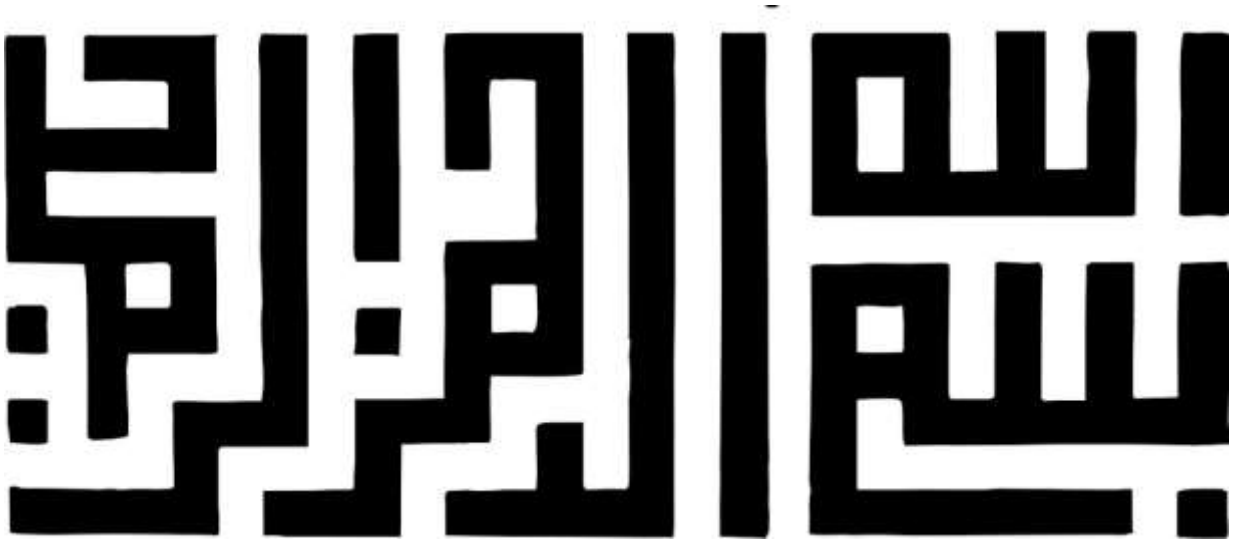
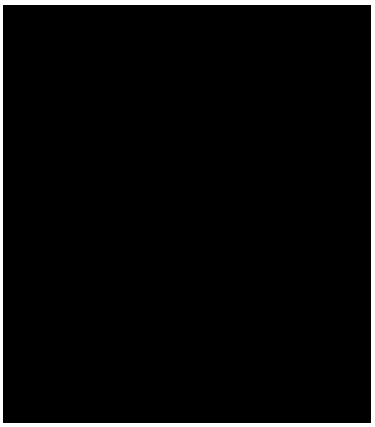




دانشکده مهندسی معدن، نفت و ژئوفیزیک

پایان نامه کارشناسی ارشد اکتشاف مواد معدنی

تحلیل رخساره لرزه‌ای چند مقیاسی در برخورد با اثر نوفه در داده‌ها

نگارنده: **میلاد برزگر**

اساتید راهنما:

دکتر محمد رداد

دکتر امین روشندل کاهو

استاد مشاور:

دکتر سعید هادیلو

اسفند ۱۴۰۰

فرم شماره (۳) صورتجلسه نهایی دفاع از پایان نامه دوره کارشناسی ارشد

با نام و یاد خداوند متعال، ارزیابی جلسه دفاع از پایان نامه کارشناسی ارشد آقای میلاد برزگر با شماره دانشجویی ۹۷۰۳۶۳۴ رشته مهندسی نفت گرایش اکتشاف تحت عنوان تحلیل رخساره‌ی لرزه‌ای چندمقیاسی در برخورد با اثر نوفه باقیمانده در داده‌ها که در تاریخ ۱۴۰۰/۱۲/۰۹ با حضور هیأت محترم داوران در دانشگاه صنعتی شاهرود برگزار شد به شرح ذیل اعلام می گردد:

الف) درجه عالی: نمره ۲۰-۱۹ ☐ ب) درجه خیلی خوب: نمره ۱۸/۹۹ - ۱۸ ☐

ج) درجه خوب: نمره ۱۷/۹۹- ۱۶ ■ د) درجه متوسط: نمره ۱۵/۹۹-۱۴ ☐

ه) کمتر از ۱۴ غیر قابل قبول و نیاز به دفاع مجدد دارد ☐

نوع تحقیق: نظری ■ عملی ☐

عضو هیات داوران	نام ونام خانوادگی	مرتبهٔ علمی	امضاء
۱ـ استاد راهنمای اول	دکتر محمد رداد	استادیار	
۲- استاد راهنمای دوم	دکتر امین روشندل کاهو	دانشیار	
۳- استاد مشاور	دکتر سعید هادیلو	—	—
۴- استاد داور اول	دکتر مهرداد سلیمانی	دانشیار	
۵- استاد داور دوم	دکتر ابوالقاسم کامکار روحانی	دانشیار	
۶- نماینده تحصیلات تکمیلی	دکتر معصومه کردی	استادیار	

نام و نام خانوادگی رئیس دانشکده: دکتر [illegible] سرشکی

تاریخ و امضاء و مهر دانشکده

به پاس زحمات شبانه روزی همسرم که در تمامی مراحل زندگی با تمام سختی‌ها در کنار من بود و دوری از من را به جان خرید تا بهانه‌ای شود برای رشد و ارتقای علمی من.

این مجموعه را به همسر عزیزم تقدیم می کنم.

تشکر و قدردانی

با تشکر از زحمات استادان عزیزم جناب آقای دکترمحمد رداد و جناب آقای دکتر امین روشندل کاهو به خاطر زحمات و راهنمایی‌های ارزنده‌شان در راستای ارتقای دانش بنده و نیز نگارش این پایان نامه کمال تشکر را دارم. از داوران محترم جناب آقای دکتر ابوالقاسم کامکار روحانی و جناب آقای دکتر مهرداد سلیمانی که داوری این پایان نامه را بر عهده داشتند قدردانی می کنم.

همچنین از تمام اساتید دانشگاه صنعتی شاهرود که از ایشان آموختم سپاسگزارم.

تعهدنامه

اینجانب میلاد برزگر دانشجوی دوره کارشناسی ارشد (دکتری) رشته اکتشاف نفت دانشکده معدن، نفت، ژئوفیزیک دانشگاه صنعتی شاهرود نویسنده پایان‌نامه تحلیل رخساره‌ی لرزه‌ای چندمقیاسی در برخورد با اثر نوفه باقیمانده در داده‌ها تحت راهنمائی دکتر محمد رداد متعهد می‌شوم.

- تحقیقات در این پایان‌نامه توسط اینجانب انجام شده است و از صحت و اصالت برخوردار است .
- در استفاده از نتایج پژوهشهای محققان دیگر به مرجع مورد استفاده استناد شده است .
- مطالب مندرج در پایان‌نامه تاکنون توسط خود یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی در هیچ جا ارائه نشده است .
- کلیه حقوق معنوی این اثر متعلق به دانشگاه صنعتی شاهرود می‌باشد و مقالات مستخرج با نام « دانشگاه صنعتی شاهرود » و یا « Shahrood University of Technology » به چاپ خواهد رسید .
- حقوق معنوی تمام افرادی که در به دست آمدن نتایج اصلی پایان‌نامه تأثیرگذار بوده اند در مقالات مستخرج از پایان‌نامه رعایت می گردد.
- در کلیه مراحل انجام این پایان‌نامه، در مواردی که از موجود زنده (یا بافتهای آنها) استفاده شده است ضوابط و اصول اخلاقی رعایت شده است .
- در کلیه مراحل انجام این پایان‌نامه، در مواردی که به حوزه اطلاعات شخصی افراد دسترسی یافته یا استفاده شده است اصل رازداری، ضوابط و اصول اخلاق انسانی رعایت شده است .

تاریخ

امضای دانشجو

چکیده

برای مدلسازی از مخزن نیاز به ارزیابی دقیق از خصوصیات سنگ و تصویرسازی از ناهمگنی آن‌ها است. به دلیل بالا بودن هزینه‌های حفاری و نبود تعداد چاه کافی در مقیاس گسترده و یا داده‌های مربوط به چاه‌پیمایی و مغزه‌ها، داده‌های لرزه‌ای ۳ بُعدی نقش بسیار مهمی را در شناسایی تغییرات مخزن ایفا می‌کنند. هرگونه تغییری در لیتولوژی، تخلخل و محتوای سیال باعث تغییر در دامنه، فرکانس، پیوستگی جانبی و دیگر نشانگرهای لرزه‌ای می‌شود. اگر این تغییرات در پارامترهای لرزه‌ای قابل شناسایی باشند به فهم زمین شناسی زیرسطحی کمک می‌کنند. هدف تحلیل رخساره‌ی لرزه‌ای، تفسیر تغییرات پارامترهای لرزه‌ای است. رخساره‌های لرزه‌ای به عنوان گروه‌های رد لرزه‌ای قابل تعریف هستند. روش‌های بسیار زیادی برای شناسایی و تشخیص الگو برای دسته‌بندی رخساره‌های لرزه‌ای تاکنون توسعه داده شده است. زمانی‌که اطلاعات زمین‌شناسی در دسترس نیستند از الگوی بدون نظارت استفاده می‌شود. وجود نوفه بخش جدانشدنی از داده‌های لرزه‌ای است که بر طبقه‌بندی رخساره‌ها و تفسیر لرزه‌ای اثر می‌گذارد و وجود آن باعث ایجاد تفسیر اشتباه می‌شود. رهیافت مورد نظر این پژوهش برای این مشکل تحلیل رخساره‌ی چند مقیاسی است. یعنی اینکه پیش از اینکه داده‌ی لرزه‌ای اصلی و یا نشانگرها وارد فرآیند تحلیل رخساره شوند، در ابتدا داده‌ی لرزه‌ای را با استفاده از یکی از روش‌های تجزیه سیگنال به مؤلفه‌های سازنده آن تجزیه و سپس تعدادی از مؤلفه‌ها را در بازسازی سیگنال شرکت می‌دهیم و برخی مؤلفه‌ها که بیانگر ماهیت نوفه باشند را کنار می‌گذاریم. آنگاه نسخه‌ی بازسازی شده سیگنال و یا نشانگرهای حاصل از آن به عنوان ورودی به الگوریتم تحلیل رخساره وارد شود. در این پژوهش قصد بر این است که با استفاده از تجزیهٔ مُدّ متغیر(VMD) داده‌های لرزه‌ای به مولفه‌های سازنده آنها تجزیه شوند و فرآیند تحلیل رخساره بر روی داده بازسازی شده و نشانگرهای حاصل از آن اعمال شود. تحلیل رخساره در این روش بدون نظارت و توسط روش K- means انجام خواهد شد. عملکرد روش برداده‌های مصنوعی و واقعی آزمایش شده و نشان داده می‌شود که می‌توان تحلیل رخساره دقیق‌تری داشت.

واژگان کلیدی: تجزیه سیگنال لرزه‌ای _ نشانگرلرزه‌ای _ تحلیل رخساره‌ی لرزه‌ای _ اثر نوفه

فهرست مطالب

فصل اول: کلیات پژوهش ۱

۱-۱ پیشگفتار ۲
۱-۲ بیان مساله، ضرورت و هدف ۲
۱-۳ سوابق پژوهش ۳
فصل دوم: روش‌های تحلیل رخساره‌ی لرزه‌ای ۷
۲-۱ پیشگفتار ۸
۲-۲ نقشه خود سازمانده (SOM) ۸
۲-۳ نقشه توپوگرافی مولد (GTM) ۱۲
۲-۴ مدل‌های ترکیبی گوسی (Gaussian mixture models) ۱۷
۲-۵ روش k-means ۲۱
۲-۶ تحلیل مولفه‌ی اصلی (PCA) ۲۴
۲ -۷ الگوریتم خوشه‌بندی سلسله مراتبی(Hierarchical clustering algorithm) ۲۷
۲-۸ روش فازی C-means ۲۸
فصل سوم: روش‌های تحلیل رخساره‌ی لرزه‌ای ۳۳
۳-۱ پیشگفتار ۳۴
۳-۲ تجزیه مد تجربی(EMD) ۳۵
۳-۳ تجزیه مُد تجربی مجموع(EEMD) ۳۸
۳-۴ روش تجزیه‌ی مد تجربی کامل(CEEMD) ۴۰
۳-۵ تجزیه مُد متغیر(VMD) ۴۱
فصل چهارم: تحلیل رخساره لرزه‌ای چند مقیاسی ۴۳
۴-۱ پیشگفتار ۴۴
۴-۲ الگوریتم روش ۴۵
۴-۳ خوشه بندی داده مصنوعی ۴۶
۴-۴ خوشه بندی داده واقعی ۵۴
فصل پنجم: نتیجه گیری و پیشنهادها ۶۱
۵-۱ نتیجه‌گیری ۶۲
۵-۲ پیشنهادها ۶۲
مراجع ۶۳

فهرست شکل ها

شکل ۲-۲ مراحل روش GTM (ژائو و همکاران، ۲۰۱۵). توضیح بخش های a و b در متن ارائه شده است... ۱۵

شکل ۲-۳ گردش کار روش GTM (ژائو و همکاران، ۲۰۱۵) ۱۶

شکل ۲-۴ یک GMM نمونه (هاردیستی و والت، ۲۰۱۷) ۱۹
شکل ۲-۵ توصیف روش خوشه‌بندی k-means (ژائو و همکاران، ۲۰۱۵) ۲۳
۲-۶ نمودار شماتیک تجزیه و تحلیل PCA (کولئو و همکاران، ۲۰۰۲) ۲۷
شکل ۲-۷ الگوریتم روش FCM (ترن، ۲۰۱۹) ۳۰
شکل ۲-۸ آزمودن فاکتورهای فازی مختلف با استفاده از مقادیر مختلف در الگوریتم FCM (سانگ و همکاران، ۲۰۱۷). ۳۱
شکل ۳-۱ نمودار جریان الگوریتم EMD (زیلر و همکاران، ۲۰۱۰) ۳۶
شکل ۳-۲ نمودار جریان الگوریتم EEMD (لی و زو، ۲۰۰۹) ۴۰
شکل ۴ـ۱. مدل رخساره برای ساخت داده مصنوعی. ۴۸
شکل ۴ـ۲. یک Inline از حجم داده لرزهای مصنوعی. افق در موقعیت زمانی ۳۰۰ میلی ثانیه نشان داده شده با فلش قرمز مورد تحلیل قرار گرفته است. ۴۸
شکل ۴ـ۳. نتیجه خوشه بندی افق لرزهای مورد نظر با استفاده از روش k-means و سه خوشه. ۴۹
شکل ۴ـ۴. یک Inline از حجم داده لرزه ای مصنوعی که نوفه تصادفی با نسبت سیگنال به نوفه ۲ اضافه شده است. ۴۹
شکل ۴ـ۵. نتیجه خوشه بندی داده لرزه ای مصنوعی آلوده به نوفه با نسبت سیگنال به نوفه ۲ با استفاده از روش k-means و سه خوشه. ۵۰
شکل ۴ـ۶ . نتیجه خوشه بندی داده لرزه ای مصنوعی (آلوده به نوفه با نسبت سیگنال به نوفه ۲) و تحلیل شده توسط روش VMD که مولفههای با فرکانس زیاد آن کنار گذاشته شده است. خوشهبندی به روش k-means و با سه خوشه انجام شده است. ۵۰
شکل ۴ـ۷. یک Inline از حجم داده لرزه ای مصنوعی که نوفه تصادفی با نسبت سیگنال به نوفه ۱ اضافه شده است. ۵۱
شکل ۴ـ۸ . نتیجه خوشه بندی داده لرزه ای مصنوعی آلوده به نوفه با نسبت سیگنال به نوفه ۱ با استفاده از روشk-means و با سه خوشه ۵۱
شکل ۴ـ۹. نتیجه خوشه بندی داده لرزه ای مصنوعی (آلوده به نوفه با نسبت سیگنال به نوفه ۱) و تحلیل شده توسط روش VMD که مولفههای با فرکانس زیاد آن کنار گذاشته شده است. خوشهبندی به روش kmean و با سه خوشه انجام شده است. ۵۲
شکل ۴ـ۱۱. نتیجه خوشه بندی داده لرزه ای مصنوعی آلوده به نوفه با نسبت سیگنال به نوفه ۰/۵ با استفاده از روش kmeans و سه خوشه. ۵۳

شکل ۴ـ۱۲. نتیجه خوشه بندی داده لرزه ای مصنوعی (آلوده به نوفه با نسبت سیگنال به نوفه ۰/۵) و تحلیل شده توسط روش VMD که مولفههای با فرکانس زیاد آن کنار گذاشته شده است. خوشهبندی به روش kmeans و با سه خوشه انجام شده است. ۵۳

شکل ۴ـ۱۳ (الف) . طیف دامنه میانگین حجم داده لرزه ای مصنوعی بدون نوفه (محنی آبی رنگ)، آلوده به نوفه با نسبت سیگنال به نوفه ۰/۵ (منحنی سیاه رنگ) و تحلیل شده توسط روش VMD به طوریکه مولفههای با فرکانس زیاد در بازسازی داده کنار گذاشته شده است (منحنی قرمز رنگ). (ب) . طیف دامنه میانگین مولفه های کنار گذاشته شده از بازسازی با رنگ قرمز نشان داده شده است. ٥٤

شکل ۴ـ۱۴ حجم لرزهای به همراه افق مورد نظر که با خط چین سبز رنگ نمایش داده شده است ۵۵

شکل۴ـ۱۵. نمایش سه بعدی نتیجه خوشه بندی افق مورد نظر در داده لرزه ای واقعی با روش kmeans و ۶ خوشه ٥٦

شکل۴ـ۱۷. گراف اندیس سیلوئت برای خوشه بندی افق نشان داده شده در شکل قبل. مقدار میانگین این اندیس برای میزان قطعیت کیفی خوشه بندی ۰/۴۲۴۱۶ برآورد شده است. ۵۷

شکل۴ـ۱۸. نمایش سه بعدی نتیجه خوشه بندی افق مورد نظر در داده لرزه ای واقعی با روش kmeans و ۶ خوشه. خوشه بندی بر روی داده تحلیل شده به روش VMD انجام شده است. به نحوی که مولفه های با فرکانس زیاد از بازسازی کنار گذاشته شده اند. ۵۸

شکل ۴ـ۱۹. نتیجه خوشه بندی افق مورد نظر در داده لرزه ای واقعی با روش kmeans و ۶ خوشه. خوشه بندی بر روی داده تحلیل شده به روش VMD انجام شده است. به نحوی که مولفه های با فرکانس زیاد از بازسازی کنار گذاشته شده اند. ۵۹

شکل۴ـ۲۰ . گراف اندیس سیلوئت برای خوشه بندی افق نشان داده شده در شکل قبل. مقدار میانگین این اندیس برای میزان قطعیت کیفی خوشه بندی ۰/۴۹۰۸ برآورد شده است. ۵۹

فصل اول

کلیات پژوهش

۱-۱ پیشگفتار

کلید مدل‌سازی مخزن ارزیابی‌های موثر از خصوصیات سنگ و تصویرسازی دقیق از ناهمگنی آن‌ها است. مدل‌های مختلف اطلاعات مثل نمودارهای چاه، نمونه‌های مغزه‌گیری شده، داده‌های مربوط به تولید و همچنین داده‌های لرزه‌ای در این مدل‌سازی استفاده می‌شوند. اما به دلایلی چون مسائل فنی و هزینه‌های بالای عملیات حفاری و همچنین نبود تعداد چاه کافی در مناطق گسترده از این نوع داده برای ساخت مدل نمی‌توان استفاده کرد. در غیاب چاه‌ها برای مدل‌سازی تنها به مغزه‌ها و یا داده‌های حاصل از چاه‌پیمایی اکتفا نمی‌توان کرد به این دلیل که برای اطمینان کافی باید تعداد چاه کافی دراختیار داشت تا بتوان قیاسی قابل اکتفا و مطمئن بین داده‌های چاه‌ها انجام داد که این باعث می‌شود تا بتوانیم تغییرات جانبی مخزن را شناسایی کنیم. در این مورد، داده‌های لرزه‌ای ۳ بُعدی نقش مهمی را در شناسایی تغییرات جانبی مخزن دارد و همچنین می‌تواند ویژگی‌های زمین‌شناسی آن‌ها را توصیف کند.

۱-۲ بیان مساله، ضرورت و هدف

هرگونه تغییری در لیتولوژی، تخلخل و محتوای سیال موجود در آن می‌تواند به تغییراتی در دامنه، فرکانس، پیوستگی جانبی و دیگر نشانگرهای لرزه‌ای منجر شود که اگر این تغییرات پارامترهای لرزه‌ای قابل شناسایی و همچنین قابل تفسیر باشند می‌توان اطلاعات با ارزش زیادی از آن استخراج کرد که همین باعث کمک به فهم بهتر نسبت به زمین‌شناسی زیرسطحی می‌شود. هدف نهایی تحلیل رخساره‌های لرزه‌ای تفسیر تغییرات پارامترهای لرزه‌ای است. رخساره‌های لرزه‌ای را می‌توان به عنوان گروه‌های رد لرزه‌ای تعریف کرد که به عنوان رخساره‌های رسوبی یا اجسام زمین‌شناسی در داده‌های لرزه‌ای دیده می‌شوند. روش‌های متنوع بسیار زیادی تاکنون برای تشخیص الگو با درجات مختلفی از موفقیت اضافه شده است. زمانی‌که اطلاعات زمین‌شناسی در دسترس نباشد از الگوی بدون نظارت می‌توان استفاده کرد. از میان این روش‌های تحلیل رخساره‌ی بدون نظارت، شکل موج لرزه‌ای و مقادیر نشانگرها ورودی‌هایی هستند که در فرآیند طبقه‌بندی استفاده می‌شوند.

پارامترهای لرزه‌ای همانطور که گفته شده است اطلاعات زیادی را به ما می‌دهد اما مساله اینجاست که این پارامترها اغلب آلوده به نوفه هستند. نوفه در واقع یک بخش جدایی ناپذیر از داده‌های لرزه‌ای است. مشکل وجود نوفه هم در این است که باعث به وجودآمدن چالش در تصمیم‌گیری و همچنین تفسیر غلط رخساره‌های لرزه‌ای برای مفسرین می‌شودکه در نهایت موجب می‌شود که تفسیری که مفسر ارائه می‌کند، بستگی کافی و مناسبی با زمین‌شناسی زیر ساختی نداشته باشد.

۱-۳ سوابق پژوهش

هوانگ و همکاران (۱۹۹۸) روش تجزیه مُد تجربی را برای تحلیل داده‌های غیرخطی و غیرثابت معرفی کردند که به وفور برای تجزیه و تحلیل سیگنال‌های لرزه‌ای استفاده شده است. دیو و همکاران (۲۰۱۵) از ترکیب و تجزیه مُد تجربی و نقشه‌های خودسازمانده برای تحلیل رخساره در برخورد با اثر نوفه استفاده کردند. سانگ و همکاران (۲۰۱۸) در این زمینه تحلیل رخساره را با طبقه‌بندی شکل موج چندگانه انجام دادند. تورس و همکاران (۲۰۱۱) الگوریتم تجزیه‌ی مُد تجربی مجموع کامل را پیشنهاد کردند، این الگوریتم نه تنها مشکل اختلاط مُد را حل می‌کند بلکه بازسازی دقیق سیگنال اصلی را نیز ارائه می‌دهد و آن را برای پردازش و تفسیر لرزه‌ای مناسب‌تر می‌کند. وو و هوانگ (۲۰۰۹) روش تجزیه مُد تجربی مجموع را برای حل مشکل اختلاط مُد معرفی کردند. دراگومیتسکی و زوسو (۲۰۱۴) تجزیه مُد متغیر را برای تجزیه داده درون مجموعه‌ای از توابع مُد ذاتی با باند محدود را پیشنهاد کردند. روی و همکاران (۲۰۱۴) از نقشه‌ی توپوگرافی مولد برای خوشه‌بندی حجم داده‌های چند نشانگری استفاده کردند. رودن و همکاران (۲۰۱۵) نشان دادند که چطور می‌توان تحلیل مولفه‌های اصلی را برای انتخاب معنی‌دارترین نشانگرهای ورودی برای نقشه‌ی خودسازمانده از نظر ریاضی ترکیب کرد. چوپرا و مارفورت (۲۰۱۴) از نقشه‌ی توپوگرافی مولد برای طبقه بندی شکل موج لرزه‌ای استفاده کردند. کی و همکاران (۲۰۱۶) طبقه‌بندی رخساره‌های لرزه‌ای را با چندین نشانگر و به شکل نیمه ناظر انجام دادند. هونوریو و همکاران (۲۰۱۷) الگوریتم تجزیه مُد تجربی کامل بهبود یافته را پیشنهاد و اثرات آن را برای تحلیل نشانگر لرزه‌ای در چارچوب تجزیه مُد تجربی استفاده کردند. در برخورد با نوفه در تحلیل رخساره، سانگ و همکاران (۲۰۱۷) یک نسخه منظم شده از روش فازی c-means را برای تحلیل رخساره بدون

نظارت مقید استفاده کردند. والت و هاردیستی (۲۰۱۹) از مدل‌های ترکیبی گوسی برای تحلیل رخساره‌های لرزه‌ای بدون نظارت استفاده کردند. سانگ و همکاران (۲۰۱۶) از طبقه‌بندی چند شکل موجی برای تحلیل رخساره‌های لرزه‌ای استفاده کردند. فریرا و همکاران (۲۰۱۹) تحلیل چند نشانگری را برای استفاده در الگوریتم طبقه‌بندی بدون نظارت برای به تصویرکشیدن رخساره‌های کربناته واقع در منطقه مرتفع بیرونی حوضه‌ی سانتوس در برزیل به کار بردند. چوپرا و مارفورت (۲۰۱۸) چندین روش بدون نظارت یادگیری ماشینی را برای طبقه‌بندی رخساره‌های لرزه‌ای دریای بارنتز استفاده کردند. چوپرا و مارفورت (۲۰۱۹) تکنیک‌های یادگیری ماشینی متعددی مانند طبقه‌بندی شکل موج، تحلیل مولفه‌های اصلی، خوشه‌بندی k-means و طبقه‌بندی با نظارت بیزین را برای حجم داده لرزه‌ای حوضه‌ی دلاویر مقایسه کردند. تروکولی و همکاران (۲۰۲۲) روش سازماندهی برچسب زدن خودکار را پیشنهاد کردند که از تجزیه و تحلیل مولفه‌های اصلی (PCA) برای سازماندهی موارد بدست آمده از الگوریتم استفاده می‌کنند و درجه‌ای از شباهت را حفظ می‌کنند. رییسی (۲۰۱۲) از شبکه‌ی عصبی مصنوعی برای تجزیه و تحلیل نشانگرهای لرزه‌ای و همچنین لاگ‌های چاه برای تعیین تغییرات رخساره‌های سنگی و ناهمگونی مخازن استفاده کرد. ساگاف (۲۰۰۳) از یک شبکه عصبی رقابتی برای شناسایی رخساره‌های مخزن استفاده کرد. تحلیل لرزه‌ای با نظارت برای تخمین رخساره‌های لرزه ای مورد نظر بهتر است، با این حال این رویکرد مستلزم دانش قبلی و تلاش قابل توجه است و به ندرت در محل اعمال می‌شود (آرشین، ۲۰۱۴). سانچت (۲۰۱۳) روش k-nearest neighbour را برای خودکارسازی تشخیص چاه بر اساس داده‌های لاگ چاه از پیش‌پردازش شده از تجزیه و تحلیل مولفه‌های مستقل و تبدیل کوسینوس گسسته استفاده کرد. وونگ (۲۰۰۵) با استفاده از روش svm خصوصیات مخزن را ارائه کرد.

۱ـ۴ ساختار پایان نامه

در این پایان نامه در فصل اول به کلیات پژوهش پرداخته شده است. در فصل دوم مروری بر روش‌های تحلیل رخساره انجام می‌شود. روش‌های تجزیه سیگنال در فصل سوم همراه با مبانی نظری ارائه شده است. در فصل

چهارم الگوریتم مورد نظر این پژوهش به همراه آزمایش بر روی داده‌های مصنوعی و واقعی ارائه می‌شود. در فصل پنجم نتیجه گیری و پیشنهادها بیان شده اند.

فصل دوم

روش‌های تحلیل رخساره‌ی لرزه‌ای

۲-۱ پیشگفتار

رخساره‌های لرزه‌ای در واقع گروه‌های رد لرزه‌ای هستند که به عنوان رخساره‌های رسوبی در داده‌های لرزه‌ای دیده می‌شوند. تحلیل رخساره‌ی لرزه‌ای اهمیت بسیار زیادی در تفسیر داده‌های لرزه‌ای و همچنین ساخت مدل مخزن دارد که این کار توسط ارائه‌ی راهکاری موثر برای شناسایی تغییرات در لیتولوژی، محتوای سیال و تخلخل در رخساره‌های زمین بین چاه‌ها انجام می‌شود. همانطور که می‌دانید این تغییرات در پارامترهای لرزه‌ای چنانچه برای مفسرین قابل تفسیرکردن باشند، از آن جهت که می‌توانند شامل اطلاعاتی با ارزش برای درک بهتر تغییرات زیرسطحی زمین شناسی باشند بسیار با ارزش هستند. روش‌های تحلیل رخساره‌ی لرزه‌ای به دو دسته با نظارت و بدون نظارت تقسیم می‌شوند. زمانی‌که اطلاعات زمین‌شناسی غیرقابل دسترس باشد از الگوی بدون نظارت می‌توان استفاده کرد در حالی‌که در حالت با نظارت، داده‌های چاه یا داده‌ی برچسب‌زده شده توسط مفسر در دسترس هستند. به همین منظور روش‌های مختلفی در طی سالیان متوالی توسعه داده شده‌اند که در ادامه به چند روش تحلیل رخساره‌ی لرزه‌ای پرداخته می‌شود.

۲-۲ نقشه خود سازمانده (SOM)

یکی از روش‌های طبقه‌بندی داده‌های لرزه‌ای نقشه خودسازمانده است که نقشه‌ی رخساره‌ی لرزه‌ای را از نشانگرهای چندگانه تولید می‌کند (چوپرا و مارفورت، ۲۰۱۸). نقشه‌ی خودسازمانده یکی از موفق‌ترین الگوریتم‌های شبکه عصبی است که به دسته بندی بدون نظارت اضافه شده است. شکل موج لرزه‌ای و مقدار نشانگرها، ورودی‌هایی هستند که در فرآیند طبقه‌بندی استفاده می‌شوند(هاوکان دو و همکاران، ۲۰۱۵). در توزیع n نشانگرهایی که در فضای داده‌ی N بُعدی قرار دارند سطحی که به بهترین وجه متناسب با داده ها است به وسیله اولین دو بردار ویژه از ماتریس کواریانس تعریف شده است. این صفحه سپس به شکل تکراری به یک سطح دو بُعدی که مینفولد نامیده می شود تغییر شکل می یابد (کولئو و همکاران، ۲۰۱۳).

۲ـ۲ـ۱ محاسبات ریاضی SOM

ماتریس‌کواریانس، مولفه اصلی و فاصله ماهالانوبیس با توجه به مجموعه‌ای از نشانگرهای N، ماتریس کواریانس به شکل زیر تعریف می شود (ژائو و همکاران، ۲۰۱۵).

$$C_{mn} = \frac{1}{J}\sum_{j=1}^{J}\left(a_{jm}(t_j, x_j, y_j) - \mu_m\right)\left(a_{jn}(t_j, x_j, y_j) - \mu_n\right) \quad (۱-۲)$$

در حالی‌که a_{jn} و a_{jm} نشانگر های nاُم و mاُم هستند، J تعداد کل بردارهای داده است، j وJ شاخص بردار نشانگر (وکسل) و تعداد وکسل‌ها هستند، K و k شاخص مینیفولد و تعداد نقاط شبکه است، a_j بردار داده‌ی jاُمین نشانگر است، P ماتریس مولفه‌های اصلی، C نشانگر ماتریس کواریانس، μ_n میانگین نشانگر nام، λ_m و v_m که m اُمین مقدار ویژه و جفت بردار ویژه هستند، m_k که k اُمین نقاط شبکه است که در منیفولد قرار دارد، U_k که k اُمین نقطه‌ی شبکه که در فضای پنهان وجود دارد، r_{jk} فاصله‌ی ماهالونوبیس بین بردار داده‌ی jاُم و k اُمین مرکز خوشه یا نقاط شبکه منیفولد است و I ماتریس همانی است.

$$\mu_n = \frac{1}{J}\sum_{j=1}^{J} a_{jn}\left(t_j, x_j, y_j\right) \quad (۲-۲)$$

اگر ما مقادیر ویژه λ_i و بردار ویژه‌ی v_i از ماتریس‌کواریانس C واقعی و مصنوعی را محاسبه کنیم، i اُمین مولفه‌ی اصلی از بردار داده‌ی j به شکل زیر تعریف می‌شود :

$$P_{ij} = \sum_{n=1}^{N} a_{in}\left(t_j\ , x_j\ , y_j\right)\ v_{ni} \quad (۳-۲)$$

که v_{ni} نشان دهنده n اُمین مولفه‌ی نشانگر ازآی اُمین بردار ویژه است. r_{jq} فاصله ماهالانوبیس از jاُمین نمونه ازq اُمین مرکز خوشه و θ_q به عنوان زیر تعریف می شود :

$$r_{jq}^{2} = \sum_{n=1}^{N}\sum_{m=1}^{N}\left(a_{jn} - \theta_{nq}\right) c_{nm}^{-1}\left(a_{jm} - \theta_{mq}\right) \quad (۴-۲)$$

در حالی‌که معکوس ماتریس کواریانس C قبل از اِلمان mn اُم اتفاق می‌افتد. علیرغم محاسبه‌ی فاصله‌ی ماهالانوبیس، SOM ابتدا داده را با استفاده از اندازه Z نرمالیزه می‌کند، اگر داده تقریبا توزیع گوسی را نشان بدهد، مقیاس z از نشانگر n اُم به وسیله‌ی تفریق میانگین و تقسیم بر انحراف استاندارد (ریشه‌ی دوم قطری ماتریس کواریانس C_{nn}) حاصل می‌شود. برای داده‌های توزیع شده غیرگوسی در مقیاس z مانند انسجام،

ابتدا باید داده‌ها را با استفاده از هیستوگرام‌هایی که گوسی را تخمین می زند تجزیه کرد. هدف از الگوریتم SOM به تصویرکشیدن نشانگرهای لرزه‌ای ورودی به منیفولد هندسی به نام نقشه‌ی خودسازمانده است. مانیفولد SOM بوسیله‌ی یک رشته از بردارهای پروتوتایپ m_k که روی سطح با بُعد پایین‌تر قرار گرفته‌اند که با داده‌ی نشانگر N بُعدی متناسب تعریف می‌شود. بردارهای پروتوتایپ m_k معمولا در نقشه‌های دو بُعدی شش ضلعی یا ساختار مستطیل شکل مرتب می‌شوند که رابطه‌ی همسایگی اصلیشان را حفظ می‌کنند، بطوریکه بردارهای نمونه اولیه (پروتوتایپ) بردارهای داده‌ی همسایه‌ی شبیه به هم را نشان می‌دهند. تعداد بردارهای نمونه اولیه در نقشه دو بُعدی اثر بخشی و عمومی سازی الگوریتم را مشخص می‌کند. بعد از محاسبه‌ی Z-scored های داده‌ی ورودی، منیفولد اولیه به عنوان صفحه‌ای تعریف می‌شود که توسط دو مولفه‌ی اصلی اول تعریف می‌شود. بردارهای نمونه اولیه m_k روی یک شبکه‌ی مستطیلی برای دو مقدار ویژه اول تعریف شده‌اند تا بین فاصله‌ی $\pm ۲(\lambda_۱)^{\frac{۱}{۲}}$ و $\pm ۲(\lambda_۲)^{\frac{۱}{۲}}$ باشند. داده نشانگر لرزه‌ای سپس با هریک از بردارهای نمونه اولیه مقایسه و نزدیکترین آن‌ها انتخاب می‌شود. نزدیکترین همسایه‌ها (آنهایی که دریک محدوده σ قرار می‌گیرند یک آشفتگی گوسی را تشریح می‌کنند) به سمت نقطه داده منتقل می‌شوند. بعد از اینکه تمام بردارهای آموزش دیده شده امتحان شد، شعاع همسایگی (σ) کاهش می یابد. تکرارها آنقدر ادامه می یابد تا σ به فاصله‌ی بین بردارهای نمونه اولیه اصلی برسد. با توجه به این پیشینه ،کوهونن (۲۰۰۱) الگوریتم آموزش SOM را با استفاده از این۵ مرحله تشریح کرد:

۱. به شکل تصادفی z-scored را از مجموعه بردارهای ورودی انتخاب کنید.

۲. فاصله‌ی اقلیدسی بین بردار a_j و تمام بردارهای نمونه‌ی اولیه m_k و k=1,2,3,……K را محاسبه کنید.

بردار نمونه اولیه که کمترین فاصله را با بردار ورودی a_j دارد به عنوان برنده یا بهترین واحد تطبیقی m_b تعریف می‌شود.

$$\|a_j - m_b\| = min_k \{\|a_j - m_k\|\} \qquad (۵-۲)$$

۳: بردار نمونه‌ی اولیه وینر و همسایگانش را به روزرسانی کنید. قانون به روزرسانی برای وزن k اُمین بردار نمونه اولیه درون و بیرون شعاع همسایگی $\sigma(t)$ از رابطه‌های زیر به دست می‌آید.

$$m_k(t+1)=m_k(t)+\propto(t)h_{bk}(t)\left[a_j-m_k(t)\right] \text{ و } \text{if}\|r_k-r_b\|\le\sigma(t) \qquad (۶-۲)$$

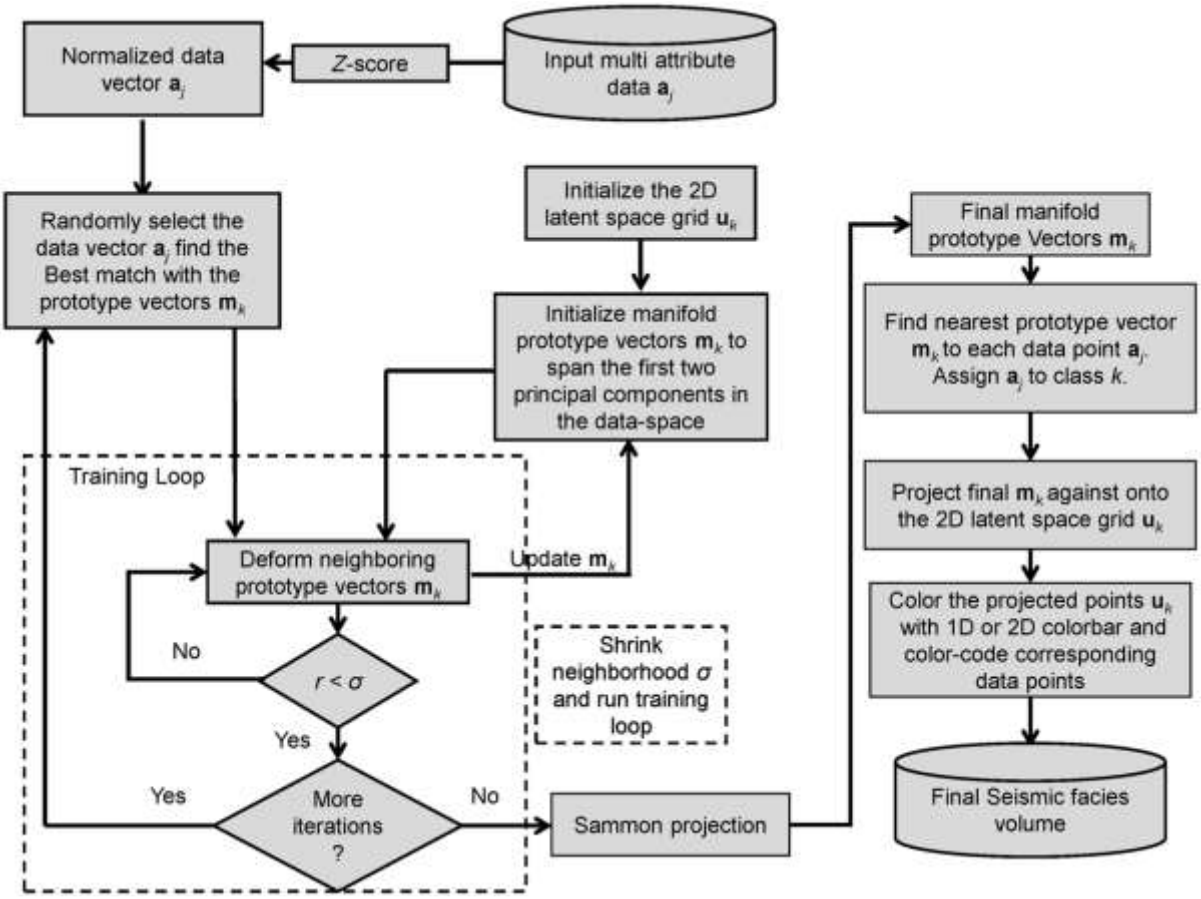
$$\text{if}\|r_k-r_b\|>\sigma(t)$$

در حالی‌که شعاع همسایگی به عنوان $\sigma(t)$ برای مشکل از پیش تعریف شده توصیف شده است و با هر تکرار t کاهش می‌یابد. اینجا r_b و r_k بردارهای موقعیت بردار نمونه اولیه برنده‌ی m_b و k اُمین بردار نمونه اولیه m_b به ترتیب هستند. همچنان تابع همسایگی $h_{bk}(t)$ تعریف شده است. تابع یادگیری نمایی $\propto(t)$ و طول یادگیری T قرار گرفته است. $h_{bk}(t)$ و $\propto(t)$ با هر تکرار در پروسه‌ی یادگیری کاهش می‌یابد و به صورت زیر تعریف شده است.

$$h_{bk}(t)=e^{-}\left(\frac{\|r_b-r_k\|^{۲}}{۲\sigma^{۲}(t)}\right) \text{ و } \propto(t)=\propto_{.}\left(\frac{۰.۰۰۵}{d_{.}}\right)^{t/T} \qquad (۷-۲)$$

۴: در هرمرحله‌ی یادگیری، مراحل ۱تا ۳ را تکرار کنید تا زمانی‌که به معیار همگرایی(که به پایین‌ترین شعاع همسایگی از پیش تعریف شده و حداقل فاصله‌ی بین بردارهای نمونه اولیه در فضای نهفته بستگی دارد) برسید.

۵: بردارهای نمونه اولیه را درون اولین دو مولفه‌ی اصلی وکد رنگی که از یک نوار رنگ دو بُعدی استفاده می‌کند طرح ریزی کنید (ماتوس و همکاران، ۲۰۰۹). شکل (۲-۱)گردش کار SOM را نشان می‌دهد.



شکل۲-۱ گردش کار روش SOM (ژائو و همکاران، ۲۰۱۵)

۲-۳ نقشه توپوگرافی مولد (GTM)

نقشه خود سازمانده در حالی‌که محبوب ترین تکنیک خوشه بندی بدون نظارت است، برای اجرا کردن آسان و از نظر محاسباتی ارزان است، ولی محدودیت‌هایی دارد و هیچ مبانی نظری برای انتخاب شعاع تمرین ندارد. تابع همسایگی و نرخ یادگیری به عنوان این پارامترها داده محور هستند (بیشاپ و همکاران، ۱۹۹۸؛ روی، ۲۰۱۳). بیشاپ و همکاران درسال ۱۹۹۸ یک راهکار تناوبی جایگزین برای نقشه‌ی خود سازمانده که بر محدودیت‌هایش فائق آید ارائه کردند که الگوریتم نقشه توپوگرافی مولد نامیده شد و تکنیک کاهش بُعد غیر خطی است که یک نمایش احتمالی از بردارهای داده در فضای مخفی است. روش GTM با یک آرایه اولیه

از نقاط شبکه که بر روی یک فضای پنهان با ابعاد پایین‌تر چیده شده‌اند آغاز می‌شود. هریک از این نقاط شبکه سپس به شکل غیر خطی درون فضای خمیده غیر اقلیدسی با بُعد مشابه به عنوان بردار مرتبط (m_k) به تصویر کشیده می‌شوند و درون فضای داده‌ی با بُعد متفاوت در نقشه توپوگرافی مولد جاسازی می‌شوند. هر بردار داده (X_k) که در این فضا نگاشت شده است به عنوان مجموعه‌ای از توابع چگالی احتمال گاوسی با محوریت این بردارهای مرجع (m_k) مدلسازی می‌شود. محتویات این مدل گاوسی سپس به صورت تکراری انجام می‌شود تا به سمت بردار داده حرکت کند که به بهترین شکل نمایان است.

۲ـ۳ـ۱ محاسبات ریاضی GTM

در جی‌تی‌ام، نقاط شبکه‌ی منیفولد تغییرشکل یافته‌ی دو بُعدی در فضای نشانگر N بُعدی، مراکز m_k را توصیف می‌کند و از توزیع‌های گوسی واریانس $\sigma^2 = \beta^{-1}$ را توصیف می‌کند. این مراکز m_k به نوبه‌ی خود بر روی یک فضای پنهان دوبعدی پیش‌بینی شده‌اند که توسط شبکه‌ای ازگره‌ها u_k و توابع پایه غیرخطی ϕ تعریف می‌شوند.

$$m_k = \sum_{m=1}^{M} W_{km}\phi_m(u_k) \qquad (۸-۲)$$

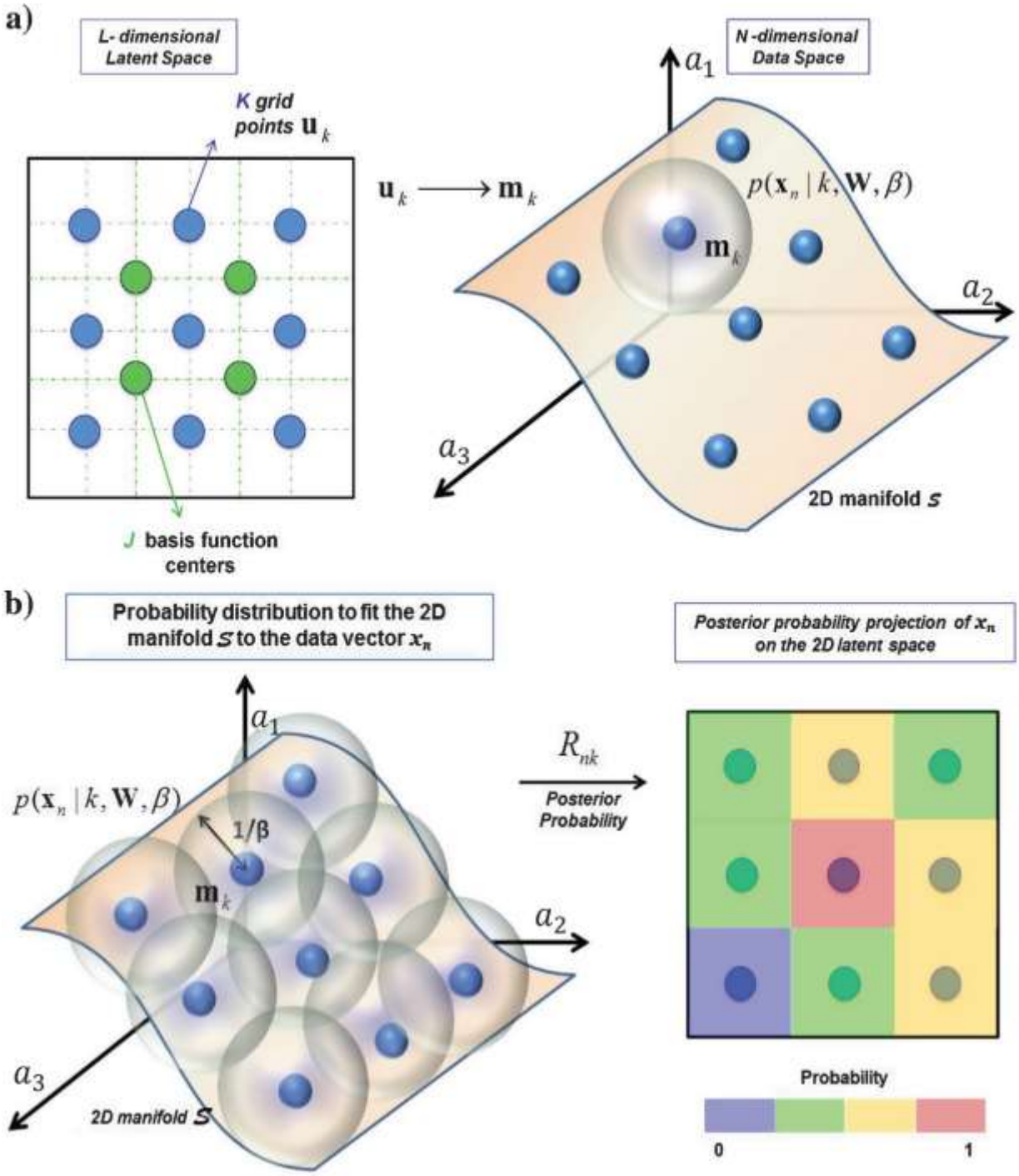
در حالی‌که W یک ماتریس $k \times M$ از وزن‌های ناشناخته است، $\phi_m(u_k)$ مجموعه‌ای از توابع پایه‌ای غیر خطی است، m_k بردارهایی هستند که منیفولد تغییر شکل یافته در فضای داده‌ی N بُعدی را تعریف می‌کند و K=1,2,… است و K تعداد نقاط شبکه مرتب شده دریک فضای مخفی با بُعد کمتراست. مدل نوفه (احتمال وجود بردار داده‌ی خاص a_j، وزن داده داده شده W و واریانس معکوس β) را برای هر بردار داده‌ی اندازه‌گیری شده معرفی می‌کند. تابع چگالی احتمال P توسط مجموعه‌ای از توابع گوسی N بُعدی شعاعی متقارن k که در m_k با واریانس $\frac{1}{\beta}$ متمرکز شده است را نشان می‌دهد.

$$P(a_j|w,\beta) = \sum_{k=1}^{k} \frac{1}{k}\left(\frac{\beta}{2\pi}\right)^{\frac{N}{2}} e^{\frac{-\beta}{2}\|m_k - a_j\|^2} \qquad (۹-۲)$$

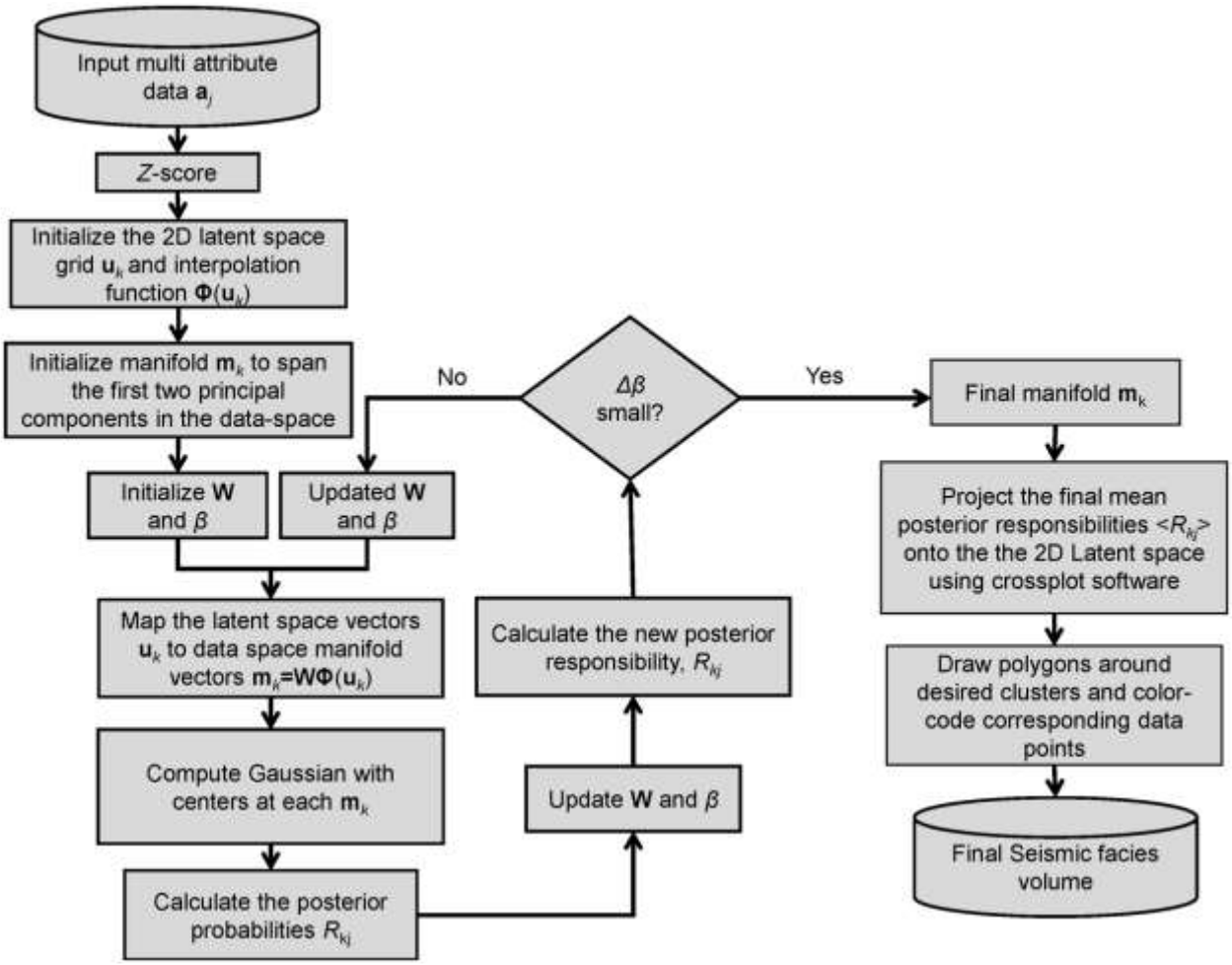
در شکل۲-۲ توضیحی شماتیک برای روش نقشه توپوگرافی مولد آمده است که در ادامه توضیح داده می‌شود.

(a) : نقاط شبکه‌ی k (u_k) بر روی شبکه‌ی فضای پنهان L بُعدی تعریف می‌شوند که بر روی نقاط شبکه‌ی K (m_k) که روی منیفولد غیر اقلیدسی در فضای داده‌ی N بُعدی قرار گرفته است به تصویر کشیده شده است. در این شکل L=۲ در نظر گرفته شده است و در مقابل یک نوار رنگ ۲ بُعدی قرار گرفته است. توابع نگاشت گاوسی مقداردهی اولیه می‌شوند تا در صفحه‌ی تعریف شده توسط دو بردار ویژه‌ی اول به طور مساوی فاصله داشته باشند.

(b): شماتیکی که آموزش شبکه‌ی فضای پنهان را نشان می‌دهد به یک بردار داده (a_j) اشاره می‌کند که نزدیک منیفولد GTM قرار گرفته است. احتمال پسین هر بردار داده برای تمام نقاط مراکز گاوسی (m_k) محاسبه می‌شود و به نقاط شبکه‌ی فضای پنهان مربوطه (u_k) اختصاص داده می‌شوند. نقاط شبکه‌ی با احتمالات بالا با رنگ‌های روشن نمایش داده شده‌اند.



شکل ۲-۲ مراحل روش **GTM** (ژائو و همکاران، ۲۰۱۵). توضیح بخش های **a** و **b** در متن ارائه شده است.



شکل ۲-۳ گردش کار روش **GTM** (ژائو و همکاران، ۲۰۱۵)

احتمالات پیشین هریک از این مولفه ها برابر با مقدار $1/k$ است که برای تمام بردارهای داده‌ی a_j در نظر گرفته شده است. شکل ۲ که GTM را ازیک فضای مخفی L=2D به فضای داده‌ی سه بُعدی نشان می‌دهد. مدل چگالی احتمالی (GTM) برای پیدا کردن پارامترهای W و β با استفاده از تخمین بیشترین احتمال، متناسب با داده a_j شده است. یکی از تکنیک‌های مطرح در تخمین پارامترها الگوریتم EM است. با استفاده کردن از قضیه‌ی تئوری بایز و مقادیر فعلی پارامترهای مدل GTM، W و β ما می‌توانیم $k \times$ J احتمال خلفی (R_{jk}) را برای هریک از اجزای k در فضای مخفی برای هر بردار داده به شکل زیر محاسبه کنیم.

$$R_{jk} = \frac{e^{\frac{-\beta}{2}\|m_k - a_j\|^2}}{\sum_i e^{\frac{-\beta}{2}\|m_i - a_j\|^2}} \qquad (۱۰-۲)$$

تساوی بالا، مرحله E یا مرحله‌ی انتظار در الگوریتم EM را شکل می‌دهد. مرحله‌ی E با بیشینه سازی یا مرحله M دنبال می‌شودکه با این کار برای به‌روزرسانی مدل برای ماتریس وزن جدید W با حل کردن مجموعه‌ای از معادلات خطی استفاده می‌کند (دپمستر و همکاران، ۱۹۷۷).

$$\left(\Phi^T G\Phi + \frac{\alpha}{\beta} I\right) W_{new}^T = \Phi^T \mathrm{RX} \qquad (۱۱-۲)$$

که $G_{kk} = \sum_{j=۱}^{j} R_{jk}$ المان‌های غیر صفر از ماتریس قطری G که $k \times k$ هستند، Φ ماتریس $K \times M$ با المان‌های $\Phi = \Phi_m(u_k)$ و α یک ثابت مقررات(regolarization) است که برای جلوگیری از تقسیم بر صفر است و I ماتریس همانی $M \times M$ است. مقدار به روز شده β توسط رابطه زیر به دست می‌آید.

$$\frac{۱}{\beta_{new}} = \frac{۱}{JN} \sum_{j=۱}^{j} \sum_{k=۱}^{k} R_{Jk} \left\| W_{km\,new} \Phi_m(u_k) - a_j \right\|^۲ \qquad (۱۲-۲)$$

مقداردهی اولیه‌ی ۷۷ به گونه‌ای انجام می‌شود که مدل اولیه‌ی GTM مولفه‌های اصلی (بزرگترین بردار ویژه) را یعنی a_j، از داده ورودی تخمین می‌زند. مقدار $\beta^{-۱}$ مقداردهی اولیه شده تا بزرگتراز L+1 اُمین مقدار ویژه از PCA باشد، درحالیکه L بُعد فضای مخفی است.

۲-۴ مدل‌های ترکیبی گوسی (Gaussian mixture models)

این مدل‌های ترکیبی، مدل‌هایی از پارامترهای توزیع احتمالات هستند که در قیاس با الگوریتم‌های خوشه‌بندی سنّتی، دقت بیشتر و همچنان انعطاف پذیری بالاتری را می‌توانند برای ما به ارمغان بیاورند. از دیدگاه نظری روش‌های خوشه‌بندی را به دو روش سخت (کلاسیک) و نرم (فازی) می‌توان تقسیم‌بندی کرد. در روش اول، هر نمونه از داده تنها در یک خوشه جای می‌گیرد و این بدین معنی است که از دیدگاه احتمالات از نظر جای‌گیری هر نمونه داده در یک خوشه‌ی خاص یا صفر یا یک می‌باشد ولی در روش دوم که همان روش نرم است، هر نمونه داده می‌تواند از نظر احتمالات در بیشتر از یک خوشه قرار بگیرد. روش مدل‌های ترکیبی گوسی نیز در دسته‌ی دوم یعنی خوشه‌بندی نَرم قرار می‌گیرند. با برجسته‌ترشدن داده های لرزه‌ای سه بُعدی و با معرفی نشانگرهای لرزه‌ای جدید هر ساله، تجزیه و تحلیل چند متغیره برای تجزیه و تحلیل رخساره‌های

لرزه‌ای رایج‌تر شده است. تکنیک‌های یادگیری ماشینی مبتنی بر رایانه ابزارهایی را برای تجزیه و تحلیل خودکار مقادیر عظیمی از داده‌های چند متغیره، شناسایی الگوهایی که در صورت تفسیر توسط یک مفسر با زمان محدود نادیده گرفته می‌شوند را فراهم می‌کنند. تکنیک‌های یادگیری ماشینی اغلب به دو دسته تقسیم می‌شوند: یادگیری بدون نظارت وبا نظارت. در مورد یادگیری با نظارت، مفسر راهنمایی‌های مهمی را برای الگوریتم در مورد وضعیت پنهان طبیعت مانند رخساره‌های لرزه‌ای زیرین ارائه می‌کند. به عنوان مثال، یک مفسر ممکن است به رایانه دستور دهد که تفاوت بین رخساره‌های مخزن و غیر مخزن را در داده های لرزه‌ای با ارائه الگوریتم با زیرمجموعه ای از داده هایی که مفسر می داند هر کدام از این دو حالت را نشان می دهد، بیاموزد. همانطور که استفاده ازنشانگرهای لرزه‌ای گسترده‌تر می‌شود، تجزیه و تحلیل لرزه‌ای چند متغیره برای تجزیه و تحلیل رخساره‌های لرزه‌ای رایج تر شده است. تکنیک‌های یادگیری ماشینی بدون نظارت، روش‌هایی را برای یافتن خودکار الگوها در داده‌ها با حداقل تعامل کاربر ارائه می‌کنند. هنگام استفاده از تکنیک‌های یادگیری ماشینی بدون نظارت، مانند k-means یا نقشه خودسازمانده Kohonen (SOM)، تعداد خوشه‌ها اغلب می‌تواند به‌طور مبهم تعریف شود و هیچ معیاری وجود ندارد که الگوریتم تا چه حد در طبقه‌بندی بردارهای داده مورد اطمینان است. فرمول‌بندی احتمالی مبتنی بر مدل مدل‌های مخلوط گاوسی (GMMs) اجازه می‌دهد تا تعداد و شکل خوشه‌ها به شیوه‌ای عینی‌تر با استفاده از چارچوب بیزین تعیین شود که احتمال و پیچیدگی یک مدل را در نظر می‌گیرد. مدل‌های ترکیبی ابزارهای آماری بر مبنای مدل برای تقریب توابع چگالی احتمال هستند. مدل‌های مخلوط گوسی در مدل‌سازی توزیع‌های چند حالتی با داشتن توزیع گوسی برای هرمُد خوب هستند. هاتاوی در سال ۱۹۸۶ روش مدل‌های ترکیبی گوسی را به عنوان تکنیک خوشه‌بندی فازی معرفی کرد. GMMها ارزیابی کمّی‌تر از خوشه‌های ذاتی را به دلیل فرمول احتمالی آنها ارائه می‌دهند. مدل‌های ترکیبی گوسی برای نمایش چگونگی عوارض به شکل خودکار با استفاده از یادگیری ماشینی تولید می‌شوند.

توزیع گاوسی چند متغیره را می توان به شرح زیر تعریف کرد.

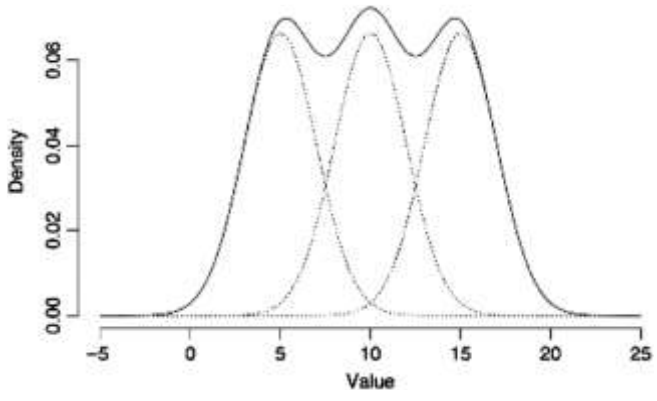
$$\phi(\chi|\mu,C)=\frac{1}{(x\pi)^{\frac{d}{2}}|C|^{\frac{1}{2}}}e^{-\frac{1}{2}(x-\mu)^{c^{-1}}(x-\mu)^T} \qquad (۱۳-۲)$$

در حالیکه d بُعد هستش، C ماتریس کواریانس است و |C| دترمینان ماتریس کواریانس است و μ میانگین توزیع است. به طورکلی GMM تابع چگالی احتمالی را با جمع کردن توزیع‌های گاوسی وزن‌دار تقریب می‌زند، همچنان GMM می‌تواند برای خوشه‌بندی و طبقه‌بندی، جایی‌که یک خوشه از GMM را می توان با توزیع گاوسی وزن‌دار آن توصیف کرد. چگالی ترکیبی گاوسی با خوشه‌های k برای یک بردار داده‌ی X توسط معادله‌ی زیر آورده شده است.

$$P(x|\psi)=\sum_{k=1}^{k}\pi_k\,\Phi(x|\mu_k,c_k) \quad \text{و} \quad \sum_{k=1}^{k}\pi_{k=1} \qquad (۱۴-۲)$$

درحالیکه ψ پارامترهای GMM را نشان می‌دهد و π_k وزن k اُمین ترم را نشان می‌دهد به طوری‌که π_k برای تمام $k \in 1\ldots K$ بزرگتر از صفر است.

تصویری از یک مدل ترکیبی گوسی ۳ ترمی در شکل ۲-۴ نشان داده شده است. نمونه ای از یک GMM با سه جزء مخلوط است. چگالی کلّی به عنوان مجموع سه جزء گاوسی تخمین زده می شود. این چگالی می‌تواند برای طبقه‌بندی بردارهای داده و همچنین ارائه معیاری از عدم قطعیت و ابهام در طبقه بندی استفاده شود.



شکل ۲-۴ یک **GMM** نمونه (هاردیستی و والت، ۲۰۱۷)

احتمال پسین $w_{i,k}$ بردار داده‌ی x_i که متعلق به ترم k است از رابطه زیر به‌دست می‌آید.

$$w_{i,k}=\frac{\pi_k \varphi(x_i|\mu_k,c_k)}{p(x_i|\psi)} \quad (۱۵-۲)$$

در حالی‌که $P=(x_i|\psi)$ در رابطه‌ی شماره ۲-۲ تعریف شده است، در مورد این خوشه‌بندی هر ترم را می‌توان به عنوان نماینده یک خوشه در نظر گرفت. دراین مورد بردار داده‌ی x_i را می‌توان متعلق به خوشه g طبقه‌بندی کرد در حالی‌که رابطه زیر برقرار است.

$$g=\text{argmax}(w_{x_i k}) \text{ و } k=\in ۱ \dots \dots \dots k \quad (۱۶-۲)$$

توجه داشته باشیم که چون ما هر خوشه‌ی $k=\in ۱ \dots \dots \dots k$ را به عنوان نماینده یک کلاس زیرین یا نهفته‌ی $G \dots \dots ۱ \in g$ مورد بررسی قرار دادیم که در آن K=G است، یک رابطه‌ی متناظر بین ks و Gs وجود دارد با این تفاوت که بیشتر معنایی است. در طبقه‌بندی رویداد بر اساس انتساب خوشه، بنزمایل و همکاران درسال ۱۹۹۷ مشاهده‌ی عدم قطعیت x_i متعلق به خوشه K را مشاهده کردند.

$$A_{xig} = ۱ - w_{xig} \quad (۱۷-۲)$$

در مورد انتساب سخت نیز مانند معادله ۲-۵ عدم قطعیت این است.

$$A_{x_i} = ۱-\square\square\square w_{x_i k} \text{ و } k\in ۱ \dots \dots . k \quad (۱۸-۲)$$

هنگامی که به یک طبقه‌بندی مرتبط می‌شود به این ابهام اشاره می‌کنیم تا آن را از عدم اطمینان GMM در توصیف یک مشاهده متمایز کنیم. توجه داشته باشید که مشاهده‌ای که بین دو جزء مدل وجود دارد می‌تواند در طبقه‌بندی آن مبهم باشد، در حالی‌که هنوز توسط GMM به خوبی توصیف می‌شود. علاوه بر این، نقاطی که در طبقه‌بندی خود مبهم هستند می توانند با توجه به GMM یا جزء GMM که از طریق طبقه‌بندی به آنها اختصاص داده شده است غیر عادی و در نتیجه نامشخص باشند. برای ارزیابی عدم قطعیت یک مشاهده جزء اختصاص داده شده k (در نتیجه خوشهg) از مشاهده‌ی x ، ما از فاصله ماهالانوبیس که در زیر آمده است استفاده می‌کنیم.

$$U_{x,k} =\sqrt{(x-\mu_k)C^{-۱}(x-\mu_k)^T} \quad (۱۹-۲)$$

درحالی‌که μ_k و C به ترتیب میانگین و کواریانس تخصیص داده شده هستند. مشاهدات با یک فاصله‌ی ماهالانوبیس بزرگ به وسیله‌ی این مدل ضعیف توضیح داده شده است و به این ترتیب غیر طبیعی یا آنومالی محسوب می‌شوند. هنگام ساختن مدل‌هایی برای خوشه‌بندی، مطلوب است مدلی داشته باشید که صرفه‌جو باشد. زیرا این مدل ساده‌ترین مدل با قدرت توضیح کافی است. همچنین می‌توانیم این مدل را پیچیده‌ترین مدل پشتیبانی شده توسط داده‌ها در نظر بگیریم. به طور معمول، این از مقایسه‌ی مدل‌هایی است با یک معیار که برای برازش داده‌ها پاداش می‌دهد و برای پیچیدگی مدل جریمه می‌کند به دست می‌آید.

۲-۵ روش k-means

k-means شاید ساده‌ترین الگوریتم خوشه‌بندی است و به‌صورت گسترده در بسته‌های نرم‌افزاری تجاری تفسیر در دسترس است (مک‌کوین، ۱۹۶۷). این روش یک الگوریتم یادگیری بدون نظارت است که در آن مفّسر هیچ اطلاعات قبلی به جز انتخاب نشانگرها و تعداد خوشه‌های مورد نظر ارائه نمی‌دهد. هدف k-means کاهش داده‌ها نیست، بلکه تقسیم‌بندی داده‌ها به زیرمجموعه‌های از هم‌گسسته است. از مزایای این روش این است که در جایی که داده‌ها بدون نوفه با ابعاد نسبی بالا باشند به خوبی کار می‌کنند. یک اشکال k-means این است که مفّسر نیاز دارد که تشریح کند چه مقدار خوشه در داده ساکن است. هنگامی که تعداد خوشه‌ها تعریف شد، میانگین خوشه‌ها یا مراکز خوشه‌ای بر روی یک شبکه یا برای شروع حلقه‌ی تکرار بصورت تصادفی تعریف می‌شوند. پارتیشین‌های خوشه‌ای k-means توزیع مشخصی از نقاط بدون برچسب را در تعداد ثابت گروه‌ها یا خوشه‌ها می‌دهد. اگر n نقطه داده‌ی x_i وجود داشته باشد، در حالیکه I=1,2,……,n است، سپس این‌ها باید در خوشه‌های k تقسیم شوند به طوری‌که یک خوشه به هر نقطه داده اختصاص یابد و نقاط درون هر خوشه شباهت بیشتری با یکدیگر نسبت به دیگری از نقاط گروه دیگر دارند.

فرآیند خوشه‌بندی با اختصاص دادن به صورت تصادفی به مراکز k شروع می‌شوندکه می‌توانند به عنوان مراکز گروه مورد نظر ما تشکیل شوند. بدین شکل هر مرکز یک خوشه را تعریف می‌کند. فاصله‌ی بین هر نقطه داده و مرکزهرخوشه اکنون محاسبه می‌شود. به طور سنتی این فاصله، فاصله‌ی اقلیدسی بین دو نقطه هست. اگر نقطه‌ای از هر مرکز دیگر به مرکز آن خوشه نزدیک‌تر باشد اکنون ممکن است در داخل یک خوشه‌ی خاص

در نظر گرفته شود. به این ترتیب نقاط در هر خوشه سازماندهی مجدد می‌شود. مرحله بعدی، محاسبه‌ی مجدد مراکز براساس نقاط تجدید سازمان یافته در هر خوشه است. این دو مرحله به شکل تکراری تا زمانی که هیچ حرکتی از مراکز وجود نداشته باشد انجام می شود، سپس همگرایی حاصل می شود. معیار فاصله‌ی اقلیدسی فرض می‌کندکه هیچ ارتباطی بین متغیرهای طبقه‌بندی وجود ندارد و به همین دلیل از آن به طور سنتی استفاده شده است. این بدان معنی است که وقتی متغیرهای طبقه‌بندی یک شکل کروی از خوشه‌ها را در فضای کراس‌پلات نمایش می‌دهند، رویکرد فاصله‌ی اقلیدسی کار خواهد کرد. بسیاری از اوقات، این نیاز به دلیل این‌که متغیرهای طبقه‌بندی باهم همبستگی دارند و خوشه‌ها شکل بیضوی را نشان می‌دهند برآورده نمی‌شوند. در چنین مواردی، این روش دراین همگرایی ممکن است حاصل نشود. در این موارد، الگوریتم خوشه‌بندی k-means اصطلاح می‌شود به این دلیل که معیار فاصله‌ی متفاوتی را به نام فاصله‌ی ماهالانوبیس به جای فاصله‌ی اقلیدسی اتخاذ می‌کند. بنابراین با انتخاب مراکز به عنوان نقاطی که داری چگالی بالا از نقاط همسایه هستند و همچنین از فاصله‌ی ماهالانوبیس به جای فاصله اقلیدسی درالگوریتم خوشه‌بندی k-means استفاده می‌کنند، اجازه می‌دهند تا آن را به درستی به عنوان خوشه‌های غیرکروی و غیر همگن دسته‌بندی کند.

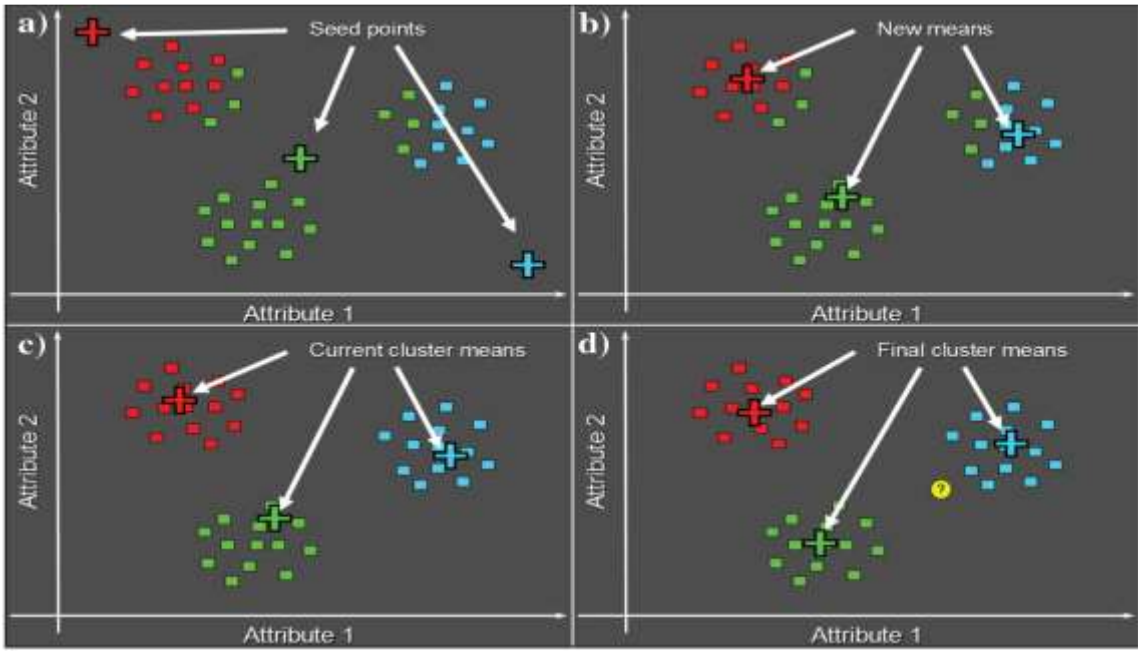
الگوریتم خوشه‌بندی k-means از الگوریتم تکراری استفاده می‌کند که مجموع فاصله‌ها را از هر نمونه تا مرکز خوشه‌ی آن بر روی همه خوشه‌ها به حداقل می‌رساند (صابر، ۱۹۸۴). این الگوریتم نمونه‌ها را بین خوشه‌ها جابجا می‌کند تا زمانی‌که میانگین نتواند بیشترکاهش یابد. نتیجه مجموعه‌ای از خوشه‌ها است که تا حد امکان فشرده و به خوبی از هم جدا شده‌اند.

الگوریتم خوشه بندی k-means با دنبال کردن مراحل زیر انجام می شود.

در مرحله‌ی اول با K مرکز خوشه تصادفی شروع کنید. در مرحله‌ی دوم هر نمونه‌ی جدید را به خوشه‌ای با نزدیکترین مرکز اختصاص دهید. بعد از اینکه همه‌ی نمونه‌ها اختصاص یافتند، مرکز جدید برای هر خوشه محاسبه می‌شود. در مرحله‌ی سوم هم مرحله‌ی ۲ را تا زمانی‌که مراکز جدید تغییر نکردند تکرار کنید.

شکل ۲-۵ نشان‌دهنده‌ی طبقه‌بندی k-means از ۳ خوشه است. (الف) ۳ نقطه‌ی تصادفی یا با فاصله‌ی مساوی اما متمایز را انتخاب کنید که به عنوان تخمین اولیه میانگین برداری هر خوشه عمل می‌کند. سپس فاصله‌ی ماهالانوبیس بین هر بردار داده و میانگین خوشه را محاسبه کنید. آنگاه کد رنگی هر بردار را به گونه‌ای برچسب گذاری کنید که متعلق به خوشه‌ای است که فاصله‌ی ماهالانوبیس کوچکتری دارد.

(ب) میانگین هر خوشه را از بردارهای داده تعریف شده قبلی مجدداً محاسبه کنید. (ج) فاصله ماهالانوبیس را از هر بردار تا میانگین خوشه جدید دوباره محاسبه کنید. هر بردار را به خوشه‌ای که کمترین فاصله را دارد اختصاص دهید. (د) این فرآیند تا زمانی ادامه می‌یابد که تغییرات در وسایل به مکان‌های نهایی خود همگرا شوند. اگر اکنون یک نقطه (زرد) جدید اضافه کنیم، از یک طبقه‌بندی کننده بیزی برای تعیین اینکه در کدام خوشه قرار می‌گیرد استفاده می‌کنیم.



شکل ۲-۵ توصیف روش خوشه‌بندی **k-means** (ژائو و همکاران، ۲۰۱۵)

فرمول k-means : یک الگوریتم خوشه‌بندی بهینه باید فاصله‌ی بین عناصر هر گروه را به حداقل برساند و در همان زمان فاصله‌ی بین خوشه‌های مختلف را حداکثر کند.

روش‌های مختلفی برای اندازه‌گیری فاصله برای سادگی وجود دارد (تئودوریدیس و کوترومراس، ۱۹۹۹). ما از نُرم اقلیدسی استفاده خواهیم کرد. برای محاسبه‌ی فاصله بین عناصر هر گروه ما از فاصله متوسط S_k بین هر عنصر x_i و مرکز گروه آن C_k استفاده می‌کنیم.

$$S_K=\frac{\sum_i \|x_i-c_k\|}{N_k} \quad (۲۰-۲)$$

در حالیکه N_k شماره عناصر (المان‌ها)در گروه است. فاصله بین گروه‌های k و l به عنوان فاصله‌ی بین مرکز آن‌ها محاسبه می‌شود.

$$d_{kl} = \|c_k - c_l\| \quad (۲۱-۲)$$

الگوریتم خوشه‌بندی جزئی، مجموعه داده‌ها را به تعدادی ازخوشه‌های از پیش تعریف شده تقسیم می‌کند و سعی می‌کند برخی از عملکردهای خطا را با تعداد گروه‌هایی که از طریق تجسم SOM انتخاب و تایید شده‌اند را به حداقل برساند (پاندیا ومِیسی، ۱۹۹۵).

برای خودکارسازی فرآیند طبقه‌بندی ما از شاخص DBI دیویس و بولدین (۱۹۷۹) به عنوان ابزاری برای ارزیابی نتایج پارتیشن بندی k-means استفاده می‌کنیم. بهترین خوشه‌بندی مربوط به حداقل DBI توسط رابطه‌ی زیر بدست می‌آید.

$$\text{DBI}=\frac{1}{K}=\sum_{k=1}^{k} \max_{l \neq k} \left\{\frac{S_k+S_l}{d_{kl}}\right\} \quad (۲۲-۲)$$

که k تعداد گروه‌ها است، S_k و S_l با رابطه‌ای که در ابتدا بیان شده است تعریف می‌شوند و d_{kl} با معادله‌ی دومی تعریف می‌شود. مقادیر DBI کوچکتر از واحد نشان‌دهنده‌ی گروه‌های جداگانه است، درحالی‌که مقادیر بزرگتر از واحد نشان‌دهنده‌ی گروه‌هایی است که ممکن است همپوشانی داشته باشند.

۲-۶ تحلیل مولفه‌ی اصلی (PCA)

تحلیل مولفه‌ی اصلی بیشترین کاربرد تکنیک آماری را برای کاهش داده‌ها و ابعاد بیش از حد داده‌ی نشانگر ورودی دارد. تحلیل مولفه‌های اصلی تکنیک طبقه‌بندی واقعی نیست، اما اغلب به عنوان یک فیلتر برای کاهش

ابعاد داده‌های لرزه‌ای قبل از طبقه‌بندی استفاده می‌شود. با کاهش نوفه و افزونگی داده‌ها می‌توان کیفیت نقشه‌ی ۲ بُعدی یا حجم‌های ۳ بُعدی حاصل را به خصوص در مناطق نوفه‌ای به میزان قابل توجهی افزایش داد. اگر چه ابعاد کم داده‌ی لرزه‌ای باعث می‌شود PCA درکاهش تعداد نمونه‌ها در داده‌هایی که باید طبقه‌بندی شوند کارآمد باشد، اندازه پیمایش‌های لرزه‌ای (از نظر تعداد رد لرزه‌ها) به دلیل فرآیندهای حافظه-فشرده یک محدودیت است. بسیاری از کاربران از PCA برای کاهش نشانگرهای زاید برای آسان‌سازی محاسبات استفاده می‌کنند، این کاهش بر مبنای این فرض است که بیشتر سیگنال‌ها در چند مولفه‌ی اصلی حفظ می‌شوند (بردار ویژه)، در حالی‌که آخرین مولفه‌های اصلی شامل نوفه‌ی غیر مرتبط هستند. اولین بردار ویژه برداری است در فضای نشانگر که به بهترین شکل الگوهای نشانگر در داده را نمایش می‌دهد.

کراس پلات کردن داده‌های N بُعدی در برابر اولین بردار ویژه درهر وکسل به ما اولین حجم مولفه اصلی را می‌دهد. اگر اولین بردار ویژه را با اولین مولفه‌ی اصلی مقایسه کنیم و آن را از بردار داده‌ی اصلی کسر کنیم یک بردار داده‌ی باقیمانده را بدست می‌آوریم. بردار ویژه‌ی دوم برداری است که به بهترین حالت نشان دهنده الگوی نشانگر در این باقی‌مانده است. بردار ویژه‌ی دوم در برابر داده‌های اصلی یا بردارداده‌ی باقیمانده در هر وکسل به ما دومین حجم مولفه‌ی اصلی را می‌دهد. این پروسه برای تمام نتایج N بُعدی که منجر به N بردار ویژه و N مولفه اصلی می‌شود ادامه پیدا می‌کند. کاهش داده با رها کردن ابعاد با کمترین تغییر پذیری به دست می‌آید. شکل ۲ که نمودار شماتیک تجزیه و تحلیل PCA را نشان می‌دهد که یک کراس پلات ۳ بُعدی به دو جز PCA کاهش می‌یابد. PCA مجموعه داده‌ی اصلی X را با استفاده از تبدیل خطی P به مجموعه داده جدید Y تبدیل می‌کند.

$$Px=y \qquad (۲۳-۲)$$

هدف از تبدیل خطی (P) حذف افزونگی از داده‌ها است (شلنس، ۲۰۱۴). این امر با قطری کردن ماتریس کوواریانس مجموعه داده‌ای جدید S_Y طبق رابطه‌ی زیر انجام می‌شود.

$$S_Y = \frac{۱}{N-۱} YY^T \qquad (۲۴-۲)$$

ما می‌توانیم S_Y را با استفاده از P طبق رابطه زیر دوباره بنویسیم و این‌که $A=XX^T$ است و بنابراین متقارن است.

$$S_Y = \frac{۱}{N-۱} PAP^T \quad (۲۵-۲)$$

یک ماتریس متقارن A از رابطه زیر به‌دست می‌آید. درحالیکه D ماتریس مورب و ماتریس بردارهای ویژه Aاست.

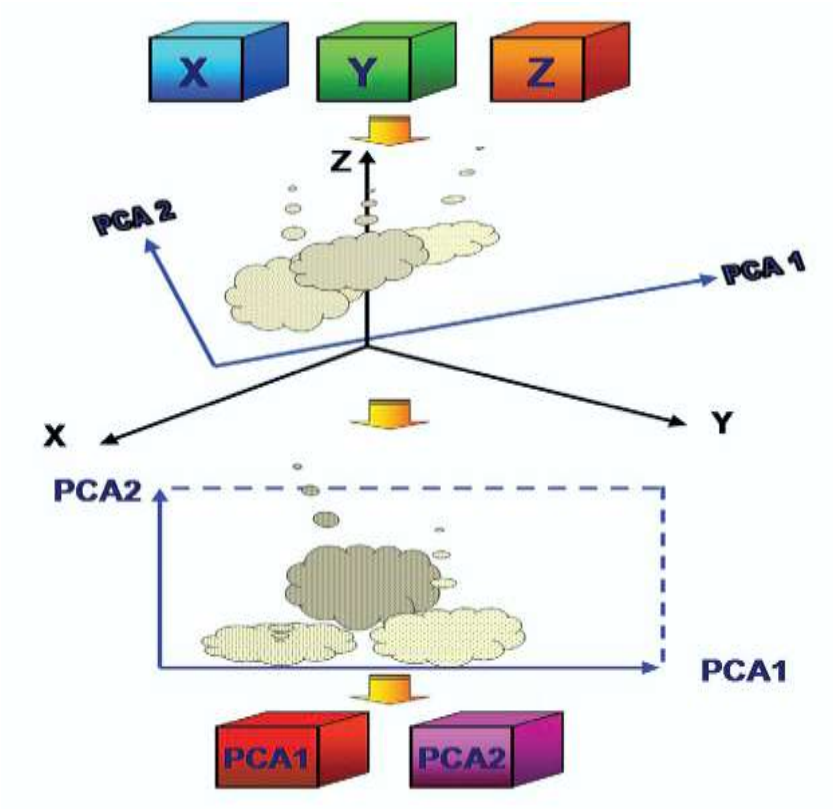
$$A=EDE^T \quad (۲۶-۲)$$

با انتخاب و جایگزین کردن رابطه ۲ـ۳ در رابطه ۲ـ۴ رابطه زیر را فراهم می‌کند.

$$S_Y = \frac{۱}{N-۱} P(P^T DP)P^T \quad (۲۷-۲)$$

$$S_Y = \frac{۱}{N-۱} D \quad (۲۸-۲)$$

این انتخاب P ماتریس کواریانس S_Y را مورب می‌کند. محتوای اصلی داده به عنوان بردارهای ویژه $A=XX^T$ و سطرهای P ظاهر می‌شود. به علاوه واریانس X در امتداد مولفه‌های اصلی مقادیر ویژه‌ی S_Y هستند. شکل ۲-۶ نمودار شماتیک تجزیه و تحلیل PCA را نشان می‌دهد که یک کراس‌پلات ۳ بُعدی به دو جز PCA کاهش می‌یابد.



۲-۶ نمودار شماتیک تجزیه و تحلیل **PCA** (کولئو و همکاران، ۲۰۰۲)

۲ -۷ الگوریتم خوشه‌بندی سلسله مراتبی(Hierarchical clustering algorithm)

در پروسه‌ی خوشه‌بندی سلسله مراتبی، داده ورودی نرمالیزه شده و به الگوریتم تحلیل مولفه‌های اصلی وارد می‌شود. تحلیل مولفه‌ی اصلی داده اضافی و نوفه‌های تصادفی را کاهش می‌دهد. سپس بر اساس ماتریس عدم تشابه، نمونه‌های زمان در خوشه‌های الگوهای مشابه ادغام می شوند. در ابتدا، هر نمونه به یک خوشه اختصاص داده می‌شود. ماتریس عدم شباهت براساس تعریف فاصله مثل فاصله اقلیدسی بین نمونه ها ساخته می‌شود. این ماتریس سپس برای خوشه‌بندی نمونه‌ها در پروسه سلسله مراتب استفاده می‌شود. در هر مرحله، خوشه‌های مشابه‌تر در خوشه‌های جدید ادغام شده و ماتریس عدم شباهت به روز می‌شود. در نهایت تمام نمونه‌ها در یک خوشه ادغام می‌شوند. در اینجا سه الگوریتم : پیوند واحد (simple linkage)، پیوند کامل و الگوریتم پیوند میانگین استفاده شده است. در الگوریتم پیوند واحد، فاصله‌ی جدید بین دو خوشه با استفاده از رابطه‌ی ۱ محاسبه شده است. الگوریتم پیوند کامل بر اساس رابطه ۲ و الگوریتم پیوند میانگین بر اساس رابطه‌ی ۳ محاسبه شده است.

$$d(r,s) = min\left[dist\left(x_{ri}, x_{sj}\right)\right] \text{ و } i \in (1, \ldots\ldots, n_r) \text{ و } j \in (1, \ldots\ldots, n_s) \quad (۲۹-۲)$$

$$d(r,s) = max\left[dist\left(x_{ri}, x_{sj}\right)\right] \text{ و } i \in (1, \ldots\ldots, n_r) \text{ و } j \in (1, \ldots\ldots, n_s) \quad (۳۰-۲)$$

$$d(r,s) = \frac{1}{n_r n_s} \sum_{i=1}^{nr} \sum_{j=1}^{ns} dist\left(x_{ri}, x_{sj}\right) \quad (۳۱-۲)$$

در حالیکه $d(r,s)$ فاصله‌ی جدید بین خوشه‌های r و s است. $dist\left(x_{ri}, x_{sj}\right)$ تمام فاصله‌های بین تمام نمونه‌های دو خوشه است و n_r و n_s به ترتیب تعداد خوشه‌های r و s است (وبو، ۲۰۰۲).

۲-۸ روش فازی C-means

تحلیل رخساره‌های لرزه‌ای به شکل خودکار با طبقه‌بندی ویژگی‌های زمین‌شناسی از طریق الگوریتم‌های تشخیص الگو پیاده‌سازی می‌شود و روشی کارآمد را برای تجزیه و تحلیل داده‌های لرزه‌ای بزرگ ارائه می‌کند (ماراکونی، ۲۰۱۳). نقشه حاصله وسعت مخزن را مشخص می‌کند و ناهمگنی آن را مشخص می‌کند. این تکنیک برای بررسی مجموعه‌ای از خطوط لرزه‌ای و برشهای زمانی(Time Slice) نیازی به مفسر ندارد. بنابراین طبقه‌بندی رخساره‌های لرزه‌ای خودکار به یک تکنیک استاندارد برای تفسیر داده های لرزه‌ای سه بُعدی تبدیل شده است(کوئلو، ۲۰۰۳؛ ماراکونی، ۲۰۰۸). مفسر می‌تواند رخساره‌های زمین‌شناسی را با استفاده از لاگ‌های چاه و اطلاعات لرزه‌ای شناسایی کند که به آن تحلیل رخساره‌های لرزه‌ای با نظارت می‌گویند. طبقه‌بندی کننده ابتدا با اطلاعات قبلی آموزش داده می‌شود و سپس برای پیش‌بینی رخساره‌های بین مکان‌های چاه استفاده می‌شود. روش دیگر برای شناسایی رخساره‌های زمین شناسی از داده‌های لرزه‌ای روش تحلیل رخساره‌های لرزه‌ای بدون نظارت است که به اطلاعات زمین‌شناسی و چاه‌ها وابسته نیست و نقشه رخساره‌های لرزه‌ای را با استفاده از تکنیک‌های خوشه‌بندی تولید می‌کند(دی ماتوس، ۲۰۰۶). نقشه‌ی به دست آمده دارای برچسب نیست و باید با استفاده از برخی اطلاعات اضافی تفسیر شود، لیو (۲۰۱۴) یک روش تقسیم‌بندی را برای انجام تحلیل رخساره لرزه‌ای پیشنهاد کرد که در آن نشانگرها و اطلاعات مکانی به طور همزمان استفاده می‌شود. الگوریتم FCM یکی از شناخته شده‌ترین روش‌ها در تجزیه و تحلیل خوشه‌بندی است و با موفقیت در تصویربرداری پزشکی، تشخیص هدف و تقسیم‌بندی تصویر و غیره استفاده شده

است(چوانگ، ۲۰۰۶؛کانگ، ۲۰۱۴). این الگوریتم فقط اجازه می‌دهد که یک عضو از داده به ۲ یا خوشه‌های بیشتر تعلق بگیرد. با توجه به مجموعه داده‌ای $\{x_1$ و x_2 و ... و $x_N\}$ در حالی‌که N تعداد نقاط داده است، الگوریتم FCM مجموعه داده را به K خوشه تقسیم می‌کند. تابع هدف به صورت زیر تعریف می شود:

$$J = \sum_{j=1}^{n} \sum_{k=1}^{k} \cup_{jk}^{q} \gamma\,(x_i\,, v_k)^2 \qquad (۳۲-۲)$$

$$\sum_{k=1}^{K} u_{ik} = 1 \qquad \forall i \qquad (۳۳-۲)$$

که u_{ik} نشان‌دهنده عضویت فازی نقطه داده‌ی i با توجه به خوشه K و v_k مرکز خوشه را نشان می‌دهد و $\gamma(x_i$ و $v_k)$ اندازه‌گیری عدم تشابه بین نقطه داده x_i و مرکز v_k است. الگوریتم FCM الگوریتمی است که حداقل مقدار محلی را پیدا می‌کند. با استفاده از روش ضریب لاگرانژ رابطه ی زیر به دست می‌آید.

$$J_l = J + \sum_{i=1}^{n} \lambda_i\,(1 - \sum_{k=1}^{K} u_{ik}) \qquad (۳۴-۲)$$

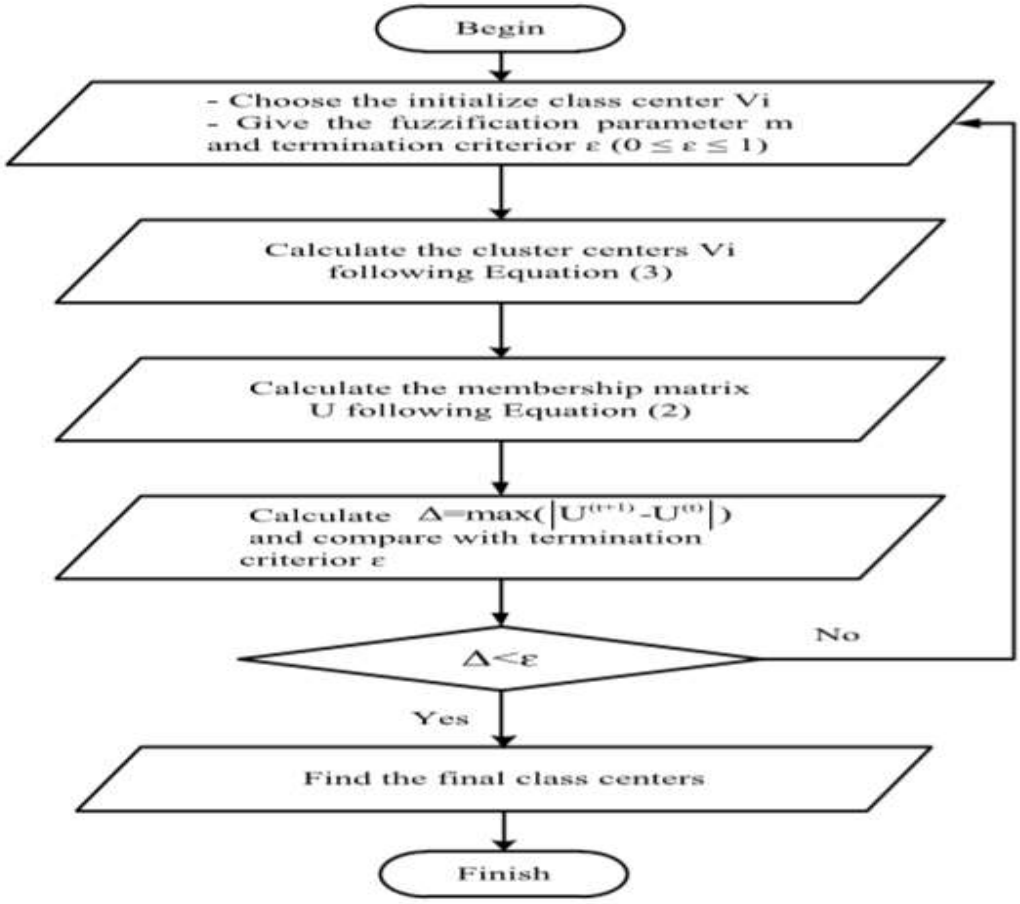
در حالیکه λ_i (j=1....N) ضرایب لاگرانژ هستند. با تنظیم مشتق با توجه به u_{ik} و v_k به صفر، ما می‌توانیم معادلات زیر را استخراج کنیم.

$$u_{ik} = \left[\sum_{j=1}^{k} \frac{(x_i - v_k)^{\frac{2}{q-1}}}{(x_i, \dots v_j)^{\frac{2}{q-1}}}\right] - 1 \qquad (۳۵-۲)$$

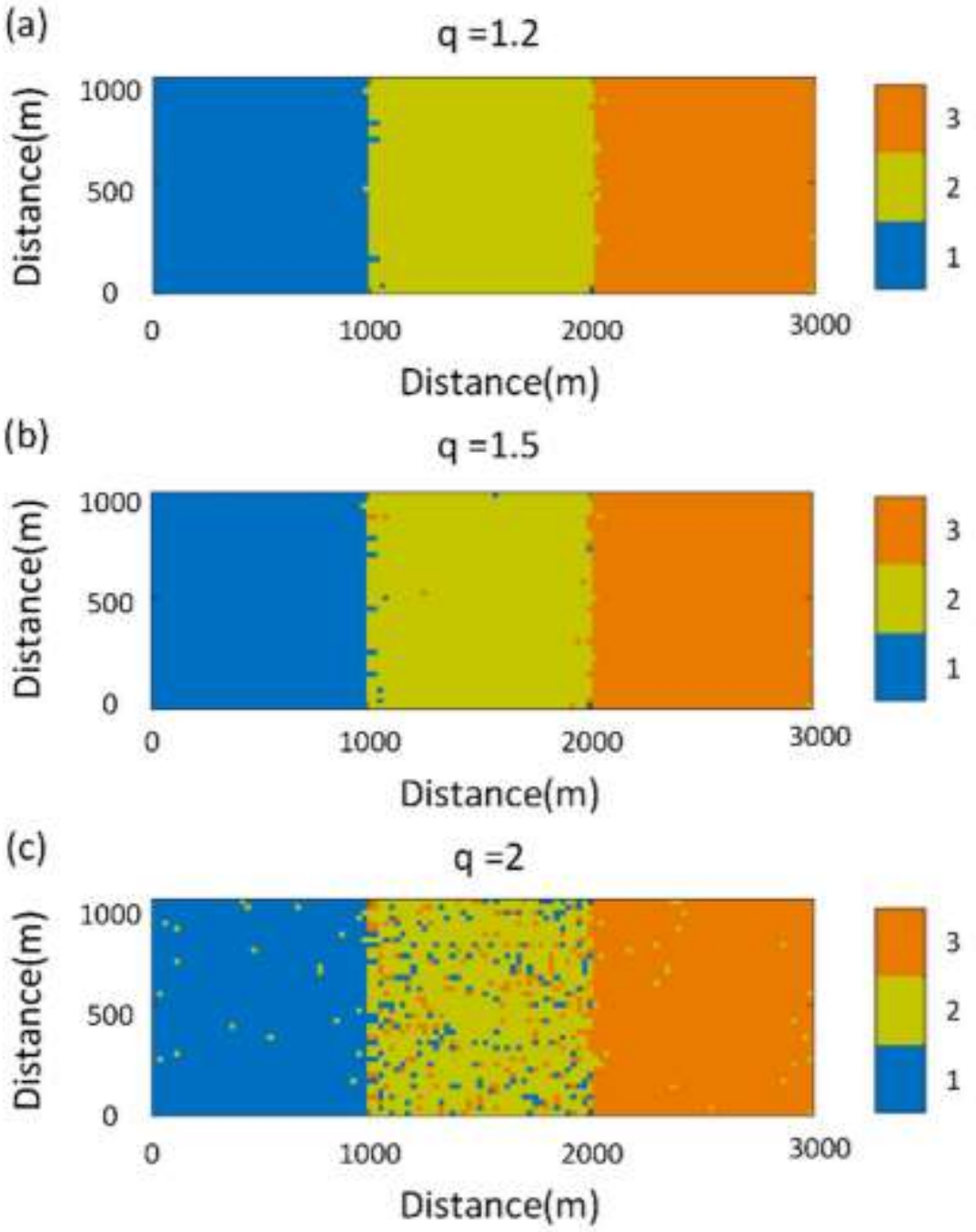
$$v_k = \frac{\sum_{i=1}^{N} (u_{ik})^q x_i}{\sum_{i=1}^{N} (u_{ik})^q} \qquad (۳۶-۲)$$

تابع عضویت u_{ik} و مرکز v_k به طور متناوب به روزرسانی می‌شود با استفاده از رابطه ۲ـ۴ و ۲ـ ۵ از مرکزهای اولیه دلخواه تا زمانی که فرآیند تکرار همگرا شود. روش بهینه سازی در واقع اجرای الگوریتم حداکثرسازی انتظار است. در فرآیند طبقه بندی نهایی نقاط داده به خوشه ای با بیشترین عضویت اختصاص داده می‌شود. در شکل به شماره ۲-۷ الگوریتم روش FCM را نشان می‌دهد.

در شکل شماره ۲-۸ مقادیر مختلفی از از فاکتور فازی را در ۳ مقدار متفاوت نشان می‌دهد . (a) فاکتور فازی q=۱/۲ است. (b) q=۱/۵ (c) q =۲ است. همانطور که ملاحظه می‌شود مقادیر کوچک از فاکتور فازی نتایج خوبی را به ارمغان می‌آورد.



شکل ۲-۷ الگوریتم روش **FCM** (ترن، ۲۰۱۹)



شکل ۲-۸ آزمودن فاکتورهای فازی مختلف با استفاده از مقادیر مختلف در الگوریتم **FCM** (سانگ و همکاران، ۲۰۱۷).

فصل سوم

روش‌های تجزیه‌ی مُد تجربی سیگنال

۳-۱ پیشگفتار

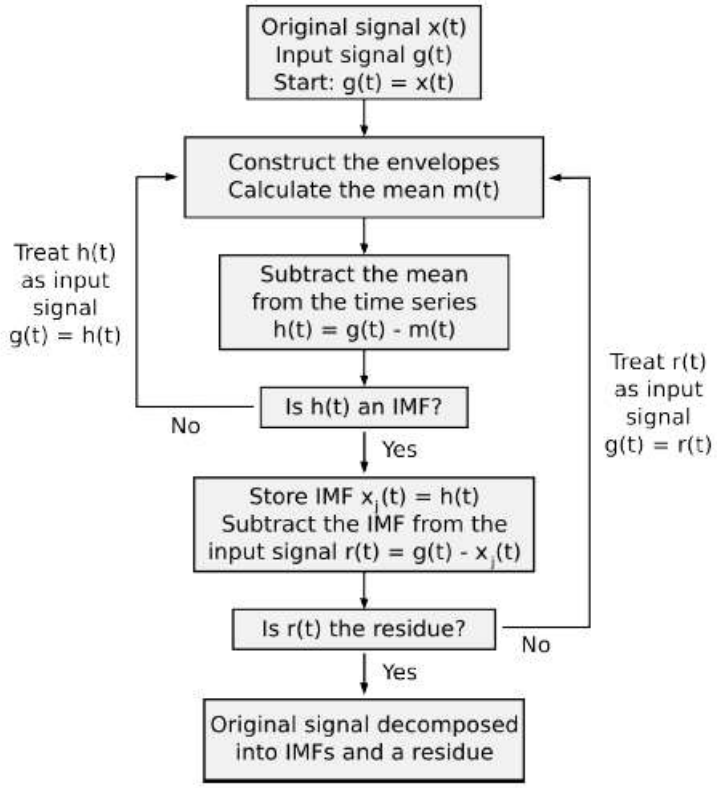
برای دستیابی به مدلی از مخزن که با واقعیت‌های زمین شناسی زیرسطحی منطقه‌ای که مخزن در آنجا واقع شده است، ابتدا باید تصویرسازی دقیق از ناهمگنی‌های آن قسمت و همچنان دانش کافی نسبت به خصوصیات سنگی آن منطقه داشته باشیم. برای مدل‌سازی از مخزن به داده‌هایی که منطقه‌ی وسیعی را بتوانند پوشش دهند نیاز داریم مانند چاه‌های حفاری شده، مغزه‌هایی که از درون چاه نمونه‌گیری شده‌اند و همچنین دسترسی به داده‌هایی که مربوط به تولید می‌باشند. اما در واقعیت معمولا تعداد چاه کافی برای مناطق وسیع که بتوان برای مدل‌سازی و شناسایی تغییرات جانبی مخزن بر روی آنها اتکا کرد به اندازه‌ی کافی وجود ندارد. بنابراین در غیاب تعداد چاه کافی، اهمیت دسترسی به داده‌های لرزه‌ای دوچندان می‌شود. این اهمیت از آن جهت است که داده‌های لرزه‌ای هرگونه تغییرات جانبی در مخزن را به خوبی نمایش می‌دهند زیرا این تغییرات شامل تغییرات در لیتولوژی، محتوای سیال و تخلل می‌شود که این تغییرات در پارامترهای لرزه‌ای قابل شناسایی است. اما این پارامترهای لرزه‌ای آلوده به نوفه هستندکه همین نوفه باعث اثرگذاری و ایجاد چالش در نتیجه‌گیری و تفسیر نهایی رخساره‌های لرزه‌ای می‌شود زیرا اغلب وجود این نوفه باعث تحلیل رخساره‌ای می‌شود که بستگی کافی با واقعیت زمین‌شناسی منطقه نداشته باشد. اهمیت وجود نوفه باعث شد که روش‌هایی برای تجزیه سیگنال به مولفه‌های سازنده‌ی آن و بازسازی سیگنال جدید و حذف نوفه‌ی آن دراین بازسازی بوجود بیاید. روش‌هایی مانند تبدیل فوریه، تبدیل موجک و تجزیه‌ی مُد تجربی را می‌توان به عنوان روش‌هایی برای تبدیل تجزیه سیگنال نام برد که برای تحلیل داده‌هایی که غیرثابت و غیرخطی هستند به کار می‌روند که در ذیل توضیحی کوتاه برای این روش‌ها داده می‌شود. روش تبدیل فوریه روشی است که درآن یک تابع ورودی را دردامنه‌ی زمانی می‌گیرد وآن را به تابعی جدید در دامنه‌ی فرکانس تبدیل می‌کند. به عبارتی می‌توان تابع اصلی را دامنه زمانی داده شده تصور کرد و تبدیل فوریه، تابع فرکانس داده شده‌ی دامنه است. تبدیل موجک روشی برای آنالیز سیگنال با طیف فرکانسی دینامیک است که هم در حوزه‌ی زمان و هم در حوزه‌ی فرکانس دارای رزولوشن بالایی است. روش تجزیه‌ی مُد تجربی روشی دیگر برای تجزیه‌ی سیگنال‌های لرزه‌ای پیچیده به یک مجموعه از توابع ذاتی در حوزه‌ی زمان است. سیگنال لرزه‌ای مُدهای

مختلفی دارد و مُدهای مختلف اطلاعات زمین شناسی مختلفی می‌تواند درونش پنهان باشد. به همین علت چون در این پایان نامه داده‌های لرزه‌ای قبل از اینکه تحلیل رخساره بر روی آنها صورت بپذیرد ما یک مروری بر روش‌های تجزیه سیگنال می‌کنیم.

۳-۲ تجزیه مد تجربی(EMD)

یک تکنیک موثر برای تجزیه و تحلیل سیگنال غیر ثابت و غیرخطی است (هوانگ و همکاران، ۱۹۹۸). EMD یک فرآیند کاملا داده محور است، سیگنال آنالیز شده را به دامنه‌ی ابتدایی هارمونیک‌های مدولار فرکانس تجزیه می کندکه توابع مد ذاتی (IMF) نامیده می‌شود.

توابع مُد ذاتی مبتنی بر استخراج مستقیم از انرژی مرتبط مقیاس‌های ذاتی مختلف است. به این معنی که هر IMF محتوای فرکانسی متفاوتی دارند . اولین IMF با بالاترین فرکانس هارمونیک در سیگنال مرتبط است که با کاهش محتوای فرکانس در توابع مُد ذاتی بعدی کاهش می‌یابد. به لحاظ تجزیه و تحلیل سیگنال‌های لرزه‌ای، چنین ویژگی بالقوه‌ای اطلاعات مختلف ساختاری و چینه‌شناسی را برجسته می‌کند. در صورت نیاز به بازسازی سیگنال می‌توان آن را با جمع کردن توابع مُد ذاتی به علاوه‌ی یک روند یکنواخت نهایی (final monotronic) انجام داد. با توجه به ماهیت محلی روش EMD می تواند نوساناتی با دامنه‌ی متفاوت در یک مُد یا با وجود محتوای فرکانس مشابه درمُدهای مختلف ایجاد کند. چنین ویژگی به عنوان اختلاط مُد (mode mixing) شناخته شده است و می‌تواند به عنوان محدودیتی از این روش آن را درنظر گرفت. برای غلبه بر این مشکل برخی از نسخه‌های EMD که با نوفه همراه هستند ارائه شده است.



شکل ۳-۱ نمودار جریان الگوریتم **EMD** (زیلر و همکاران، ۲۰۱۰)

EMD به صورت بازگشتی IMF ها را از نوسانی‌ترین تا روند یکنواخت نهایی استخراج می‌کند. طرح تجزیه مبتنی بر شناسایی بیشینه و کمینه‌ی محلی سیگنال تجزیه و تحلیل شده است که در آن یک spline برای تعریف پاکت‌های بالا و پایین به ترتیب متناسب شده است. سپس پاکت میانگین از سیگنال اولیه کم می‌شود و این فرآیند تا زمانی که پاکت میانگین به اندازه کافی به صفردرکل سری‌های زمانی نزدیک شود تکرار می‌شود. این روش غربال کردن نامیده می‌شود و اولین IMF را تعریف می‌کند. سپس اولین IMF از سیگنال اصلی کسرمی‌شود وهمان فرآیند غربالگری مشابه بر روی سیگنال باقیمانده برای تعریف IMF بعدی اعمال می‌شود. معیارهای توقف زمانی حاصل می‌شود که IMF استخراج شده دامنه‌ی کوچکی داشته باشد یا یکنواخت شود (هوانگ و همکاران، ۱۹۹۸). همانطور که قبلاً اشاره شد EMD همچنان از محدودیت‌های خاص خود نیز رنج می برد که شامل اختلاط مُدها (هر IMF شامل مقیاس های مختلف است) و تقسیم مُدها (گسترش یک مقیاس برروی IMF های مختلف) است (مانتیک و همکاران، ۲۰۰۹). مراحل روش تجزیه

مُد تجربی به این شکل است که اگر سیگنال اصلی را x(n) در نظر بگیریم EMD آن را به تعدادی IMF تجزیه می‌کند که از رابطه زیر به دست می‌آید.

$$x(n) = \sum_{i=1}^{n} c_i(x) + r_n(x) \qquad (۱-۳)$$

$r_n(x)$ مقدار باقیمانده بعد از n تعداد توابع مُد ذاتی و $c_i(x)$ تابع مُد ذاتی است. میدانیم که در یک زمان داده‌ی ما نیز می‌تواند دارای چند IMF باشد. دو شرطی که باید برقرار باشد تا یک مولفه به عنوان IMF در نظر گرفته شود به قرار زیر است. شرط اول این است که نقاطی که برابر صفر هستند و مجموع فرین‌ها باید برابر باشند یا حداکثر اختلاف کمی داشته باشند. شرط دوم آن است که مقدار میانگین پوش برازش داده شده بر نقاط حداقل و حداکثر محلی باید برابر با صفر باشد. باید در نظر داشته باشید که توابع مُد ذاتی دارای بسامدهایی متفاوت با دامنه‌هایی متفاوت هستند. پروسه‌ی تجزیه مُد تجربی یا غربالگری به شرح زیر است. ابتدا نقاط حداقل و حداکثر محلی سیگنال $x(n)$ را تعیین کنید. سپس با استفاده از برازش نقاط حداقل و حداکثری محلی با استفاده از روش cube spiline پوش بالا و پایین سیگنال را محاسبه و آن را $m_1(n)$ نام‌گذاری کنید. در ادامه اختلاف بین میانگین پوش بالا وپایین را از طریق رابطه زیر بدست بیاورید. اگر $h_1(n)$ شرایطی را که باید برای تابع IMF داشته باشد را داشت، آن را IMF اول که با $Imf_1(n)$ نشان می‌دهیم در نظر می‌گیریم و محاسباتمان را به مرحله‌ی بعدی الگوریتم انتقال می‌دهیم. اگر $h_1(n)$ شرایط لازم و مربوطه را نداشت آنگاه مراحل اول تا چهارم را با این تفاوت که الگوریتم به جای $x_1(n)$ که همان سیگنال اولیه بود روی $h_1(n)$ اعمال می‌شود.

$$h_1(n) = x(n) - m_1(n) \qquad (۲-۳)$$

از طریق رابطه بعدی که در زیر می‌آید نیز میتوان میزانی که باقی‌مانده است را محاسبه کرد.

$$r_1(n) = x(n) - Imf_1(n) \qquad (۳-۳)$$

در آخر نیز اگر مقدار باقی‌مانده حداقل ۲ فرین داشته باشد مرحله‌های ۱ تا ۵ را تکرار کنید ولی اگر باقی‌مانده حداقل ۲ فرین نداشته باشد الگوریتم متوقف و آخرین باقی‌مانده همان باقی‌مانده‌ی سیگنال در نظر گرفته می‌شود.

۳-۳ تجزیه مُد تجربی مجموع(EEMD)

وو و هوانگ (۲۰۰۹) EMD مجموع یا EEMD را معرفی‌کردند. EEMD اساسا EMD است که با تثبیت نوفه ترکیب شده است که تجزیه را روی یک گروه از نسخه‌های نوفه ای از سیگنال اصلی انجام می‌دهد. سپس به طور همزمان به عنوان میانگین تمام IMF های طرف معادله را تخمین می‌زند. اگر چه EEMD به جداسازی بهتر مُدها کمک می‌کند، تعداد متفاوتی از مُدها برای تحقق بخشیدن متفاوت (درک بهتر) به سیگنال نوفه‌ای ایجاد می‌شوند و سیگنال بازسازی شده نمی‌تواند به شکل کاملی سیگنال اصلی را باز تولید کند. همانطور که ذکر شد نقص عمده‌ی EMD اثر اختلاط مُد است.این پدیده زمانی رخ می‌دهد که نوسانات با مقیاس‌های زمانی مختلف دریک IMF حفظ می‌شوند یا نوسانات با مقیاس‌های زمانی یکسان درتوابع مُد ذاتی مختلف غربال می‌شوند. EEMD شامل اضافه کردن سری‌های مختلف نوفه‌ی سفید به سیگنال درچندین آزمایش است. از آنجایی که نوفه‌ی اضافه شده در هر آزمایش متفاوت است، توابع مُد ذاتی به‌دست آمده هیچ ارتباطی با توابع مُد ذاتی مربوطه از یک آزمایش به آزمایش دیگر را نشان نمی‌دهند. اگر تعداد آزمایش‌ها کافی باشد، نوفه‌ی اضافه شده را می‌توان با میانگین‌گیری مجموعه‌ای از توابع مُد ذاتی به دست آمده‌ی مربوط به آزمایش‌های مختلف حذف کرد.

مراحل خاص الگوریتم EEMD به شرح زیر است.

مرحله‌ی۱: در آزمایش n اُم، یک سری زمانی جدید با اضافه کردن سری زمانی نوفه‌ی سفید U_n (t) به سیگنال داده شده‌ی X(T) برای $n = ۱,۲, \ldots, N$ با تعداد مجموع N تولید می شود.

$$Y_n(t) = X(t) + U_N(\mathrm{T}) \qquad (۴-۳)$$

مرحله‌ی۲: بر اساس EMD اصلی، سیگنال آلوده به نوفه $Y_n(T)$ در یک مجموعه از IMF ها و یک باقی‌مانده تجزیه می‌شود.

$$Y_n(T) = \sum_{M=1}^{M-1} IME_m^{(n)}(t) + r_M^{(n)}(t) \qquad (۵-۳)$$

در حالیکه M-1 تعداد کل IMF های نتیجه شده از هر تجزیه‌ی $Y_N(t)$ است، $IMF_m^{(n)}$ ، m اُمین IMF است و $r_m^{(n)}$ باقیمانده‌ی حاصل از N اُمین آزمایش است. به منظور تعداد برابر از IMF ها در هر تجزیه، ثابت غربالگری عدد ۱۰ در نظر گرفته شده است.

مرحله ۳ : مرحله ۱ و ۲ برای n تعداد آزمایش تکرار شده و در هر آزمایش سری‌های نوفه‌ی سفید متفاوت $U_n(t)$ به سیگنال اصلی اضافه می‌شود.

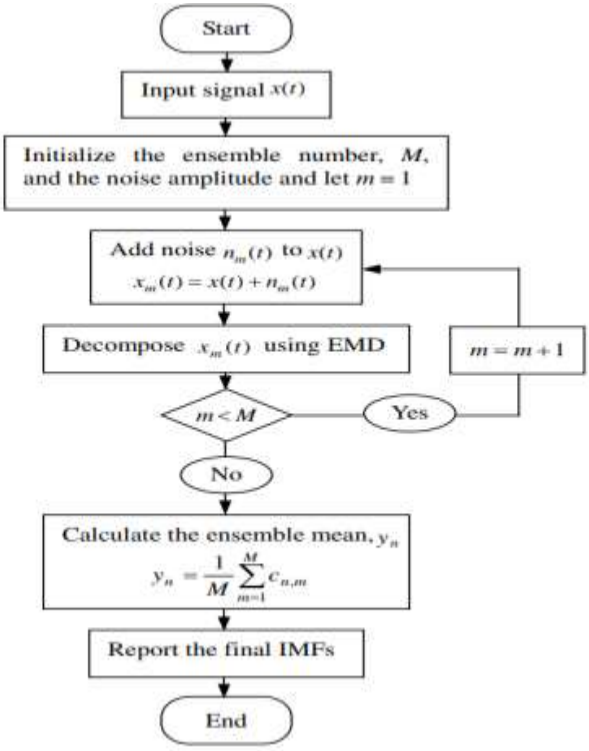
مرحله ۴ : IMF نهایی ازEEMD (IMF_m^{ave}) با میانگین‌گیری از کل IMF m مربوط به N تعداد آزمایش بدست آمد.

$$IMF_m^{ave}(t) = \frac{1}{N}\sum_{n=1}^{(N)} IMF_m^n(t) \qquad (۶-۳)$$

نتایج به دست آمده از EEMD بستگی به تعداد مجموع (N) و دامنه‌ی نوفه‌ی اضافه شده است که رابطه زیر باید حاصل شود:

$$\varepsilon = \frac{A}{\sqrt{N}} \qquad (۷-۳)$$

ε که انحراف استاندارد نهایی خطای محاسباتی است که به عنوان تفاوت بین سیگنال اصلی و مجموع IMF های حاصل از EEMD محاسبه می‌شود.



شکل ۳-۲ نمودار جریان الگوریتم **EEMD** (لی و زو، ۲۰۰۹)

۳-۴ روش تجزیه‌ی مد تجربی کامل(CEEMD)

همانطورکه هدف اصلی ما ارزیابی و مقایسه‌ی CEEMD با نسخه‌های بهبود یافته‌ی آن است. ما بیشتر می‌خواهیم در مورد جزئیات و ویژگی‌های EMD و EEMD بحث کنیم. برتری CEEMD بر EMD و EEMD برای تجزیه و تحلیل داده‌ی لرزه‌ای توسط هان و واندربان در سال ۲۰۱۳ نشان داده شده است. رابطه‌ی زیر را در نظر بگیرید.

$$x^i = x + \omega^i \qquad (۸-۳)$$

x^i را به عنوان نسخه‌ی نوفه‌ای از x تحت تاثیر i اُمین تحقق (realizations) نوفه‌ی گوسی سفید ω^i و $E_K(\cdot)$ به عنوان عملگری که مُد k ام رابه وسیله‌ی EMD تولید می‌کند. بنابراین اولین مُد EEMD به شکل زیر تخمین زده می‌شود.

$$IMF_1 = \frac{1}{I}\sum_{i=1}^{I} E_1(x + \varepsilon\omega_i) \quad (۹-۳)$$

در حالیکه I تعداد تحقق‌ها (realizations) و ε درصد ثابت از نوفه‌ی سفید گاوسی تزریق شده است. سپس اولین باقی مانده‌ی r_i به شکل زیر محاسبه می‌شود.

$$r_1 = x - IMF_1 \quad (۱۰-۳)$$

حالا r_1 و تحقق نوفه‌های مختلف (different noise realizations) به عنوان مجموعه‌ای از سیگنال‌های جدید تلقی می‌شوند و اولین مُد EMD از هرنوفه (r_1) استخراج می‌شود. سپس دومین مُد CEEMD به عنوان میانگین این مُدها تخمین زده می‌شود.

$$IMF_2 = \frac{1}{I}\sum_{i=1}^{I} E_1\left(r_1 + \varepsilon E_1\left(\omega^{(i)}\right)\right) \quad (۱۱-۳)$$

باقی‌مانده‌ی دوم به صورت $r_2 = r_1 - IMF_2$ محاسبه می‌شود. به این ترتیب ، IMFها بر خلاف EEMD به شکل متوالی استخراج می‌شوند. این رویه تا رسیدن به معیار توقف ادامه پیدا می‌کند. معمولا زمانی‌که آخرین باقی مانده R بیشتر از دو فرین نداشته باشد به عنوان یک روند یکنواخت می‌تواند در نظر گرفته شود. بنابراین، بازسازی سیگنال بوسیله K _ IMF ها و باقیمانده نهایی R انجام می‌شود.

$$x = \sum_{k=1}^{k} IMF_K + \mathrm{R} \quad (۱۲-۳)$$

۳-۵ تجزیه مُد متغیر(VMD)

این روش یک روش تجزیه مدرن است که جایگزین روش EMD شده است. روش تجزیه متغیر توسط دراگومیرتسکی و زولو پیشنهاد و مورد توسعه قرار گرفت. روش VMD به شکلی خاص به این موضوع بستگی دارد که تعداد مُدها از قبل تعیین شده باشد. VMD به حل مشکل اختلاط مُد در نتیجه‌ی تجزیه با تغییر در رویکرد پروسه‌ی غربالگری با استفاده از روش رویکرد ضریب کمک می‌کند. همچنین بنیانگذار VMD به اهمیت تعیین تعداد مُد برای جلوگیری از مشکل کم تجزیه شدن یا بیش تجزیه شدن اشاره کرد. یکی از دلایل برتری روش تجزیه مُد متغیر به روش EMD در این است که در VMD توابع مُد ذاتی را در یک زمان

یکسان، محاسبه می‌کند. بایستی دانست که هر یک از توابع مُد ذاتی بیشتر در اطراف مرکز فشرده و متمرکز هستند. VMD روش خوبی برای نمونه برداری و نوفه‌ی سیگنال است. ارتباط تنگاتنگ با فیلتر وینر باعث می‌شود این روش توانایی بهینه برای مقابله با نوفه را در سیگنال داشته باشد. VMD برتری خودش را در طیف وسیعی از کاربردها مثل تشخیص ماشینی، سیگنال گفتاری، پیش‌بینی نفت خام، پیش‌بینی سرعت باد و پردازش تصویر نشان داده است. تئوری فرآیند تجزیه‌ی روش VMD در زیر شرح داده شده است. مرحله اول:

(۱۳-۳) مقدار دهی کنید. $n \leftarrow ۰$ و $\lambda^{۱}$ و $\{\hat{u}_k^{۱}\}$ و $\{\widehat{U}_k^{۱}\}$

مرحله دوم: مقدار u_k، ω_k و λ را با توجه به رابطه زیر به‌روزرسانی کنید.

$$u\hat{}_k^{n+۱} \leftarrow \frac{f\hat{}(\omega) - \sum_{i<k} u\hat{}_i^{n+۱}(\omega) - + \frac{\lambda^n(\omega)}{۲}}{۱ + ۲\square(\omega - \omega_k^n)^{۲}} \quad (۱۴-۳)$$

$$\omega_k^{n+۱} \leftarrow \frac{\int_{۰}^{+\infty} \omega \left|\hat{u}_k^{n+۱}(\omega)\right|^{۲} d\omega}{\int_{۰}^{+\infty} \omega \left|\hat{u}_k^{n+۱}(\omega)\right|^{۲} d\omega} \quad (۱۵-۳)$$

$$\hat{\lambda}^{n+۱}(\omega) \leftarrow \hat{\lambda}^n(\omega) + \tau \left[\hat{f}(\omega) - \sum_{k} \hat{u}_k^{n+۱}(\omega)\right] \quad (۱۶-۳)$$

مرحله ۳. فرآیند تکراری را از مرحله ۲ تکرار کنید تا زمانی که تابع بر اساس معیارهای همگرایی که شرط زیر را برآورده می‌کند همگرا شود.

$$\sum_k \left\|\hat{u}_k^{n+۱} - \hat{u}_k^n\right\|_{۲}^{۲} / \left\|\hat{u}_k^n\right\|_{۲}^{۲} < \epsilon$$

فصل چهارم

تحلیل رخساره لرزه‌ای چند مقیاسی

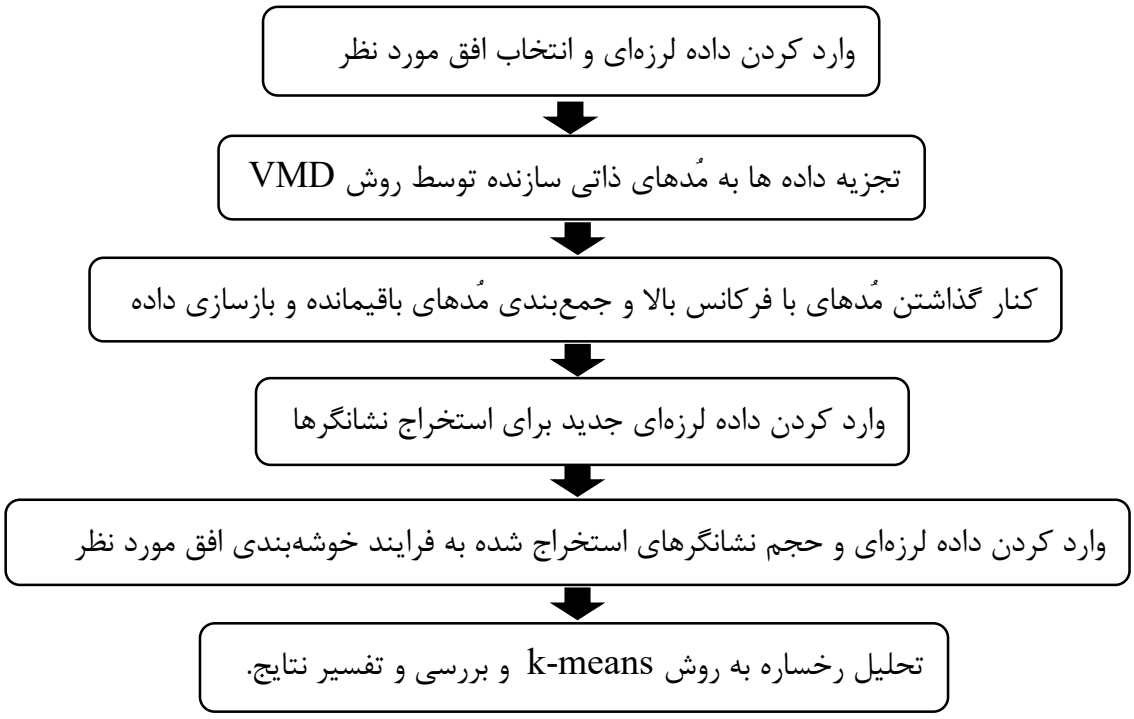
۴-۱ پیشگفتار

اگر چه نوفه در داده‌های لرزه‌ای در مراحل پردازشی فیلتر شده است اما به هر حال یک سری از نوفه‌های تصادفی ممکن است بر روی داده‌های لرزه‌ای باقی بمانند و اشکالات و ایراداتی که به علت عملیات داده برداری ممکن است وجود داشته باشد باعث می‌شود یک سری از نوسانات با فرکانس بالا بر روی داده لرزه‌ای باقی بماند. همانطور که میدانیم همیشه نوفه بخش جدایی ناپذیر از داده‌های لرزه‌ای است و ممکن است که این نوفه منشاء زمین‌شناسی نداشته باشد و این مولفه‌ها ممکن است پس از پردازش هم باقی بماند، به هر حال در مرحله بعد از پردازش که مرحله تفسیر است باید مقاطع لرزه‌ای را برای اهداف مورد نظر تفسیر شود زیرا وجود این نوفه‌ها باعث تفسیر اشتباه می‌شود. یکی از کارهایی که عملیات تفسیری که بر روی داده‌های لرزه‌ای انجام می‌شود بحث تحلیل رخساره لرزه‌ای است که در این بخش باید رخساره‌های لرزه‌ای را بشناسیم و بتوانیم خصوصیات لرزه‌ای اهدافی مانند تخلخل و اشباع را از روی رخساره‌های لرزه‌ای تشخیص بدهیم که معمولاً در واقع افق‌های مخزنی را در رخساره‌های لرزه‌ای مورد نظر قرار می‌دهند که با اهداف اکتشافی این کار صورت می‌گیرد. اگر مشکلی که اشاره شد یعنی وجود نوفه موجود باشد ممکن است که طرح‌های لرزه‌ای را نتوانیم به‌خوبی خوشه‌بندی کنیم. روشی که در این پایان نامه انجام شده است این است که ابتدا داده‌های لرزه‌ای توسط یکی از روش‌های تجزیه مُد تجربی تحت عنوان تجزیه مُد متغیر به مُدهای سازنده آن تجزیه شده است. مُدهای اول با فرکانس بالا هستند و به عبارتی بیشترین انرژی بر روی این مُدها است اگرچه مشکل اختلاط مُد وجود دارد و شاید نتوانیم این مُدها را به شکل کامل از هم تفکیک کنیم اما تلاش خود را می‌کنیم تا از این طریق آن دسته از مُدهایی را که دربرگیرنده انرژی نوفه هستند را حتی الامکان کنار گذاشته شود و داده بازسازی شود، سپس این داده بازسازی شده را خوشه بندی کرده و نتایج را بررسی کنیم.

۴-۲ الگوریتم روش

در ابتدا داده‌ی لرزه‌ای را وارد می‌کنیم و افق مورد نظر را انتخاب می‌کنیم. سپس با استفاده از روش تجزیه‌ی مُد متغیر داده‌های لرزه‌ای مربوط به افق مورد نظر را به مُدهای ذاتی سازنده‌ی آن(IMF) تجزیه می‌کنیم. در مرحله بعدی مُدهایی که بیشترین انرژی بر روی آن‌ها هست یا به عبارت دیگر فرکانس بالایی دارند و نوفه‌ای هستند را شناسایی کرده و از مرحله بازسازی سیگنال کنار گذاشته و بازسازی داده را با جمع کردن مُدهای باقیمانده انجام می‌دهیم. سپس این داده‌ی بازسازی شده که از کنارگذاشتن مُدهایی با فرکانس بالا و جمع مُدهای باقیمانده حاصل شده است را به عنوان داده‌ی لرزه‌ای جدید برای استخراج نشانگرهای مورد نیاز به کار می‌بریم. این داده لرزه‌ای جدید و حجم نشانگرهای استخراج شده را به عنوان ورودی در فرآیند خوشه‌بندی برای افق موردنظر به کار می‌بریم و در انتها تحلیل رخساره را به روش k-means انجام داده و نتایج حاصل را مورد تفسیر و بررسی قرار می‌دهیم.

مراحل الگوریتم به شرح زیر است:



۴-۳ خوشه بندی داده مصنوعی

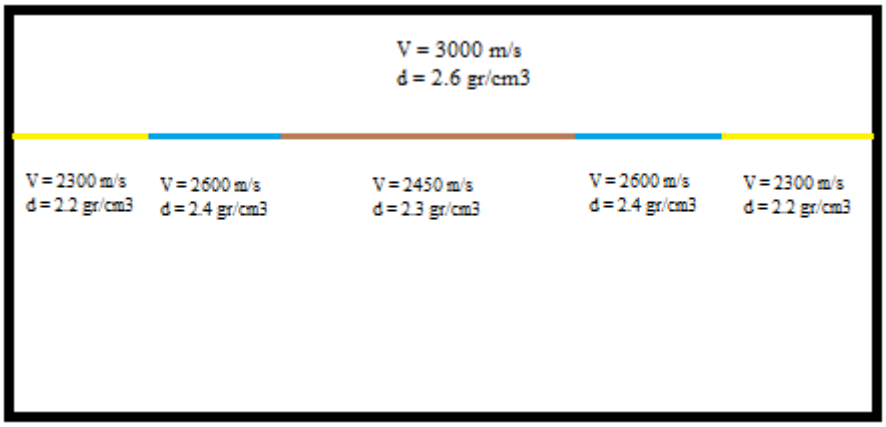
هدف این پایان نامه تحلیل رخساره چندمقیاسی بوده است تا نقش نوفه‌ی باقیمانده بررسی شود. به همین منظور در آزمایش مصنوعی این موضوع را به این شکل بررسی کرده‌ایم که میزانی نوفه‌ی تصادفی به داده اضافه کرده اما با توجه به این موضوع که دامنه و فاز نوفه‌ی تصادفی به حالت تصادفی تغییر می‌کند این موضوع دقیقاً نقشش را در تغییر رخساره می‌گذارد و ما این تغییر رخساره و نقش نوفه‌ی باقیمانده را بدین حالت مدل کرده‌ایم و هدف مورد نظر ما با استفاده از نوفه‌ی تصادفی میسر می شود. اگر چه حذف نوفه بر روی داده‌های لرزه‌ای انجام می شود ولی همچنان مقادیری از نوفه باقیمانده و ناپیوستگی‌هایی که می‌تواند منشأهای دیگری هم داشته باشد باقی می‌ماند. این تغییرات شدید رخساره را ما با استفاده از نوفه تصادفی مدل‌سازی کردیم که باعث ایجاد تغییرات دامنه و فاز بر داده‌های لرزه‌ای می‌شود.

برای آزمودن الگوریتم مورد نظر پایان نامه داده لرزه‌ای مصنوعی تولید شده است. این داده مصنوعی به شکلی است که یک حجم داده لرزه‌ای مصنوعی شامل ۲۰۰ inline و ۳۰۰ crossline و ۲۵۰ نمونه زمانی تولید شده است. یک افق لرزه‌ای مطابق با مشخصات نمایش داده شده در شکل ۴ـ۱ در نظر گرفته شده است که شامل سه رخساره متفاوت است که مشخصات سرعت و چگالی آن در شکل ۴ـ۱ نشان داده شده است. برای ساخت داده لرزه‌ای یک موجک ریکر با فرکانس غالب ۳۰ هرتز استفاده شده است و بازه نمونه برداری زمانی هم ۴ میلی‌ثانیه در نظر گرفته شده است. افق مورد نظر در زمان ۳۰۰ میلی ثانیه در شکل ۴ـ۲ نشان داده شده است که این شکل یک inline استخراج شده از حجم داده لرزه‌ای مصنوعی تولید شده است که افق مورد نظر در موقعیت زمانی ۳۰۰ میلی ثانیه بر روی شکل با فلش قرمز رنگ نشان داده شده است و این افق مورد تحقیق قرار گرفته است. در مورد داده لرزه‌ای مصنوعی ما فقط دامنه‌های لرزه‌ای را وارد بحث خوشه‌بندی کردیم و نتیجه خوشه بندی به روش k-means و با در نظر گرفتن سه خوشه در شکل ۴ـ۳ نشان داده شده است که مشخص می شود که این سه خوشه به چه ترتیبی با رنگ‌های مختلف نشان داده شده‌اند که هر رنگ بیانگر یک خوشه منحصر به فرد است. برای هدف مورد نظر این پایان نامه ما نیاز به اضافه کردن نوفه به داده داریم. به همین منظور در سه سطح مختلف نوفه به داده اضافه شد و نتیجه خوشه‌بندی بررسی شد.

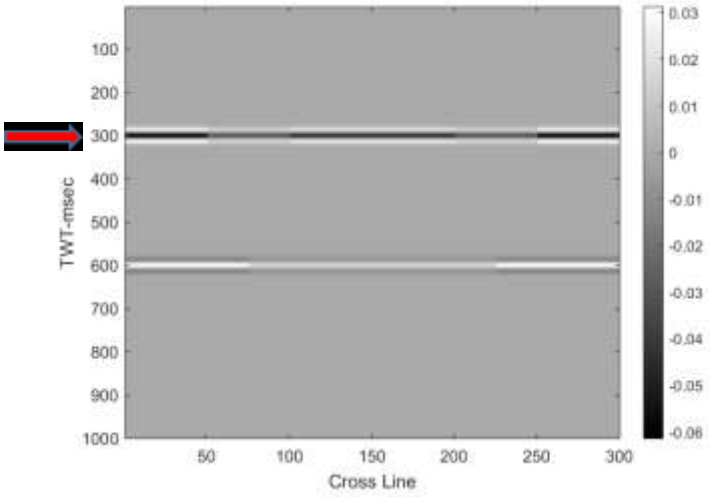
شکل ۴ـ۴ یک inline مستقل از داده لرزه‌ای است که نقطه تصادفی با نسبت سیگنال به نوفه ۲ به آن اضافه شده است را نشان می دهد. این حجم داده‌ی لرزه‌ای آلوده به نوفه وارد روند خوشه بندی برای افق مورد نظر شد که نتیجه‌ی آن در شکل ۴ـ ۵ دیده می‌شود. مجددا روش k-mean و سه خوشه در نظر گرفته شده که اثر وجود نوفه در خوشه های نارنجی رنگ خودش را نشان داده که قابل مشاهده است. بعد از این مرحله حجم داده‌های لرزه‌ای آلوده به نوفه وارد فرآیند تجزیه مُد متغیر شد و مولفه های فرکانس بالا از روی آن برداشته شد. نتایج خوشه بندی و پس از آن داده بازسازی شده وارد مرحله خوشه‌بندی شد که نتیجه آن در شکل ۴ـ۶ قابل مشاهده است که مشاهده می کنیم در مقایسه با شکل ۵ خوشه بندی با دقت بیشتری ویژگی‌های واقعی زمین‌شناسی در نظر گرفته شده را نشان می‌دهد.

در مرحله بعدی آزمایش میزان نوفه را می‌افزاییم و نسبت میزان نوفه تصادفی در قسمت بعدی افزایش داده شد و داده مصنوعی آلوده به نوفه با نسبت سیگنال به نوفه ۱ تولید شد که در شکل ۴ـ۷ یک inline این داده نشان داده می شود. نتیجه خوشه‌بندی این داده نوفه‌ای را در شکل ۴ـ۸ مشاهده می‌کنید و این داده لرزه‌ای وارد مرحله تجزیه مُد متغیر شد و مولفه‌های فرکانس بالا از روی آن برداشته و در بازسازی کنار گذاشته شد و داده لرزه‌ای که بازسازی شد وارد فرایند خوشه بندی شد که نتیجه را در شکل ۴ـ۹ مشاهده می‌کنید که در آن خوشه‌بندی با دقت بیشتری انجام شده است. میزان نوفه مجددا افزایش داده شد و یک حجم داده لرزه‌ای مصنوعی با نوفه تصادفی با نسبت سیگنال به نوفه ۰/۵ تولید شد که در شکل ۴ـ۱۰یک inline استخراجی آن دیده می‌شود و نتیجه خوشه‌بندی این داده هم در شکل ۴ـ۱۱ مشاهده می‌شود که ملاحظه می‌کنیم اثر نوفه چقدر مشهود شده است. سپس این داده لرزه‌ای نوفه مجدداً وارد مرحله تجزیه به روش مد متغیر نتیجه خوشه‌بندی داده بازسازی شده که مولفه‌های فرکانس بالای آن کنار گذاشته شده‌اند در شکل ۴ـ۱۲ دیده می‌شود که دقت خوشه‌بندی افزایش پیدا کرده است و آن خوشه‌های بیانگر رخدادهای واقعی زمین‌شناسی بهتر دیده می‌شوند. اثر مولفه‌های با فرکانس بالا که بیانگر نوفه هستند در شکل ۴ـ۱۳دیده می‌شود جایی که طیف دامنه میانگین داده لرزه‌ای بدون نوفه همراه با طیف میانگین داده لرزه‌ای آلوده به نوفه با نسبت سیگنال به نوفه ۰/۵ و همینطور طیف دامنه میانگین داده بازسازی شده به روش تجزیه مد

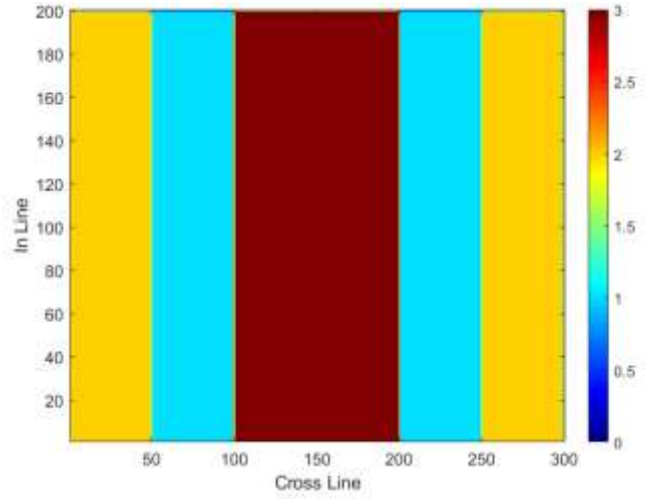
تجربی که نوفه آن را کنار گذاشته باشد در شکل الف قابل مقایسه هستند و در شکل ۴ـ۱۳ قسمت ب منحنی قرمز رنگ را می‌توانیم ببینیم که بیانگر طیف دامنه همان مولفه‌های فرکانس بالا هستند که از بازسازی سیگنال کنار گذاشته شده‌اند.



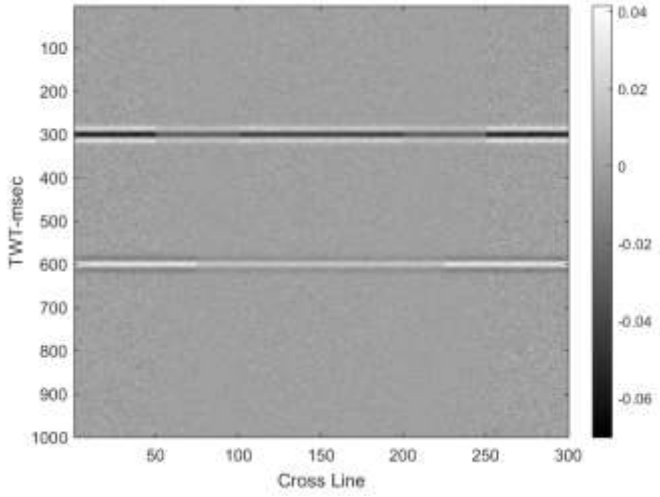
شکل ۴ـ۱. مدل رخساره برای ساخت داده مصنوعی.



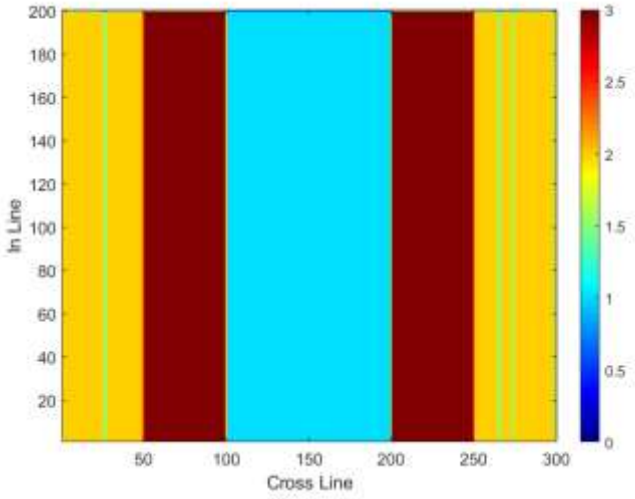
شکل ۴ـ۲. یک **Inline** از حجم داده لرزه‌ای مصنوعی. افق در موقعیت زمانی ۳۰۰ میلی ثانیه نشان داده شده با فلش قرمز مورد تحلیل قرار گرفته است.



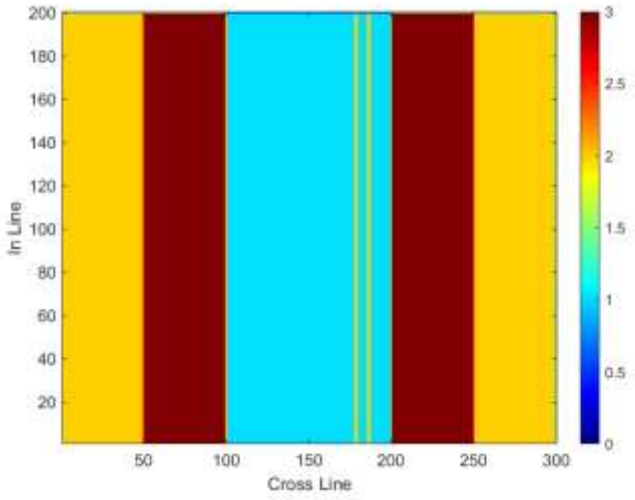
شکل ۴ـ۳. نتیجه خوشه بندی افق لرزه‌ای مورد نظر با استفاده از روش **k-means** و سه خوشه.



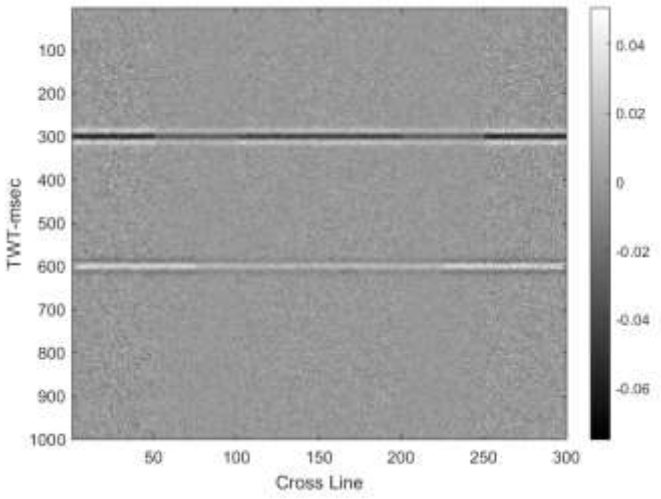
شکل ۴ـ۴. یک **Inline** از حجم داده لرزه ای مصنوعی که نوفه تصادفی با نسبت سیگنال به نوفه ۲ اضافه شده است.



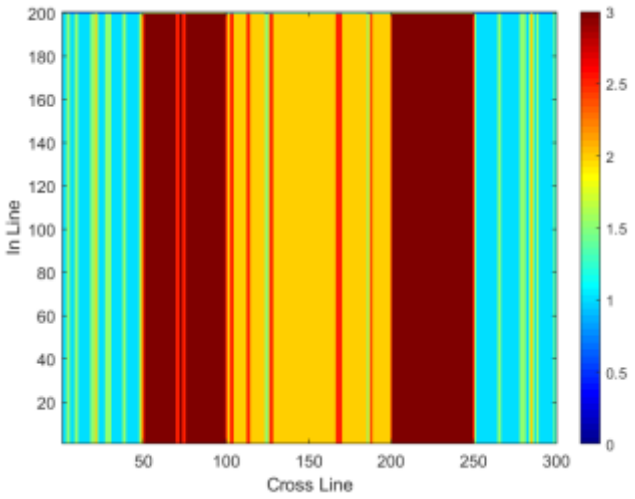
شکل ۴ـ۵. نتیجه خوشه بندی داده لرزه ای مصنوعی آلوده به نوفه با نسبت سیگنال به نوفه ۲ با استفاده از روش **k-means** و سه خوشه.



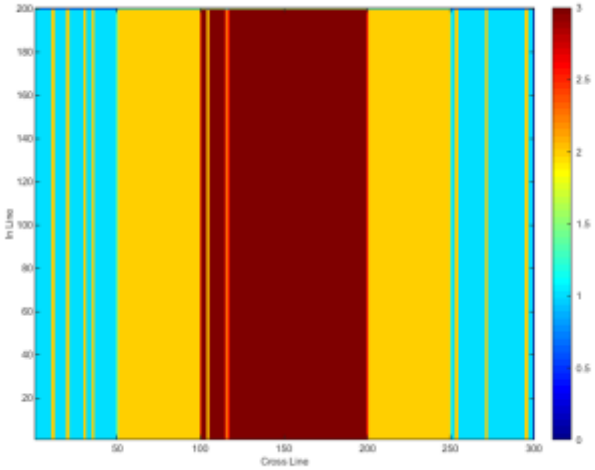
شکل ۴ـ۶ . نتیجه خوشه بندی داده لرزه ای مصنوعی (آلوده به نوفه با نسبت سیگنال به نوفه ۲) و تحلیل شده توسط روش **VMD** که مولفه‌های با فرکانس زیاد آن کنار گذاشته شده است. خوشه‌بندی به روش **k-means** و با سه خوشه انجام شده است.



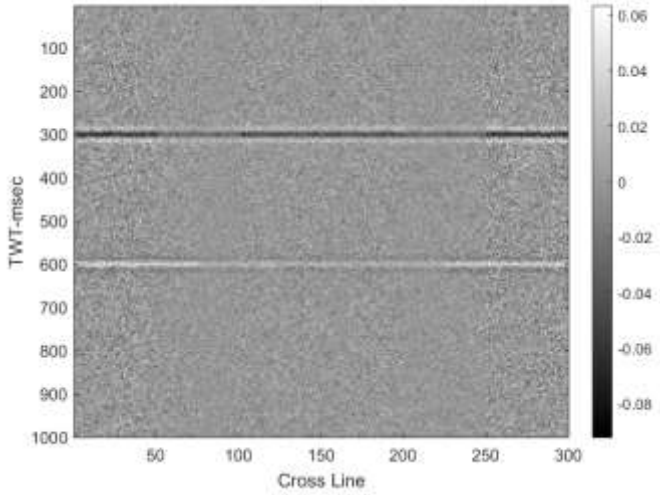
شکل ۴ـ۷. یک **Inline** از حجم داده لرزه ای مصنوعی که نوفه تصادفی با نسبت سیگنال به نوفه ۱ اضافه شده است.



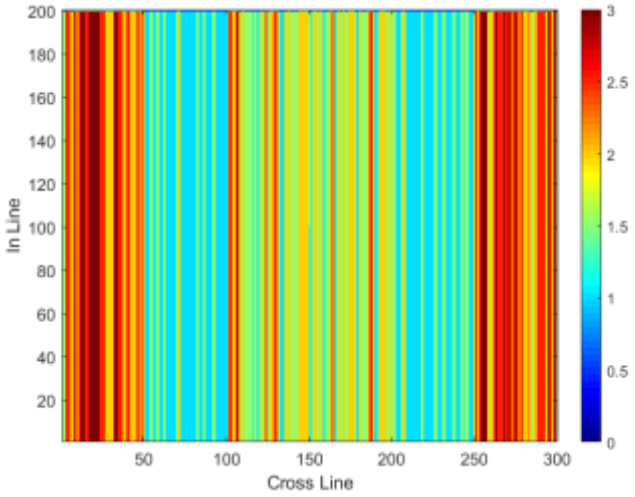
شکل ۴ـ۸ . نتیجه خوشه بندی داده لرزه ای مصنوعی آلوده به نوفه با نسبت سیگنال به نوفه ۱ با استفاده از روش**k-means** و با سه خوشه



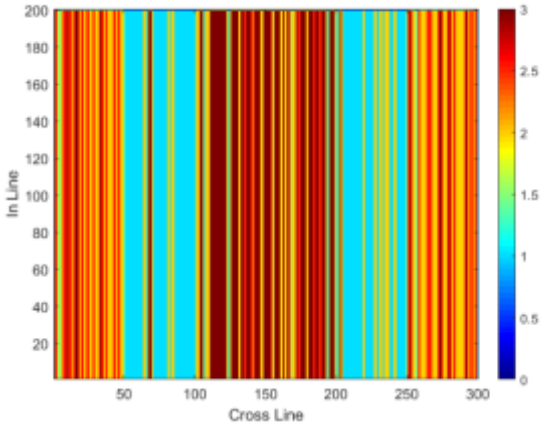
شکل ۴ـ۹. نتیجه خوشه بندی داده لرزه ای مصنوعی (آلوده به نوفه با نسبت سیگنال به نوفه ۱) و تحلیل شده توسط روش **VMD** که مولفه‌های با فرکانس زیاد آن کنار گذاشته شده است. خوشه‌بندی به روش **kmean** و با سه خوشه انجام شده است.



شکل ۴ـ۱۰. یک Inline از حجم داده لرزه ای مصنوعی که نوفه تصادفی با نسبت سیگنال به نوفه ۰/۵ اضافه شده است.

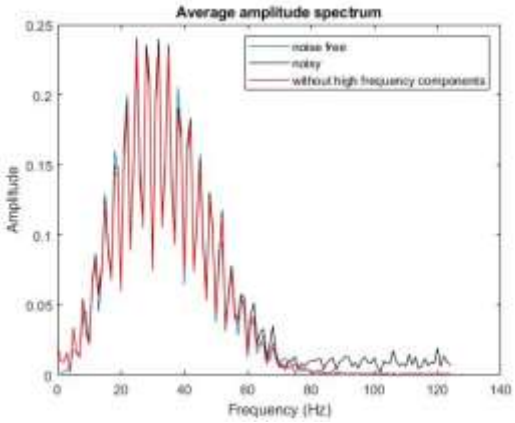


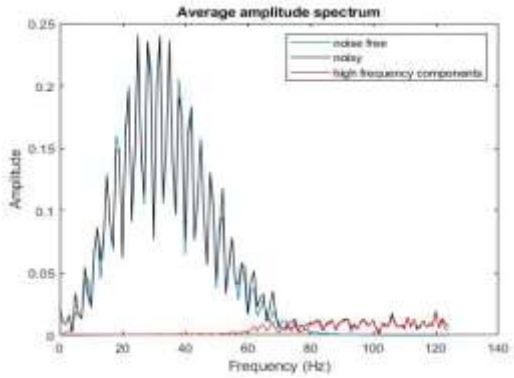
شکل ۴ـ۱۱. نتیجه خوشه بندی داده لرزه ای مصنوعی آلوده به نوفه با نسبت سیگنال به نوفه ۰/۵ با استفاده از روش **kmeans** و سه خوشه.



شکل ۴ـ۱۲. نتیجه خوشه بندی داده لرزه ای مصنوعی (آلوده به نوفه با نسبت سیگنال به نوفه ۰/۵) و تحلیل شده توسط روش **VMD** که مولفه‌های با فرکانس زیاد آن کنار گذاشته شده است. خوشه‌بندی به روش **kmeans** و با سه خوشه انجام شده است.

(الف)

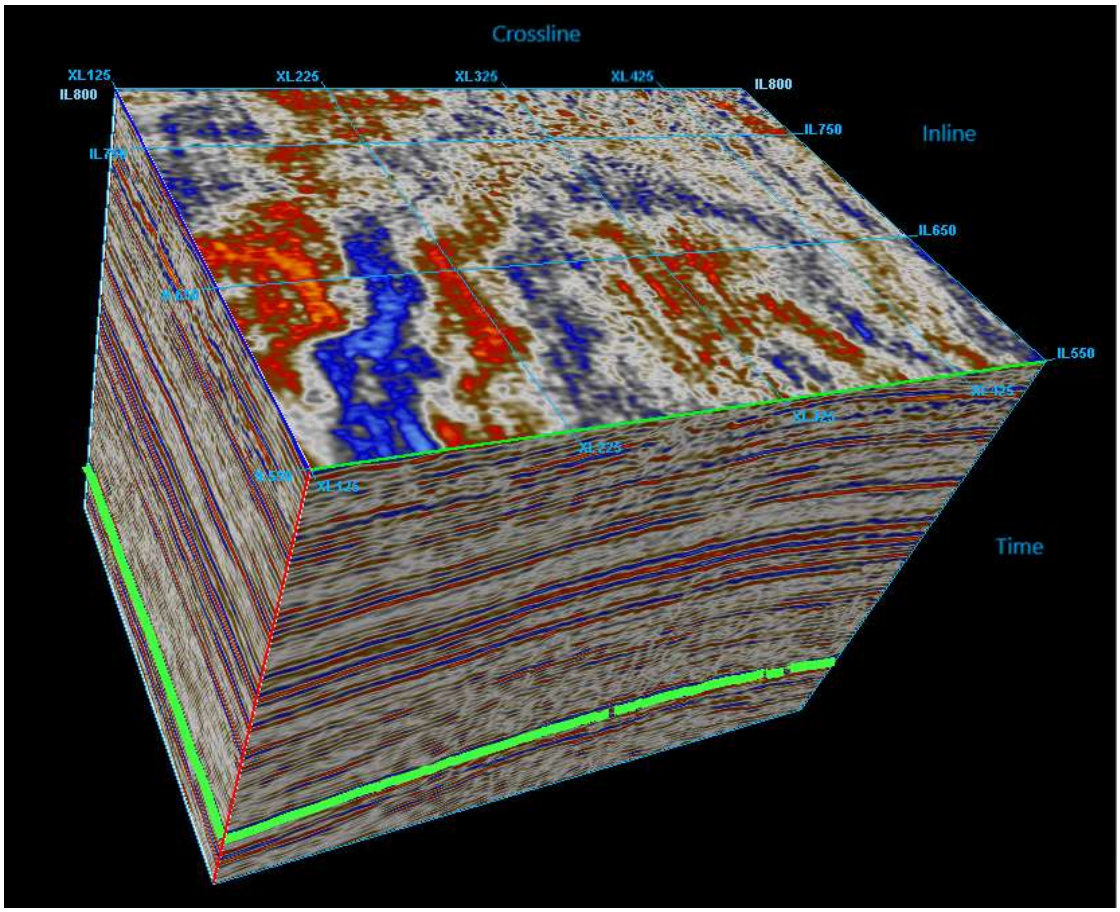




شکل ۴ـ۱۳ (الف) . طیف دامنه میانگین حجم داده لرزه ای مصنوعی بدون نوفه (محنی آبی رنگ)، آلوده به نوفه با نسبت سیگنال به نوفه ۰/۵ (منحنی سیاه رنگ) و تحلیل شده توسط روش **VMD** به طوریکه مولفه‌های با فرکانس زیاد در بازسازی داده کنار گذاشته شده است (ب) قرمز رنگ). (ب) . طیف دامنه میانگین مولفه های کنار گذاشته شده از بازسازی با رنگ قرمز نشان داده شده است.

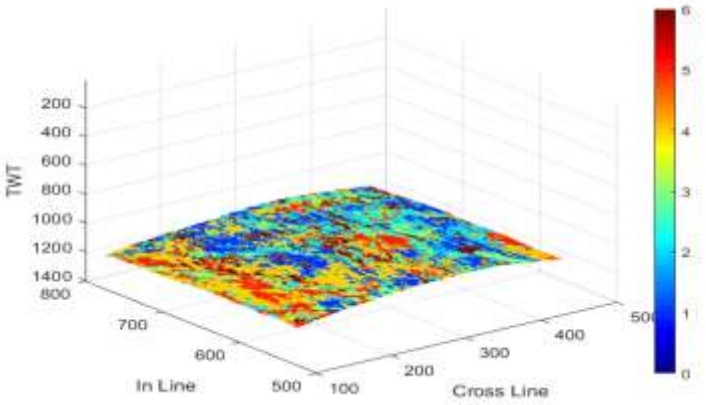
۴-۴ خوشه بندی داده واقعی

با توجه به اینکه در قسمت داده‌های مصنوعی مشاهده شد که اثر نوفه چه مقدار بر خوشه بندی تاثیر دارد و به روش تجزیه مُد متغیر می‌توان این اثر را کمتر کرد و در واقع دقت خوشه‌بندی را با بازسازی سیگنال بدون مولفه‌های فرکانس بالا افزایش داد. در این بخش روش موردنظر روی داده‌ی واقعی اعمال و نتیجه بررسی می‌شود. داده لرزه‌ای واقعی مربوط به یکی از میادین هیدروکربنی جنوب ایران است. حجم لرزه‌ای به همراه افق مورد نظر در شکل ۴ـ۱۴ نشان داده شده است.



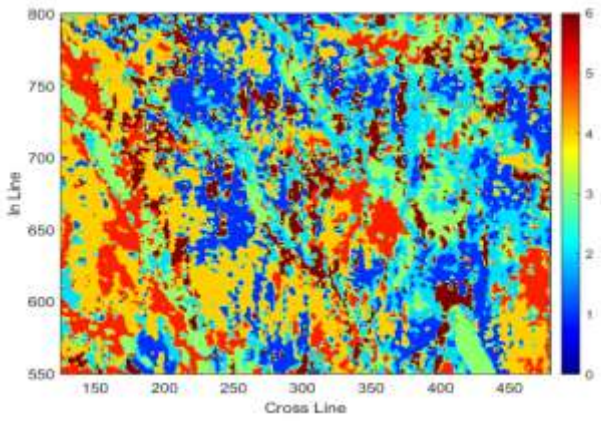
شکل ۴ـ۱۴ حجم لرزه‌ای به همراه افق مورد نظر که با خط چین سبز رنگ نمایش داده شده است

خوشه‌بندی روی افق مورد نظر انجام شد و نتیجه خوشه‌بندی در شکل ۴ـ۱۵ و ۴ـ۱۶ نشان داده می‌شود. برای انجام خوشه‌بندی در یکی از مراحل چند نشانگر لرزه‌ای را از داده‌های لرزه‌ای استخراج کردیم. این نشانگرها شامل دامنه لحظه‌ای، فرکانس لحظه‌ای و فاز لحظه‌ای می‌باشد و شکل ۴ـ۱۵ نمایش سه بُعدی نمایش خوشه‌بندی است و شکل ۴ـ۱۶ نمایش دو بُعدی خوشه بندی افق مورد نظر را نشان می‌دهد.



شکل۴ـ۱۵. نمایش سه بعدی نتیجه خوشه بندی افق مورد نظر در داده لرزه ای واقعی با روش **kmeans** و ۶ خوشه

خوشه‌بندی به روش k-means با ۶ خوشه انجام شده و همانطور که مشاهده می‌شود در نتیجه خوشه‌بندی الگوی پراکنده خوشه‌ها را مشاهده می‌کنیم که یک روند مناسبی از رخساره نشان داده نمی شود و خوشه‌بندی ضعیف به نظر می رسد و پراکندگی رنگ‌ها و مجاورت خوشه‌های مختلف در یک محدوده کوچک به چشم می‌آید. ما این داده لرزه‌ای را وارد تجزیه به روش مُد متغیر کردیم و تعدادی ازمُدهای ذاتی آن را از مرحله بازسازی کنار گذاشتیم که فرض کردیم که این مُدهای ذاتی بیانگر نوفه‌ی باقی مانده باشند و نتیجه خوشه‌بندی را در شکل‌های ۴ـ۱۷ و ۴ـ۱۸ مشاهده می‌کنید. برای بررسی قطعیت خوشه‌بندی با توجه به این که داده‌ی حقیقی از خوشه‌بندی هم به روش بدون ناظر استفاده شده است و همچنین اطلاعات چاه در اختیار ما نبوده که بررسی شود چه میزان قطعیت خوشه‌بندی مشخص شود که چقدر است، از اندیس سیلوئت استفاده می‌شود.



شکل۴ـ۱۶. نتیجه خوشه بندی افق مورد نظر در داده لرزه ای واقعی با روش kmeans و ۶ خوشه

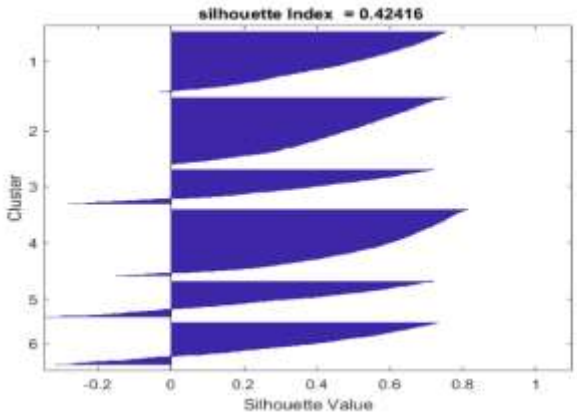
یکی از روش‌های بررسی قطعیت خوشه‌بندی استفاده از اندیس سیلوئت است. اندیس سیلوئت (Silouette Index) معیاری برای بررسی میزان قطعیت خوشه‌بندی است.

شاخص سیلوئت فاصله هر مشاهده i اُمین یک خوشه را با میانگین فواصل برون خوشه‌ای مقایسه می‌کند و به عبارت دیگر شاخص سیلوئت درجه اطمینان را در انتساب خوشه‌بندی یک مشاهدات معین را اندازه‌گیری می‌کند. برای مشاهده i شاخص سیلوئت به صورت Rawashdeh و Ralescu تعریف می‌شود که در آن a_i میانگین عدم تشابه i به تمام مشاهدات دیگر از خوشه خودش است. b_i عدم شباهت i به همه مشاهدات در خوشه c است. شاخص میانگین سیلوئت با میانگین sli به دست می‌آید.

$$SI_i = \frac{(b_i - a_i)}{\max(a_{i,} b_i)} \qquad (۱-۴)$$

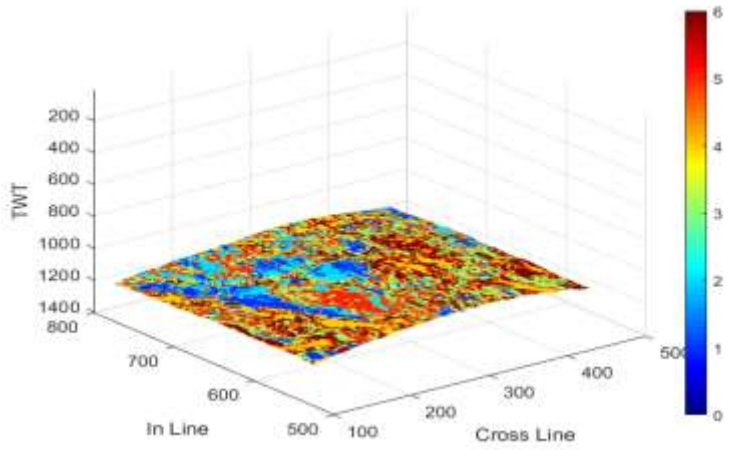
$$\overline{SI} = \frac{۱}{N}\sum_{i=۱}^{N} Si_i \qquad (۲-۴)$$

بنابراین نه تنها شاخص سیلوئت امکان ارزیابی تعداد بهینه خوشه را ارائه می دهد بلکه امکان ارزیابی ازعضویت خوشه را فراهم می‌کند. همانطور که اشاره شد برای قطعیت خوشه بندی چون به شیوه بدون نظارت است از یک اندیس سیلوئت استفاده کردیم که گراف اندیس سیلوئت برای خوشه‌های مختلف و مقدار میانگین آن در شکل ۴ـ۱۷ نشان داده شده است و مقدار میانگین اندیس سیلوئت برابر ۰/۴۲۴۱۶ محاسبه شده است.

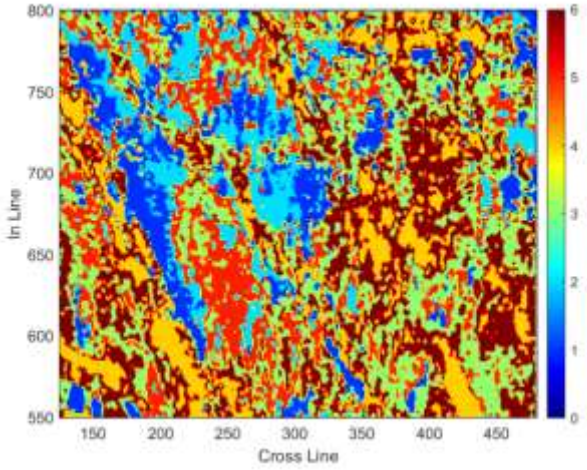


شکل۴ـ۱۷. گراف اندیس سیلوئت برای خوشه بندی افق نشان داده شده در شکل قبل. مقدار میانگین این اندیس برای میزان قطعیت کیفی خوشه بندی ۰/۴۲۴۱۶ برآورد شده است

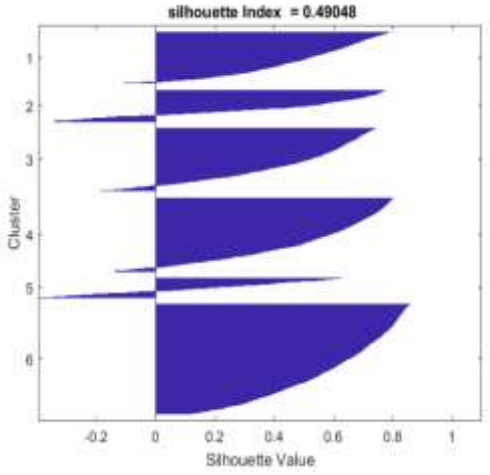
الگوریتم مورد نظر برای داده واقعی تجزیه شده پیاده سازی شد. به این شکل که داده لرزه‌ای ابتدا وارد مرحله تجزیه مُد متغیر شد و چند مُد ذاتی اول که فرض شد به‌عنوان نوفه باقی‌مانده باشند کنار گذاشته شد و داده بازسازی شده و وارد مرحله خوشه‌بندی شد. نتیجه خوشه‌بندی در شکل های ۴ـ۱۸ و ۴ـ۱۹ نشان داده می‌شود که شکل ۴ـ۱۸ نمایش سه بُعدی نتیجه خوشه‌بندی افق مورد نظر است و شکل ۴ـ۱۹ نمایش دو بُعدی خوشه‌بندی این افق را نشان می دهد. همانطور که مشاهده می‌شود روندهای زمین‌شناسی بر روی نتیجه خوشه بندی به شکل واضح تری دیده می شوند و پراکندگی خوشه‌ها کمتر شده و مشاهده می شود که با این ایده که مولفه‌های فرکانس بالا بیانگر نوفه باشند و کنار گذاشته شوند نتیجه خوشه‌بندی بسیار واضح‌تر و مشخص‌تر صورت گرفته است اما همانطور که اشاره شد برای قطعیت خوشه‌بندی چون به شیوه بدون ناظر است ما از یک اندیس سیلوئت استفاده کردیم که گراف اندیس سیلوئت و مقدار آن در شکل ۴ـ۲۰ نشان داده شده است که مقدار این اندیس برای این خوشه بندی ۰/۴۹۰۸ است که بیشتر بودن این اندیس خوشه بندی داده لرزه‌ای اصلی بیانگر دقت بیشتر خوشه بندی است.



شکل۴ـ۱۸. نمایش سه بعدی نتیجه خوشه بندی افق مورد نظر در داده لرزه ای واقعی با روش **kmeans** و ۶ خوشه. خوشه بندی بر روی داده تحلیل شده به روش **VMD** انجام شده است. به نحوی که مولفه های با فرکانس زیاد از بازسازی کنار گذاشته شده اند.



شکل ۴ـ۱۹. نتیجه خوشه بندی افق مورد نظر در داده لرزه ای واقعی با روش **kmeans** و ۶ خوشه. خوشه بندی بر روی داده تحلیل شده به روش **VMD** انجام شده است. به نحوی که مولفه های با فرکانس زیاد از بازسازی کنار گذاشته شده اند.



شکل۴ـ۲۰ . گراف اندیس سیلوئت برای خوشه بندی افق نشان داده شده در شکل قبل. مقدار میانگین این اندیس برای میزان قطعیت کیفی خوشه بندی ۰/۴۹۰۸ برآورد شده است.

فصل پنجم

نتیجه گیری و پیشنهادها

۵-۱ نتیجه‌گیری

همانطور که گفته شد داده‌های لرزه‌ای همواره به دلایل مختلفی حاوی نوفه هستند و در مراحل پردازشی سعی بر حذف نوفه می‌شود اما همواره مقداری نوفه باقی می‌ماند که باعث ایجاد مشکلات بسیاری در تفسیر داده‌ی لرزه‌ای می‌شود که این اشکالات باعث ایجاد عدم انطباق با واقعیت‌های زمین‌شناسی می‌شود که یکی از کارهایی که برای رفع این مشکل وجود دارد تحلیل رخساره‌ی لرزه‌ای است. به دلیل وجود نوفه که پیشتر به آن و مشکلاتی که به وجود می‌آورد اشاره شد، در این پایان‌نامه تصمیم گرفته شد که ابتدا داده‌های لرزه‌ای را توسط روش VMD به مُدّهای سازنده آن تجزیه و مُدّهای اول را که حاوی فرکانس بالا هستند به عنوان نوفه تشخیص داده و آن‌ها را از داده‌های لرزه‌ای کنار گذاشته و سپس داده بازسازی شده وارد بحث تحلیل رخساره شد. اگر چه همواره بحث اختلاط مُدّ وجود داشت.

روی داده‌ی مصنوعی و روی داده‌ی واقعی این الگوریتم پیاده‌سازی شد و نشان داده شد که با کنار گذاشتن مولفه‌های فرکانس بالا، خوشه‌بندی با دقت بیشتری حاصل می‌شود که برای داده‌ی حقیقی هم میزان اندیس سیلوئت بالاتر بیانگر قطعیت و دقت بیشتر خوشه بندی است.

۵-۲ پیشنهادها

برای ادامه این پژوهش پیشنهادهای زیر ارائه می‌شود:

۱: بررسی روش‌های خوشه‌بندی دیگر مثل SOM ، FCM و یا GMM

۲: انجام خوشه‌بندی با ناظر و فراهم بودن داده‌ی چاه که بتوانیم با قطعیت بیشتری از میزان بررسی عملکرد الگوریتم مورد نظر داشته باشیم.

۳: روش‌های تجزیه سیگنال فراوان هستند و در این پایان نامه از یکی از نسخه‌های تجزیه مُد تجربی استفاده شد و پیشنهاد می‌شود که از روش‌هایی مثل تبدیل موجک و یا روش‌های بر مبنای فوریه هم برای بررسی نتایج استفاده شود.

مراجع

Arshin, B.M., Ghazali, A.R., Amin, Y.K. and Barnes, A.E., 2014, December. Hybrid waveform classification applied to delineate compartments in a complex reservoir in the Malay Basin. In *International Petroleum Technology Conference*. OnePetro.

Battista, B.M., Knapp, C., McGee, T. and Goebel, V., 2007. Application of the empirical mode decomposition and Hilbert-Huang transform to seismic reflection data. *Geophysics*, *72*(2), pp.H29-H37.

Chopra*, S. and Marfurt, K.J., 2014. Seismic facies analysis using generative topographic mapping. In *SEG Technical Program Expanded Abstracts 201۴* (□□. ۱۳۹۰-۱۳۹۴). □□□□□□□ □□ □□□□□□□□□□□ □□□□□□□□□□□□□□.

Chopra, S. and Marfurt, K.J., 2018. Seismic facies classification using some unsupervised machine-learning methods. In *SEG Technical Program Expanded Abstracts 2018* (pp. 2056-۲۰۶۰). □□□□□□y of Exploration Geophysicists.

Chopra, S., Marfurt, K. and Sharma, R., 2019, September. Unsupervised machine learning facies classification in the Delaware Basin and its comparison with supervised Bayesian facies classification. In *SEG International Exposition and Annual Meeting*. OnePetro.

Coléou, T., Poupon, M. and Azbel, K., 2003. Unsupervised seismic facies classification: A review and comparison of techniques and implementation. The Leading Edge, 22(10), pp.942-953.

Correa, C., Valero, C., Barreiro, P., Diago, M.P. and Tardáguila, J., 2011, November. A comparison of fuzzy clustering algorithms applied to feature extraction on vineyard. In *Proceedings of the XIV Conference of the Spanish Association for Artificial Intelligence*.

de Matos, M.C., Osorio, P.L. and Johann, P.R., 2007. Unsupervised seismic facies analysis using wavelet transform and self-organizing maps. *Geophysics*, *72*(1), pp.P9-P21.

Du, H.K., Cao, J.X., Xue, Y.J. and Wang, X.J., 2015. Seismic facies analysis based on self-organizing map and empirical mode decomposition. *Journal of Applied Geophysics*, *112*, pp.52-61.

Ferreira, D.J.A., Lupinacci, W.M., Neves, I.D.A., Zambrini, J.P.R., Ferrari, A.L., Gamboa, L.A.P. and Olho Azul, M., 2019. Unsupervised seismic facies classification applied to a presalt carbonate reservoir, Santos Basin, offshore Brazil. *AAPG Bulletin*, *103*(4), pp.997-1012.

Gaci, S., 2016. A new ensemble empirical mode decomposition (EEMD) denoising method for seismic signals. *Energy Procedia*, *97*, pp.84-91.

Han, J. and van der Baan, M., 2012, November. Complete ensemble empirical mode decomposition for seismic time-frequency analysis. In *2012 SEG Annual Meeting*. OnePetro.

Hardisty, R. and Wallet, B., 2017. Unsupervised seismic facies from mixture models to highlight channel features. In *SEG Technical Program Expanded Abstracts 2017* (pp. 2289-2293). Society of Exploration Geophysicists.

Honório, B.C.Z., de Matos, M.C. and Vidal, A.C., 2017. Progress on empirical mode decomposition-based techniques and its impacts on seismic attribute analysis. *Interpretation*, *5*(1), pp.SC17-SC28.

Isham, M.F., Leong, M.S., Lim, M.H. and Ahmad, Z.A., 2018. Variational mode decomposition: mode determination method for rotating machinery diagnosis. *Journal of Vibroengineering*, *20*(7), pp.2604-2621.

John, A.K., Lake, L.W., Torres-Verdin, C. and Srinivasan, S., 2005, October. Seismic-based facies identification and classification using simple statistics. In *SPE Annual Technical Conference and Exhibition*. OnePetro.

Kuroda, M.C., Vidal, A.C. and de Carvalho, A.M.A., 2012. Interpretation of seismic multiattributes using a neural network. *Journal of Applied Geophysics*, *85*, pp.15-24.

Lei, Y. and Zuo, M.J., 2009. Fault diagnosis of rotating machinery using an improved HHT based on EEMD and sensitive IMFs. *Measurement Science and Technology*, *20*(12), p.125701.

Li, F., Zhao, T., Qi, X., Marfurt, K. and Zhang, B., 2016. Lateral consistency preserved variational mode decomposition (VMD). In *SEG Technical Program Expanded Abstracts 2016* (pp. 1717-1721). Society of Exploration Geophysicists.

Li, J. and Castagna, J., 2004. Support vector machine (SVM) pattern recognition to AVO classification. *Geophysical Research Letters*, *31*(2).

Liu, J., Dai, X., Gan, L., Liu, L. and Lu, W., 2018. Supervised seismic facies analysis based on image segmentation. *Geophysics*, *83*(2), pp.O25-O30.

Pham, D.T., Dimov, S.S. and Nguyen, C.D., 2005. Selection of K in K-means clustering. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, *219*(1), pp.103-119.

Qi, J., Lin, T., Zhao, T., Li, F. and Marfurt, K., 2016. Semisupervised multiattribute seismic facies analysis. *Interpretation*, *4*(1), pp.SB91-SB106.

Roy, A., Marfurt, K.J. and de Matos, M.C., 2010, October. Applying self-organizing maps of multiple attributes, an example from the Red-Fork Formation, Anadarko Basin. In *2010 SEG Annual Meeting*. OnePetro.

Sabeti, H. and Javaherian, A., 2009, May. Seismic facies analysis based on K-means clustering algorithm using 3D seismic attributes. In *Shiraz 2009-1st EAGE International Petroleum Conference and Exhibition* (pp. cp-125). European Association of Geoscientists & Engineers.

Sabeti, H. and Nadjar, B., 2011. Seismic facies classification using 2-D and 3-D multi-attribute hierarchical clustering algorithms. In *SEG Technical Program Expanded Abstracts 2011* (pp. 1160-1164). Society of Exploration Geophysicists.

Saggaf, M.M., Toksöz, M.N. and Marhoon, M.I., 2003. Seismic facies classification and identification by competitive neural networks. *Geophysics*, *68*(6), pp.1984-1999.

Singh, V.B., Subrahmanyam, D., Negi, S.P.S., Baid, V.K., Kumar, A. and Biswal, S., 2004. Facies classification based on seismic waveform-a case study from Mumbai High North. In *5th Conf. and Exposition on Petroleum Geophysics* (pp. ۴۵۶-۴۶۲).

Song, C., Liu, Z., Cai, H., Wang, Y., Li, X. and Hu, G., 2017. Unsupervised seismic facies analysis with spatial constraints using regularized fuzzy c-means. *Journal of Geophysics and Engineering*, *14*(6), pp.1535-1543.

Song, C., Liu, Z., Wang, Y., Xu, F., Li, X. and Hu, G., 2018. Adaptive phase k-means algorithm for waveform classification. *Exploration Geophysics*, *49*(2), pp.213-219.

Song, C., Liu, Z., Wang, Y., Xu, F., Li, X. and Hu, G., 2018. Adaptive phase k-means algorithm for waveform classification. *Exploration Geophysics*, *49*(2), pp.213-219.

Tran, T.A., 2019. A study of the energy efficiency management for bulk carriers considering navigation environmental impacts. *Journal of Intelligent & Fuzzy Systems*, *36*(3), pp.2871-2884.

Troccoli, E.B., Cerqueira, A.G., Lemos, J.B. and Holz, M., 2022. K-means clustering using principal component analysis to automate label organization in multi-attribute seismic facies analysis. *Journal of Applied Geophysics*, p.104555.

Wallet, B.C. and Hardisty, R., 2019. Unsupervised seismic facies using Gaussian mixture models. *Interpretation*, *7*(3), pp.SE93-SE111.

Wen, X., He, Z. and Huang, D., 2009. Reservoir detection based on EMD and correlation dimension. *Applied Geophysics*, *6*(1), pp.70-76.

Wrona, T., Pan, I., Bell, R.E., Fossen, H. and Gawthorpe, R., 2020. Principal component analysis of seismic facies.

Xue, Y.J., Cao, J.X. and Tian, R.F., 2014. EMD and Teager–Kaiser energy applied to hydrocarbon detection in a carbonate reservoir. *Geophysical Journal International*, *197*(1), pp.277-291.

Zeiler, A., Faltermeier, R., Keck, I.R., Tomé, A.M., Puntonet, C.G. and Lang, E.W., 2010, July. Empirical mode decomposition-an introduction. In The 2010 international joint conference on neural networks (IJCNN) (pp. 1-8). IEEE.

Zhao, T., Jayaram, V., Roy, A. and Marfurt, K.J., 2015. A comparison of classification techniques for seismic facies recognition. *Interpretation*, *3*(4), pp. SAE29-SAE58.

Zhao, T., Li, F. and Marfurt, K.J., 2017. Constraining self-organizing map facies analysis with stratigraphy: An approach to increase the credibility in automatic seismic facies classification. *Interpretation*, *5*(2), pp.T163-T171.

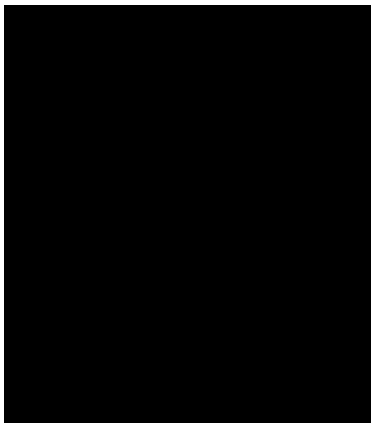
طهماسبی، خ.، روشندل کاهو، ا. و نجاتی کلاته، ع.، ۲۰۱۷. تضعیف نوفه‌های زمین‌غلت در داده‌های لرزه‌ای با استفاده از تجزیه مُد متغیر. پژوهش های ژئوفیزیک کاربردی, ۳(۲), pp.177-188.

آریان نژاد، ا.، رداد, م. و هادیلو, س.، ۲۰۲۰. طبقه‌بندی غیرنظارتی داده‌های لرزه‌ای با استفاده از مدل‌های ترکیبی گوسی. *پژوهش نفت*, ۳۰(۹۹-۳), pp.129-144.

Abstract

To model a reservoir, it is necessary to accurately evaluate the properties of rocks and visualize their heterogeneity. Due to high drilling costs and the lack of sufficient large-scale wells or data related to wells and cores, 3D seismic data play a very important role in detecting reservoir changes. Any change in lithology, porosity, and fluid content causes changes in amplitude, frequency, lateral cohesion, and other seismic markers. If these changes are detectable in seismic parameters, they will help to understand subsurface geology. The purpose of seismic facies analysis is to interpret changes in seismic parameters. Seismic facies can be defined as seismic rejection groups. Many methods have been developed to identify and detect patterns for seismic facies. An unsupervised pattern is used when geological information is not available. The presence of noise is an integral part of seismic data that affects the classification of facies and seismic interpretation, and its presence causes misinterpretation. The research approach for this problem is multi-resolution seismic facies analysis. This means that before the main seismic data or markers enter the facies analysis process, we first decompose the seismic data into its constituent components using one of the signal decomposition methods and then participate in a number of components in signal reconstruction and some components. We leave out the ones that express the nature of noise. Then a reconstructed version of the signal or the resulting markers is entered as input to the facies analysis algorithm. In this research, the intention is to decompose seismic data into their constituent components using variational mode decomposition (VMD) and to reconstruct the facies analysis process on the data and apply the resulting markers. Facies analysis in this method will be performed without supervision and by K-means method. The performance of artificial and real data methods is tested and it is shown that more accurate facies analysis can be performed.

Keywords: Signal decompotion _ Seismic attribute _ Seismic facies analysis _ Noise effect



Shahrood University of Technology

Faculty of Mining, Petroleum and Geophysics Engineering

M.Sc. Thesis in Mineral Exploration

Multi-resolution seismic facies analysis for tackling the residual noise on data

By: Milad Barzegar

Supervisor:

Dr. Mohammad Radad

Dr. Amin Roshandel Kahoo

Advisor:

Dr. Saeed Hadiloo

February 2022