربنی، در اکتشاف و توسعه میادین هیدروکربنی دارای اهمیت فراوانی می باشند. در سال های اخیر، حجم داده های لرزه ای و همچنین تعداد نشانگرهای لرزه ای ارائه شده افزایش چشمگیری داشته است که کار مفسرین را برای تفسیر خط به خط داده های لرزه ای با مشکل مواجه کرده است. برای برطرف نمودن این مشکلات، الگو شناسی و استفاده از نشانگرهای چندگانه به عنوان ابزاری کارآمد در تفسیر و دسته بندی رخساره های لرزه ای معرفی شده اند. روش های k-میانگین (k-means)، نقشه های خودسازمان ده (Self-organizing maps) و نقشه های توپوگرافی مولد (Generative topographic maps) از روش های غیرنظارتی می باشند که توانسته اند برای دسته بندی رخساره های لرزه ای مورد استفاده قرار بگیرند. در این مطالعه، توانایی دو الگوریتم k-میانگین و نقشه های خودسازمان ده برای تشخیص کانال های مدفون در داده های لرزه ای انجام و مقایسه شده است، از روش تحلیل مؤلفه اصلی (Principal component analysis) نیز برای به تصویر کشیدن رخساره های لرزه ای موجود در داده لرزه ای مورد استفاده که مربوط به تنگه هرمز در خلیج فارس است، استفاده شده است. پس از مشخص نمودن نشانگرهای مناسب و اعمال روش های مورد اشاره، روش تحلیل مؤلفه اصلی به عنوان روشی مناسب جهت تعیین تقریبی تعداد رخساره های لرزه ای موجود در محدوده مورد مطالعه و شناسایی رخساره های کانالی تشخیص داده شد. اگرچه اعمال این الگوریتم ها بر روی پنجره محاسباتی باعث کاهش تفکیک پذیری داده های لرزه ای می گردد ولی نسبت به اعمال آن ها بر یک برش زمانی خاص کیفیت بهتری ارائه می کند. با توجه به مطالعات انجام گرفته مشخص گردید یک حوضچه توریدایتی در شرق و جنوب شرقی منطقه مورد مطالعه وجود دارد که به دلیل شیب منطقه که ناشی از بالآمدگی رخساره نمکی در سمت غرب این منطقه است، سیستم کانالی موجود رسوبات را از سمت غرب، به این حوضچه وارد می نمایند.

تحلیل مؤلفه اصلی  
تشخیص کانال  
تنگه هرمز  
خوشه بندی  
یادگیری غیرنظارتی

## ۱- مقدمه

کانال‌ها یکی از انواع رخساره‌های زمین‌شناسی هستند که مطالعه آن‌ها دارای اهمیت فراوانی در صنعت نفت است چرا که این نوع رخساره زمین‌شناسی می‌تواند محلی مناسب برای تجمع سیالات هیدروکربنی باشد. در صورتی که کانال‌ها از سنگ منشأ مناسبی تغذیه گردند و پوش سنگ مناسبی نیز از خروج سیالات هیدروکربنی موجود در آن‌ها جلوگیری نماید، می‌توانند به عنوان یکی از انواع تله‌های نفتی قابل اهمیت، هدف اکتشافی محسوب گردند. از طرف دیگر شناسایی کانال‌های تراوا جهت انجام مطالعات مربوط به ازدیاد برداشت و بررسی سازوکار تزریق، اهمیت شناسایی این رخساره‌های زمین‌شناسی را نشان می‌دهد. به همین دلیل در سال‌های اخیر، مطالعات فراوانی در زمینه شناسایی بهتر کانال‌ها با استفاده از داده‌های لرزه‌ای انجام گرفته است.

Robinson and Davis (2012) از تجزیه طیفی برای شناسایی رخساره کانالی موجود در داده لرزه‌ای سه‌بعدی برداشت‌شده از میدان Delhi در لوئیزیانا استفاده نمودند تا بتوانند بازدهی تزریق گاز را جهت ازدیاد برداشت افزایش دهند. Roy et al. (2013) از الگو شناسی جهت تشخیص رخساره‌های گوناگون از جمله کانال موجود در داده‌های لرزه‌ای سه‌بعدی برداشت‌شده از حوضه آنادارکو (Anadarko) استفاده نمودند. Junhuai et al. (2014) با استفاده از نمودارهای مقاطع نشانگرهای لرزه‌ای به شناسایی کانال‌ها در داده‌های لرزه‌ای برداشت‌شده از حوضه Junggar در کشور چین پرداختند. (Al-Dossary (2015 با مقایسه سه روش تضعیف نوفه، به اهمیت مراحل پردازش در شناسایی کانال‌ها و تفاوت سه روش مطرح‌شده در این زمینه پرداخته است. Ashtari and Arzani (2016) با استفاده از تجزیه طیفی و ترکیب برش لرزه‌ای به‌دست آمده از سه فرکانس ۲۵، ۲۷ و ۳۰ هرتز به شناسایی کانال مدفون در سازند آسماری پرداخته‌اند.

با توجه افزایش حجم داده‌های لرزه‌ای در سال‌های اخیر، تفسیر خط به خط این داده‌ها برای مفسرین دچار مشکلاتی بوده است. اگرچه با افزایش تعداد نشانگرهای لرزه‌ای، دقت مطالعات انجام‌گرفته در تشخیص رخساره‌های لرزه‌ای گوناگون افزایش یافته است، ولی افزایش حجم اطلاعات تولیدی باعث افزایش زمان مورد نیاز جهت دسته‌بندی این رخساره‌ها می‌گردد. برای برطرف کردن مشکلات مطرح‌شده، روش‌های گوناگونی از الگو شناسی شامل k-میانگین (k-means)، نقشه‌های خودسازمان‌ده (Self-organizing maps) و نقشه‌های توپوگرافی مولد (Generative topographic mapping) از انواع دسته‌بندی غیرنظارتی رخساره‌های لرزه‌ای و شبکه عصبی (Neural network) و سامانه بردار پشتیبان (Support vector machine) برای دسته‌بندی نظارت‌شده رخساره‌های لرزه‌ای مورد توجه مفسران قرار گرفته‌اند. در این مطالعه، الگوریتم‌های آموزش

غیرنظارتی (Unsupervised learning) جهت شناسایی کانال‌های مدفون در داده‌های لرزه‌ای با یکدیگر مقایسه شده‌اند. داده لرزه‌ای استفاده‌شده در این مطالعه، مربوط به تنگه هرمز در خلیج فارس است.

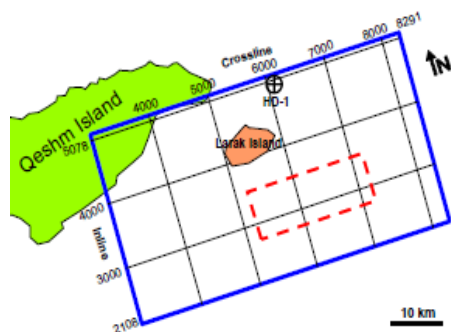
پس از معرفی و توسعه نشانگرهای لرزه‌ای (Balch, 1971; Taner et al., 1979)، دسته‌بندی اطلاعات در لرزه‌شناسی توسط رایانه‌ها با مطالعات (Sonneland (1983 و Justice et al., (1985) مورد توجه قرار گرفت. الگوریتم خوشه‌بندی k-میانگین (Forgy, 1966; Jancey, 1965) از اولین الگوریتم‌های خوشه‌بندی محسوب می‌شود که به سرعت توسط شرکت‌های مختلف مورد استفاده قرار گرفت و امروزه نیز در بسیاری از نرم‌افزارهای تفسیری کارایی دارد. Barnes and Laughlin (2002) تعدادی از روش‌های خوشه‌بندی غیرنظارتی شامل k-میانگین و نقشه‌های خودسازمان‌ده را مورد تحلیل و بررسی قرار دادند و به این نتیجه رسیدند که اهمیت انتخاب نشانگرهای مناسب به مراتب از نوع الگوریتم استفاده‌شده بیشتر است. آن‌ها در بین الگوریتم‌های مورد بررسی، نقشه‌های خودسازمان‌ده را بیشتر مورد توجه قرار دادند چرا که در این روش، از روابط چندجانبه بین خوشه‌ها نیز اطلاعات کسب می‌گردد. Strecker and Uden (2002) را می‌توان از اولین افرادی دانست که از الگوریتم نقشه‌های خودسازمان‌ده دوبعدی برای خوشه‌بندی داده‌های لرزه‌ای استفاده نمودند. Gao (2007) از نقشه‌های خودسازمان‌ده برای تهیه نقشه رخساره‌های لرزه‌ای داده‌های دریایی آنگولا استفاده نمود. Roden et al. (2015) از نقشه‌های خود سازمان‌ده برای شناسایی الگوهای موجود در پدیده‌های زمین‌شناسی استفاده نمودند. Zhao et al. (2015) از روش‌های k-میانگین و نقشه‌های خودسازمان‌ده برای خوشه‌بندی رخساره‌های موجود در داده لرزه‌ای برداشت‌شده از منطقه دریایی نیوزیلند استفاده نمودند و برای منطقه مورد مطالعه چهار نوع رخساره پیشنهاد دادند.

در این مقاله، ابتدا زمین‌شناسی منطقه مورد مطالعه قرار می‌گیرد و سپس خلاصه‌ای در رابطه با روش تحلیل مؤلفه اصلی (Principal component analysis) که یکی از روش‌های یادگیری غیرنظارتی برای آماده‌سازی مجموعه داده‌های در دسترس است ارائه می‌گردد. در ادامه خلاصه‌ای از روش‌های خوشه‌بندی استفاده‌شده در تفکیک رخساره‌های لرزه‌ای گوناگون شامل روش خوشه‌بندی k-میانگین و نقشه‌های خودسازمان‌ده مورد بحث قرار می‌گیرند. بعد از معرفی داده لرزه‌ای مورد مطالعه، نشانگرهای لرزه‌ای مورد استفاده معرفی می‌گردند و الگوریتم‌های اشاره‌شده با استفاده از نشانگرهای انتخابی اجرا می‌گردند و پاسخ دریافتی از الگوریتم‌های گوناگون جهت تشخیص بهتر کانال‌های منطقه مورد مقایسه قرار می‌گیرند.

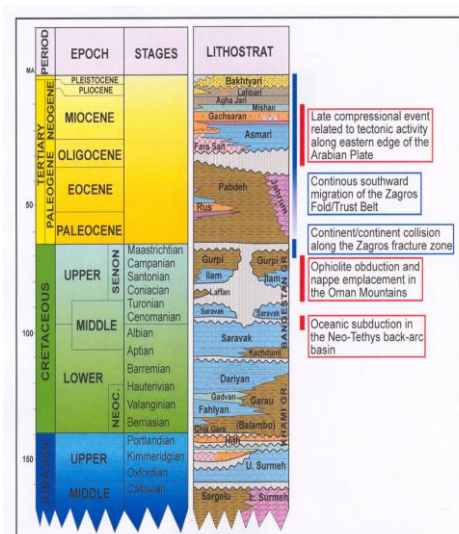
$$C = \frac{1}{m} A^* A, \quad (1)$$

که  $m$  تعداد بردار نشانگرها است. سپس برای کاهش ابعاد این ماتریس از ۳ بعد به ۲ بعد، ماتریس داده‌های لرزه‌ای در دو بردار ویژه اول از ماتریس کواریانس ضرب می‌گردد و بنابراین ماتریس نشانگرها در این فضای دوبعدی تصویر می‌گردد و ماتریس حاصل دارای  $m$  سطر و دو ستون خواهد بود. باید توجه شود که دو مشخصه فعلی حاصل ترکیب سه نشانگری هستند که به عنوان ورودی تحلیل مؤلفه اصلی مورد استفاده قرار گرفته‌اند و بنابراین برابر با هیچ یک از این سه نشانگر نمی‌باشند.

برای تعیین تعداد مؤلفه‌های اصلی مناسب جهت انجام مطالعات بعدی از طیف ویژه (Eigenspectrum) استفاده می‌گردد تا مشخص شود چه تعداد از مؤلفه‌های اصلی بیشترین اطلاعات را ارائه می‌نمایند (Roden et al, 2015). در این مطالعه از تحلیل مؤلفه اصلی به عنوان مرحله اول از اجرای دیگر الگوریتم‌ها استفاده شده است تا ابعاد داده‌های ورودی به این الگوریتم‌ها را به دو بعد کاهش دهد. روند نمای اجرای این الگوریتم را می‌توان در شکل ۳ مشاهده کرد.



شکل ۱: موقعیت جغرافیایی منطقه مورد مطالعه، این مطالعه در محدوده مشخص شده با خط چین قرمز انجام شده است.



شکل ۲: جدول چینه‌شناسی منطقه مورد مطالعه (Wynd, 1965).

## ۲- زمین‌شناسی منطقه مورد مطالعه

منطقه تنگه هرمز در جنوب شرقی استان هرمزگان و در حد فاصل خلیج فارس و دریای عمان قرار دارد. منطقه مورد مطالعه بخش شرقی خلیج فارس و در جنوب شرقی جزیره لارک در تنگه هرمز است که در شکل ۱ قابل مشاهده است. این منطقه در محل تلاقی چند پهنه زمین‌شناسی قرار دارد. در واقع، کمربند چین خورده زاگرس در شمال غربی، سکوی عربی در جنوب غربی، حوضه مکران در شمال شرقی، کوه‌های عمان و شبه جزیره مسندم در جنوب شرقی، این پهنه‌های زمین‌شناسی را تشکیل می‌دهند. کوهزایی عمان (کرتاسه بالایی و پالتوسن زیرین) و مراحل اولیه کوهزایی زاگرس (الیگوسن-میوسن زیرین) در جریان رو به بالای نمک هرمز کامبرین تأثیر زیادی گذاشته است (Alsouki et al., 2008). تحت تأثیر فاز چین‌خوردگی صفحه عربی با صفحه اوراسیا (الیگوسن) و گسترش آن به طرف جنوب شرقی در منطقه (میوسن زیرین)، دگرشیبی مهمی در بخش گوری (میوسن میانی) قاعده سازند میشان ایجاد شده است که به خوبی در داده‌های لرزه‌ای منطقه قابل مشاهده است. تأثیرات کوهزایی عمان در الیگوسن، بر ناپیوستگی و فاز فرسایشی متعاقب آن در لایه‌های زیرین سطح دگرشیبی گوری تأثیر زیادی گذاشته است. چین‌خوردگی و راندگی در زیر سطح ناپیوستگی قاعده گوری حاکی از عملکرد مراحل اولیه فاز کوهزایی زاگرس (میوسن زیرین) است (Alsouki et al., 2008). شکل ۲ جدول چینه‌شناسی منطقه هرمز را نشان می‌دهد.

## ۳- مروری بر روش‌های یادگیری غیرنظارتی

یادگیری غیرنظارتی نوعی از یادگیری سامانه (Machine learning) است که در آن، داده‌های ورودی به سامانه بدون برچسب می‌باشند.

### ۳-۱- تحلیل مؤلفه اصلی

یکی از روش‌های یادگیری غیرنظارتی، تحلیل مؤلفه اصلی است که روشی برای خوشه‌بندی اطلاعات نیست، بلکه جهت کاهش ابعاد داده‌ها، استخراج مشخصه، کاهش افزونگی اطلاعات (Redundancy) و تجسم اطلاعات مورد استفاده قرار می‌گیرد (Bishop, 2006). در رابطه با کاهش افزونگی داده‌ها، اکثر اطلاعات مفید و مورد نیاز در چند مؤلفه اصلی اول حفظ می‌شوند و مؤلفه‌های دیگر دارای نوفه بیشتری می‌باشند. مؤلفه‌های اصلی یک ماتریس از تجزیه مقدار تکین کواریانس آن ماتریس محاسبه می‌گردد. سپس با تصویر نمودن تعداد مورد نظر از بردارهای ویژه به دست آمده از تجزیه تکین ماتریس کواریانس بر ماتریس داده‌های لرزه‌ای، ابعاد این ماتریس کاهش می‌یابد. فرض نمایید از سه نشانگر لرزه‌ای استفاده گردد، در این صورت ماتریس داده‌های لرزه‌ای (A) مورد مطالعه ۳ بعدی خواهد بود. برای محاسبه ماتریس کواریانس از رابطه (۱) استفاده می‌گردد:

## ۳-۲- روش k-میانگین

روش خوشه‌بندی k-میانگین (MacQueen, 1967) را احتمالاً می‌توان ساده‌ترین الگوریتم خوشه‌بندی دانست که در بسیاری از نرم‌افزارهای تفسیری نیز مورد استفاده قرار گرفته است. بعد از تشخیص تعداد خوشه‌ها، برای شروع تکرار، مرکز خوشه‌ها به صورت شبکه‌ای و یا اتفاقی تعیین می‌شود. سپس فاصله هر نقطه از داده‌های ورودی با مراکز خوشه محاسبه می‌گردد و هر نمونه به نزدیک‌ترین خوشه تعلق می‌گیرد. در این مرحله از نمونه‌های موجود در هر خوشه میانگین گرفته می‌شود و مرکز خوشه‌ها به این نقاط منتقل می‌گردند و تکرار بعد برای تعلق دادن دوباره نمونه‌ها به خوشه‌ها برحسب فاصله آن‌ها با مراکز جدید انجام می‌گیرد. اگر k خوشه وجود داشته باشد، این فرآیند تقریباً با k تکرار همگرا می‌شود (Zhao et al., 2015). محاسبه فاصله نمونه‌ها با مراکز هر خوشه با استفاده از فاصله اقلیدسی انجام می‌گیرد.

اجرای الگوریتم k-میانگین سریع و ساده است ولی این نوع خوشه‌بندی بدون ساختار است و اطلاعاتی از رابطه بین یک خوشه با خوشه‌های دیگر ارائه نمی‌گردد. روند نمای اجرای این الگوریتم در شکل ۴ قابل مشاهده است.



شکل ۳: روند نمای استخراج مؤلفه‌های اصلی داده‌های لرزه‌ای.

## ۳-۳- نقشه‌های خودسازمان‌ده

نقشه‌های خودسازمان‌ده (Kohonen, 1982) یکی از روش‌های یادگیری غیرنظارتی است که ابتدا برای تشخیص ژن مورد استفاده قرار گرفت و امروزه به یکی از روش‌های رایج در خوشه‌بندی رخساره‌های لرزه‌ای تبدیل شده است و حداقل در چهار بسته نرم‌افزاری برای دسته‌بندی رخساره‌های لرزه‌ای مورد استفاده قرار می‌گیرد (Zhao et al., 2015). مهم‌ترین مزیت الگوریتم نقشه‌های خودسازمان‌ده نسبت به روش k-میانگین، به نقشه درآمدن خوشه‌های به دست آمده از داده‌ها در یک فضای مخفی (Latent

space) با شبکه‌بندی (Gridding) منظم است که در واقع ساختار و ارتباط هر خوشه را نسبت به خوشه‌های دیگر ارائه می‌دهد. محدودیت‌های الگوریتم نقشه‌های خودسازمان‌ده عدم وجود یک معیار کمی برای نشان دادن میزان دقت و صحت خوشه‌بندی انجام‌شده است (Zhao et al., 2015). برای اجرای روش خوشه‌بندی نقشه‌های خودسازمان‌ده پس از آنکه داده‌های موجود مانند استفاده در دو الگوریتم قبل نرمال شدند، طبق مراحل زیر عمل می‌گردد که روند نمای آن در شکل ۵ قابل مشاهده است.

(الف) انتخاب اتفاقی یک بردار نشانگر ورودی (aj) از مجموعه داده‌های ورودی.

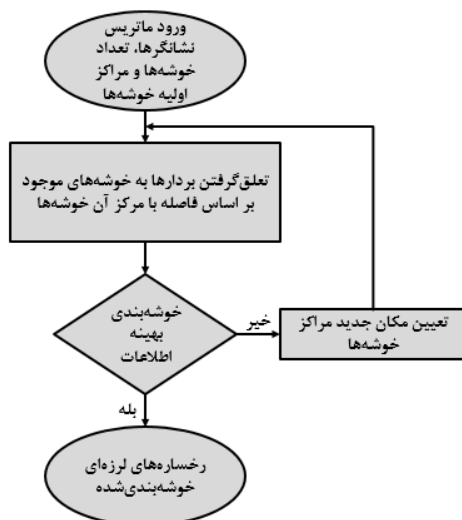
(ب) فاصله بین نمونه انتخابی (aj) با بردارهای اولیه (mk) با استفاده از رابطه (۲) محاسبه گردد. بردار اولیه‌ای که کمترین فاصله را با بردار ورودی (aj) دارد، بردار برنده (mb) نامیده می‌شود،

$$\|a_j - m_b\| = \min\{\|a_j - m_k\|\}. \quad (2)$$

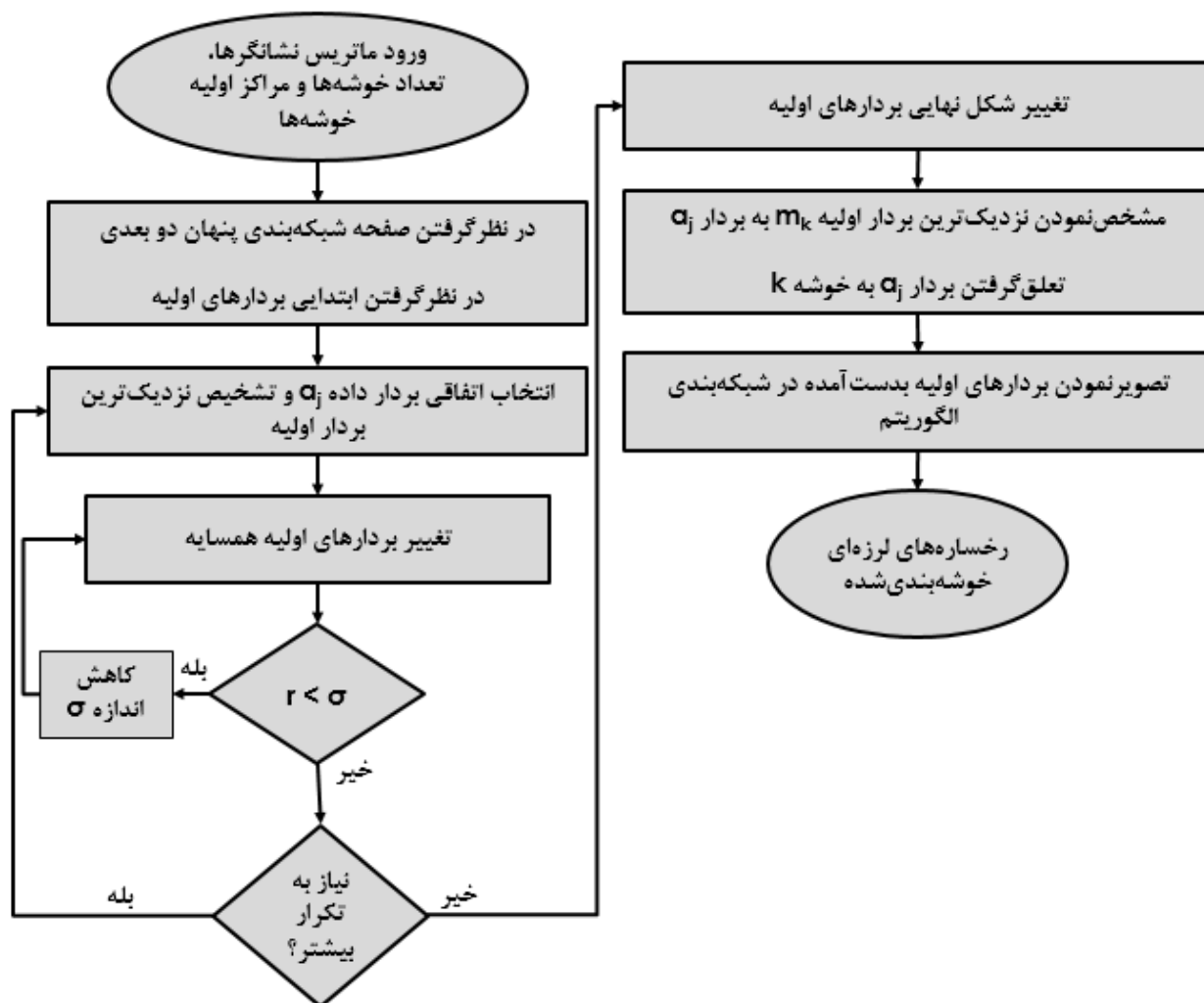
(ج) به‌روز شدن بردار اولیه برنده و همسایه‌های آن. قانون به‌روز شدن برای وزن بردار اولیه k ام در داخل و خارج از شعاع همسایگی ( $\sigma(t)$ ) به صورت رابطه (۲) است.

$$m_k(t+1) = \begin{cases} m_k(t) + \alpha(t) h_{bk}(t) [a_j - m_k(t)], & \text{if } \|r_k - r_b\| \leq \sigma(t) \\ m_k(t), & \text{if } \|r_k - r_b\| > \sigma(t) \end{cases} \quad (3)$$

که شعاع همسایگی در هر تکرار (t) کاهش می‌یابد. پارامترهای rb و rk مکان بردار اولیه برنده و بردار اولیه k ام می‌باشند. تابع همسایگی  $h_{bk}(t)$  تابع نمایی یادگیری  $\alpha(t)$  از دیگر پارامترهای مؤثر در رابطه (۳) می‌باشند که به صورت روابط (۴) و (۵) ارائه می‌گردند.



شکل ۴: روند نمای خوشه‌بندی رخساره‌های لرزه‌ای با استفاده از الگوریتم k-میانگین.



شکل ۵: روند نمای خوشه‌بندی رخساره‌های لرزه‌ای توسط الگوریتم نقشه‌های خودسازمان‌ده.

مناسب مهم‌ترین مرحله در خوشه‌بندی رخساره‌های لرزه‌ای به وسیله رایانه‌ها است. به طور کلی نشانگرهای دامنه و فرکانس، تشخیص‌دهنده جنس سنگ‌ها می‌باشند و نشانگرهای هندسی از قبیل شیب و انحنا برای تشخیص ویژگی‌های مربوط به شکل بازتابنده‌ها هستند و این در حالی است که نشانگرهای بافتی مانند GLCM (Grey-level co-occurrence matrix) در رابطه با توزیع داده‌ها، اطلاعات ارائه می‌کنند (Chopra and Marfurt, 2007).

در خوشه‌بندی به وسیله رایانه باید تا حد امکان از تعداد کمتری نشانگر لرزه‌ای استفاده گردد و نشانگرهای استفاده‌شده باید از یکدیگر مستقل باشند (Zhao et al., 2015). در این مطالعه پس از بررسی و تحلیل‌های گوناگون، سه نشانگر تجزیه طیفی، میزان انحنا (Curvedness) و همگنی GLCM به عنوان نشانگرهای لرزه‌ای ورودی انتخاب شدند. فرکانس به نازک‌شدگی لایه‌ها در مجاورت کانال‌های موجود و جنس آن‌ها حساس است. انحنا یک نشانگر مرتبط با سطح است که با مشتق دوم سطح ارتباط دارد و به همین دلیل بر روی تغییرات شکل ساختارها متمرکز است و از شکل کانال

$$h_{bk}(t) = e^{-\|r_b - r_k\|^2 / 2\sigma^2(t)} \quad (4)$$

$$\alpha(t) = \alpha_0 \left( \frac{0.005}{\alpha_0} \right)^{t/T} \quad (5)$$

که  $T$  طول یادگیری است.

(د) تکرار مراحل یادگیری (مراحل الف) تا ج) تا رسیدن به

ملاک همگرایی که از قبل تعریف شده است.

(ه) تصویر بردارهای اولیه بر دو مؤلفه اصلی خروجی از تحلیل

مؤلفه اصلی.

تفاوت و نحوه کارایی سه روش یادگیری غیرنظارتی مطرح‌شده

را می‌توان در شکل ۶ که یک شکل آموزشی است، مشاهده نمود.

#### ۴- نشانگرهای لرزه‌ای

(Barnes and Laughlin, 2002) با مقایسه تعدادی از روش‌های

یادگیری غیرنظارتی به این نتیجه رسیدند که انتخاب نشانگرهای

دست آمده با اطلاعات قبلی موجود از پدیده‌های شاخص منطقه، چهار عدد تشخیص داده شد.

در نهایت با استفاده از قابلیت نمایشی تحلیل مؤلفه اصلی و نمایش تعداد مؤلفه‌های انتخابی در سیستم RGB، به شناخت کانال‌های نهفته در محدوده برش زمانی ۱۱۵۲ میلی‌ثانیه پرداخته شد که در شکل ۱۰ (ج) قابل مشاهده است. مزایای این روش نسبت به دو روش قبل، عدم نیاز به مشخص نمودن تعداد رخساره‌های لرزه‌ای (تعداد خوشه‌ها)، موجود در محدوده مورد مطالعه است. همان‌طور که در شکل ۱۰ (ج) می‌توان دید، بدون تعیین تعداد خوشه‌های موجود در پنجره محاسباتی مورد مطالعه، تحلیل مؤلفه اصلی، رخساره‌های لرزه‌ای موجود در این پنجره محاسباتی را به صورت چهار نوع رخساره عمده نشان داده است که با تعداد خوشه‌های مناسب تشخیص داده شده برای این محدوده تطابق کامل دارد. باید دقت شود که تعداد خوشه‌ها را نمی‌توان همیشه با این دقت توسط تحلیل مؤلفه اصلی تخمین زد ولی این روش تقریب مناسبی برای تعداد رخساره‌ها ارائه می‌نماید.

همان‌طور که از شکل‌های ۱۰ (الف) تا ۱۰ (ج) مشخص است، تحلیل مؤلفه اصلی برای شناسایی کانال‌های دقت بیشتری دارد. همچنین نتایج به دست آمده از دو الگوریتم  $k$ -میانگین و نقشه‌های خودسازمانده نشان می‌دهد که پاسخ به دست آمده از این دو الگوریتم شباهت بسیار زیادی به یکدیگر دارند.

آنچه از مطالعات مشخص است، عدم توانایی روش‌های مطرح‌شده در تفکیک گسل‌های موجود در منطقه با رسوبات کانالی است که شاید این امر بتواند به گونه‌ای اهمیت انتخاب نشانگرهای لرزه‌ای مورد استفاده را بیش از پیش نشان دهد.

با توجه به مطالعات انجام‌گرفته در این منطقه مشخص گردید که به دلیل بالاآمدگی رخساره‌های نمکی در کف حوضه، شیب منطقه به سمت شرق و جنوب شرقی است و کانال‌های مشخص‌شده، رسوبات را از سمت غرب به این حوضه توریدیتی وارد می‌کنند.

آنچه که از مقایسه پاسخ‌های حاصل از اجرای سه روش یادگیری غیرنظارتی مشخص گردید نشان می‌دهد ابتدا برای تعیین تقریبی تعداد رخساره‌های نهفته در محدوده مورد مطالعه بهتر است از روش تحلیل مؤلفه اصلی استفاده گردد. سپس با توجه به هدف مطالعه، در صورتی که شناسایی سیستم کانالی و یا گسلی منطقه مورد نیاز باشد، تحلیل مؤلفه اصلی پاسخ مناسبی ارائه می‌دهد و در صورتی که هدف خوشه‌بندی رخساره‌های لرزه‌ای نهفته در محدوده مورد مطالعه باشد می‌توان از پاسخ روش‌های  $k$ -میانگین و نقشه‌های خودسازمانده استفاده نمود و جهت افزایش اطمینان از صحت خوشه‌بندی، پاسخ دریافتی با پاسخ حاصل از تحلیل مؤلفه اصلی مقایسه گردد زیرا همان‌طور که اشاره شد، تحلیل مؤلفه اصلی توانایی خوشه‌بندی ندارد و نمی‌تواند داده‌ها را به عنوان مثال به صورت خوشه‌های شماره ۱، شماره ۲ و ... ارائه نماید و تنها باعث بهبود

اثر می‌گیرد و همگنی GLCM که یک نشانگر بافتی است که از نحوه توزیع داده‌ها اطلاعات ارائه می‌کند و در حالت شباهت ردلرزه‌های مجاور، بیشترین مقدار را نشان می‌دهد.

هدف اصلی از این مطالعه، مقایسه سه الگوریتم  $k$ -میانگین، نقشه‌های خودسازمانده و تحلیل مؤلفه اصلی جهت شناسایی کانال‌های موجود در داده‌های لرزه‌ای است که به صورت موردی بر محدوده برش زمانی ۱۱۵۲ میلی‌ثانیه داده لرزه‌ای برداشت‌شده از منطقه هرمز مورد استفاده قرار گرفته است. دامنه لرزه‌ای این برش زمانی در شکل ۷ (الف) قابل مشاهده است. شکل ۷ (ب) همان برش زمانی را بعد از اعمال نشانگر تجزیه طیفی نشان می‌دهد. در شکل ۷ (ج) نیز نشانگرهای ترکیبی (Marfurt, 2015) میزان انحنا و شاخص شکل (Shape index) در برش مورد مطالعه را می‌توان مشاهده نمود. پاسخ دریافتی از اعمال نشانگر همگنی GLCM نیز در شکل ۷ (د) قابل مشاهده است.

مقطع گرفته‌شده از خط AA' که در شکل ۷ (الف) رسم شده است را می‌توان در شکل ۸ شکل ۸ مشاهده نمود. رخساره‌های کانالی با پیکان سیاه‌رنگ مشخص شده‌اند.

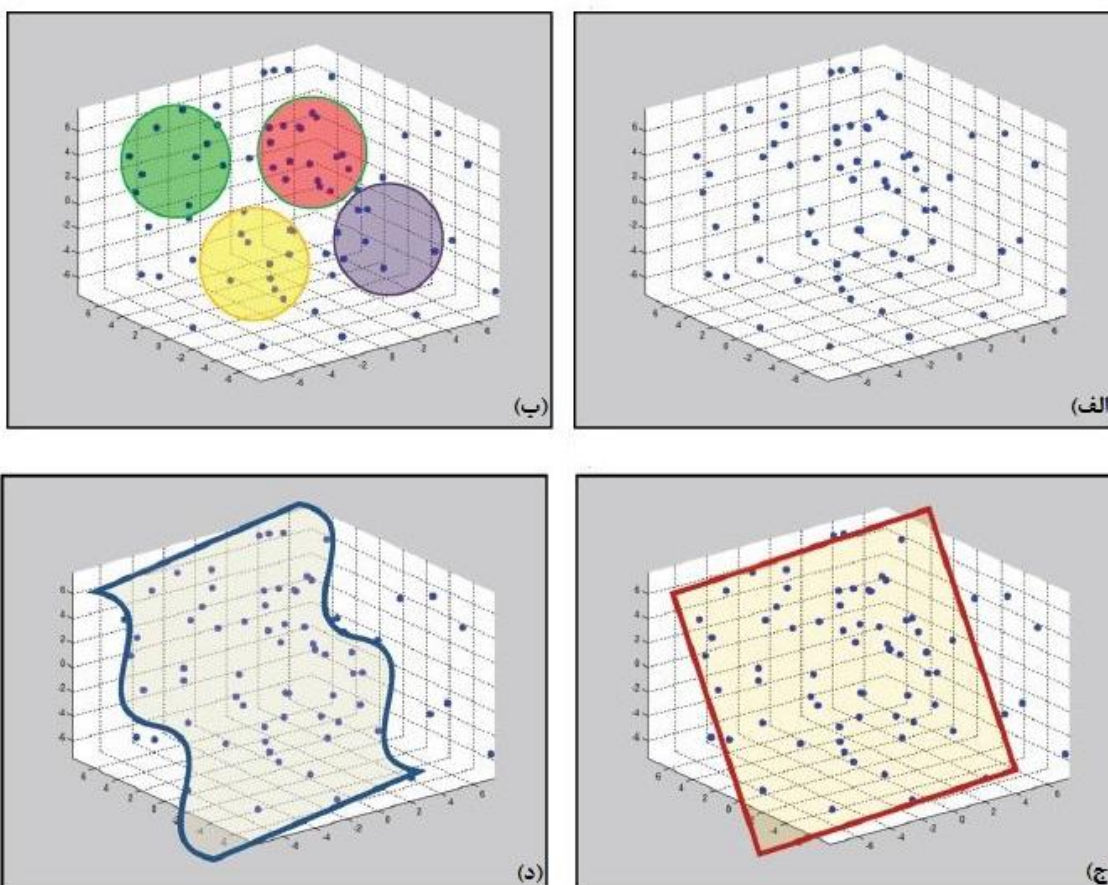
## ۵- تحلیل و بحث

ابتدا الگوریتم‌های  $k$ -میانگین و نقشه‌های خودسازمانده برای برش زمانی ۱۱۵۲ میلی‌ثانیه اجرا شدند. سپس این دو الگوریتم برای یک پنجره محاسباتی در اطراف برش زمانی مورد مطالعه، اجرا شدند و با مقایسه پاسخ‌های دریافتی از این دو فرآیند، مشاهده شد که اجرای الگوریتم‌ها در یک پنجره محاسباتی پاسخ مناسب‌تری دارد؛ بنابراین پس از بررسی و تحلیل ضخامت‌های گوناگون، پنجره محاسباتی با ضخامت ۷ نمونه زمانی با نرخ نمونه‌برداری ۴ میلی‌ثانیه مناسب تشخیص داده شد. باید دقت شود که با افزایش این ضخامت، تفکیک‌پذیری کاهش می‌یابد.

بنابراین با توجه به انتخاب ۷ برش زمانی و محاسبه ۳ نشانگر مطرح‌شده، داده‌های ورودی ۲۱ بعدی خواهند بود. ابتدا با استفاده از تحلیل مؤلفه و با توجه به طیف ویژه ارائه‌شده در شکل ۹، این داده‌ها به ۲ بعد منتقل می‌شوند و در واقع بر روی یک صفحه ۲ بعدی تصویر می‌شوند، سپس داده‌های منتقل‌شده به ۲ بعد، به عنوان ورودی به الگوریتم‌های  $k$ -میانگین و نقشه‌های خودسازمانده وارد می‌شوند. در شکل ۱۰ (الف) و شکل ۱۰ (ب) خروجی این دو الگوریتم قابل مشاهده است که کانال‌های مشخص‌شده در شکل ۸ شکل ۸، با پیکان سیاه‌رنگ مشخص شده‌اند. پیکان‌های قرمز و سبز نیز در این تصاویر به ترتیب گسل موجود در منطقه و یک حوضه توریدیتی در بخش شرق و جنوب شرقی آن را نشان می‌دهد. لازم به ذکر است با استفاده از این مطالعات، تعداد خوشه‌های مناسب برای تحلیل رخساره‌های منطقه برحسب تغییر تعداد خوشه‌ها و مقایسه پاسخ به

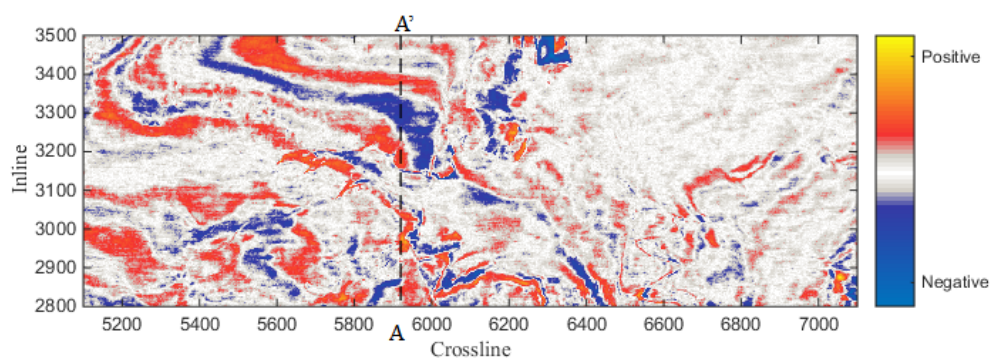
تعداد خوشه‌ها نیز بر همین اساس و برحسب تغییر تعداد آن‌ها و مقایسه پاسخ به دست آمده با اطلاعات مختصری است که از پدیده‌های شاخص منطقه داریم. تحلیل مؤلفه اصلی نیز به دلیل عدم خوشه‌بندی اطلاعات می‌تواند کمک شایانی را در جهت انتخاب تعداد خوشه‌ها ارائه نماید. به عنوان مثال با تغییر تعداد خوشه‌ها در الگوریتم‌های  $k$ -میانگین و نقشه‌های خودسازمان‌ده، بخشی از کانال‌های قابل مشاهده در پاسخ حاصل از تحلیل مؤلفه اصلی به خوشه مربوط به کانال‌ها تعلق نگردید که با توجه به واقعیت ساختار منطقه، به نظر می‌رسد این بخش به اشتباه خوشه‌بندی شده است.

قدرت تجسم مفسر برای کسب اطلاعات بیشتر از داده‌ها می‌گردد. با توجه به مطالعات مختلف انجام‌شده برای خوشه‌بندی رخساره‌های مختلف با استفاده از مطالعات چندنشانگری (Rondon and Banchs, 2007; Roy and Marfurt, 2010; Roy et al., 2013; Chengyin et al., 2014; Zhao et al., 2015)، ذکر این نکته ضروری است که برخلاف الگوریتم‌های یادگیری نظارت‌شده مانند شبکه عصبی و سامانه بردار پشتیبان، در استفاده از الگوریتم‌های غیرنظارتی، شاخص کمی برای اعتبارسنجی خوشه‌بندی وجود ندارد و تنها ملاک، مقایسه پاسخ حاصل از چند روش و همچنین مقایسه خروجی این الگوریتم‌ها با اطلاعات قبلی در مورد یک رخساره خاص موجود در منطقه است.

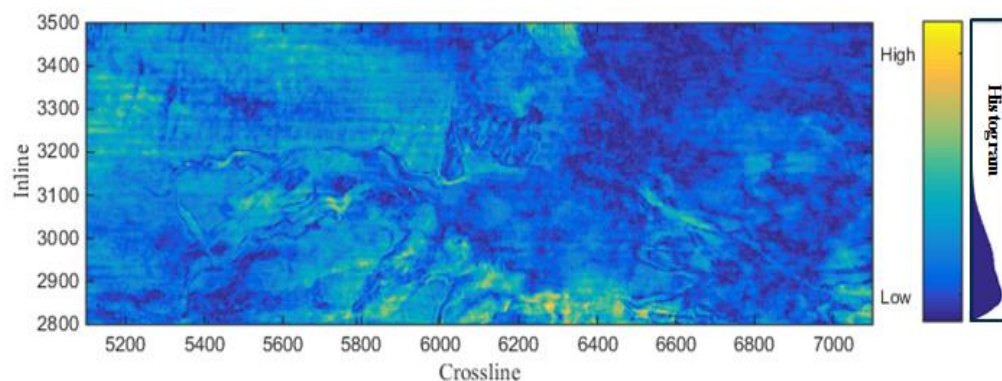


شکل ۶: الف) توزیع نقاط آموزشی در فضای سه‌بعدی نشانگرها. ب) الگوریتم  $k$ -میانگین می‌تواند داده‌ها را در تعداد خوشه مشخص شده توسط کاربر (۴ عدد در این مثال) قرار دهد. ج) صفحه‌ای مسطح که به بهترین نحو بر داده‌ها منطبق می‌شود. این صفحه از دو بردار ویژه اول کوواریانس ماتریس داده‌های لرزه‌ای یا در واقع از روش تحلیل مؤلفه اصلی مشخص می‌گردد. در واقع با تصویر کردن نقاط ارائه‌شده در فضای سه‌بعدی بر این صفحه، داده‌ها به دو بعد منتقل می‌شوند. د) صفحه دوبعدی که برای تطابق بهتر با نقاط تغییر شکل یافته است. الگوریتم نقشه‌های خودسازمان‌ده با تغییر شکل صفحه مشخص شده در قسمت (ج) و تعیین این صفحه تغییر شکل یافته، خوشه‌بندی انجام می‌دهد (Zhao et al., 2015).

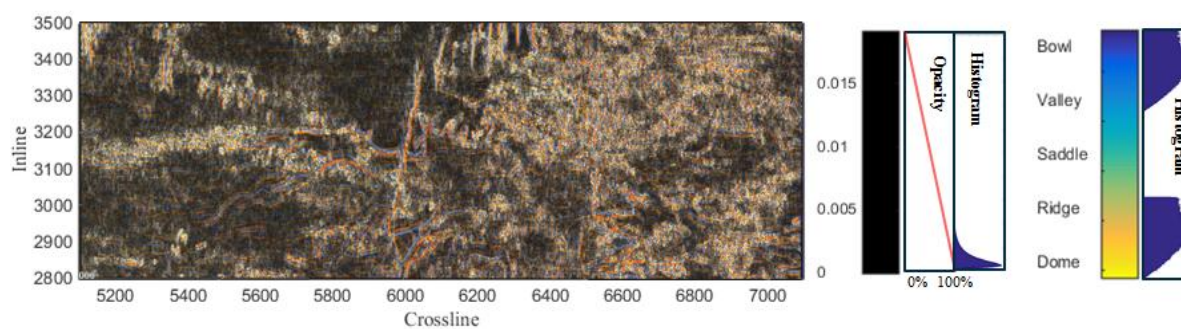




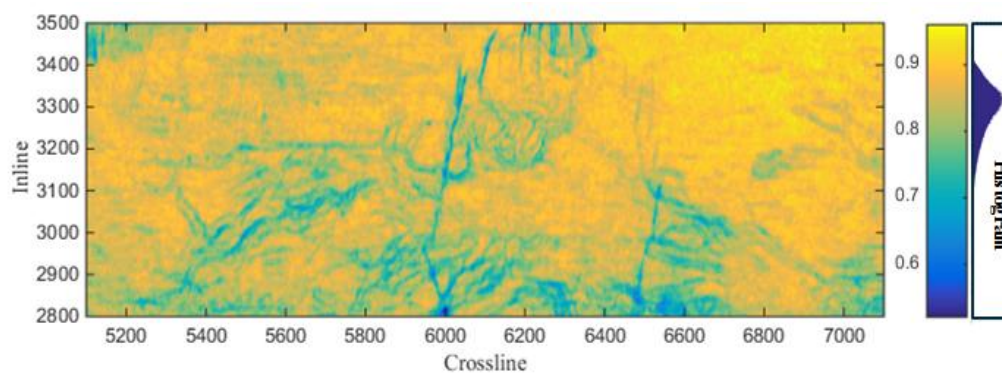
(الف)



(ب)



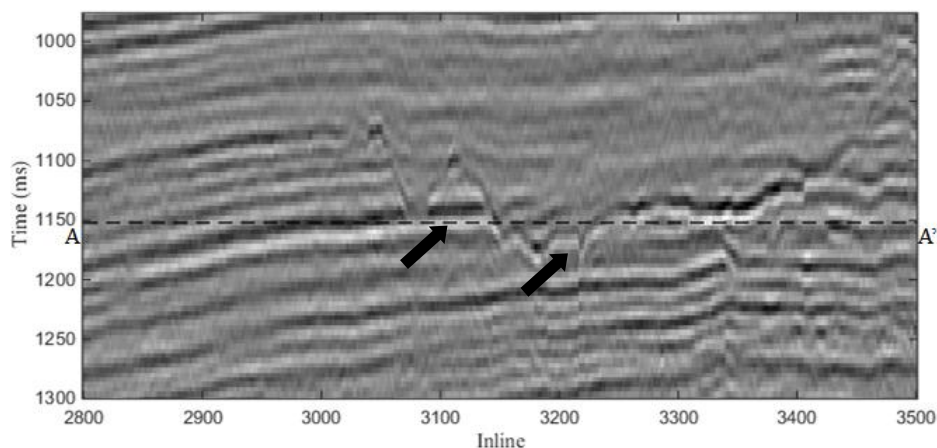
(ج)



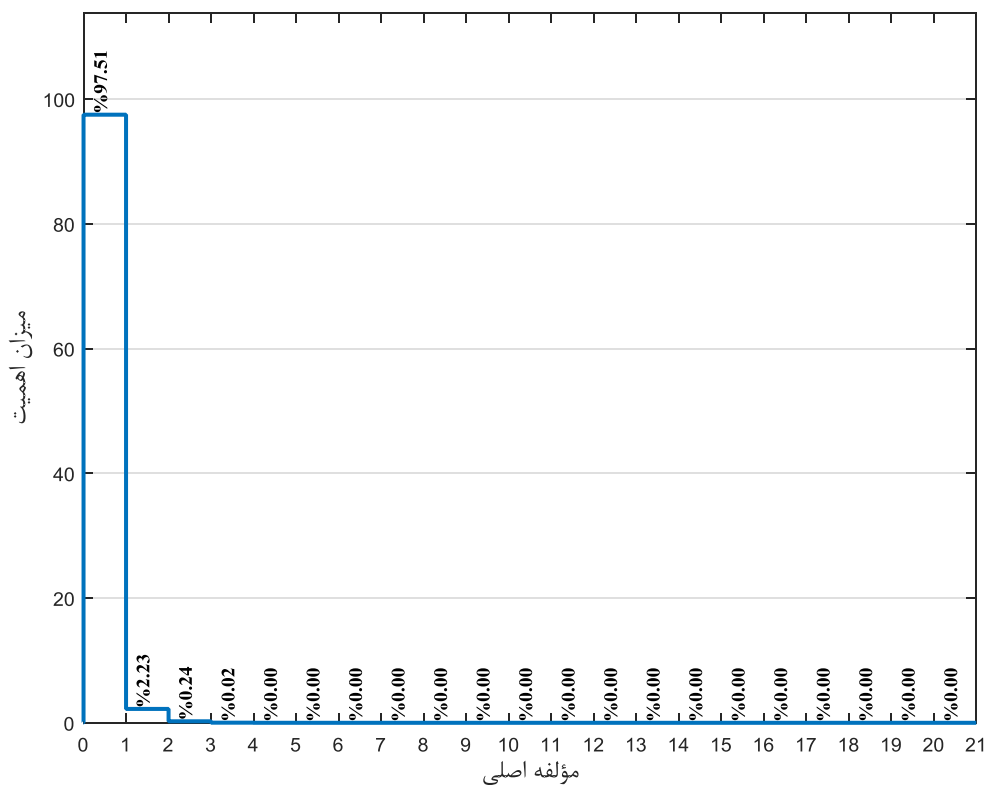
(د)

شکل ۷: الف) دامنه لرزه‌ای در برش زمانی ۱۱۵۲ میلی‌ثانیه. خط AA' محل برخط ۵۹۲۰ است که در شکل ۸ قابل مشاهده است. ب) اعمال نشانگر تجزیه طیفی بر برش زمانی ۱۱۵۲ میلی‌ثانیه. ج) اعمال نشانگر ترکیبی شاخص شکل و میزان انحنا بر برش زمانی ۱۱۵۲ میلی‌ثانیه. د) اعمال نشانگر همگنی GLCM بر برش زمانی ۱۱۵۲ میلی‌ثانیه.

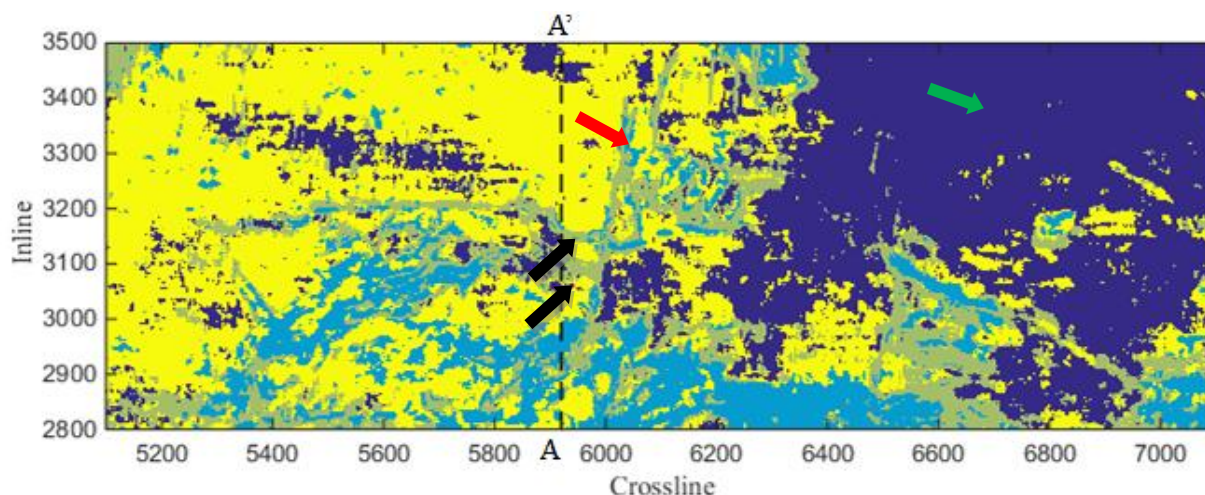




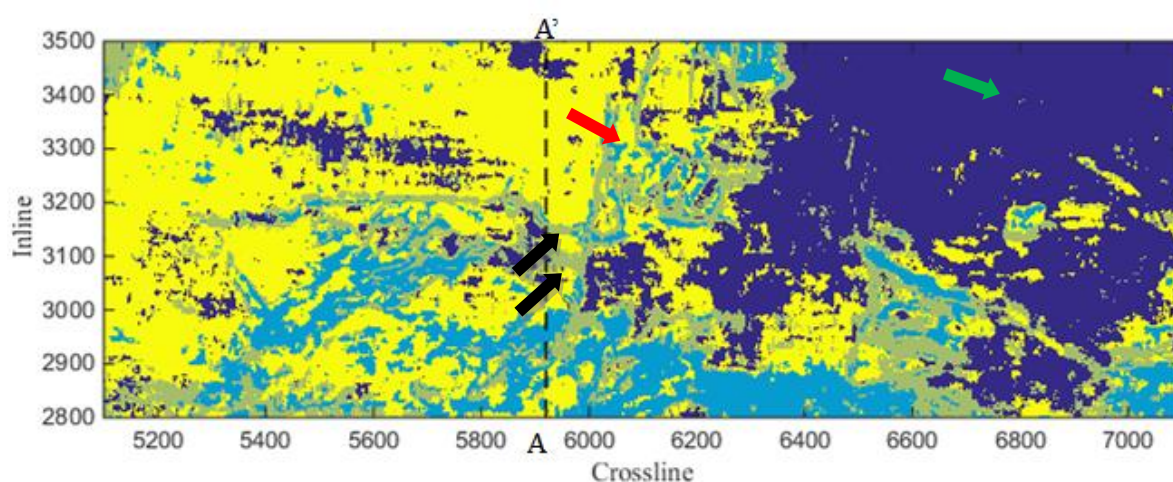
شکل ۸: دامنه لرزه‌ای در راستای خط AA' که در شکل ۷ (الف) رسم شده است. پیکان‌های سیاه‌رنگ، رخساره‌های کانالی موجود در راستای این خط را نشان می‌دهند.



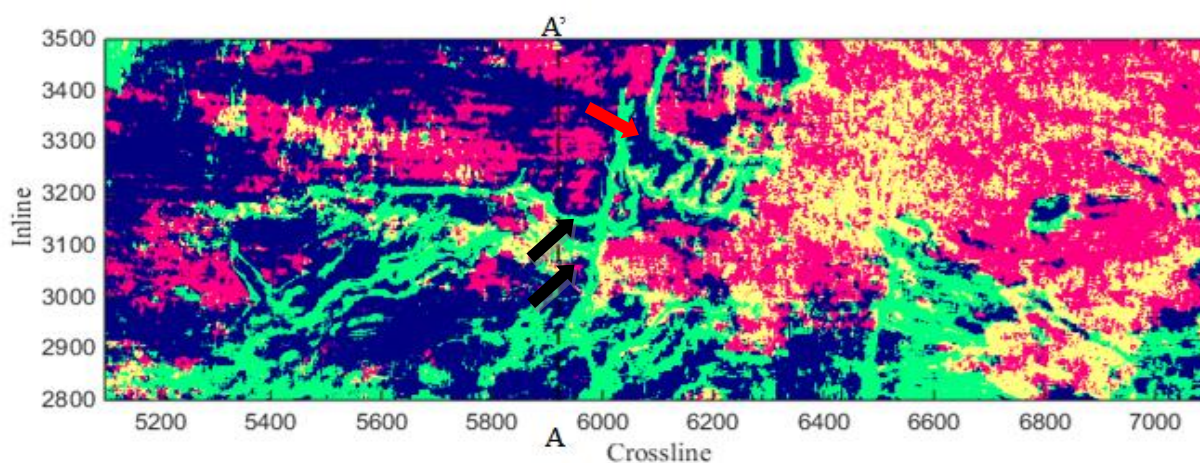
شکل ۹: طیف ویژه جهت تعیین تعداد مؤلفه‌های اصلی مورد استفاده در ادامه مطالعات. همان‌طور که مشاهده می‌شود دو مؤلفه اول به تنهایی بیش از ۹۹ درصد از اطلاعات قابل دست‌یابی در کل نشانگرها را ارائه می‌کنند و بنابراین با استفاده از تحلیل مؤلفه اصلی نشانگرها را از فضای ۲۱ بعدی به ۲ بعد منتقل می‌کنیم.



(الف)



(ب)



(ج)

شکل ۱۰: (الف) خروجی الگوریتم  $k$ -میانگین برای  $k = 4$ . (ب) خروجی الگوریتم نقشه‌های خودسازمان‌ده برای ۴ خوشه. (ج) استفاده از قابلیت نمایش روش تحلیل مؤلفه اصلی برای تشخیص بهتر رخساره کانالی منطقه تنگه هرمز در برش زمانی ۱۱۵۲ میلی‌ثانیه. پیکان‌های سیاه‌رنگ نشان‌دهنده رخساره‌های کانالی مشخص‌شده در شکل ۸ می‌باشند و پیکان قرمز گسل موجود در منطقه را نشان می‌دهد. پیکان سبز نیز یک حوضه توریدایتی کوچک در بخش شرق و جنوب شرقی منطقه مورد مطالعه را نشان می‌دهد.

methods for unsupervised classification of seismic data, 72<sup>nd</sup> Annual International Meeting, SEG, Expanded Abstracts, 2221-2224.

Bishop, C.M., 2006, Pattern recognition and machine learning, Springer, 738 P.

Chengyin, C., Xiaodong, X. and Jinsong, L., 2014, Seismic attributes clustering based on artificial bee colony, CPS/SEG Beijing 2014 International Geophysical Conference, 918-921.

Chopra, S. and Marfurt, K.J., 2007, Seismic attributes for prospect identification and reservoir characterization, SEG, 464 P.

Forgy, E.W., 1965, Cluster analysis of multivariate data: Efficiency vs interpretability of classifications, Biometrics, 21, 768-769.

Gao, D., 2007, Application of three-dimensional seismic texture analysis with special reference to deep-marine facies discrimination and interpretation: An example from offshore Angola, West Africa, AAPG Bulletin, 91, 1665-1683.

Jancey, R. C., 1966, Multidimensional group analysis, Australian Journal of Botany, 14, 127-130.

Junhuai, W., Lei, P., Zhijie, N., Gangqiang, C. and Baoli, B., 2014, Application of seismic Attributes crossplot technology in identification of channel sands, CPS/SEG International Geophysical Conference, Beijing, 837-841.

Justice, J.H., Hawkins, D.J. and Wong, D.J., 1985, Multidimensional attribute analysis and pattern recognition for seismic interpretation, Pattern Recognition, 18, 391-399.

Kohonen, T., 1982, Self-organized formation of topologically correct feature maps, Biological Cybernetics, 43, 59-69.

MacQueen, J., 1967, Some methods for classification and analysis of multivariate observations: in L. M. Le Cam, and J. Neyman, eds., Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, 281-297.

Marfurt, K.J., 2015, Techniques and best practices in multiattribute display, Interpretation, 3 (4), B1-B23.

Robinson, H. and Davis, T., 2012, Seismic reservoir characterization of distributary channel sandstones in the Lower Cretaceous Paluxy reservoir, Delhi Field, Louisiana, SEG Annual Meeting, Las Vegas, 1-6.

Roden, R., Smith, T. and Sacrey, D., 2015, Geologic pattern recognition from seismic attributes: Principal component analysis and self-organization maps, Interpretation, 3 (4), SAE59-SAE83.

Rondon, O. and Banchs, R., 2007, Seismic similarity

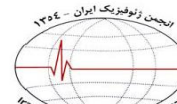
## ۶- نتیجه‌گیری

در این مطالعه نتایج حاصل از اجرای سه روش یادگیری غیرنظارتی شامل روش‌های خوشه‌بندی k-میانگین و نقشه‌های خودسازمان‌ده به همراه روش تحلیل مؤلفه اصلی برای شناسایی رخساره‌های کانالی تنگه هرمز با یکدیگر مقایسه شدند. با اجرای روش‌های خوشه‌بندی در پنجره محاسباتی مورد مطالعه برای تعداد خوشه‌های مختلف، مشخص گردید چهار رخساره اصلی در این محدوده محاسباتی وجود دارند که تحلیل مؤلفه اصلی به خوبی این واقعیت را تأیید نمود. در تشخیص تعداد رخساره‌های موجود در محدوده محاسباتی مورد اشاره، روش تحلیل مؤلفه اصلی که در سیستم رنگی RGB ارائه شد کمک شایانی به مطالعات انجام داد چرا که در دو روش خوشه‌بندی ذکرشده، تعداد خوشه‌ها توسط مفسر مشخص می‌گردد، بنابراین علاوه بر زمان‌بر بودن تحلیل منطقه و داده‌های موجود، نظرات مفسر در این فرآیند می‌تواند ناخواسته باعث انحراف پاسخ از واقعیت گردد ولی تحلیل مؤلفه اصلی برخلاف دو الگوریتم k-میانگین و نقشه‌های خودسازمان‌ده، بدون دریافت تعداد رخساره‌های لرزه‌ای نهفته در محدوده محاسباتی، این مقدار را به خوبی تشخیص می‌دهد. پاسخ دریافتی از اجرای دو الگوریتم k-میانگین و نقشه‌های خودسازمان‌ده تفاوت زیادی در شناسایی رخساره‌های کانال ندارند و آنچه بدیهی است، تحلیل مؤلفه اصلی دارای دقت بیشتری (با توجه به پیوستگی خطوط کانالی در مقاطع خروجی از سه الگوریتم مورد مطالعه) نسبت به دو روش خوشه‌بندی k-میانگین و نقشه‌های خودسازمان‌ده در تشخیص رخساره کانالی و سیستم‌های گسلی دارد ولی هیچ‌یک از روش‌های استفاده‌شده، نتوانستند گسل و رخساره‌های کانالی منطقه را از یکدیگر جدا نمایند. رخساره‌های کانالی در منطقه مورد مطالعه، به‌خوبی شناسایی شدند که از سمت جنوب و جنوب غربی به حوضه توربیدیتی کوچکی واقع در شرق و جنوب شرقی منطقه وارد می‌شوند.

## ۷- منابع

- Al-Dossary, S., 2015, Preconditioning seismic data for channel detection, Interpretation, 3 (1), T1-T4.
- Alsouki, M., Hassani, H., Irannajad, M. and Riahi, M.A., 2008, The structural imaging in offshore area of Strait of Hormuz based on 3D seismic Data, Journal of Applied Sciences, 8, 15, 2725-2731.
- Ashtari, A. and Arzani, A., 2016, Seismic spectral decomposition and inversion for buried channels delineation: a case study from the Asmari Reservoir, southwestern Iran, First Break, 34 (1), 49-53.
- Balch, A.H., 1971, Color sonagrams: A new dimension in seismic data interpretation, Geophysics, 36, 1074-1098.
- Barnes, A.E. and Laughlin, K.J., 2002, Investigation of

- poststack attribute volumes using Kohonen self-organizing maps, *The Leading Edge*, 21, 1032-1037.
- Taner, M.T., Koehler, F. and Sheriff, R.E., 1979, Complex seismic trace analysis, *Geophysics*, 44, 1041-1063.
- Wynd, J.G., 1965, Biofacies of the Iranian oil consortium agreement area, Iranian Oil Operating Companies Geological and Exploration Division, Report no. 1082, unpublished.
- Zhao, T., Jayaram, V., Roy, A. and Marfurt, K.J., 2015, A comparison of classification techniques for seismic facies recognition, *Interpretation*, 3 (4), SAE29-SAE58.
- analysis through self-organising maps, *Exploration Geophysics*, 38, 184-188.
- Roy, A. and Marfurt, K.J., 2010, Applying self-organizing maps of multiple attributes, an example from the Red-Fork Formation, Anadarko Basin, SEG, Expanded Abstracts, 1591-1595.
- Roy, A., Jayaram, V. and Marfurt, K.J., 2013, Active learning algorithms in seismic facies classification, SEG Annual Meeting, Houston, 1467-1471.
- Sonneland, L., 1983, Computer aided interpretation of seismic data, 53<sup>rd</sup> Annual International Meeting, SEG, Expanded Abstracts, 546-549.
- Strecker, U. and Uden, R., 2002, Data mining of 3D



## A comparison of unsupervised learning techniques for channel detection in 3-d seismic data acquired over the Strait of Hormuz

Amir Hossein Mardan<sup>1</sup>, Abdolrahim Javaherian<sup>2\*</sup> and Marzieh Mirzakhania<sup>3</sup>

1- M.Sc. Student, Department of Petroleum Engineering, Amirkabir University of Technology, Tehran, Iran

2- Professor, Department of Petroleum Engineering, Amirkabir University of Technology, Formerly Institute of Geophysics, University of Tehran, Tehran, Iran

3- Senior Reservoir Geophysicist, Geophysics Office, Exploration Directorate of National Iranian Oil Company, Tehran, Iran

Received: 2 March 2016; Accepted: 21 May 2016

Corresponding author: javaherian@aut.ac.ir

### Keywords

Principal Component Analysis  
Channel Detection  
Strait of Hormuz  
Clustering  
Unsupervised Learning

### Extended Abstract

#### Summary

In recent years, due to increasing the size of three-dimensional (3-D) seismic data and the number of seismic attributes, it is difficult for interpreters to examine every seismic line and time slice. Pattern recognition techniques as first-hand interpretation tools are used to both address the problem of large size of seismic data and to provide an initial guidance when working on a

new seismic data where previous studies and data are limited. Different types of unsupervised learning techniques have recently been used for seismic facies clustering and object detection in seismic data. Among unsupervised learning techniques k-means, self-organizing maps (SOM), generative topographic mapping, and principal component analysis (PCA) are used for facies analysis. In this study, we have applied k-means, SOM, and PCA on a 3-D seismic data volume acquired over the Strait of Hormuz to detect buried channels. Not surprisingly, the most important parameter in this study is the choice of appropriate seismic attributes. Although the PCA is not a clustering technique, it can detect channels in 3-D seismic data more efficiently than the k-means and SOM. According to the dip of the structure, the detected channels are prolonged from the west to the east and the southeast where there is a mini basin within the Mishan Formation.

### Introduction

One important class of machine learning tasks is the unsupervised learning. In the unsupervised learning, no labels are given to the learning algorithm, leaving it on its own to find structure in its input. The main task of this learning method is data clustering, but some different tasks such as dimensionality reduction and density estimation are belonged to this category. PCA is a dimensionality reduction technique, which can be used for better visualization of data. After explaining the geology of the study area, we discuss the learning methods, and their workflows. In the next step, we present the chosen attributes, and the learning algorithms applied to the data.

### Methodology and Approaches

We have used OpendTect for attribute measurements. After preparing the data, we have applied three unsupervised learning techniques of k-means, SOM, and PCA, on attributes of the 3-D seismic data volume acquired over the Strait of Hormuz. The chosen attributes in this study are spectral decomposition, curvedness, and gray-level co-occurrence matrix (GLCM) homogeneity. First, we have applied the PCA to reduce the dimension of attribute data to 2-D, and then, the k-means and SOM are applied on the data. Next, we have presented the two first principal components of attributes to the RGB color system, and consequently, we found that this method is superior than the k-means and SOM in the illumination of the channels.

### Results and Conclusions

Although the PCA method is not a clustering technique, it can detect channels in 3-D seismic data more efficiently than the k-means and SOM clustering methods. According to the dip of structure, these channels are prolonged from the west to the east and to the southeast where there is a mini basin within the Mishan Formation.