

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



دانشکده مهندسی معدن، نفت و ژئوفیزیک
پایان نامه کارشناسی ارشد اکتشاف مواد معدنی

بررسی تاثیر بسته بودن داده‌ها بر تحلیل چند متغیره داده‌های ژئوشیمیایی

نگارنده: جمال الدین کلته

استاد راهنما

دکتر سید رضا قوامی ریابی

استاد مشاور

مهندس حمید معینی

خرداد ۱۳۹۶

تقدیر و تشکر:

سپاس یگانه‌ای که آموزش را آموخت تا از آموخته‌هایم راز آفرینش را فراگیرم.

وظیفه خود می‌دانم از تمام عزیزانی که با صبر و حوصله، همکاری و حمایت خود را از اینجانب دریغ نداشته‌اند و از استاد

راهنمای محترم جناب آقای دکتر قوامی و استاد مشاور جناب مهندس معینی که با راهنمایی‌های خود من را در انجام

این پروژه یاری نمودند کمال قدردانی را بنمایم.

تعهد نامه

این جانب **جمال الدین کلتی** دانشجوی دوره کارشناسی ارشد رشته **مهندسی معدن** دانشکده معدن، نفت و ژئوفیزیک دانشگاه صنعتی شاهرود نویسنده پایان نامه **بررسی تاثیر بسته بودن داده ها بر آنالیز چندمتغیره داده های ژئوشیمیایی** تحت راهنمایی **دکتر سید رضا قوامی ریایی** متعهد می شوم.

- تحقیقات در این پایان نامه توسط این جانب انجام شده است و از صحت و اصالت برخوردار است.
- در استفاده از نتایج پژوهش های محققان دیگر به مرجع مورد استفاده استناد شده است.
- مطالب مندرج در پایان نامه 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 می باشد 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 پایان نامه تأثیرگذار بوده اند در مقالات مستخرج از پایان نامه رعایت می گردد.
- در کلیه مراحل انجام این پایان نامه، در مواردی که از موجود زنده (یا بافت های آن ها) استفاده شده است ضوابط و اصول اخلاقی رعایت شده است.
- در کلیه مراحل انجام این پایان نامه، 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

مالکیت نتایج و حق نشر

- کلیه حقوق معنوی این اثر و محصولات آن (مقالات مستخرج، کتاب، برنامه های رایانه ای، نرم افزار ها و تجهیزات ساخته شده است) متعلق به دانشگاه صنعتی شاهرود می باشد. این مطلب باید به نحو مقتضی در تولیدات علمی مربوطه ذکر شود.
- استفاده از اطلاعات و نتایج موجود در پایان نامه بدون ذکر مرجع مجاز نمی باشد.

* متن این صفحه نیز باید در ابتدای نسخه های تکثیر شده پایان نامه وجود داشته باشد

چکیده

داده‌های ژئوشیمیایی از نوع داده‌های ترکیبی هستند به این معنی که اجزای آن جزئی از یک کل و معمولاً دارای مجموع ثابت هستند. این داده‌ها فضای نمونه‌ای متفاوتی نسبت به داده‌های غیر ترکیبی دارند و استفاده از روش‌های آماری معمول برای بررسی داده‌ها ممکن است با خطا همراه باشد. برای بررسی داده‌های ترکیبی ابتدا باید داده‌ها را با استفاده از تبدیل‌های مناسب باز کرد. یکی از مناسب‌ترین تبدیل‌ها، نسبت‌های لگاریتمی است. به منظور بررسی تأثیری که باز کردن داده‌ها بر تفکیک هاله‌های ژئوشیمیایی و شناسایی کانی‌سازی پنهان در این تحقیق داده‌های ماسه بادی‌های موجود از یک کانی-سازی سولفید توده‌ای آتشفشانی (VHMS) به صورت باز و بسته مورد تحلیل قرار گرفته است. داده‌های این تحقیق از ماسه بادی‌های معدن مس-روی Areachap از آفریقای جنوبی بوده است.

یکی از روش‌های متداول برای بررسی چندمتغیره داده‌های ژئوشیمیایی، تحلیل مولفه‌های اصلی (PCA) است. روش تحلیل مولفه‌های اصلی به علت استفاده از ماتریس کوواریانس به شدت در برابر وجود مقادیر خارج از ردیف داده‌ها حساس است. برای حل این مشکل روش تحلیل مولفه‌های اصلی مقاوم (RPCA) معرفی شده است. داده‌های بسته به وسیله تحلیل PCA و داده‌های باز با روش RPCA مورد بررسی قرار گرفت. سپس با استفاده از روش مدل‌سازی نمودار احتمال حد آنومال مولفه کانی‌سازی محاسبه گردیده است. با بررسی روند ایجاد شده بر روی نمودار مولفه‌های مربوط به کانی‌سازی مشخص شد که روش (RPCA) بر روی داده‌های باز شده با نسبت‌های لگاریتمی، به خوبی زون کانی‌سازی را می‌تواند مشخص کند و روش (PCA) بر روی داده‌های بسته محدود سنگ در بر گیرنده کانی‌سازی را به خوبی مشخص کند و با استفاده از نمودارهای PCA و RPCA در کنار یکدیگر به خوبی می‌توان به ساختار کانی‌سازی در منطقه پی برد. علاوه بر آن از روش (PCA) می‌توان به عنوان راهنمای محدود مناسب برای حفاریات معدنی استفاده کرد.

کلمات کلیدی: ماسه‌بادی، داده‌های ترکیبی، تحلیل مولفه‌های اصلی معمول و مقاوم، باز کردن داده‌ها

فهرست مطالب

فصل اول:	۱
کلیات	۱
۱-۱ مقدمه	۲
۱-۲ ضرورت و هدف انجام تحقیق	۲
۳-۱ روش انجام تحقیق	۵
۴-۱ مجموعه داده‌ها و نوع کانی‌سازی	۶
۵-۱ ساختار پایان نامه	۶
فصل دوم:	۷
معرفی منطقه	۷
۱-۲ مقدمه	۸
۲-۲ موقعیت مکانی معدن	۸
۳-۲ زمین‌شناسی ناحی‌های	۱۰
۴-۲ زمین‌شناسی محلی	۱۲
۱-۴-۲ توالی سنگ‌شناسی در معدن Areachap	۱۲
۲-۴-۲ دگرسانی و دگرگونی در معدن Areachap	۱۴
۳-۴-۲ کانی‌شناسی ذخیره Areachap	۱۴
۵-۲ محیط‌های کالکریتی	۱۵
۴-۲ پیشینه کارهای اکتشافی در منطقه	۱۷
۵-۲ کانسارهای سولفید توده‌ای	۱۸
فصل سوم:	۲۱
پیشینه تحقیق و تئوری روش‌ها	۲۱
۱-۳ مقدمه	۲۲
۲-۳ تحلیل مولفه‌های اصلی	۲۲
۳-۳ تحلیل مولفه‌های اصلی به روش مقاوم	۲۸

۳۰	۴-۳ نمودار احتمال
۳۱	۶-۳ روند تکامل بررسی آماری داده‌های ترکیبی
۳۲	۵-۳ اصول بررسی داده‌های ترکیبی
۳۳	۵-۳ هندسه سیمپلکس
۳۶	۱-۵-۳ تبدیل نسبت لگاریتمی
۳۹	۶-۳ اصول تفسیر نمودار دوتایی ترکیبی
۳۹	۱-۶-۳ نمودار دوتایی
۳۹	۲-۶-۳ تفسیر یک نمودار دوتایی ترکیبی
۴۳	۷-۳ بررسی تابع توزیع داده‌ها در داخل سیمپلکس
۴۴	۱-۳-۷ آزمون شعاع
۴۶	۷-۳ جمع بندی
۴۷	فصل چهارم:
۴۷	تجزیه و تحلیل داده‌ها منطقه مورد مطالعه
۴۸	۱-۴ مقدمه
۴۸	۲-۴ مجموعه داده‌ها
۵۰	۳-۴ بررسی داده‌های خارج از ردیف به صورت چندمتغیره
۵۲	۴-۴ بررسی هیستوگرام عنصرهای کانی‌سازی در منطقه
۵۳	۳-۴ آنالیز همبستگی
۵۵	۱-۳-۴ بررسی نرمال بودن داده‌ها
۶۱	۳-۴ تحلیل مولفه‌های اصلی به روش معمولی و مقاوم
۶۲	۱-۳-۴ روش مقاوم در برابر روشهای کلاسیک تحلیل مولفه‌های اصلی
۶۲	۲-۳-۴ بررسی داده‌ها برای انجام PCA
۶۴	۴-۴ نتایج PCA و RPCA
۶۵	۴-۴-۴ نمودار دوتایی
۷۱	۵-۴ نمودار احتمال

۸۱ فصل پنجم:
۸۱ نتیجه گیری و پیشنهاد
۸۲ ۱-۵ نتیجه گیری
۸۳ ۲-۵ پیشنهاد
۸۵ پیوست
۸۹ منابع و مآخذ

فهرست اشکال

- شکل ۱-۱: موقعیت منطقه Areachap ۶
- شکل ۱-۲: موقعیت مکانی منطقه مورد مطالعه ۹
- شکل ۲-۲: زمین‌شناسی ناحیه‌ای منطقه ۱۱
- شکل ۳-۲: زمین‌شناسی منطقه Areachap ۱۳
- شکل ۴-۲: نمایی شماتیک از کانسار سولفید توده‌ای ۲۰
- شکل ۱-۳: نمودار پراکندگی داده اصلی ۲۶
- شکل ۲-۳: قرار دادن داده‌ها در راستای مولفه‌های اصلی ۲۷
- شکل ۳-۳: داده‌ها بعد از قرار گرفتن بر روی مولفه اصلی ۲۸
- شکل ۴-۳: نمودار احتمال یک نمونه دو جامعه‌ای ۳۰
- شکل ۳-۵: تصویری از اصول تفسیر نمودار دوتایی مربوط به یک داده ترکیبی ۴۰
- شکل ۱-۴: نقشه زمین‌شناسی و موقعیت پروفیل نمونه‌برداری ۴۹
- شکل ۲-۴: هیستوگرام داده‌های بسته ۵۲
- شکل ۳-۴: هیستوگرام داده‌های باز ۵۳
- شکل ۴-۴: نمودار پراکندگی داده بسته مس با داده‌های باز ۵۵
- شکل ۴-۵: نمودار دوتایی و توضیحات مربوط به محورهای آن ۶۷
- شکل ۴-۶: نمودار دوتایی برای الف) تحلیل مولفه‌های اصلی معمولی و ب) مقاوم ۶۹
- شکل ۴-۷: نمودار منطقه آنومال برای عنصر مس ۷۵
- شکل ۴-۸: نمودار منطقه آنومال برای عنصر سرب ۷۶
- شکل ۴-۹: نمودار منطقه آنومال برای عنصر روی ۷۷

فهرست جداول

- جدول ۳-۱: مقادیر بحرانی برای سطح اعتمادهای مختلف برای تست شعاع..... ۴۵
- جدول ۳-۲: مقادیر بحرانی برای سطح اعتمادهای مختلف برای آزمون‌های Marginal و Bivariate angle..... ۴۶
- جدول ۴-۱: فاصله مایلانابیوس برای داده‌های باز..... ۵۱
- جدول ۴-۲: نتایج آزمون شعاع..... ۵۶
- جدول ۴-۳: نتایج آزمون Marginals..... ۵۷
- جدول ۴-۴: مقادیر شاخص شاپیرو-ویلک برای داده‌های باز شده به روش ilr..... ۵۸
- جدول ۴-۵: ضرایب همبستگی اسپیرمن برای داده‌های بسته..... ۵۹
- جدول ۴-۶: ضرایب اسپیرمن برای داده‌های باز شده به روش clr..... ۶۰
- جدول ۴-۷: جدول قضاوت درباره ضریب KMO..... ۶۳
- جدول ۴-۸: مقادیر ویژه برای PCA و RPCA..... ۶۵
- جدول ۴-۹: ماتریس واریانس داده‌ها..... ۷۰
- جدول ۴-۱۰: واریانس داده‌های بسته..... ۷۱
- جدول ۴-۱۱: بار فاکتوری RPCA..... ۷۲
- جدول ۴-۱۲: بار فاکتوری PCA..... ۷۳
- جدول ۴-۱۳: نتایج نمودار احتمال..... ۷۴

فصل اول:

کلیات

۱-۱ مقدمه

ذخایری که در سطح دارای رخنمون بوده و یا دارای عیار و ذخیره زیاد می‌باشند، اکثراً تاکنون کشف شده‌اند. در آینده محققان باید به دنبال روش‌هایی برای شناسایی آنومالی‌های ژئوشیمیایی ضعیف یا پنهان با ارزش اقتصادی باشند. امروزه علاوه بر تفکیک آنومالی‌ها از زمینه، باید آنومالی‌ها را از آنومالی‌های ناشی از فعالیت‌های انسانی نیز تفکیک نمود. یکی از این روش‌های رسیدن به این هدف استفاده از روش‌های آماری برای پردازش داده‌های ژئوشیمیایی است.

در دهه‌های اولیه رواج استفاده از رایانه‌ها در پروژه‌های اکتشافی، بررسی میانگین، میانه و انحراف معیار برای تعیین محدوده‌های زمینه و آنومالی و نمودار توزیع داده‌ها برای تعیین نوع توزیع داده‌ها مورد بررسی قرار می‌گرفت. در این دوران، داده‌های ژئوشیمیایی معمولاً به صورت تک متغیره مورد تحلیل قرار می‌گرفت. امروزه با استفاده از پردازش داده‌ها و مدل‌سازی، می‌توان چندین متغیر را به صورت هم‌زمان مورد بررسی قرارداد (روش‌های چند متغیره). از جمله روش‌های چند متغیره مورد استفاده برای پردازش داده‌های ژئوشیمیایی می‌توان به آنالیز فاکتوری و آنالیز خوشه‌ای اشاره کرد.

۲-۱ ضرورت و هدف انجام تحقیق

داده‌های ژئوشیمیایی معمولاً به صورت درصد وزنی و به صورت قسمتی از یک کل گزارش می‌شوند. استفاده از این روش به این دلیل است که بتوان آنالیز نمونه‌ها را با یکدیگر مقایسه کرد در ژئوشیمی به این ماهیت داده‌ها، داده‌های ترکیبی^۱ اطلاق می‌گردد که دارای مجموع ثابت یک یا صد است [۱]، [۲]، [۳].

از نظر هندسی داده‌های ترکیبی متعلق به یک زیرمجموعه با یک بعد کمتر از مجموعه داده‌های اصلی است زیرا هر داده ترکیبی را می‌توان به صورت ترکیب خطی $n-1$ بردار مستقل مجموعه داده‌ی

^۱ - Compositional data

اصلی نمایش داد. در حالت کلی می توان داده های ترکیبی را جزئی از یک کل و یا یک ساختار صفحه ای در نظر گرفت. داده های ترکیبی تنها اطلاعات نسبی را نشان می دهند، یا به عبارتی، تنها نسبت بین اجزا (برای مثال مختصات یک داده ترکیبی) اطلاعات کامل را نشان می دهد [۴].

داده های ژئوشیمیایی علاوه بر ترکیبی بودن، چندمتغیره نیز هستند. بنابراین برای بررسی این داده ها باید از روش هایی استفاده کرد که بتوانند تغییرات چندین متغیر را به صورت همزمان اندازه گیری کنند. روش های پردازش داده های ژئوشیمیایی به دو دسته تقسیم می شوند: ۱- روش های براساس فراوانی و ۲- روش های براساس مکان. روش های آماری، روش های پرکاربردی هستند که در دسته اول قرار می گیرند. استفاده از روش های آماری برای داده های ژئوشیمیایی که ماهیت ترکیبی دارند، دارای چندین اشکال اساسی است.

بنا به تعریف، کار با داده های ترکیبی دارای چندین اشکال ذاتی است. برای مشخص شدن این موضوع مثال زیر را در نظر بگیرید. توزیع تعداد دانشجویان پسر و دختر یک کلاس فرضی با ۱۱۰ دانشجو را به عنوان مثال مورد بررسی قرار می دهیم. تعداد دانشجویان پسر برابر تعداد دانشجویان دختر است. روش متداول این است که این مقادیر (تعداد دانشجویان دختر و پسر) را به صورت درصدی از یک کل نشان می دهند. برای این کار باید داده ها بسته و یا نرمالیزه شوند. عملگر بستن به صورت زیر تعریف می شود [۲]:

$$C(z) = \left[\frac{k.z_1}{\sum_{i=1}^D z_i}, \frac{k.z_2}{\sum_{i=1}^D z_i}, \dots, \frac{k.z_D}{\sum_{i=1}^D z_i} \right] \quad (1-1)$$

با به کار بردن عملگر بستن داریم:

$$C(\text{female}, \text{male}) = \frac{100}{55+55} [55, 55] = [50, 50]\% \quad (2-1)$$

بعد از بستن داده ها، مجموع دو جزء برابر ثابت $k=100$ است. در این روش تعداد دانشجویان پسر و دختر برابر ۵۰٪ دختر و پسر قرار می گیرد زیرا محتوای نسبی کلاس (یعنی تعداد دانشجویان دختر و پسر) مستقل از واحد مورد استفاده برای بیان آن است [۵].

اگر از تعداد دانشجویهای دختر ده نفر و از تعداد دانشجویهای پسر ۵ نفر را کم کنیم، جمعیت کلاس برابر نودوپنج نفر خواهد شد. با تکرار رویه بالا برای تبدیل به درصد داریم:

$$C(\text{female}, \text{male}) = \frac{100}{50+45} [50,45] = [52.6,47.4]\% \quad (3-1)$$

از معادله بالا چنین برمی آید که تعداد دانشجویهای پسر افزایش و تعداد دانشجویهای دختر کاهش پیدا کرده است (همبستگی منفی). درحالی که تعداد دانشجویهای دختر و پسر هر دو کاهش یافته است ولی شیب این کاهش برای هر دو یکسان نیست (همبستگی مثبت). این موضوع مسئله اصلی در مورد مشکلی است که با عنوان همبستگی‌های غیرواقعی از آن یاد می‌شود. بنابراین قید مجموع ثابت برای این داده‌ها باعث می‌شود که همبستگی بین اجزا واقعی نبوده و غیرواقعی باشد [۴].

علاوه بر مشکل همبستگی‌های غیرواقعی، فضای نمونه‌ای محدود داده‌های ترکیبی بنا به تعریف نیز مشکلاتی برای بهره‌گیری از روش‌های آماری ایجاد می‌کند. با توجه به بسته بودن داده‌های ترکیبی، فضای نمونه‌ای داده‌های ترکیبی محدود و زیرمجموعه‌ای از R است. روش‌های آماری استاندارد مورد استفاده برای بررسی داده‌های ژئوشیمیایی از فضای هندسی اقلیدسی و فضای نمونه‌ای R استفاده می‌کنند. فضای نمونه‌ای داده‌های ترکیبی سیمپلکس^۱ S^D نام دارد و به صورت زیر تعریف می‌شود [۴]:

$$S^D = \{x = (x_1, x_2, \dots, x_D) : x_i > 0 (i = 1, 2, \dots, D), \sum_{i=1}^D x_i = k\} \quad (4-1)$$

k یک ثابت با مقدار مثبت است که مقدار آن به نوع اندازه‌گیری عناصر (درصد وزنی، ppm و...) وابسته است. می‌توان نشان داد که ساختار هندسی سیمپلکس، اقلیدسی نیست. بنابراین استفاده از روش‌های آماری استاندارد برای بررسی داده‌های ژئوشیمیایی خالی از اشکال نیست.

روش‌هایی برای بررسی اکتشافی داده‌ها در درون سیمپلکس در حال گسترش است و تبدیل‌هایی نیز برای تصویر داده‌های ترکیبی به درون فضای R گسترش یافته است. برای بررسی این موضوع که

¹ Simplex

استفاده از روش‌های پیشنهادی برای داده‌های ترکیبی چه تاثیری بر تفکیک هاله‌های ژئوشیمیایی و متعاقب آن مناطق شناسایی کانی‌سازی پنهان دارد، از ماسه بادی‌های مربوط یک کانی‌سازی سولفید توده‌ای آتش‌فشانی هم به صورت باز و هم به صورت بسته مورد بررسی قرار گرفت.

۱-۳ روش انجام تحقیق

برای بررسی داده‌های ژئوشیمیایی می‌توان از روش‌های چندمتغیره استفاده کرد. یکی از روش‌های معمول برای بررسی داده‌های چند متغیره داده‌های ژئوشیمیایی استفاده از روش تحلیل مولفه‌های اصلی (PCA^۱) است.

روش تحلیل مولفه‌های اصلی، روش بسیار کارآمدی برای کاهش ابعاد داده‌ها است، ولی نسبت به داده‌های خارج از ردیف حساس است و باعث ایجاد خطا خواهد شد. به طور معمول برای حل این مشکل در پیش‌پردازش داده‌ها، داده‌های خارج از ردیف حذف می‌شوند. اما داده‌های خارج از ردیف داده‌های اشتباه نیستند و حذف آنها باعث کاهش واریانس طبیعی داده‌ها می‌شود. برای حل این موضوع روش تحلیل مولفه‌های اصلی مقاوم معرفی (RPCA^۲) شد. در روش RPCA به جای استفاده از ماتریس کوواریانس داده‌ها، از تخمین‌گر کم‌ترین میزان دترمینان کوواریانس استفاده می‌کند تا اثر داده‌های خارج از ردیف را کاهش دهد [۶].

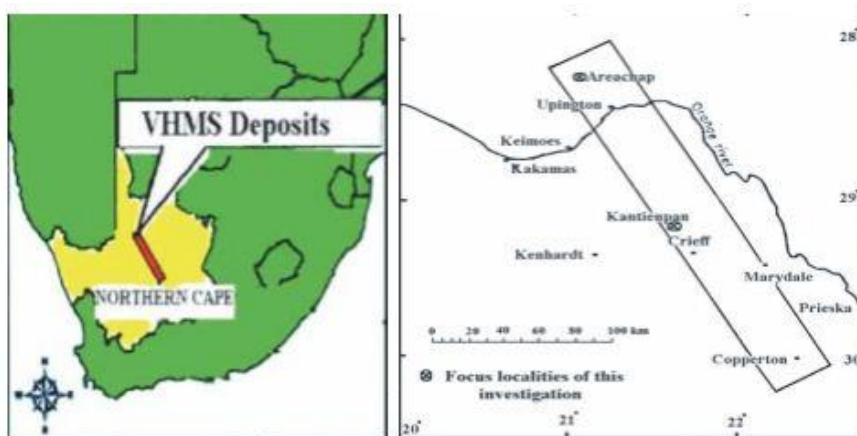
در این پایان‌نامه برای بررسی اثر بسته یا باز بودن داده‌ها بر تحلیل و نتیجه بررسی‌های چند متغیره، ابتدا داده‌ها را بدون در نظر گرفتن اثر بسته‌بودن و به شکل معمول با استفاده از PCA مورد بررسی قرار داده و سپس نتایج مولفه اول تحلیل مولفه‌های اصلی در برابر مولفه دوم آن رسم شده و نمودار آنومالی-های PC های مربوط به عناصر کانی سازی در منطقه رسم خواهد شد. سپس با باز کردن داده‌ها به روش نسبت لگاریتمی ایزومتریک همین مراحل تکرار و نتایج با هم مقایسه خواهد شد.

¹ -Principal Component Analysis

² - Robust Principal Component Analysis

۴-۱ مجموعه داده‌ها و نوع کانی‌سازی

داده مورد استفاده در این پایان‌نامه از ماسه بادی‌های محدوده متروک معدنی Areachap در آفریقای جنوبی برداشت شده است. نوع کانی‌سازی در این محدوده از نوع سولفید توده‌ای آتش‌فشانی است. کل منطقه Areachap پس از تشکیل کانسارهای سولفید توده‌ای دچار دگرگونی شدید شده است. سنگ‌های موجود در منطقه، عموماً رسوبی و دگرگونی است. داده‌ها از یک پروفیل نمونه‌برداری (T2) و از بخش دانه‌ریز و از عمق ۱۰-۱۵ سانتیمتری برداشت شده است. شکل (۱-۱) موقعیت مکانی منطقه Areachap را نشان می‌دهد.



شکل (۱-۱): موقعیت منطقه Areachap [۷]

۵-۱ ساختار پایان‌نامه

این پایان‌نامه دارای ۵ فصل است. در فصل اول به کلیات مربوط به بحث پرداخته شد. در فصل‌های دوم و سوم به ترتیب معرفی منطقه مورد مطالعه و پیشینه اکتشافی آن و تئوری روش‌های مورد استفاده بیان خواهد شد. فصل چهارم به شرح روند انجام روش‌های مورد استفاده بر روی داده‌ها اختصاص خواهد یافت. فصل پنجم نیز در برگیرنده نتایج و پیشنهادات است.

فصل دوم:

معرفی منطق

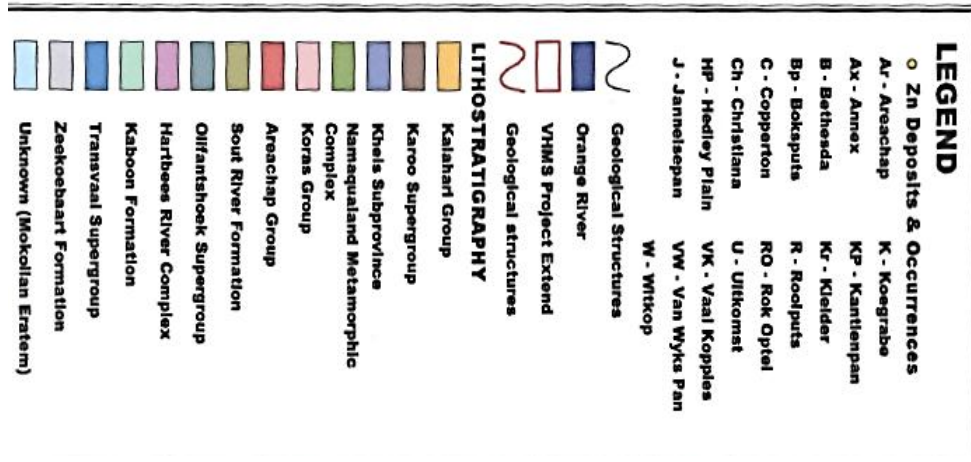
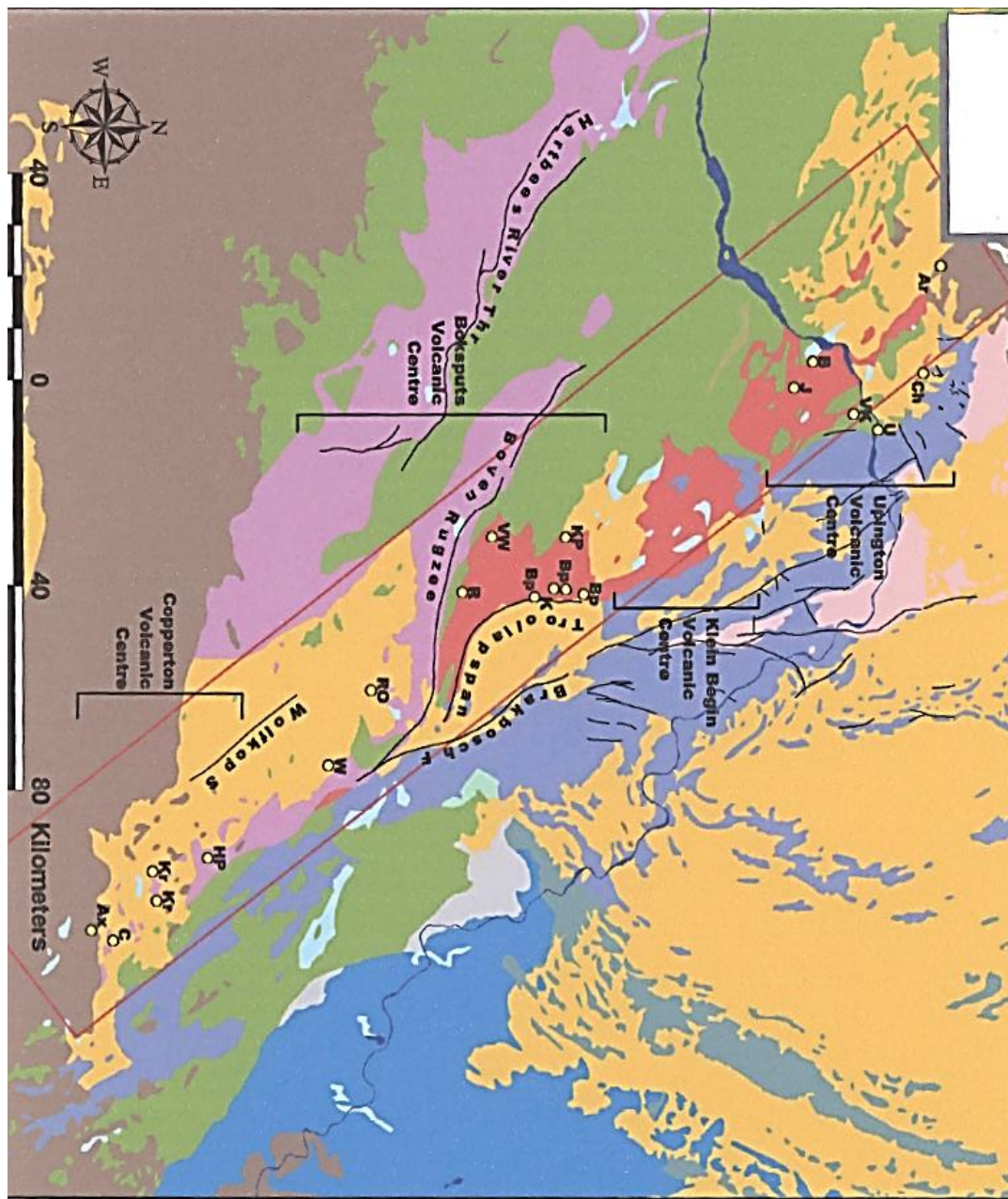
۲-۱ مقدمه

گروه Areachap، در قسمت شرقی ایالت زمین‌شناسی Namaqua قرار دارد و شامل تعدادی از ذخایر سولفید توده‌ای (VHMS) از قبیل Copperton، Kantienpan و Areachap است. در این فصل از پایان‌نامه مراحل تکامل پوسته شرقی Namaqua بررسی شده و سپس توالی زمین‌شناسی ناحیه‌ای و محلی در مورد منطقه مورد مطالعه با تمرکز بر روی توالی زمین‌شناسی نزدیک کانسار آورده شده است. [۷].

چون در این پایان‌نامه محیط ثانویه (Regolith) مورد بررسی قرار گرفته است، بنابراین تکامل سطح فرسایش کنونی مورد بحث قرار گرفته است. برای بررسی این موضوع، حوادث ریخت‌شناسی متفاوت که قاره آفریقا را از زمان مزوزوئیک تا زمان حاضر تحت تاثیر قرار داده است، مورد بررسی قرار خواهد گرفت. فرآیند ایجاد کالکریت اهمیت دو چندان دارد، زیرا تاثیر بسزایی بر روی پراکندگی ثانویه عناصر تشکیل دهنده کانی‌سازی داشته است. در بخش دوم این فصل کارهای اکتشافی در منطقه مرور شده است. بررسی کانسارهای سولفید توده‌ای قسمت آخر این فصل را تشکیل می‌دهد.

۲-۲ موقعیت مکانی معدن

معدن متروکه مس-روی Areachap در مزرعه ۴۲۶ Areachap، در ۲۲ کیلومتری شمال غربی Upington از ناحیه Gordonia قرار دارد. این منطقه دارای توپوگرافی مسطح و ارتفاع متوسط آن از سطح آزاد دریا ۹۱۵ تا ۹۲۲ متر است. منطقه مورد مطالعه به راحتی توسط راه شوسه قابل دسترسی است. یک جاده شوسه به طول ۱۲ کیلومتر معدن Areachap را به جاده Upington-Nambia وصل می‌کند. شکل (۱-۲) به صورت کلی موقعیت مکانی منطقه مورد مطالعه را نشان می‌دهد [۷].



شكل (٢-١): موقعيت مكانى منطقه مورد مطالعه [٨]

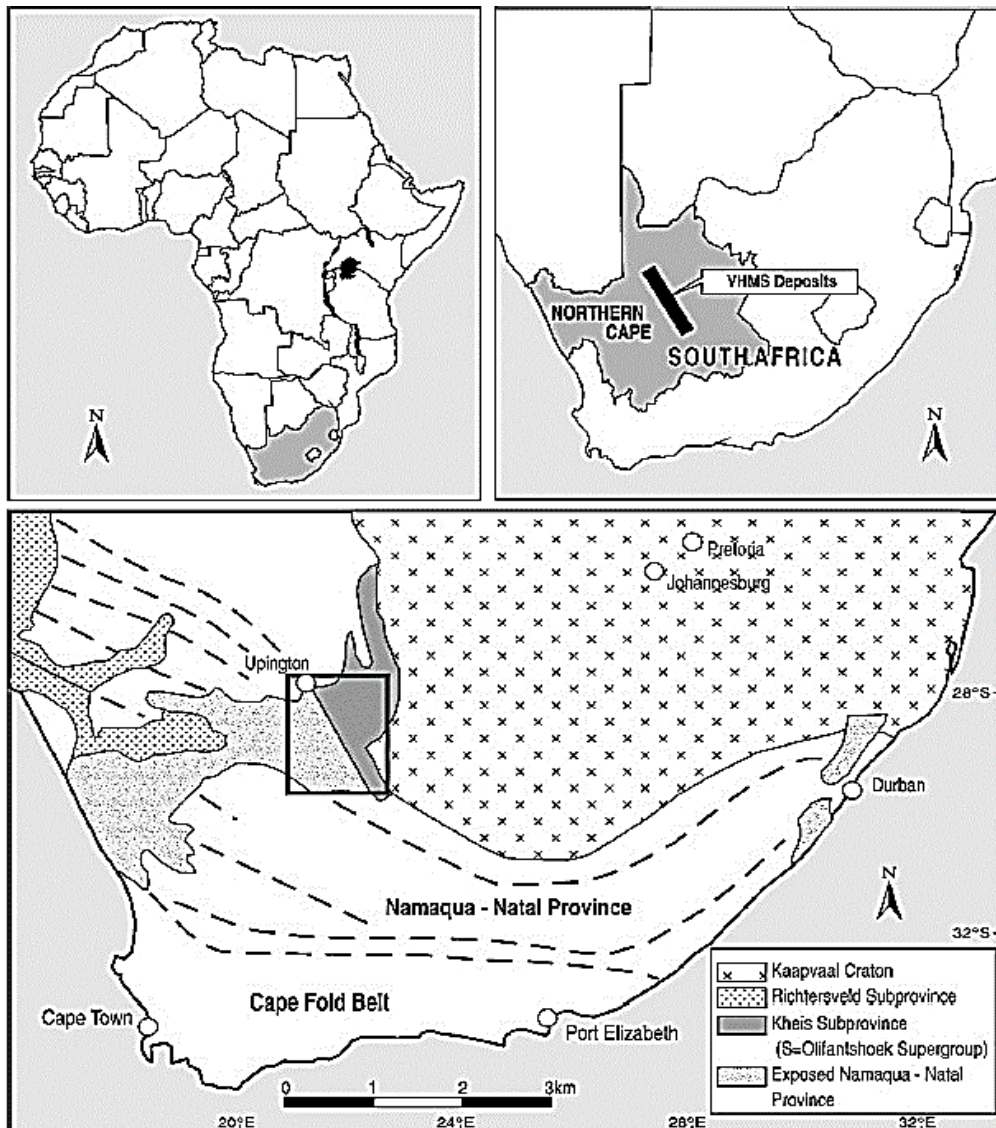
۲-۳ زمین‌شناسی ناحیه‌ای

سکانس فوق پوسته‌ای Kiabaran (۱۳۰۰-۱۶۰۰ Ma) بر روی حوضه Eburnian (۲۰۰۰ Ma) در غرب Namaqua ایجاد شد. سنگ‌شناسی این سکانس ممکن است نشان‌دهنده وجود یک حوضه اقیانوسی بین Kaapvaal و قسمت‌های قدیمی‌تر ایالت Namaqua در حدود ۱۶۰۰ Ma قبل باشد. این حوضه اقیانوسی در ناحیه‌ای که امروزه در شرق مرکز ایالت Namaqua-Natal قرار دارد (شکل ۲-۲)، تحت تاثیر فعالیت‌های آتش‌فشانی کالک‌آلکالن و تولییتی قرار گرفته است. سنگ‌های آذرین بیرونی و رسوبات مرتبط با این رخدادها در گروه Areachap در شرق ایالت Namaqua و در گروه Mfongosi در ایالت Natal حفظ شده است. در این دوره و در طول فعالیت‌های آتش‌فشانی، سولفید توده‌ای در نتیجه واکنش بین سیالات هیدروترمال و آب دریا ایجاد شد. سپس فرورانش صفحات، بالاراندگی، افزایش ضخامت و تغییر شکل شدید به دلیل وجود یک نظام تنش در راستای شمال غرب- جنوب شرق، اتفاق افتاد. گروه‌های Koras و Sinclair که شامل سنگ‌های کالک‌آلکالن آتش‌فشانی و رسوبی است در طول رخداد برخورد بین صفحات در حدود ۱۱۵۰ Ma تشکیل شد [۹]. برخورد صفحات کراتون و حوادث مرتبط باعث دگرگونی درجه بالا و ذوب گسترده و تشکیل ماگمای باتولیتی گرانیتوئید (گرانیت نوع I) بین ۱۲۰۰ Ma تا ۱۰۰۰ Ma سال قبل شد.

ناحیه در نظر گرفته شده در این پایان‌نامه، درون زیر ایالت Gordonia از ایالت Namaqua قرار دارد. سنگ‌های فوق پوسته‌ای در این زیر ایالت متعلق به گروه Areachap است. گروه Areachap متشکل از تشکیلات زمین‌شناسی Bokspits، Jannelspan و Copperton است. Middelton (۱۹۷۶) و بعدتر Geriner (۱۹۹۴) به این نتیجه رسیدند که تشکیلات گروه زمین‌شناسی Areachap در یک بازه زمانی و در نتیجه فعالیت‌های آتش‌فشانی مناطق متفاوت ایجاد شده‌اند [۱۰][۱۱].

تشکیلات زمین‌شناسی Jannelspan که ذخیره Areachap در آن واقع شده است، از چهار عضو زمین‌شناسی تشکیل شده است. این چهار عضو شامل Swartkop، Quarry، Dokerkoekspruit و

Skietbaan است. Swartkop از نظر سنگ‌شناسی از دیوپسیدهای پرفیروبلاستیک حاوی آمفیبولیت تشکیل شده است. Quarry متشکل از گنایس‌های بیوتیتی لایه‌ای است. Dockerkoekspruit از گنایس‌های هورنبلند-بیوتیت و بیوتیت-هورنبلند تشکیل شده و Skietbaan شامل گنایس کوارتز-فلدسپات تشکیل شده است [۶]. شکل (۲-۲) نقشه زمین‌شناسی عمومی منطقه را نشان می‌دهد.



شکل (۲-۲): زمین‌شناسی ناحیه‌ای منطقه [۱۲]

۲-۴ زمین شناسی محلی

۲-۴-۱ توالی سنگ‌شناسی در معدن Areachap

واحدهای زمین‌شناسی نزدیک کانسار سولفید توده‌ای در نقشه زمین‌شناسی کانسار Areachap در شکل ۲-۳ نشان داده شده است. سنگ‌های مهم تشکیلات Jannelsepan در معدن Areachap توسط Voet و King (۱۹۸۶) مورد بررسی قرار گرفته است. این دو محقق بر این باورند که توالی سنگ‌ها به دلیل چین‌خوردگی هم‌خواب دارای تکرار شونده است [۱۳].

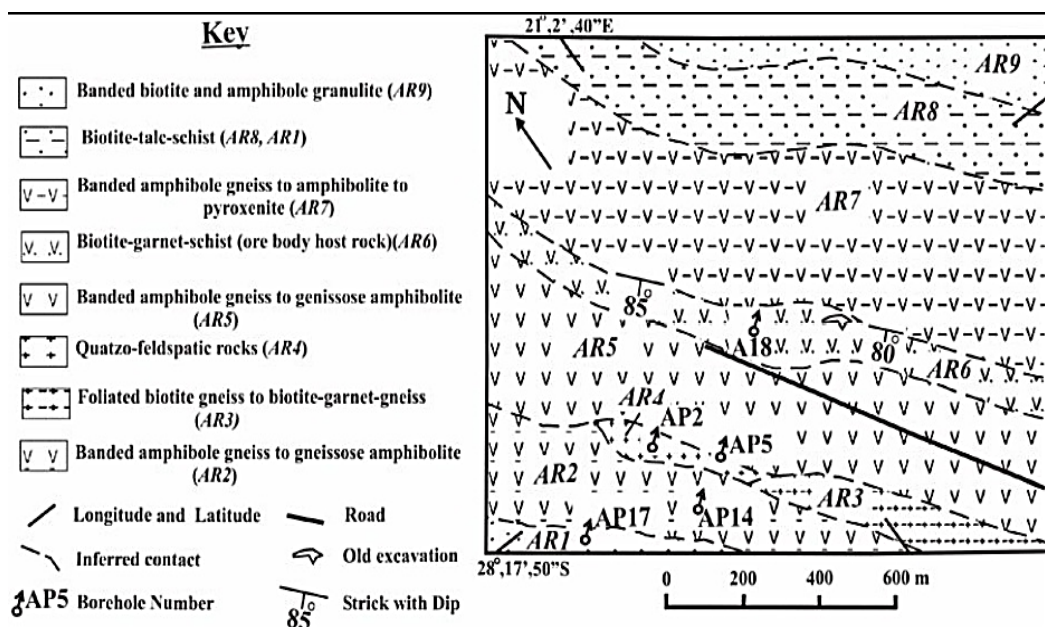
واحد زمین‌شناسی بیوتیت-گارنت گنایس (واحد AR₃ در نقشه زمین‌شناسی در شکل ۲-۳)، از کوارتز، پلاژیوکلازهای به شدت اشباع شده، میکروکلین، بیوتیت، گارنت، کوردیریت، سیلیمانیت و کانی‌های ثانویه از قبیل بیوتیت سبز، کلریت، شیست سبز تشکیل شده است. سنگ‌ها یک حالت ورقه‌شده شدن زیر میکروسکوپ از خود نشان می‌دهند که این حالت بیشتر در کوارتز و بیوتیت دیده می‌شود. واحد زمین‌شناسی گنایس کوارتز-فلدسپات (واحد AR₄ در نقشه زمین‌شناسی در شکل ۲-۳)، از ریولیت و یا لایه آرکوز دگرگون تشکیل شده است.

در گنایس هورنبلند-بیوتیتی لایه‌ای (واحد زمین‌شناسی AR₅)، لایه‌بندی ناشی از دگرسانی هورنبلند سبز به پلاژیوکلاز در آمفیبولیت و دگرسانی لایه‌های بیوتیت به پلاژیوکلاز و زیاد بودن مقدار کوارتز در گنایس‌های غنی از بیوتیت است. کانی‌شناسی شامل هورنبلند سبز، پلاژیوکلاز (لابرادوریت) و اسفن است. لایه‌های غنی از بیوتیت شامل کوارتز، پلاژیوکلاز، بیوتیت، هورنبلند، آپاتیت و اپیدوت است. کانی‌شناسی نواحی دارای کانی‌های روشن پلاژیوکلاز، کوارتز و میکا به همراه کانی‌های ثانویه از قبیل میکای دانه ریز، کلریت و اپیدوت است.

گنایس‌های آمفیبولیتی لایه‌ای تا آمفیبولیت واحد زمین‌شناسی AR₂ را مشخص می‌کند. Theart (۱۹۸۹) در بررسی‌های خود واحدهای AR₂ و AR₅ را یکی در نظر گرفته است [۱۴]. این واحدها را

می‌توان یک نوع از گنایس داسیتی اسموسپاین^۱ از تشکیلات Copperton در نظر گرفت. گنایس نشان-دهنده کمر پایین اصلی در معدن مس-روی Prieska است.

گنایس گارنت-بیوتیت-کوردیریت واحد زمین‌شناسی AR₆ مشخص کرده است. این واحد غنی از آلومینیم، سنگ میزبان کانسار سولفید توده‌ای حاوی بیوتیت به شدت دگرسان شده است. سنگ‌های غنی از کوردیریت در قسمت جنوب غربی کانسار نسبت به قسمت جنوب شرقی آن فراوان‌تر است. ناحیه کانی‌سازی به وسیله دو توده عدسی شکل مشخص شده‌اند. لایه‌های سولفید توده ای شامل کانی‌های سولفیدی از قبیل پیروتیت، پیریت و به مقدار کم تر اسفالریت، کالکوپیریت و گالن است.



شکل (۲-۳): زمین‌شناسی منطقه Arechap [v]

دومین واحد از گنایس هورنبلند-بیوتیت لایه‌ای واحد زمین‌شناسی AR₇ را مشخص می‌نماید. این واحد سنگ‌شناسی در نزدیکی گنایس گارنت-کوردیریت-بیوتیت در قسمت جنوب شرقی کانسار سولفید توده ای قرار دارد. در مقایسه این واحد سنگ‌شناسی با گنایس‌های هورنبلند-بیوتیت قسمت جنوب

¹ - Dacitic Smouspan Gneiss

غربی کانسار مشاهده شد که از نظر گارنت و آمفیبولیت غنی تر است. کانی شناسی لایه‌های آمفیبولیت شامل هورنبلند، پلاژیوکلاز، بیوتیت، کوارتز و اسفن است.

گنایس فلیتیک لایه‌ای واحد زمین‌شناسی AR₈ و AR₁ هستند. این واحد سنگ‌شناسی، تنها گنایس حاوی بیوتیت در این ناحیه است که هورنبلند ندارد. گنایس فلیتیک لایه‌ای مشابه سنگ‌های میزبان توده کانسار سولفیدی است. کانی پیریت به صوت پراکنده درون بعضی از لایه‌های بیوتیت شایسته این واحد قرار دارد. کانی‌شناسی این واحد شامل کوارتز، پلاژیوکلاز، بیوتیت، فیبرولتیک، سیلیمانیت، گارنت، کامینگتونیت و پیریت است.

گنایس بیوتیت ورقه‌ای واحد زمین‌شناسی AR₉ است. این واحد سنگ‌شناسی در قسمت شمال شرق توده معدنی قرار دارد. کانی‌شناسی این واحد شامل آندزیت، کوارتز، بیوتیت و هورنبلند بدون حضور پتاسیم فلدسپار است [۷].

۲-۴-۲ دگرسانی و دگرگونی در معدن Areachap

مجموعه کانی‌های سنگ‌های Areachap مانند فلدسپار، گارنت، کوردیریت و سیلیمانیت در واحد AR₆ نشان‌دهنده دگرسانی درجه بالا است. وجود شکستگی‌های فراوان نفوذپذیری سنگ‌ها را بالا برده و باعث شده است که سیالات آبی به طور گسترده کانی‌های فاز اولیه را به مسکویت-کلریت و اپیدوت دگرسان کنند. این سیالات باعث دگرگونی پس‌گرا (قهقرایی) شده و تا قسمتی مجموعه کانی‌های درگرسان شده اولیه را از بین برده‌اند. کانی‌های شاخص برای دگرگونی پس‌گرا می‌تواند شامل تبدیل فینیت به کوردیریت (AR₆) و کلریت به بیوتیت (در واحدهای AR₅ و AR₇) باشد [۷].

۲-۴-۳ کانی‌شناسی ذخیره Areachap

همانگونه که قبلاً اشاره شد منطقه مورد مطالعه دچار دگرگونی شده است. کم‌رپاین کانسار Prieska از هورنبلند-بیوتیت-گنایس تشکیل شده است. با در نظر گرفتن ویژگی‌های سنگ‌شناسی، ژئوشیمی

عناصر اصلی و ردیاب و بررسی‌های ایزوتوپی به این نتیجه رسیده‌اند که سنگ اولیه این کانسار داسیت می‌باشد که دارای شاخصه‌های آلتراسیون ذخایر سولفید توده‌ای است. کم‌پایین نواحی دگرسانی در کانسار Areachap مشابه به کانسار Prieska است. این نواحی دگرسانی با کاهش مقدار کلسیم و افزایش مقدار منیزیوم همراه بوده که معادل‌های دگرگونی سنگ‌های این دگرسانی مجموعه‌ای از کانی‌های فلوگوپیت، آمفیبولیت و یا انستاتیت، کوردیریت‌های غنی از منیزیوم، آلماندیت و گارنت‌های غنی از پیروپ تشکیل شده‌اند. سنگ‌های در برگیرنده این نواحی دگرسانی از سنگ‌های رسوبی - آتشفشانی دگرگون شده، دارای بیوتیت، گارنت‌های غنی از گروسولار، هورنبلند و یا اوزیت و کوردیریت‌های غنی از آهن هستند [۱۲].

کانی‌سازی سولفیدی در کم‌پایین این ذخایر از کالکوپیریت، پیروتیت، پیریت و اسفالریت تشکیل شده است. نواحی کانی‌سازی دارای پیریت، اسفالریت، پیروتیت، کالکوپیریت، باریت، انیدرید و مقادیر کم گالن است. مگنتیت، پیریت و اسفالریت در توالی کمر بالای این ذخایر حضور دارند [۱۲].

۲-۵ محیط‌های کالکریتی

کالکریت یک سنگ رسوبی و سیمان طبیعی سخت شده از کربنات کلسیم است که مواد دیگری از قبیل گراول، ماسه، رس و سیلت را به هم متصل می‌کند. کالکریت در مناطق دارای آب و هوای خشک تا نیمه خشک تشکیل می‌شود. کالکریت می‌تواند به عنوان یک سد ژئوشیمیایی تاثیر به سزایی در پراکندگی ثانویه عناصر داشته باشد.

کالکریت عمدتاً جزء کلسیم کربنات است که در نزدیکی سطح زمین نهشته می‌شود. کالکریت یا در نتیجه فعالیت‌های مربوط به تشکیل خاک و یا فعالیت‌های مربوط به آب‌های زیرزمینی و یا ترکیبی از این دو عامل ایجاد می‌شود. منبع کربنات ممکن است که گرد و غبار مربوط به دوره ائولین، بارش باران، گیاهان و یا هوازدگی مواد دارای کربنات باشد. کالکریت ممکن است در آب و هوای خشک (ریزش سالیانه باران بین ۵۰-۱۰۰ میلیمتر) تا نیمه مرطوب (ریزش سالیانه باران بین ۵۰۰-۷۰۰ میلیمتر)

تشکیل شود. وجود کالکریت در یک منطقه نشان‌گر دوره‌های خشکسالی است که در آن تبخیر از ریزش باران بیشتر بوده است. رسوبات کالکریتی که بیشتر از کوارتز تشکیل شده باشند، براساس درجه دیاژنز ماسه کالکریتی یا ماسه سنگ نامیده می‌شود.

کالکریت‌ها به طور عمده از کلسیت، کوارتز و فلدسپار تشکیل شده‌اند. میزان سیلیکای موجود در کالکریت بر سختی آن تاثیر چشمگیری دارد. هر چه مقدار سیلیکا بیشتر باشد، کالکریت سخت‌تر است. آراگونیت و کانی‌های کلسیم کربنات ممکن است در مراحل اولیه تشکیل کالکریت حضور داشته باشند. Kkhadkikar از حضور کانی‌های رسی و مقدار کم ژپس و باریت در کالکریت خبر می‌دهد.

در طول بارش باران آب اسیدی به همراه دی اکسید کربن حل شده از آب رودخانه و گرد و غبار به درون خاک نفوذ می‌کند. در صورتی که یک توده معدنی در نزدیک سطح حضور داشته باشد، آب حاوی کربنات با توده معدنی واکنش می‌دهد و در نتیجه عناصر متحرک در توده معدنی در آب حل می‌شوند. سپس در طول فرآیند تبخیر و در نتیجه آن، محلول به دلیل خاصیت موئینگی به طرف سطح زمین بالا می‌آید. آب و دی اکسید کربن کم باعث افزایش pH محلول می‌شود. ممکن است pH به حد مطلوب برای تشکیل کلسیم کربنات برسد و در نتیجه کلسیت به همراه دولومیت، کوارتز، اکسیدهای آهن و دیگر عناصر در سطح زمین نهشته شود.

قسمت‌های بزرگی از مناطق خشک مرکزی شمال-غرب ایالت Northern Cape در آفریقای جنوبی با لایه‌های کالکریتی پوشیده شده است. لایه‌های کالکریت موجود در زیر گروه Kalahari در طول یک دوره خشکسالی و در اواخر دوره ترشیاری وجود داشته‌اند. لایه کالکریتی تشکیل شده در نزدیکی سطح در طول ناحیه مورد مطالعه از Areachap در شمال تا Copperton در جنوب گسترده شده است [۷].

۲-۴ پیشینه کارهای اکتشافی در منطقه

کانسار سولفید توده‌ای مس-روی Areachap و معدن روی-مس Prieska در روی کمر بند مرکزی منطقه Arechap است. کانسار Areachap در حدود سال‌های ۱۸۸۵ به دلیل تشخیص گوسان در این منطقه کشف شد. حفاری‌ها در سال ۱۹۶۰ نشان داد که یک توده سولفید توده‌ای با ذخیره ۱/۶ میلیون تن در اعماق ۲۰۰ تا ۳۰۰ متری وجود دارد. ادامه پروژه اکتشافی در این ناحیه به دلیل کشف معدن Prieska به تاخیر افتاد [۸].

در سال ۱۹۹۵ کل منطقه Arechap برای اکتشاف کانسارهای سولفید توده‌ای مورد کاوش قرار گرفت. در این پی‌جویی از روش‌های لیتوژئوشیمیایی، ژئوشیمی خاک و از روش‌های مغناطیس هوابرد و زمینی و رادیومتریک و گرانی‌سنجی استفاده شد [۸].

کارهای اکتشافی در منطقه Areachap در دهه‌های ۱۹۷۰ و ۱۹۸۰ توسط شرکت‌های مختلفی مانند Phelps Dodge و Anglo American انجام شده است. علی‌رغم این فعالیت‌های اکتشافی، کمر بند مرکزی منطقه کم‌تر مورد کاوش قرار گرفته است [۸].

هدف‌های اکتشافی در کارهای اکتشافی در این دوره زمانی همگی کانسارهای سطحی و نیمه عمیق به وسیله پی‌جویی‌های ژئوشیمیایی خاک/گوسان و سپس پی‌جویی‌های ژئوفیزیکی (مغناطیس، IP، EM و گرانی‌سنجی) بوده است. حفاری‌های انجام شده همگی با عمق کم بوده است. پی‌جویی‌های الکترومغناطیس و مغناطیس هوابرد در بعضی مواقع به دلیل برداشت موازی و نه عمود بر توده ماده معدنی، موفق به ثبت آنومالی نشده‌اند. اگرچه در بعضی از کانسارها، مانند کانسار Prieska این روش‌ها موفقیت‌آمیز بوده است [۱۵].

کارهای پیشین اکتشافی در کانسار Areachap توسط شرکت Iscor با هدف استخراج مقادیر فراوان پیریت از توده Areachap برای تولید اسید سولفوریک انجام شده است [۸].

۲-۵ کانسارهای سولفید توده‌ای

کانسارهای سولفید توده‌ای به کانسارهایی گفته می‌شود که در اثر فعالیت‌های آتش‌فشانی در کف اقیانوس‌ها ایجاد می‌شوند. بیش از ۶۰٪ حجم این کانسارها را کانی‌های سولفیدی تشکیل می‌دهد. کانسارهای سولفید توده‌ای را در حالت کلی به دو دسته تقسیم می‌کنند:

۱- آتش‌فشانی

۲- رسوبی اگزهالتیو (نوع سدکس)

تفاوت‌های بین این دو نوع کانسار سولفید توده‌ای در این است که کانسارهای نوع اول معمولاً مرتبط با کانی‌زایی مس-روی است. سنگ میزبان آن از نوع سنگ‌های آتش‌فشانی است و با فعالیت‌های آتش‌فشانی در زیر سطح دریا ارتباط دارند. در حالی که کانسارهای نوع سدکس اکثراً با کانی‌زایی روی-قلع مرتبط بوده و سنگ میزبان رسوبی دارند و برای ایجاد سیستم هیدروترمال به حرارت ایجاد شده ناشی از عمق متکی هستند [۱۶].

بین کانسارهای سولفید توده‌ای و سنگ‌های آتشفشانی زیردریایی نوعی همزمانی وجود دارد به همین علت بصورت ذخیره‌های عدسی شکل و یا صفحه‌ای در بین واحدهای ولکانیکی و یا واحدهای ولکانیک-رسوبی قرار گرفته‌اند. این کانسارها عمدتاً بافت توده‌ای دارند و دارای کانی‌شناسی ساده شامل: پیریت، کالکوپیریت، اسفالریت \pm گالن \pm پیروتیت است [۱۷].

کانسارهای سولفید توده‌ای از نظر محیط تکتونیکی که در آن تشکیل می‌شوند، به سه دسته تقسیم می‌شوند [۱۷]:

۱- کانسارهای ماسیو سولفید نوع کروکو

۲- کانسارهای ماسیو سولفید نوع قبرس

۳- کانسارهای ماسیو سولفید نوع بشی

کانسارهای سولفید توده‌ای تیپ کروکو در کمربندهای آتش‌فشانی زیردریایی مرتبط با زون فرورانش جزایر قوسی یا زون بازشدگی پشت کمر بند جزایر قوسی تشکیل می‌شوند. محیط تکتونیکی تشکیل کانسارهای سولفید توده‌ای تیپ قبرس مناطق زون گسترش پوسته‌های اقیانوسی است. این نوع ذخایر ابتدا در بخش فوقانی افیولیت‌ها کشف گزارش شدند. سولفیدهای توده‌ای تیپ بشی همراه با مجموعه‌ای از سنگ‌های رسوبی آواری و سنگ‌های آتش‌فشانی و کربنات‌ها در محیط تکتونیکی زون کششی پشت جزایر اقیانوسی تشکیل می‌شوند.

کانسارهای سولفید توده‌ای آتش‌فشانی از دوره پرکامبرین تا عهد حاضر، در زمانی که فعالیت‌های آتش‌فشانی زیردریایی گسترده بوده و شرایط لازم از قبیل pH-Eh مناسب در آب دریا، عمق، توپوگرافی و ترکیب شیمیایی مناسب آب و سنگ فراهم بوده، تشکیل شده‌اند. سولفیدهای توده‌ای منبع بسیار مهمی برای فلزات بنیادی Cu,Zn,Pb و عناصر با ارزشی نظیر Au,Ag هستند.

کانسارهای سولفید توده‌ای دارای دو بخش اصلی هستند:

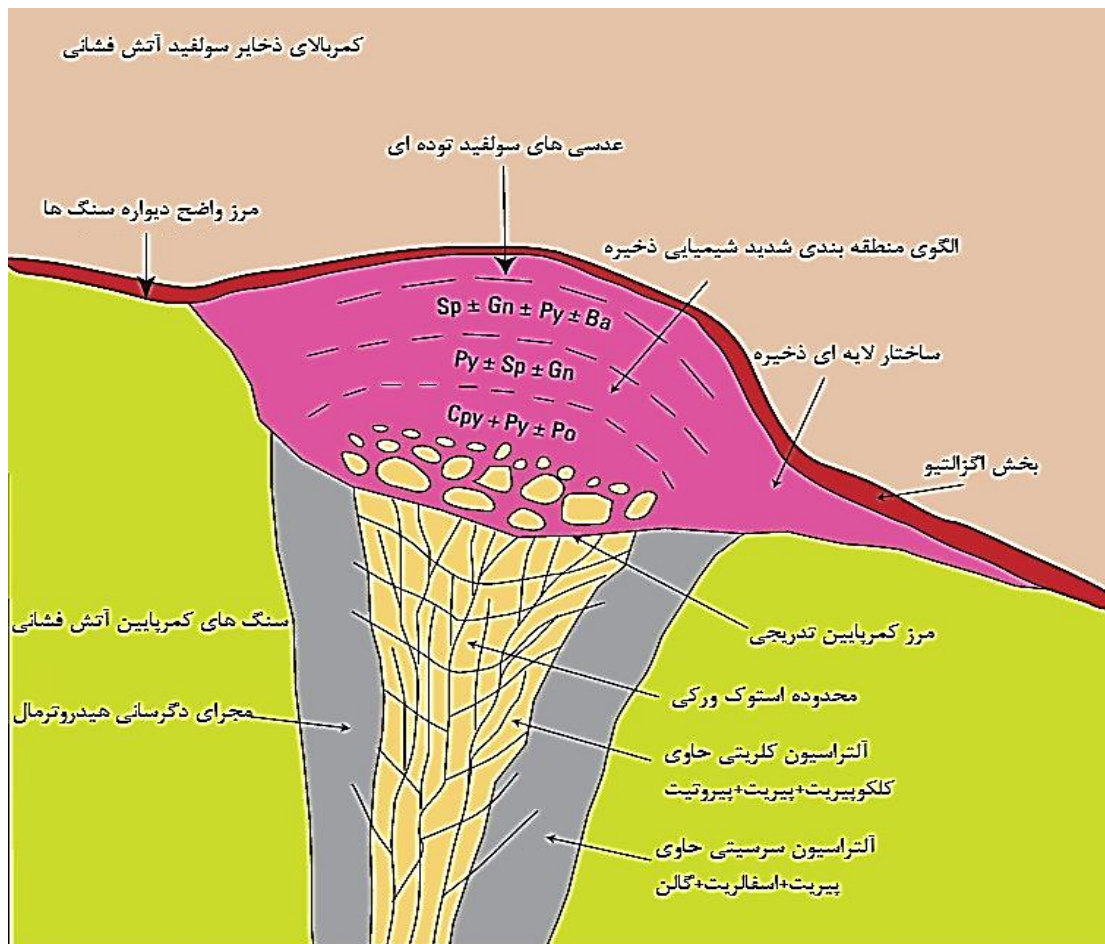
۱- بخش زیرین کانی‌سازی به حالت استوک‌ورک و افشان که در داخل سنگ‌های آتش‌فشانی تشکیل می‌شود. به این بخش از کانی‌سازی زون استرینگ (Stringer Zone) نیز گفته می‌شود (شکل ۲-۴).

۲- کانی‌سازی بخش توده‌ای که شامل نازک لایه‌های سولفید و سایر کانی‌هاست. سولفیدها بسیار ریزدانه هستند و گاهی اشکال گل کلمی دارند. بخش اعظم کانی‌سازی را زون توده‌ای تشکیل می‌دهد.

در بخش عدسی شکل ذخیره، یک زون‌بندی شیمیایی مشاهده می‌شود. در بخش بیرونی اسفالریت ± گالن ± پیریت ± باریت، در بخش میانی پیریت ± اسفالریت ± گالن و در بخش درونی

کالکوپیریت + پیریت ± پیروتیت مشاهده می‌شود (شکل ۲-۴). زون‌های آلتراسیون اطراف ذخایر سولفید

توده‌ای عبارت از زئولیتی، سرسیتی-کلریتی، کلریتی است (شکل ۲-۴) [۱۷].



شکل (۲-۴): نمایی شماتیک از کانسار سولفید توده‌ای [۱۶]

فصل سوم:

پیشینه تحقیق و تئوری روش‌ها

۳-۱ مقدمه

روش‌های بسیاری برای پردازش داده‌های ژئوشیمیایی تاکنون مورد استفاده قرار گرفته است. این روش‌ها را به طور کلی می‌توان به دو دسته تقسیم کرد.

۱- روش‌هایی که توزیع فراوانی داده‌ها را با استفاده از ابزارهایی مانند هیستوگرام و یا نمودارهای احتمال مورد بررسی قرار می‌دهند (روش‌هایی بر پایه فراوانی).

۲- روش‌هایی که علاوه بر فراوانی، مکان نمونه‌ها را نیز در نظر می‌گیرند (روش‌هایی بر پایه فراوانی- مکان). روش‌های آمار فضایی U و روند سطح نمونه‌های بسیار معروفی از این دسته است [۱۸].

داده‌های ژئوشیمیایی، داده‌هایی چندمتغیره و با ابعاد بسیار بزرگ هستند. قبل از اعمال روش‌های یاد شده در بالا، معمولاً با استفاده از روش‌های مانند تحلیل مولفه‌های اصلی و یا تحلیل فاکتوری ابعاد داده‌ها را کاهش می‌دهند و سپس این روش‌ها را بر روی داده‌ها اعمال می‌کنند.

در این پایان‌نامه از روش تحلیل مولفه‌های اصلی برای کاهش ابعاد داده‌ها و از روش نمودار احتمال برای مدل‌سازی آن استفاده شده است. در این فصل شرح مختصری از تئوری این روش‌ها به اضافه‌ی توضیحی مختصر از هندسه سیمپلکس و تبدیل‌های نسبت لگاریتمی^۱ آورده شده است.

۳-۲ تحلیل مولفه‌های اصلی

داده‌های ترکیبی در اصل داده‌های چندمتغیره هستند، بنابراین روش‌هایی که برای بررسی این داده‌ها استفاده می‌شود، باید بتواند تغییرات در چند متغیر را به صورت همزمان مورد مطالعه قرار دهد. برای این تحقیق از یکی از روش‌های معمول آنالیز چند متغیره یعنی تحلیل مولفه‌های اصلی استفاده شده است. در این روش داده‌ها در جهت بیشترین مقدار واریانس تصویر می‌شوند.

¹ - Logratio

برای یک مجموعه داده مفروض، واریانس به صورت میانگین مربع انحراف تمامی داده‌ها از میانگین حسابی تعریف می‌شود [۱۹]. به عبارت دیگر واریانس معیاری برای نشان دادن پراکندگی داده‌ها حول میانگین است و مطابق معادله (۳-۱) محاسبه می‌شود:

$$s(x) = \sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} \quad (۱-۳)$$

که در آن \bar{x} میانگین داده‌ها توسط رابطه (۳-۲) به دست می‌آید:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad (۲-۳)$$

و σ نشان‌دهنده انحراف از معیار است.

مفهوم میانگین حسابی را برای یک مجموعه داده چندمتغیره می‌توان به صورت زیر تعریف کرد:

$$\bar{x} = \begin{bmatrix} \frac{1 \sum x_i}{n} \\ \frac{1 \sum y_i}{n} \\ \frac{1 \sum z_i}{n} \end{bmatrix} \quad (۳-۳)$$

به همین ترتیب می‌توان مفهوم واریانس را برای یک مجموعه داده چندمتغیره به ماتریس واریانس گسترش داد که در آن قطر ماتریس نشان‌دهنده واریانس مختصات می‌باشد.

کوواریانس، پراکندگی مشترک دو متغیر حول میانگین‌هایشان است. برای تعریف کوواریانس ابتدا فرمول واریانس را براساس رابطه (۳-۴) می‌نویسیم:

$$var(x) = \frac{\sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})}{n-1} \quad (۴-۳)$$

با استفاده از تعریف بالا کوواریانس مطابق معادله (۳-۵) تعریف می‌شود:

$$cov(x, y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n-1} \quad (۵-۳)$$

ساختار کوواریانس یک مجموعه داده چندمتغیره به وسیله ماتریس کوواریانس بیان می‌شود. برای

مثال برای یک مجموعه داده با سه متغیر ماتریس کوواریانس مطابق رابطه (۳-۶) بیان می‌شود:

$$\text{cov}(X) = \begin{bmatrix} \text{var}(x, x) & \text{cov}(x, y) & \text{cov}(x, z) \\ \text{cov}(y, x) & \text{var}(y, y) & \text{cov}(y, z) \\ \text{cov}(z, x) & \text{cov}(z, y) & \text{var}(z, z) \end{bmatrix} \quad (۶-۳)$$

در حالت کلی یک تبدیل خطی داده‌ها را منتقل و می‌چرخاند. برای تمام تبدیلات خطی جهت‌های خاصی وجود دارد که داده‌ها تنها منتقل می‌شوند و دوران پیدا نمی‌کنند. بردارهایی که هم‌راستای چنین جهت‌هایی باشند، پس از اعمال تبدیل خطی تنها در مقدار بزرگی بردار تغییر حاصل می‌شود و تغییری در جهت آن ایجاد نمی‌شود. به چنین بردارهایی، بردارهای ویژه گفته می‌شود. این ویژگی را مطابق معادله (۷-۳) نشان داد:

$$f(x) = \lambda x \quad (۷-۳)$$

که در آن x بردار ویژه و کمیت اسکالر λ مقدار ویژه مربوط به آن است. هنگامی که تحلیل مولفه‌های اصلی بر روی داده‌ها اعمال می‌شود، داده‌ها به گونه‌ای تصویر می‌شوند که محور جدید موازی بردار ویژه ماتریس کوواریانس باشد [۱۹]. به عبارت دیگر داده‌ها در جهتی تصویر می‌شوند که واریانس‌ها بیشینه است. در این جهت خاص، مقدار ویژه متناسب با بزرگی واریانس است. بنابراین بردار ویژه با بزرگترین مقدار ویژه در جهت بیشترین مقدار واریانس قرار دارد و به همین ترتیب بردار ویژه با مقدار ویژه کمتر در جهت مقدار واریانس کمتر قرار دارد.

به عنوان مثالی از این فرآیند، مجموعه داده غیرترکیبی X را در نظر بگیرید (معادله ۸-۳):

$$X = \begin{bmatrix} 3 & 2 \\ 4 & 10 \\ 6 & 5 \\ 6 & 8 \\ 6 & 10 \\ 7 & 2 \\ 7 & 13 \\ 8 & 9 \\ 9 & 5 \\ 9 & 8 \\ 9 & 14 \\ 10 & 7 \\ 11 & 12 \end{bmatrix} \quad (۸-۳)$$

اولین قدم در مرکز قرار دادن مجموعه داده‌ها با کم کردن میزان میانگین هر یک از متغیرها از داده‌های مربوط به آن متغیر می‌باشد. میانگین مجموعه داده‌های بالا به صورت یک بردار بیان شده است:

$$x = [7.3077 \quad 8.0679] \quad (9-3)$$

با کم کردن میانگین از سطرهاى بردار x داریم:

$$X_c = \begin{bmatrix} -4.3077 & -6.0769 \\ -3.3077 & 1.9231 \\ -1.3077 & -3.0769 \\ -1.3077 & -0.0769 \\ -1.3077 & 1.9223 \\ -1.3077 & -6.0769 \\ -1.3077 & 4.9231 \\ 0.6923 & 0.9231 \\ 1.6923 & -3.0769 \\ 1.6923 & -0.769 \\ 1.6923 & 5.9231 \\ 2.6923 & -1.0769 \\ 3.6923 & 3.9231 \end{bmatrix} \quad (10-3)$$

این مجموعه داده‌ها در مرکز قرار گرفته است، زیرا میانگین آنها برابر $(0,0)$ است. با توجه به توضیحات داده شده در بالا، ساختار کوواریانس داده‌ها مطابق معادله $(11-3)$ محاسبه می‌شود:

$$\text{cov}(X_c) = \begin{bmatrix} 5.3974 & 3.2244 \\ 3.2244 & 14.7436 \end{bmatrix} \quad (11-3)$$

در ماتریس کوواریانس که در معادله $(11-3)$ نشان داده شده است، x_{11} و x_{22} نشان‌دهنده واریانس متغیرهاست در حالی که x_{12} و x_{21} نشان‌دهنده کوواریانس (x,y) است. مشاهده می‌شود کوواریانس x متقارن است زیرا $\text{cov}(x,y) = \text{cov}(y,x)$.

بردار ویژه ماتریس کوواریانس را می‌توان با حل معادله $(12-3)$ به دست آورد:

$$(X_c - \lambda I)x = 0 \quad (12-3)$$

که در آن I ماتریس همانی است.

با حل این معادله داریم:

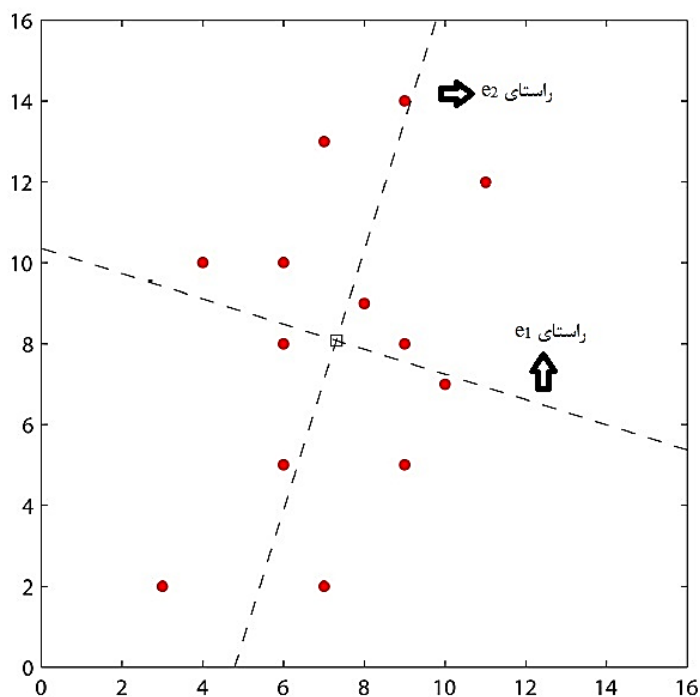
$$e_1 = \begin{bmatrix} -0.9547 \\ 0.2974 \end{bmatrix} \quad e_2 = \begin{bmatrix} 0.2974 \\ 0.9547 \end{bmatrix} \quad (13-3)$$

e_1 و e_2 بردارهای ویژه X است. مقادیر ویژه متناظر با این بردارها در زیر آورده شده است.

$$\lambda_1 = 4.3930$$

$$\lambda_2 = 15.4780$$

شکل (۱-۳) نشان‌دهنده یک اسکتر پلات داده‌های اصلی است که جهت‌یابی بردارهای ویژه را نیز نشان می‌دهد. توجه کنید خطی که دارای جهت تقریباً قائم است در راستای بیشترین مقدار واریانس و هم‌جهت با بردار ویژه e_2 با مقدار ویژه λ_2 برابر با ۱۵/۴۷۸۰ است.

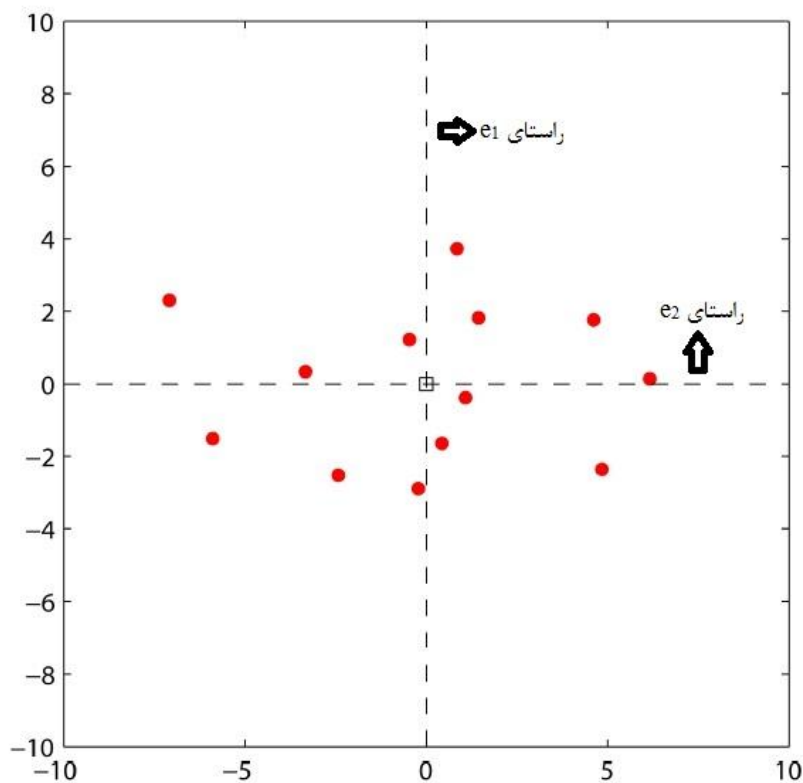


شکل (۱-۳): نمودار پراکندگی داده اصلی [۲۰]

تبدیل داده‌ها به صورت یک ضرب ماتریسی که خواص تبدیل‌های خطی را دارد، انجام می‌شود. در شکل (۲-۳) داده‌ها در مرکز قرار گرفته و در درون یک محور مختصات قرار گرفته‌اند که محورهای آن بردارهای ویژه هستند.

یکی از مهمترین مزایای استفاده از PCA آن است که می‌توان میزان سهم هر یک از داده‌ها را در واریانس کلی به صورت کمی محاسبه کرد. برای مثال واریانس در جهت بردار ویژه e_1 کوچکتر از مقدار

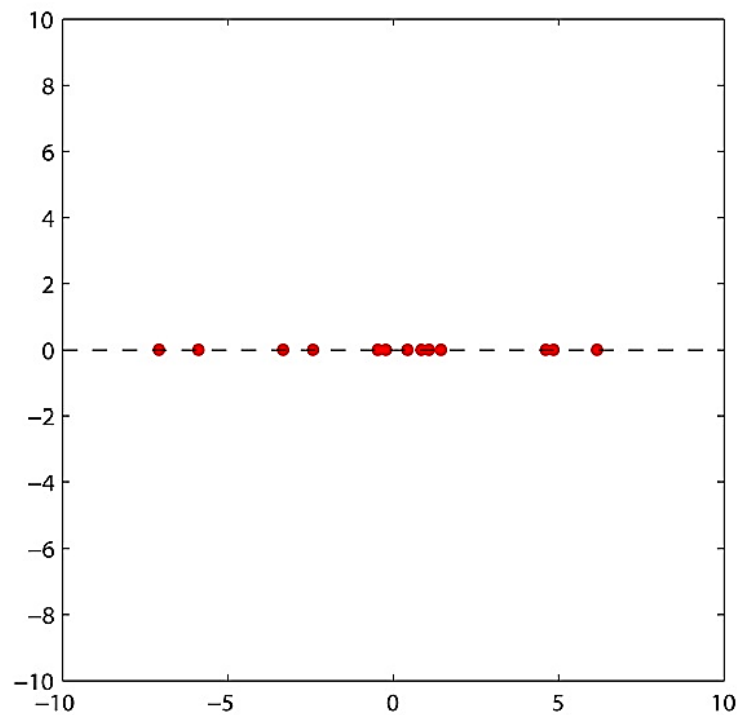
واریانس در جهت بردار ویژه e_2 است، همانگونه که در مقادیر ویژه متناظر با هر یک از این بردارهای ویژه مشاهده می‌شود. بنابراین با نادیده گرفتن جهت‌هایی که دارای مقدار ویژه کوچکتر هستند، می‌توان تعداد و یا ابعاد داده‌ها را کوچکتر کرد. برای مثال مجموعه داده‌های مثال بالا را می‌توان تنها با در نظر گرفتن راستای بیشترین مقدار واریانس e_2 با ضرب ماتریس داده‌ها در ماتریس بردار ویژه e_2 ، تصویر کرد (شکل ۳-۳).



شکل (۳-۲): قرار دادن داده‌ها در راستای مولفه‌های اصلی [۲۰]

شکل (۳-۳) نشان می‌دهد که داده‌ها کاملاً در راستای بردار ویژه e_2 که راستای بیشترین مقدار واریانس است، قرار گرفته‌اند. بردار ویژه دوم نادیده گرفته شده و یک بعد از داده‌ها حذف شده است. با این که در این فرآیند اطلاعات کاهش پیدا می‌کنند، ولی آنچه که حفظ می‌شود به لحاظ پراکندگی (یا به عبارت دیگر بیشترین واریانس) بسیار مهم است. تعمیم دادن این فرآیند بسیار ساده و برای ترسیم داده‌های با ابعاد بزرگ بسیار مناسب است. برای این کار، یک داده با n بعد را می‌توان در راستاهای

کمتری که دارای مقادیر واریانس بزرگتر هستند و با نادیده گرفتن راستاهایی که مقدار واریانس کوچکتری دارند، ترسیم کرد. در عمل برای تجسم داده‌ها، آن را در دو یا سه بعد ترسیم می‌کنند.



شکل (۳-۳): داده‌ها بعد از قرار گرفتن در روی مولفه اصلی [۲۰]

۳-۳ تحلیل مولفه‌های اصلی به روش مقاوم^۱

در هنگام استفاده از تحلیل مولفه‌های اصلی، اگر داده‌های اولیه دارای مقادیر خارج از ردیف باشند، PCهای محاسبه شده دارای خطا خواهد بود. در روش‌های معمول برای جلوگیری از این خطا، در پیش‌پردازش داده‌ها، به روش‌های مختلف داده‌های خارج از ردیف شناسایی شده و حذف یا جایگزین می‌شوند. باید در نظر گرفت که با حذف این داده‌ها، واریانس طبیعی داده‌ها کاهش پیدا می‌کند و مقداری از اطلاعات موجود در آن از بین می‌رود.

¹ - Robust

برای پرهیز از این امر می‌توان از روش‌های آماری مقاوم که در برابر داده‌های حاوی مقادیر خارج از ردیف مقاوم هستند، استفاده کرد [۶]. در مورد روش تحلیل مولفه‌های اصلی، روش تحلیل مولفه‌های اصلی مقاوم معرفی شد تا بتوان داده‌های دارای مقادیر خارج از ردیف را بدون دستکاری مورد بررسی قرار داد.

در روش معمول تحلیل مولفه‌های اصلی از ماتریس کوواریانس داده‌ها و میانگین حسابی داده‌ها به ترتیب به عنوان تخمین‌گر پراکندگی و مکان استفاده می‌شود. در روش مقاوم مینیمم دترمینان کوواریانس (MCD^1) یا تخمین‌گر S مورد استفاده قرار می‌گیرد. در روش MCD یک زیرمجموعه از داده‌ها با حداقل h عضو که مقدار دترمینان کوواریانس آن برابر کمینه باشد، استفاده می‌شود. سپس ماتریس کوواریانس این زیرمجموعه از داده‌ها ضربدر یک مقدار ثابت و میانگین حسابی این زیرمجموعه از داده‌ها به ترتیب به عنوان تخمین‌گر پراکندگی و تخمین‌گر مکان مورد استفاده قرار می‌گیرد [۶].

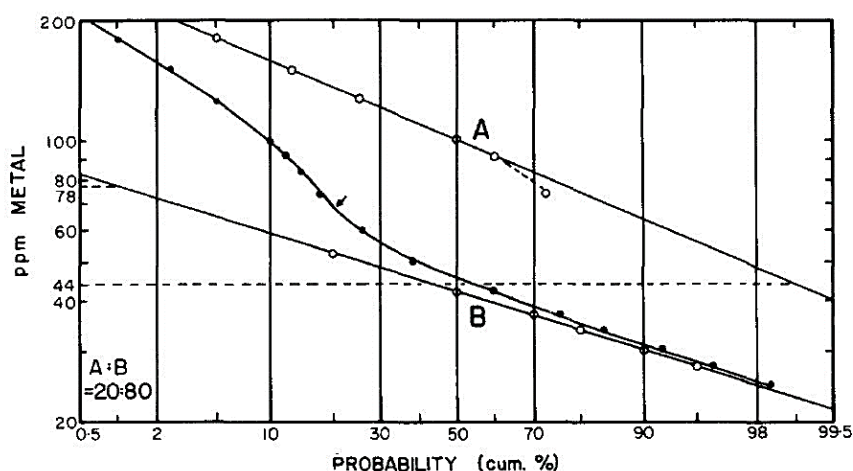
برای انتخاب اندازه h باید در نظر گرفت که با افزایش h میزان مقاوم بودن تخمین‌گر پراکندگی و مکان کاهش پیدا می‌کند ولی کارایی آن افزایش پیدا می‌کند و با کاهش میزان h این موارد به طور عکس تغییر پیدا می‌کنند. بنابراین برای انتخاب میزان h باید هر دو مورد مقاوم بودن در برابر داده‌های خارج از ردیف و کارایی را در نظر گرفت. معمولاً این مقدار را $\frac{3}{4}$ تعداد کل داده‌ها در نظر می‌گیرند. بنابراین با این روش می‌توان داده‌هایی را که تا $\frac{1}{4}$ کل داده‌ها دارای مقادیر خارج از ردیف باشند، مورد بررسی قرار داد. اگر میزان مقادیر خارج از ردیف از این کسر بیشتر شود، نتایج به دست آمده قابل اطمینان نخواهد بود [۶].

¹ - Minimum Covariance Determinant

۳-۴ نمودار احتمال

نمودار احتمال داده‌های یک نمونه را به صورت توزیع تجمعی احتمال نمایش می‌دهد. در این نمودار، محور افقی متغیر x (معمولاً غلظت عنصر مورد نظر) را به صورت خطی و یا لگاریتمی نشان می‌دهد. محور عمودی، یک مقیاس خاص احتمال است که از عکس تابع توزیع نرمال مشتق شده است. مزیت این مقیاس آن است که نمونه‌هایی که از تابع توزیع نرمال پیروی می‌کنند، به صورت یک خط مستقیم در نمودار احتمال به نمایش در می‌آیند. اگر مقیاس محور افقی لگاریتمی باشد، نمونه‌هایی که از توزیع لاگ-نرمال پیروی می‌کنند، به صورت یک خط مستقیم نمایش داده می‌شوند.

تنت و وایت^۱ در سال ۱۹۵۹ بیان کردند که رسم داده‌های ژئوشیمیایی بر روی نمودارهای احتمال باعث نمایش موجز و مفید داده‌های ژئوشیمیایی به صورت گرافیکی می‌شود [۲۱]. در سال ۱۹۷۴ سینکلر^۲ با انتشار مقاله‌ای، به صورت مفصل روشی را برای تعیین حد آستانه‌ای برای داده‌های ژئوشیمیایی توضیح می‌دهد. در این روش در نمودار احتمال، با جدا کردن جامعه‌های مختلف می‌توان حد آستانه‌ای را مشخص کرد [۲۲]. شکل (۳-۴) نمودار احتمال ایده‌آل برای یک نمونه با دو جامعه A و B را نشان داده شده است.



شکل (۳-۴): نمودار احتمال یک نمونه دو جامعه‌ای [۲۲]

¹ Tennant & White

² Sinclair

۳-۶ روند تکامل بررسی آماری داده‌های ترکیبی

بررسی داده‌های ترکیبی تقریباً از چهارمرحله تاریخی عبور کرده است. مرحله اول سال‌های قبل از ۱۹۶۰ را شامل می‌شود. بیشتر طول این دوره در اوج گسترش روش‌های آنالیز چند متغیره استاندارد سپری شد. روش‌های آنالیز چند متغیره استاندارد، روش‌های مناسبی برای بررسی مسائلی است که در فضای نمونه‌ای حقیقی می‌گذرند. در این دوره، با وجود علم به این که داده‌های ترکیبی، که اجزای آن قسمتی از یک کل هستند، دارای مجموع ثابت و بنابراین به طور کلی با بردار غیر ترکیبی روش‌های آنالیز چند متغیره استاندارد متفاوت هستند، محققین همه این روش‌های معمول، به خصوص روش‌های بررسی همبستگی را برای داده‌های ترکیبی به کار می‌بردند. با این که Pearson در سال ۱۹۸۷ مقاله بسیار شاخصی درباره همبستگی‌های غیرواقعی منتشر کرد و در آن به مشکلات تفسیر نتایج روش‌های آنالیز چند متغیره استاندارد برای داده‌های ترکیبی اشاره کرده بود [۲۳]. با وجود این هشدارها تا حوالی سال‌های ۱۹۶۰ این روش‌های استاندارد به طور گسترده برای بررسی آماری داده‌های ترکیبی مورد استفاده قرار می‌گرفت.

در دوره دوم، منتقد اصلی استفاده از روش‌های آماری چند متغیره استاندارد برای داده‌های ترکیبی Chayes در سال ۱۹۶۰ بود. انتقاد اصلی Chayes درباره تفسیر گشتاوری همبستگی بود. با وجود مشکل انحراف داده‌ها به سمت منفی، که معمولاً به آن مشکل بسته بودن داده‌ها گفته می‌شود، ساختار ماتریس همبستگی به هم می‌ریزد و نتایج گمراه کننده است [۲۴]. Visteius & Sarmanov در سال ۱۹۵۹ وجود این مشکلات را در کاربردهای زمین‌شناسی تصدیق کردند [۲۵] و Mossiman در سال ۱۹۶۲ توجه بیولوژیست‌ها را به این موضوع جلب کرد [۲۶]. با این وجود حتی محققین آگاه به این موضوع نیز به جای تلاش برای ایجا یک چهارچوب مناسب برای آنالیز داده‌های ترکیبی به استفاده از روش‌های معمول ادامه دادند.

مرحله سوم با کشف این موضوع برای Aitchison همراه بود که اجزای یک داده ترکیبی، اطلاعات نسبی و نه مطلق به دست می‌دهند. بنابراین تمام اطلاعات مربوط به داده ترکیبی را می‌توان به وسیله نسبت بین اجزای داده ترکیبی بیان کرد [۱،۲،۳]. این موضوع که بررسی نسبت‌های لگاریتمی از نظر ریاضی نسبت به نسبت‌های معمولی آسان‌تر است و یک رابطه یک به یک بین اجزای داده ترکیبی و فضای واقعی وجود دارد، باعث ایجاد چندین روش تبدیل نسبت لگاریتمی شد. با استفاده از این تبدیل-های نسبت لگاریتمی می‌توان روش‌های آماری چند متغیره استاندارد را بر روی داده‌های تبدیل شده اعمال کرد و نتایج را دوباره به صورت یک داده ترکیبی بیان کرد.

مرحله چهارم در اواخر دهه ۹۰ و اوایل دهه اول ۲۰۰۰ شروع شد. چندین محقق به صورت جداگانه به این نتیجه رسیدند که با استفاده از عملگرهای Powering و Perturbation که به ترتیب با علامت‌های \oplus و \odot نمایش داده می‌شوند، می‌توان یک فضای برداری متریک ایجاد کرد [۲۷]. بنابراین می‌توان مسائل مربوط به داده‌های ترکیبی را درون این فضا، با ساختار هندسی-جبری خاص آن بررسی کرد. بسیاری از مسائل مربوط به داده‌های ترکیبی را می‌توان داخل این فضا حل کرد. با استفاده از این فضای برداری می‌توان داده‌های ترکیبی را به وسیله مختصات^۱ بیان کرد و روابط بین اجزای داده‌های ترکیبی را درون سیمپلکس بررسی و تفسیر کرد. با این دیدگاه فضای نمونه‌ای یک داده ترکیبی تصادفی سیمپلکس است که دارای ساختار هندسی-جبری متفاوتی نسبت به فضای R دارد.

۳-۵ اصول بررسی داده‌های ترکیبی

دو اصل اساسی در بررسی داده‌های ترکیبی، استقلال از واحد اندازه‌گیری و انسجام زیر مجموعه‌های داده‌های ترکیبی است. استقلال از واحد اندازه‌گیری به این مورد اشاره دارد که اطلاعات موجود در داده‌های ترکیبی نسبی است و به مقدار مطلق آن بستگی ندارد. بنابراین تمام اطلاعات موجود در داده‌های ترکیبی را می‌توان به صورت نسبت‌هایی از اجزای بین این داده‌ها بیان کرد. چون این اطلاعات

¹ - Coordinates

به صورت نسبت بین اجزای داده‌ها بیان می‌شود، وابسته به واحد اندازه‌گیری نیست. انسجام زیرمجموعه داده‌های ترکیبی بیان می‌کند که اگر یک محقق از تمام اجزای یک داده ترکیبی استفاده کند و محقق دیگری فقط از بخشی از داده‌ها استفاده کند، نتایج این دو محقق در مورد نسبت بین اجزای مشترک نباید با یکدیگر متناقض باشد [۲].

۵-۳ هندسه سیمپلکس

داده‌های ترکیبی را می‌توان به صورت یک بردار سطری با D جزء (معادله ۳-۱۴) در نظر گرفت:

$$x = [x_1, x_2, \dots, x_D] \quad (۳-۱۴)$$

که در یک مجموعه $S^D \subset R^n$ قرار دارد. برای مثال:

$$S^D = \{ x = (x_1, x_2, \dots, x_D) : x_i > 0 (i = 1, 2, \dots, D), \sum_{i=1}^D x_i = 1 \} \quad (۳-۱۵)$$

این مجموعه نشان‌دهنده یک سیمپلکس با D جزء است [۲۸].

در یک فضای واقعی، مانند R^n ، می‌توان بردارها را با هم جمع کرد، در یک ثابت ضرب کرد، ویژگی‌هایی همچون تعامد را از آنان انتظار داشت و یا فاصله بین دو نقطه را محاسبه کرد. استفاده از این عملگرهای سنتی در یک فضای واقعی امکان‌پذیر است که دارای فضای برداری با هندسه اقلیدسی باشد. به طور هم‌ارز، هندسه اقلیدسی، هندسه‌ای است که در آن جمع و ضرب اسکالر معمولی ساختار فضای برداری را در R^n تشکیل می‌دهد. این کار معادل آن است که برای اندازه‌گیری فاصله بین دو بردار از معیار اقلیدسی (معادله ۳-۱۶) استفاده کنیم. برای مثال:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (۳-۱۶)$$

برای این که تنها نسبت بین اجزا نشان‌دهنده اطلاعات کامل است، ساختار هندسی سیمپلکس اقلیدسی نیست. به همین دلیل استفاده از روش‌های معمول برای بررسی داده‌ها، به خصوص روش‌های آماری امکان‌پذیر نیست زیرا این روش‌ها برای فضای اقلیدسی طراحی شده‌اند [۴].

ساختار فضای برداری، یک مجموعه با دو عملگر جمع (+) و ضرب عددی (.) است که دارای خواص

زیر است:

اگر u, v, w اجزایی از V (بردار) و a, b, c اجزایی از R (عدد) باشند، داریم:

۱- خاصیت داخلی

$$u, v \in V \rightarrow (u + v) \in V \quad (۱۷-۳)$$

۲- خاصیت خارجی

$$c \in R, u \in V \rightarrow (c \cdot u) \in V \quad (۱۸-۳)$$

۳- اشتراک‌پذیری: برای هر u, v, w متعلق به V داریم:

$$(u + v) + w = u + (v + w) \quad (۱۹-۳)$$

۴- عضو صفر: یک عضو متعلق به V که با \cdot نمایش داده می‌شود، وجود دارد که برای تمام u های

متعلق به V داریم:

$$0 + u = u + 0 = u \quad (۲۰-۳)$$

۵- قرینه‌پذیری: برای تمام u های متعلق به V یک عضو یکتای V مانند u' وجود دارد که:

$$u + u' = 0 \quad (۲۱-۳)$$

۶- جابجایی‌پذیری: برای تمام عضوهای $u, v \in V$ داریم:

$$u + v = v + u \quad (۲۲-۳)$$

۷- اگر $c \in V$ باشد داریم:

$$c \cdot (u + v) = (c \cdot u) + (c \cdot v) \quad (۲۳-۳)$$

علامت (+) در اینجا نشان‌دهنده جمع برداری است.

۸- اگر $a, b \in V$ باشند داریم:

$$(a + b).v = (a.v) + (b.v) \quad (24-3)$$

در اینجا $a+b$ جمع معمولی و $(a.v)+(b.v)$ جمع برداری است.

۹- اگر $a, b \in R$ باشند داریم:

$$(ab).v = a.(b.v) \quad (25-3)$$

۱۰- برای تمام u های متعلق به V داریم:

$$1.u = u \quad (26-3)$$

برای مثال ضرب اسکالر عدد یک در هر برداری همان بردار است.

با در نظر گرفتن $S^D \subset R^D$ ، با انجام عمل‌های جمع برداری و ضرب اسکالر معمول نتایجی به دست می‌آید که خواص ۱ و ۲ را نقض می‌کنند زیرا نتایج به دست آمده در خارج از سیمپلکس قرار می‌گیرند.

برای برطرف کردن مشکلات مرتبط با عملگرهای سنتی (برای مثال هندسه اقلیدسی) باید از عملگرهای جایگزینی که در چارچوب هندسی عمل می‌کنند، استفاده کرد. اولین عملگر که با نام اختلال^۱ شناخته می‌شود، جایگزین عملگر جمع برداری می‌شود. اختلال مطابق معادله (۲۷-۳) و (۳-۳) تعریف می‌شود:

اگر $x = [x_1, x_2, \dots, x_D]$ و $y = [y_1, y_2, \dots, y_D]$ متعلق به S^D باشند، داریم:

$$x \oplus y = C[x_1y_1, x_2y_2, \dots, x_Dy_D] \quad (27-3)$$

عملگر معمول ضرب عددی با عملگری به نام توان^۲ جایگزین می‌شود:

¹ Perturbation

² Powering

$$c \odot x = C[x_1^c, x_2^c, \dots, x_D^c] \quad (28-3)$$

که در آن $c \in R$ و $x \in S^D$ م است.

با جایگزین کردن اختلال و توان با عملگرهای معمول، می توان به راحتی نشان داد که سیمپلکس تحت این چارچوب هندسی دوباره دارای ساختار فضای برداری می شود.

اگرچه (S^D, \oplus, \odot) یک فضای برداری غیراقلیدسی است، می توان آن را با استفاده از تبدیل خطی بر روی فضای اقلیدسی نگارش کرد. این ویژگی پایه و اساس تبدیل نسبت لگاریتمی است [28]، [29].

۳-۵-۱ تبدیل نسبت لگاریتمی^۱

تبدیل نسبت لگاریتمی سیمپلکس را درون فضای برداری اقلیدسی تصویر می کند. به عبارت ساده، تبدیل نسبت لگاریتمی شامل محاسبه نسبت های تمام اجزای یک مجموعه داده و محاسبه لگاریتم این نسبت ها است. هر یک از این تبدیل های نسبت لگاریتمی که در ادامه معرفی می شود، دارای یک تبدیل معکوس نیز هست که داده های باز شده را به درون سیمپلکس تصویر می کند. با استفاده از تبدیل های نسبت لگاریتمی می توان داده ها را به فضای حقیقی تصویر کرد و پس از انجام آنالیزهای آماری، نتایج را با استفاده از تبدیل های معکوس دوباره به درون سیمپلکس برد.

تبدیل لگاریتمی افزایشی^۲ (alr) ساده ترین تبدیل نسبت لگاریتمی است که به صورت زیر تعریف شده است: [۲]

$$y = alr(x) = alr(x_1, x_2, \dots, x_n) = \left[\ln \frac{x_1}{x_n}, \ln \frac{x_2}{x_n}, \dots, \ln \frac{x_{n-1}}{x_n} \right] \quad (29-3)$$

توجه کنید که با استفاده از این تبدیل یک جزء را از دست می دهید و حاصل تبدیل دارای $n-1$ جزء است.

¹ Log-ratio transformation

² Additive log-ratio

تبدیل معکوس alr براساس معادله (۳۰-۳) تعریف می‌شود:

$$x = alr^{-1}(y) = C[\exp(y_1), \exp(y_2), \dots, \exp(y_{D-1}), 1] \quad (30-3)$$

تبدیل نسبت لگاریتمی بعدی تبدیل نسبت لگاریتمی مرکزی^۱ و (clr) است که توسط معادله (۳-۳)

(۳۱) تعریف می‌شود [۲۷]:

$$z = clr(x) = clr(x_1, x_2, \dots, x_n) = \left[\ln \frac{x_1}{(\prod_{i=1}^n x_i)^{1/n}}, \ln \frac{x_2}{(\prod_{i=1}^n x_i)^{1/n}}, \dots, \ln \frac{x_{n-1}}{(\prod_{i=1}^n x_i)^{1/n}} \right] \quad (31-3)$$

که در آن $(\prod_{i=1}^n x_i)^{1/n}$ میانگین هندسی بردار x است. تبدیل معکوس آن در معادله (۳۲-۳)

تعریف می‌شود:

$$x = clr^{-1}(z) = C[\exp(z_1), \exp(z_2), \dots, \exp(z_D)] \quad (32-3)$$

مزیت استفاده از این تبدیل در آن است که بعد از تبدیل n جزء باقی می‌ماند و جزیی از بین

نمی‌رود. اما هر دوی این تبدیل‌های معرفی شده دارای یک نقص ذاتی می‌باشند. داده تبدیل شده را

نمی‌توان در یک پایه متعامد (برای مثال پایه‌ایی که در آن بردارهای پایه بر هم عمود باشند) نمایش

داد.

برای حل این مشکل تبدیل نسبت لگاریتمی ایزومتريک^۲ معرفی شد. این تبدیل بردارهای تبدیل

شده را درون یک فضای متعامد نشان می‌دهد. این تبدیل در رابطه (۳۳-۳) تعریف می‌شود [۲۷]:

$$ilr(x) = ilr(x_1, x_2, \dots, x_n) = \left[\frac{1}{\sqrt{1(1+1)}} \ln \frac{x_1 \dots x_1}{x_{1+1}^1}, \dots, \frac{1}{\sqrt{i(i+1)}} \ln \frac{x_1 \dots x_i}{x_{i+1}^i} \right] \quad (33-3)$$

که در آن $i = (1, 2, \dots, n-1)$ است.

مختصات و یا ضرایبی که با استفاده از این سه فرمول به دست می‌آیند، دارای ویژگی‌های متفاوتی

هستند:

¹ Centered log-ratio

² Isometric log-ratio

۱- نتایجی که از تبدیل نسبت لگاریتمی افزایشی به دست می‌آیند، دارای تفاسیر آسانی هستند. D-1 جزء از داده‌های ترکیبی بر جزء باقی‌مانده تقسیم می‌شوند و از هر یک لگاریتم گرفته می‌شود. باید توجه داشت که داده‌های باز شده به روش alr را نمی‌توان درون محورهای متعامد مورد استفاده قرار داد، زیرا زاویه بین محورهای مختصات alr برابر ۶۰ درجه می‌باشد. این موضوع ممکن است گاهی اوقات باعث ایجاد مشکل شود، برای مثال از ضرب داخلی بین دو بردار نمی‌توان برای محاسبه بین زاویه بین دو بردار و یا فاصله Aitchison آنها استفاده کرد. برای این کار باید از روش‌های مناسب دیگر استفاده کرد.

۲- داده‌های باز شده به روش clr را می‌توان با تقسیم اجزا بر میانگین هندسی اجزا و لگاریتم گرفتن از آنها محاسبه کرد. تفسیر نتایج به دست آمده از تبدیل clr به آسانی مختصات alr نیست ولی این ضرایب بردارهای واقعی هستند و مزایای مختص خود را دارند. فرمول مورد استفاده برای محاسبه clr متقارن است و برای محاسبه فاصله بین دو مشاهده ترکیبی به جای استفاده از فاصله آپسون می‌توان از فاصله معمولی استفاده کرد و برای محاسبه نمودارهای دوتایی^۱ بسیار مفید هستند. مهمترین ضعف داده‌های باز شده به clr آن است که مجموع آنها برابر صفر می‌باشد. این بدان معناست که داده‌های باز شده به روش clr درون یک صفحه در فضای D بعدی قرار می‌گیرند. آنالیز این گونه داده‌ها باید با احتیاط همراه باشد، زیرا ماتریس همبستگی و کوواریانس آنها تکینه است (دترمینان آنها صفر می‌شود).

۳- داده‌های باز شده با استفاده از تبدیل نسبت لگاریتمی ایزومتريک را مختصات می‌نامند. فرمولی که برای محاسبه مختصات ilr مورد استفاده قرار می‌گیرد، نسبت به دوتای قبلی پیچیده‌تر است و قوانین متفاوتی برای محاسبه آنها حاکم است. علاوه بر این، تفسیر آنها نیز چندان ساده نیست. مزیت عمده تبدیل ilr آن است که مختصات به دست آمده، درون یک سیستم متعامد قرار دارند و بنابراین تمام روش‌های آنالیز آماری چندمتغیره را می‌توان به صورت مستقیم بر روی آنها اعمال کرد.

¹ Biplot

با توجه به مزایای ذکر شده در بالا استفاده از تبدیل نسبت لگاریتمی ایزومتریک در این پایان نامه از این تبدیل استفاده شده است [۲۸].

۳-۶ اصول تفسیر نمودار دوتایی ترکیبی

۳-۶-۱ نمودار دوتایی

Gabriel در سال ۱۹۷۱ برای نمایش همزمان سطرها و ستون‌های یک ماتریس به صورت تخمین درجه دو از نمودار دوتایی استفاده کرد. [۳۰] در سال ۱۹۹۷ Aitchison نمودار دوتایی را برای داده‌های ترکیبی گسترش داد و نشان داد که می‌توان از این روش به عنوان یک ابزار اکتشافی قدرتمند بهره برد [۲۹]. قواعد مربوط به تفسیر نمودارهای دوتایی برای داده‌های ترکیبی تا حدودی با قواعد مربوط به داده‌های غیرترکیبی متفاوت است. در این بخش اصول مربوط به تفسیر نمودارهای ترکیبی دوتایی آورده می‌شود.

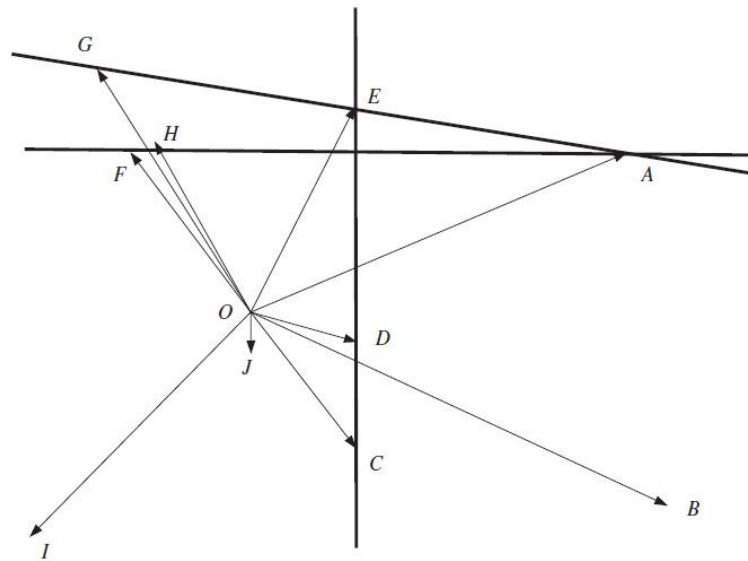
۳-۶-۲ تفسیر یک نمودار دوتایی ترکیبی

نمودار دوتایی ترکیبی یک تخمین درجه دو از داده‌های باز شده به روش تبدیل لگاریتمی مرکزی است که به وسیله تجزیه مقادیر منفرد انجام شده است. یک نمودار دوتایی داده‌های ترکیبی از اجزای زیر تشکیل شده است:

- ۱- مرکز O ، که نشان‌دهنده مرکز مجموعه داده ترکیبی است.
- ۲- یک نقطه در موقعیت b_j برای هر یک از داده‌های باز شده به روش تبدیل لگاریتمی مرکزی.
- ۳- یک علامت در موقعیت a_i برای هر یک از نقاط نمونه‌برداری.
- ۴- خط واصل $\overline{Ob_j}$ که نقطه b_j را به مرکز O وصل می‌کند و به آن شعاع نیز گفته می‌شود.

۵- خط واصل بین $\overline{b_j b_k}$ که دو نقطه b_j و b_k را به هم وصل می کند و لینک نامیده می شود.

خواص اصلی اجزای نام برده شده در ادامه آورده خواهد شد. این اجزا ویژگی های یک نمودار دوتایی را برای تفسیر تغییرپذیری داده های ترکیبی در خود دارند. شمایی از نمودار دوتایی و اجزای آن در شکل (۵-۳) آورده شده است.



شکل (۵-۳): تصویری از اصول تفسیر نمودار دوتایی مربوط به یک داده ترکیبی [۲۹]

۱- O، مرکز نمودار، مرکز ثقل نقطه های b_1, b_2, \dots, b_D است (نقاط مربوط به نوک پیکان متغیرها در نمودار دوتایی). O نشان دهنده میانگین هندسی اجزای مورد استفاده در تبدیل لگاریتمی مرکزی است.

۲- لینکها و شعاعها درباره تغییرپذیری نسبت لگاریتمی در یک مجموعه داده، اطلاعات به دست می دهد. اگر X_j و X_i دو متغیر از یک داده ترکیبی باشند و b_j و b_i نقاط متناظر دو متغیر در نمودار دوتایی باشند و $g_m(x)$ نشان دهنده مرکز هندسی مجموعه داده ترکیبی باشد، می توان نوشت: [۲۹]

$$|b_j b_k|^2 \approx \text{var} \left(\ln \frac{X_i}{X_k} \right) \quad (3-34)$$

$$|Ob_j|^2 \approx \text{var} \left(\ln \frac{X_i}{g_m(x)} \right) \quad (3-35)$$

در هنگام تفسیر شعاع‌ها باید احتیاط کرد. نباید از شعاع‌ها برای بررسی تغییرپذیری یک جزء تنها استفاده کرد. در شکل (۳-۵) ضریب ClI مرتبط با متغیر J تغییرپذیری کوچکی دارد. در حالی که اجزای مربوط به A, B, G و I بزرگترین تغییرپذیری را دارند. با توجه به این که این اجزای دارای شعاع‌هایی در جهت‌های متفاوت از یکدیگر هستند، بنابراین لینک بین این اجزا نیز از بسیاری از اجزای دیگر مجموعه داده ترکیبی بزرگتر خواهد بود. به همین دلیل نسبت لگاریتمی بین آنها بیشترین سهم را در تغییرپذیری کل دارد.

۳- زاویه‌ها اطلاعاتی از همبستگی بین زیر مجموعه‌های داده‌های ترکیبی در اختیار می‌گذارد. اگر

لینک‌های $\overline{b_j b_k}$ و $\overline{b_l b_l}$ در نقطه M یکدیگر را قطع کنند، خواهیم داشت: [۲۹]

$$\cos(\overline{b_j b_k} \overline{b_l b_l}) \simeq \text{corr}(\ln \frac{X_j}{X_k}, \ln \frac{X_l}{X_l}) \quad (۳-۳۶)$$

بنابراین اگر دو لینک دارای زاویه عمود بر یکدیگر باشند، زاویه کسینوس بین دو لینک برابر صفر می‌شود و نسبت‌های لگاریتمی بین این متغیرها همبستگی با یکدیگر ندارد. برای مثال زیر مجموعه‌های داده‌های ترکیبی $\{A, F, H\}$ و $\{C, D, E\}$ را در شکل (۳-۵) در نظر بگیرید. نسبت‌های لگاریتمی این زیر مجموعه از داده ترکیبی احتمالاً دارای همبستگی صفر هستند.

۴- در هنگام تشکیل زیر مجموعه از داده‌های ترکیبی، نسبت بین اجزای این زیر مجموعه‌ها ثابت

باقی می‌ماند. بنابراین می‌توان نمودار دوتایی هر زیر مجموعه از داده‌های ترکیبی را با جدا کردن نقاط مربوط به نوک پیکان و قرار دادن مرکز ثقل کل داده‌ها به عنوان مرکز نمودار دوتایی، ایجاد کرد. اما باید به یاد داشت که نمودار دوتایی تصویری از داده‌هاست و ممکن است نمودار دوتایی ایجاد شده با محاسبه عددی مقادیر با نمودار دوتایی به دست آمده با این روش متفاوت باشد. این موضوع به این بستگی دارد که زیر مجموعه انتخاب شده در نمودار دوتایی کلی چقدر خوب معرفی شده باشد.

۵- اگر نقاط نوک پیکان b_j و b_k بر هم منطبق باشد و یا به هم بسیار نزدیک باشد. واریانس نسبت لگاریتمی بین این دو متغیر برابر صفر و یا به صفر بسیار نزدیک است. بنابراین نسبت بین این دو متغیر ثابت و یا تقریباً ثابت است. بنابراین دو عنصر را در بررسی تغییرپذیری می‌توان زائد در نظر گرفت. اگر نسبتی از واریانس که به وسیله نمودار دوتایی دو مولفه اصلی توجیه می‌شود، کوچک باشد، می‌توان نتیجه گرفت که نسبت لگاریتمی بین این دو متغیر بر صفحه نمودار دوتایی عمود است و این موضوع می‌تواند نشان‌گر آن باشد که نسبت لگاریتمی بین این دو جزء نسبت به راستای دو مولفه اصلی تحلیل مولفه‌های اصلی مستقل هستند. H و F در شکل (۳-۵) این موضوع را نشان می‌دهند.

۶- اگر نقاط نوک پیکان یک زیر مجموعه از داده‌های ترکیبی موازی هم باشند، این موضوع می‌تواند نشان‌گر آن باشد که نمودار دوتایی این زیر مجموعه از داده‌های ترکیبی یک بعدی است. این موضوع ممکن است به این معنی باشد که اجزای داده ترکیبی در طول یک خط ترکیبی گسترده شده‌اند. شکل ۳-۵ سه نمونه از این موضوع را نشان می‌دهد: $\{A, E, G\}, \{C, D, E\}, \{A, F, H\}$.

نقاط نشان‌دهنده نمونه‌ها، مرتبط با نقاط نمونه‌برداری (مشاهدات) هستند. تمام اجزای داده ترکیبی برای آن نمونه را می‌توان با توجه به موقعیت نسبی نقطه نسبت به کل مجموعه نمودار دوتایی تخمین زد.

با توجه به مطالب گفته شده، می‌توان نتیجه گرفت که پایه اصلی نمودارهای دوتایی ترکیبی لینک است. در حالی که شعاع‌ها پایه اصلی در نمودارهای تغییرپذیری برای داده‌های غیر ترکیبی است. طول لینک‌ها (تقریباً) متناسب با واریانس نسبت‌های لگاریتمی ساده بین اجزای داده‌های ترکیبی است (همان‌گونه که در ماتریس واریانس خود را نشان می‌دهد). مجموعه کامل لینک‌ها اطلاعاتی در مورد ساختار کوواریانس ترکیبی نسبت‌های لگاریتمی ساده در اختیار می‌گذارد و می‌توان از روی آن درباره استقلال و تغییرپذیری زیرمجموعه‌های داده ترکیبی اطلاعاتی به دست آورد. تفسیر نمودار دوتایی

داده‌های ترکیبی تحت تاثیر هندسه درونی خود آن قرار دارد و با دوران آن تغییری در آن حاصل نمی‌شود.

یک تفاوت دیگر در نمودارهای دوتایی ترکیبی و غیر ترکیبی استفاده از هم‌مقیاس کردن است. برای داده‌های غیر ترکیبی اگر تفاوت‌های قابل ملاحظه‌ای بین واریانس اجزا وجود داشته باشد، تخمین نمودار دوتایی ممکن است کاملاً بر روی توضیح تغییرپذیری اجزایی با بیشترین تغییرپذیری متمرکز شود و هیچ گونه توضیحی در مورد تغییرپذیری اجزای دیگر در اختیار نگذارد. از آن جایی که این تفاوت در واریانس ممکن است در نتیجه مقیاس اندازه‌گیری باشد، یک روش معمول در این نوع نمودارهای دوتایی استفاده از یک نوع هم‌مقیاس کردن بر روی اجزای داده‌های غیر ترکیبی قبل از انجام هر گونه عملیاتی است. در هنگام استفاده از روش‌های نسبت لگاریتمی احتیاجی به این گونه هم‌مقیاس کردن وجود ندارد.

۳-۷ بررسی تابع توزیع داده‌ها در داخل سیمپلکس

بررسی تابع توزیع داده‌های ژئوشیمیایی معمولاً برای تعیین نرمال بودن تابع توزیع داده‌ها انجام می‌شود. زیرا بسیاری از روش‌های آماری استفاده شده برای تحلیل و پردازش داده‌ها، فرض می‌کنند که داده‌ها از توزیع نرمال پیروی می‌کنند. در مورد داده‌های ژئوشیمیایی به دلایل مختلف از قبیل وجود مقادیر خارج از ردیف که باعث چولگی در شکل تابع توزیع می‌شود و یا محدودیت‌های مربوط به آنالیز شیمیایی نمونه‌ها، این داده‌ها معمولاً از توزیع نرمال پیروی نمی‌کنند. علاوه بر این با بررسی‌های انجام شده توسط محققین توزیع نرمال کلاسیک غالباً جایی در سیمپلکس ندارد و این داده‌ها از توزیع‌هایی مانند Dirichlet پیروی می‌کند [۲۸]. توزیع Dirichlet در نظریه احتمال و آمار یک توزیع پیوسته است. این توزیع بطور کلی حالت گسترش یافته توزیع بتا برای توابع چندمتغیره است. معمولاً از توزیع Dirichlet به عنوان توزیع پیشین در مدل سازی بیزی استفاده می‌شود؛ چرا که توزیع Dirichlet مزدوج پیشین برای توزیع چندجمله‌ای و توزیع دسته‌ای است.

برای بررسی نرمال بودن تابع توزیع داده‌ها درون سیمپلکس روش‌های متنوعی وجود دارد. در این جا از روش آزمون شعاع^۱ استفاده شده است. فرض اولیه در این بررسی‌ها این است که اگر داده‌ها درون سیمپلکس دارای توزیع نرمال باشند، نسبت‌های لگاریتمی این داده‌ها نیز از توزیع نرمال چند متغیره پیروی می‌کنند. برای بررسی این موضوع از آزمون اماری با فرض‌های زیر استفاده می‌شود:

H_0 : داده‌ها از یک توزیع نرمال در سیمپلکس پیروی می‌کنند.

H_1 : داده‌ها از یک توزیع نرمال در سیمپلکس پیروی نمی‌کنند.

این فرضیات را برای نسبت‌های لگاریتمی این داده‌ها نیز صادق است.

۷-۳-۱ آزمون شعاع

این آزمون بر این قاعده استوار است که در حالتی که داده‌ها از یک توزیع نرمال پیروی می‌کنند، فاصله ماهالانابیوس نمونه‌ها از میانگین نمونه‌ها از توزیع Chi پیروی می‌کند. به همین دلیل نتایج در مرکز قرار دارند و ماتریس کوواریانس آن برابر ماتریس همانی است. هنگامی که ماتریس کوواریانس ضریبی از ماتریس همانی است، شکل تابع توزیع کروی است. آزمون شعاع براساس بررسی تابع توزیع تجربی یا EDF استوار است. از بین تمام آزمون‌های EDF ممکن، از روش‌های Anderson-Darling (Q_a)، Cramer-von Mises (Q_c) و Watson (Q_w) استفاده شده است.

اجرای یک آزمون شعاع شامل مراحل زیر است:

۱- بهترین تخمین ممکن از $\hat{\mu}^*$ و $\hat{\Sigma}$ را محاسبه کنید (میانگین نمونه‌ها و ماتریس کوواریانس)

۲- مقدار فاصله ماهالانابیوس را محاسبه کنید:

¹ - Radii test

$$u_r = (x_r^* - \widehat{\mu}^*) \widehat{\Sigma}^{-1} (x_r^* - \widehat{\mu}^*)^t \quad r=1,2,\dots,n \quad (37-3)$$

۳- مقدار $p_r = F(u_r)$ که در آن $r=1,\dots,n$ و F تابع توزیع χ^2_{D-1} است.

۴- مقادیر به دست آمده در مرحله ۳ را به صورت نزولی مرتب کنید.

۵- مقادیر آماره‌های Q_i را محاسبه و سپس توسط روابط (۳۸-۳) تا (۴۰-۳) زیر مقدار آنها را

تصحیح کنید.

$$Q_a^* = Q_a \quad (38-3)$$

$$Q_c^* = (Q_c - \frac{0.4}{n} + \frac{0.6}{n})(1 + \frac{1}{n}) \quad (39-3)$$

$$Q_w^* = (Q_w - \frac{0.1}{n} + \frac{0.1}{n^2})(1 + \frac{0.8}{n}) \quad (40-3)$$

۶- سپس نتایج را با مقادیر بحرانی که در جدول (۱-۳) آورده شده است، مقایسه کنید.

۷- فرض H_0 رد خواهد شد، اگر مقدار آماره Q_i از مقدار بحرانی بیشتر باشد.

جدول (۱-۳): مقادیر بحرانی برای سطح اعتمادهای مختلف برای تست شعاع [۲۹]

mod. statistic	signif. level (%)			
	10	5	2.5	1
Q_a^*	1.933	2.492	3.070	3.857
Q_c^*	0.347	0.461	0.581	0.743
Q_w^*	0.152	0.187	0.221	0.267

واضح است هنگامی که نرمال بودن یک مجموعه داده مورد بررسی قرار می‌دهیم، مطلوب آن است که هر سه آماره بالا جواب یکسان داشته باشند. متأسفانه در اکثر موارد این گونه نیست. هنگامی که این اتفاق می‌افتد، کارشناس باید یک تصمیم شخصی و بحرانی بگیرد. خوشبختانه این آزمون‌ها را می‌توان به همراه نمودارهای Q-Q معمولی انجام داد. راه دیگر استفاده از آماره شاپیرو-ویلک است. این آزمون باید بر روی مختصات داده‌ها (داده‌های باز شده به روش نسبت لگاریتمی ایزومتریک) انجام شوند.

برای بررسی این که چه متغیرهایی بیشترین تاثیر را بر روی انحراف از توزیع نرمال داشته‌اند، می‌توان از آزمون‌های حاشیه‌ای^۱ و آزمون‌های دوتایی^۲ همان‌گونه که توسط Aitchison در سال ۱۹۸۶ بیان شده است، استفاده کرد [۲]. در این آزمون‌ها نیز مانند آزمون شعاع از تابع توزیع تجربی EDF استفاده می‌شود. در جدول (۲-۳) مقادیر بحرانی برای این دو تست آورده شده است. اگر مقادیر به دست آمده از مقدار بحرانی بیشتر باشد، فرض نرمال بودن تابع توزیع داده رد می‌شود.

جدول (۲-۳): مقادیر بحرانی برای سطح اعتمادهای مختلف برای آزمون‌های Marginal و Bivariate

[۲]angle

Significance level (per cent)	Anderson-Darling	Cramer-von Mises	Watson	Bivariate angle and radius tests			
	Marginal tests			Q_A	Q_C	Q_W	
	$Q_A \left[1 + \frac{4}{N} - \frac{25}{N^2} \right]$	$Q_C \left[1 + \frac{1}{2N} \right]$	$Q_W \left[1 + \frac{1}{2N} \right]$	$\left[Q_C - \frac{0.4}{N} + \frac{0.6}{N^2} \right]$	$\left[Q_W - \frac{0.1}{N} + \frac{0.1}{N^2} \right]$		
				$\times \left[1 + \frac{1}{N} \right]$	$\times \left[1 + \frac{0.8}{N} \right]$		
10	0.656	0.104	0.096	10	1.933	0.347	0.152
5	0.787	0.126	0.116	5	2.492	0.461	0.187
2.5	0.918	0.148	0.136	2.5	3.070	0.581	0.221
1	1.092	0.178	0.163	1	3.857	0.743	0.267

۳-۷ جمع بندی

کار کردن با داده‌های ژئوشیمیایی به دلیل ابعاد بزرگ داده‌ها معمولاً دشوار است. با استفاده از PCA و یا RPCA می‌توان ابعاد داده‌ها را کاهش داد به گونه‌ای که کمترین اطلاعات از دست برود. مزیت بسیار مهم این روش‌ها آن است که می‌توان سهم هر متغیر را در واریانس کلی داده‌ها حساب کرد.

روش نمودار احتمال، روشی بسیار کارآمد برای مدل‌سازی و تعیین حد زمینه و آنومالی است. این روش بر پایه فراوانی داده‌ها عمل می‌کند. قبل از اعمال این روش‌ها برای در نظر گرفتن فضای نمونه‌ای محدود داده‌های ژئوشیمیایی و هندسه غیراقلیدسی آنها می‌توان از تبدیل‌های نسبت لگاریتمی استفاده کرد. با استفاده از تبدیل ایزومتریک می‌توان داده‌ها را در یک پایه متعامد نگاشت و از روش‌های مقاوم برای پردازش آنها استفاده کرد.

¹ -Marginal test

² - Bivariate test

فصل چهارم:

تجزیه و تحلیل داده‌ها منطبقه مورد مطالعه

۴-۱ مقدمه

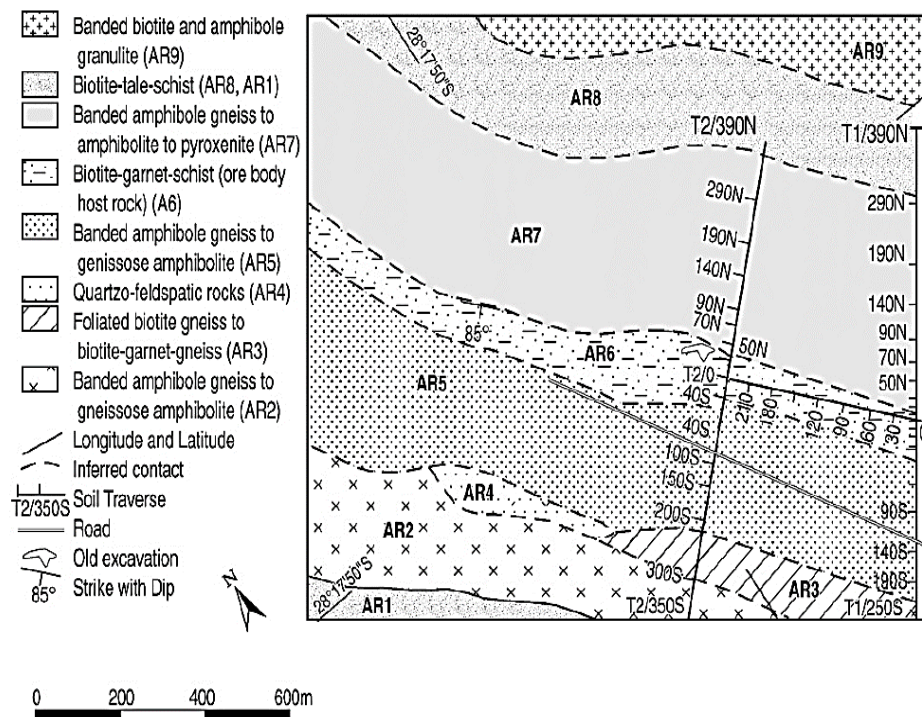
همان‌گونه که شرح داده شد که بسته‌بودن داده‌های ژئوشیمیایی می‌تواند تاثیر به‌سزایی در بررسی آماری این داده‌ها داشته باشد. استفاده از روش‌های چند متغیره مانند تحلیل مولفه‌های اصلی، تحلیل فاکتوری و یا آنالیزهای همبستگی برای بررسی ارتباط بین عناصر روشی بسیار معمول است. این روش‌ها همگی براساس بررسی ساختار ماتریس واریانس و کوواریانس عمل می‌کنند. بسته‌بودن داده‌های ژئوشیمیایی باعث به هم ریختن ساختار این ماتریس می‌شود. محققین نسبت به استفاده از این روش‌ها بدون در نظر گرفتن ماهیت بسته داده‌ها هشدار داده‌اند. برای بررسی این موارد مقایسه‌ای بین نتایج روش معمول و روش پیشنهادی جدید انجام شد.

روش تحلیل مولفه‌های اصلی، روشی چندمتغیره متداولی در بررسی داده‌ها است که براساس ماتریس واریانس-کوواریانس عمل می‌کند. روش مقاوم در سال‌های اخیر معرفی شده است. روش مقاوم تحلیل مولفه‌های اصلی در مقابل داده‌های خارج از ردیف حساس نیست. در این پایان‌نامه از روش تحلیل مولفه‌های اصلی برای مقایسه بین روش‌های ترکیبی و روش معمول استفاده شده است. در روند انجام این کار یک بار داده‌ها به روش معمول تحلیل مولفه‌های اصلی انجام و سپس داده‌ها با در نظر گرفتن بسته‌بودن داده‌ها و به روش مقاوم مورد بررسی قرار گرفته‌اند. سپس با استفاده از روش نمودار احتمال حد زمینه و آنومال برای نتایج روش معمول و ترکیبی محاسبه شده است. نتایج روی نمودار ترسیم گردید و با یکدیگر مقایسه شد.

۴-۲ مجموعه داده‌ها

داده‌های مورد استفاده در این پایان‌نامه از ماسه بادی در منطقه Areachap آفریقای جنوبی برداشته شده است. این برداشت‌ها به منظور بررسی پراکندگی هاله‌های ژئوشیمیایی ثانویه انجام شده است. شکل (۴-۱) نقشه زمین‌شناسی و موقعیت پروفیل‌های برداشت T_1 ، T_2 و T_3 را نشان می‌دهد. در این پایان‌نامه تنها از داده‌های پروفیل T_2 به تعداد ۲۵ نمونه استفاده شده است. نمونه‌برداری از عمق ۱۰ تا

۱۵ سانتیمتری از بخش ریزدانه ($>75\mu$) انجام شده است. برای بررسی گسترش محیطی هاله‌های ژئوشیمیایی، نمونه‌برداری در نزدیکی ماده معدنی فشرده‌تر و به صورت ۱۰ متری و با دور شدن از آن فواصل افزایش یافته و به ۲۰، ۵۰ و ۱۰۰ متری رسیده است. نمونه‌های برداشت شده توسط روش اشعه فلورسانس X (XRF) آنالیز شده است [۸].



شکل (۴-۱) نقشه زمین‌شناسی و موقعیت پروفیل‌های نمونه‌برداری [۸]

قبل از انجام هرگونه عملیاتی بر روی داده‌ها باید داده‌ها پیش‌پردازش شوند. پیش‌پردازش داده‌های ژئوشیمی، شامل جایگزینی داده‌های سنسورد و جایگزینی مقادیر خارج از ردیف است. این داده‌ها با توجه به مقادیر بسیار زیاد و یا بسیار کم عناصر و با عنایت به حد تشخیص دستگاه گزارش می‌شوند [۳۱]. روش‌های مختلفی برای شناسایی مقادیر خارج از ردیف وجود دارد از آن جمله می‌توان به نمودار جعبه‌ای، Q-Q Plot و نمودار دورفل اشاره کرد یکی از مزیت‌های روش مقاوم همان‌گونه که قبلاً نیز بیان شد، این است که در برابر داده‌های خارج از ردیف مقاوم است. بنابراین در روش مقاوم اگر تعداد داده‌های خارج از ردیف از ۲۵ درصد کل داده‌ها کمتر باشد، نیازی به اصلاح نیست.

۳-۴ بررسی داده‌های خارج از ردیف به صورت چندمتغیره

به طور کلی، داده خارج از ردیف به داده‌ای گفته می‌شود که با بقیه داده‌ها فرق دارد. با استفاده از روش‌های آماری می‌توان میزان تفاوت داده خارج از ردیف را کمی کرد. در آمار با استفاده از تفاوت‌ها، شیوه‌های اندازه‌گیری اختلاف، فاصله و واگرایی می‌توان تفاوت یک داده را با بقیه داده‌ها کمی کرد.

در حالت کلی، یک داده خارج از ردیف چندمتغیره به نمونه‌ای گفته می‌شود که دارای مقادیر اکستریم (خیلی بزرگ و یا خیلی کوچک) در یک جز و یا چند جز باشد. به عبارت دیگر، مقداری از متغیرهای یک داده خارج از ردیف چندمتغیره، داده خارج از ردیف تک‌متغیره نیز هستند. اگرچه در برخی موارد خارج از ردیف بودن یک داده چندمتغیره ناشی از مقادیر اکستریم نسبت‌های بین متغیرهای آن داده است.

اکثر روش‌های تشخیص داده‌های خارج از ردیف چندمتغیره براساس مکان و یا گسترش (شکل) توزیع نمونه‌ها عمل می‌کنند. در واقع مکان داده‌هایی با مقادیر بزرگ (و یا کوچک) فاصله بیشتری از مرکز مجموعه داده‌ها دارد. در بررسی شکل و یا گسترش داده‌ها، داده‌هایی با مقادیر میانی ممکن است خارج از ردیف باشند. [۲۹]

مراحل شناسایی داده‌های خارج از ردیف در یک داده ترکیبی به شرح زیر است:

۱- برای هر متغیر، مقدار مختصات ilr را محاسبه کنید.

۲- فاصله ماهالانابیوس معمولی را با استفاده از رابطه (۱-۴) محاسبه کنید:

$$d_M(ilr x, \mu) = (ilr x - \mu)\Sigma^{-1}(ilr x - \mu)^t \quad (1-4)$$

با فرض توزیع نرمال لگاریتم نسبتی، فاصله ماهالانابیوس داده‌ها تقریباً از توزیع مربع کای با درجه

آزادی D-1 پیروی می‌کند. داده‌های خارج از ردیف احتمالی داده‌هایی هستند که فاصله ماهالانابیوس

بزرگتری دارند، به عنوان مثال بزرگتر از ۹۹٪ چندک توزیع مربع کای. [۲۹]

برای بررسی داده‌های خارج از ردیف ابتدا داده‌ها با استفاده از تبدیل ایزومتریک باز شد سپس فاصله ماهالانابیوس برای داده‌های باز شده محاسبه شد. جدول (۴-۱) فاصله ماهالانابیوس را برای داده‌های باز شده نشان می‌دهد.

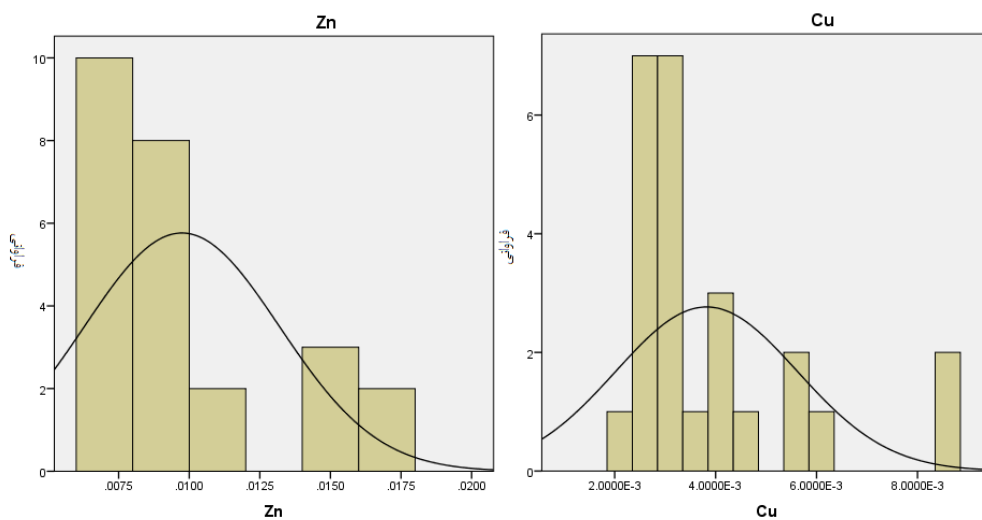
جدول (۴-۱): فاصله ماهالانابیوس برای داده‌های باز شده

فاصله	نمونه	فاصله	نمونه	فاصله	نمونه
۴۱۹۰۵/۴۳	۱۹	۴۱۸۳۸/۱	۱۰	۴۱۳۱۰/۵۷	۱
۴۱۵۰۲/۵۹	۲۰	۴۲۴۸۹/۵۷	۱۱	۴۱۵۲۸/۲۵	۲
۴۱۱۵۶/۵۸	۲۱	۴۱۲۹۲/۹۵	۱۲	۴۱۹۶۲/۵	۳
۴۱۵۲۷/۴۱	۲۲	۴۱۶۸۶/۹۲	۱۳	۴۱۱۱۸/۱۲	۴
۴۱۹۵۶/۹۱	۲۳	۴۱۴۲۵/۶۸	۱۴	۴۱۱۳۳/۵	۵
۴۱۰۵۹/۱۱	۲۴	۴۱۰۵۳/۳۵	۱۵	۴۱۱۴۹/۷۸	۶
۴۱۲۳۸/۴	۲۵	۴۱۹۳۹/۸	۱۶	۴۱۰۸۱/۱۹	۷
		۴۲۰۰۰/۲۳	۱۷	۴۱۴۶۷/۳	۸
		۴۱۷۷۱/۹۱	۱۸	۴۱۹۴۰/۵۲	۹

برای بررسی کمی نمونه‌های بزرگتر از ۹۹٪ چندک توزیع کای دو از نرم‌افزار CoDapack استفاده شد. براساس نتایج این نرم‌افزار تنها نمونه ۱۱ بزرگتر از ۹۹٪ چندک توزیع کای دو است و داده خارج از ردیف احتمالی است. با بررسی پروفیل نمونه‌برداری مشخص شد که نمونه ۱۱ از روی کانه‌سازی برداشت شده است. به همین دلیل این نمونه به عنوان نمونه آنومال در نظر گرفته شد و به عنوان نمونه خارج از ردیف از محاسبات حذف نشد.

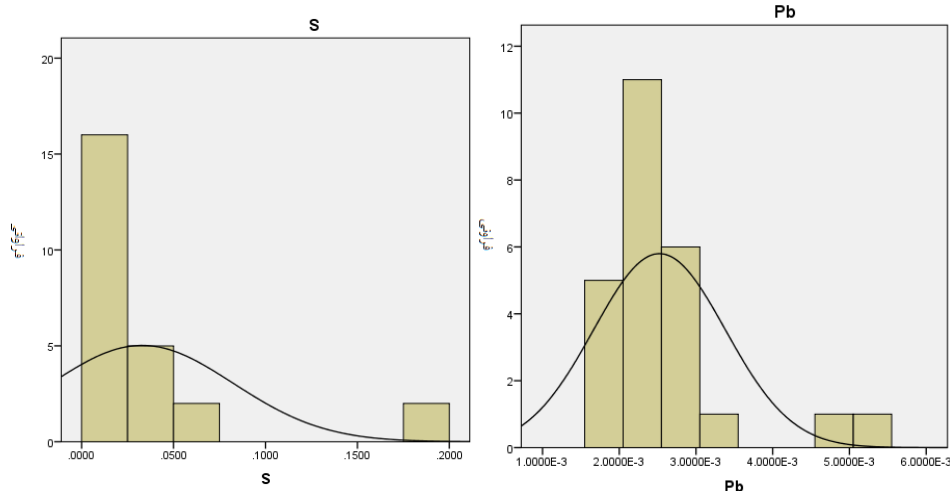
۴-۴ بررسی هیستوگرام عنصرهای کانی سازی در منطقه

شکل‌های (۲-۴) و (۳-۴) نشان دهنده هیستوگرام عنصرهای کانی سازی در منطقه برای داده‌های بسته و داده‌های باز شده است. همان‌گونه که در این شکل‌ها مشخص است، شکل تابع توزیع عنصرها برای داده‌های بسته با کشیدگی به سمت راست همراه است ولی با باز شدن داده‌ها شکل تابع توزیع داده‌ها متقارن‌تر شده و از میزان کشیدگی به سمت راست کاسته شده است.



(ب)

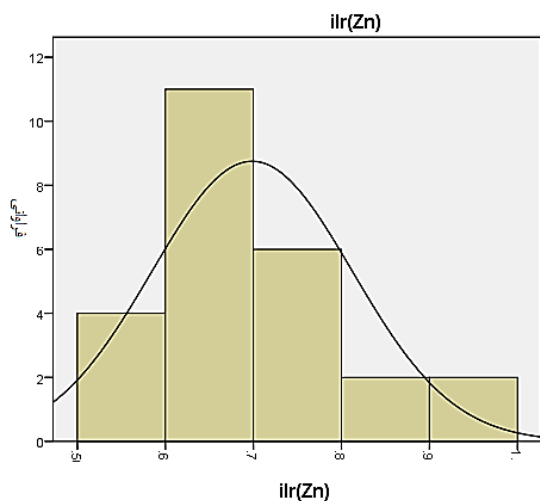
(الف)



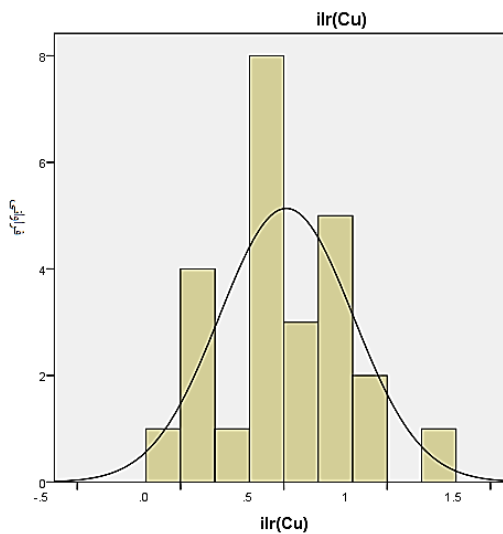
(د)

(ج)

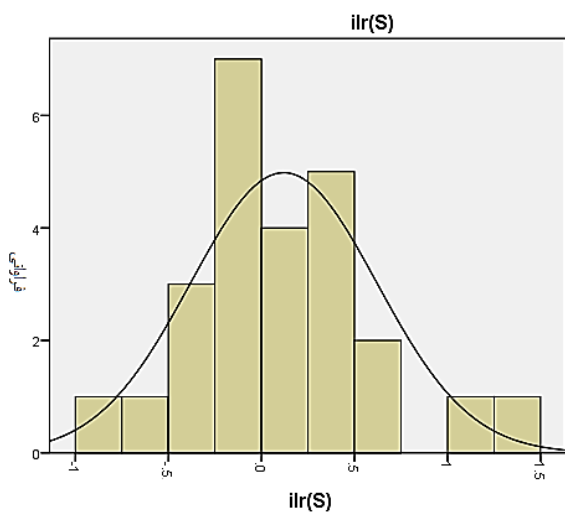
شکل (۲-۴): هیستوگرام داده‌های بسته عناصر الف) مس (ب) روی (ج) سرب (د) گوگرد



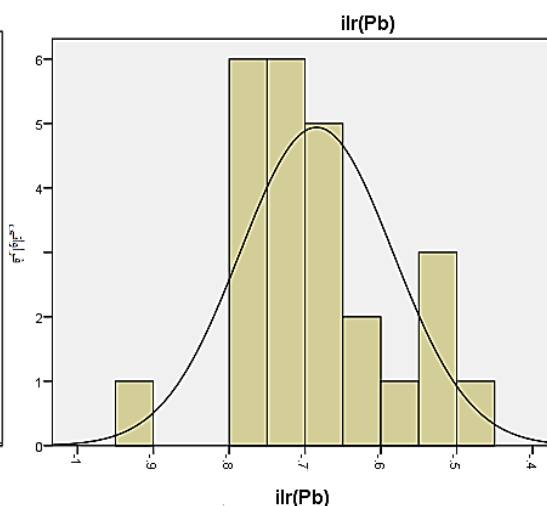
(ب)



(الف)



(د)



(ج)

شکل (۳-۴): هیستوگرام داده‌های باز شده برای عناصر الف(مس ب) روی ج(سرب د) گوگرد

۳-۴ آنالیز همبستگی

اولین قدم در هر تحلیل چند متغیره‌ای بررسی همبستگی است. آنالیز همبستگی میزان ارتباط بین دو متغیر را بررسی می‌کند. کوواریانس معیاری برای اندازه‌گیری این رابطه است که به تغییرپذیری هر یک از متغیرها بستگی دارد. چون کوواریانس هر عددی را می‌تواند اختیار کند، تنها علامت (+ و -) قابل

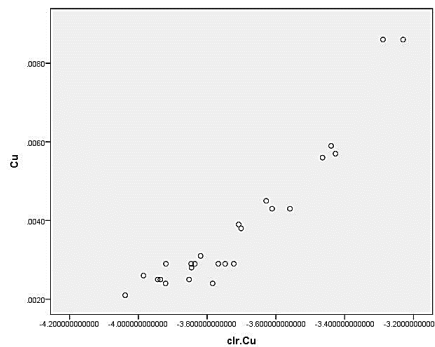
تفسیر است و شدت رابطه بین متغیرها را نمی‌توان به صورت کمی اندازه گرفت. برای بیان کمی رابطه بین متغیرها از ضرایب همبستگی استفاده می‌شود. ضرایب پیرسون، اسپیرمن و کندال معروف‌ترین ضرایب همبستگی هستند. در این ضرایب همبستگی اعداد در بازه (۱-،۱) تغییر می‌کنند. ± 1 بیشترین رابطه و عدد صفر نشان‌دهنده عدم وجود رابطه است.

برای بررسی همبستگی بین متغیرها باید داده‌ها با استفاده از تبدیل نسبت لگاریتمی افزایشی (alr) و یا تبدیل نسبت لگاریتمی مرکزی (clr) باز شوند. در شکل (۴-۴) نمودار پراکندگی عنصر مس با داده‌های باز شده این عنصر به روش‌های alr، clr و ilr نشان داده شده است. برای باز کردن داده به روش alr از Ni به عنوان مخرج معادله (۲۹-۳) استفاده شده است.

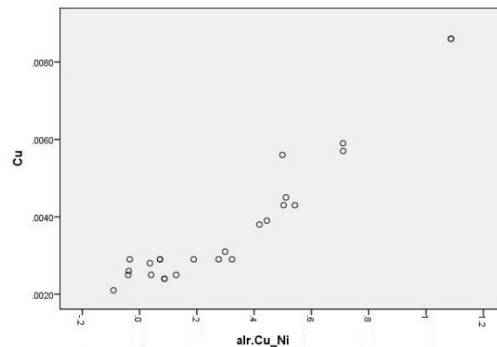
$$y = alr(x) = alr(x_1, x_2, \dots, x_n) = \left[\ln \frac{x_1}{x_n}, \ln \frac{x_2}{x_n}, \dots, \ln \frac{x_{n-1}}{x_n} \right] \quad (29-3)$$

در شکل (۴-۴) الف محور افقی نشان‌دهنده داده باز شده به روش alr است که برای مخرج از Ni استفاده شده است. Ni کم‌ترین واریانس clr را دارد به همین علت از این عنصر به عنوان مخرج معادله (۲۹-۳) استفاده شده است. در قسمت ج این شکل نیز محور افقی نشان‌دهنده داده باز شده به روش ilr است و به همین ترتیب محور افقی در قسمت ب شکل (۴-۴) داده باز شده به روش clr را نشان می‌دهد. در این شکل یک رابطه تقریبی بین داده‌های باز شده به روش alr و clr وجود دارد. ولی هیچ رابطه‌ای بین داده باز شده به روش ilr و داده اولیه مشاهده نمی‌شود، به همین دلیل در قسمت ج شکل (۴-۴) از علامت ilr-8 استفاده شده است تا بر نبود یک رابطه خطی بین داده‌های بسته و باز شده به روش ilr تاکید شود. با این تفاسیر استفاده از تبدیل ilr برای همبستگی بین متغیرها مناسب نیست. با استفاده از هر دوی تبدیل‌های alr و clr، ضرایب همبستگی بدست آمده متعلق به خود متغیرها نبوده و مربوط به نسبت بین آنها می‌باشد. این روش ممکن است نتایج بهتری برای تفسیر رابطه بین متغیرها به دست دهد. در این تحقیق از تبدیل نسبت لگاریتمی مرکزی برای باز کردن داده‌ها استفاده شد، زیرا نتایج به دست آمده از روش alr به مخرج انتخاب شده برای معادله (۲۹-۳) بستگی دارد و با انتخاب

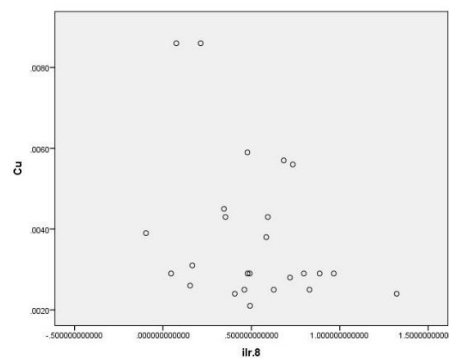
مخرج متفاوت ممکن است نتایج تا حدودی فرق کند. نتایج بررسی همبستگی داده‌های بسته و باز شده محاسبه شده و با یکدیگر مقایسه می‌گردد.



(ب)



(الف)



(ج)

شکل (۴-۴) نمودار پراکندگی داده بسته مس با الف) نتایج روش alr (ب) نتایج روش clr (ج) نتایج روش ilr

۴-۳-۱ بررسی نرمال بودن داده‌ها

بسیاری از روش‌های آماری مورد استفاده برای بررسی داده‌های ژئوشیمیایی از جمله آنالیز فاکتوری، رگرسیون چندمتغیره و ضریب همبستگی پیرسون و ... دارای پیش‌فرض نرمال بودن توزیع داده‌ها هستند. داده‌ها در داخل سیمپلکس تقریباً هیچ‌گاه از توزیع نرمال کلاسیک پیروی نمی‌کنند و معمولاً از توزیع‌های دیگر مانند Dirichlet پیروی می‌کنند [۲۹].

یک بردار ترکیبی تصادفی x درون سیمپلکس از توزیع نرمال پیروی می‌کند، اگر و فقط اگر بردار داده‌های باز شده آن به روش ilr از توزیع نرمال چندمتغیره پیروی کند. بنابراین اگر یک داده ترکیبی درون سیمپلکس از توزیع نرمال پیروی کند، داده‌های باز شده آن نیز توزیع نرمال چند متغیره دارد. روش‌های مختلفی برای بررسی نرمال بودن توزیع درون سیمپلکس وجود دارد. روش آزمون شعاع یکی از روش‌های مناسب برای این کار است. با استفاده از روش‌های آزمون Marginal و Bivariate angle می‌توان سهم هر یک از داده‌ها را در انحراف از توزیع نرمال به صورت جداگانه بررسی کرد. نتایج مربوط به آزمون شعاع در جدول (۲-۴) و نتایج مربوط به آزمون‌های Marginal در جدول (۳-۴) آورده شده است.

جدول (۲-۴): نتایج آزمون شعاع

Watson (Q_w)	Cramer-von Mises (Q_c)	Anderson- Darling(Q_a)
۰/۰۲۰۶	۰/۰۲۰۲	۰/۲۷۴۴

مقادیر بحرانی آزمون شعاع برای آماره‌های Q_a ، Q_c و Q_w در جدول (۱-۳) آورده شده است. به همین ترتیب مقادیر بحرانی این آماره‌ها برای آزمون Bivariate Angle و Marginal در جدول (۲-۳) آورده شده است. نتایج به دست آمده برای این سه آماره برای هر عنصر در هر آزمون با مقدار بحرانی مربوط به آن آماره در آزمون مورد نظر مقایسه می‌شود. چنانچه مقدار به دست آمده از مقدار بحرانی بیشتر باشد، فرض نرمال بودن داده رد می‌شود. با مقایسه نتایج به دست آمده در جدول (۲-۴) با جدول (۱-۳) می‌توان متوجه شد، که هیچ‌یک از مقادیر محاسبه شده بیشتر از مقدار بحرانی نیست. ولی با مقایسه مقادیر جدول (۳-۴) با مقادیر بحرانی در جدول (۲-۳) می‌توان مشاهده کرد، عناصر TiO_2 ، CaO ، K_2O ، Pb ، Zn ، Cu و As از مقادیر بحرانی بیشتر است. برای بررسی بیشتر از روش بررسی تک-متغیره شاپیرو-ویلک برای داده‌های باز شده با روش نسبت لگاریتمی ایزومتریک استفاده شد.

جدول (۴-۳): نتایج آزمون Marginals

Watson (Q_w)	Cramer-von Mises (Q_c)	Anderson-Darling(Q_a)	متغیرهای باز شده
۰/۰۵۰۵	۰/۰۵۰۵	۰/۴۰۴۲	alr(SiO ₂ ,Ni)
۰/۱۲۲۶	۰/۱۳۹۰	۰/۹۴۵۶	alr(TiO ₂ ,Ni)
۰/۰۳۳۱	۰/۰۳۳۳	۰/۲۱۷۹	alr(Al ₂ O ₃ ,Ni)
۰/۰۶۶۷	۰/۰۶۷۸	۰/۴۹۱۸	alr(MgO,Ni)
۰/۲۶۱۳	۰/۲۸۰۹	۱/۵۴۵۳	alr(CaO,Ni)
۰/۰۸۷۳	۰/۰۸۷۳	۰/۵۱۴۴	alr(Na ₂ O,Ni)
۰/۰۹۴۲	۰/۱۱۰۳	۰/۸۶۵۳	alr(K ₂ O,Ni)
۰/۱۴۲۲	۰/۱۶۴۲	۱/۱۸۴۸	alr(Cu,Ni)
۰/۱۴۲۲	۰/۱۶۴۲	۱/۱۸۴۸	alr(Zn,Ni)
۰/۱۶۰۲	۰/۱۹۵۸	۱/۵۰۸۷	alr(Pb,Ni)
۰/۰۵۹۲	۰/۰۶۴۹	۰/۴۷۳۹	alr(S,Ni)
۰/۱۲۱۶	۰/۱۲۳۵	۰/۷۹۳۹	alr(As,Ni)

آزمون شاپیرو-ویلک: معمولاً اگر شاخص معناداری ارائه شده در این آزمون از ۰/۰۵ بیشتر باشد، داده‌ها از توزیع نرمال پیروی می‌کنند [۳۲]. نتایج این آزمون نیز در جدول (۴-۴) آورده شده است. در این جدول مقادیر 1 ilr تا 12 ilr نشان‌دهنده داده‌های باز شده با روش ilr است. مقادیر 1 ilr، 2 ilr و 3 ilr از مقدار بحرانی ۰/۰۵ کم‌تر هستند. یک بردار k-متغیره دارای توزیع نرمال چندمتغیره است، اگر هر ترکیب خطی ممکن از k متغیر آن دارای توزیع نرمال باشد. لذا برای این که یک مجموعه داده چندمتغیره دارای توزیع نرمال باشد، تمام متغیرهای آن باید دارای توزیع نرمال یک متغیره باشد. با توجه به جدول (۴-۴) متوجه می‌شویم که داده‌های باز شده به روش ilr دارای توزیع نرمال چندمتغیره نیستند. بنابراین داده‌ها درون سیمپلکس از توزیع نرمال پیروی نمی‌کنند.

جدول (۴-۴): مقادیر شاخص شاپیرو-ویلیک برای داده‌های باز شده به روش ilr

مقدار شاخص شاپیرو-ویلیک	متغیر	مقدار شاخص شاپیرو-ویلیک	متغیر
۰/۰۰۲	ilr.7	۰/۰۲۰	ilr.1
۰/۹۴۹	ilr.8	۰/۸۸۱	ilr.2
۰/۱۸۶	ilr.9	۰/۲۰۶	ilr.3
۰/۲۴۰	ilr.10	۰/۱۲۷	ilr.4
۰/۳۳۶	ilr.11	۰/۲۶۶	ilr.5
۰/۱۳۰	ilr.12	۰/۰۱۵	ilr.6

اگر توزیع داده‌ها نرمال نباشد برای محاسبه ضریب همبستگی می‌توان از روش‌های ناپارامتری استفاده کرد. این روش‌ها نسبت به توزیع داده‌ها حساس نمی‌باشد. یکی از این روش‌های ناپارامتری محاسبه ضریب همبستگی، ضریب همبستگی رتبه‌ای اسپیرمن است که با r نمایش داده می‌شود [۳۱]. نتایج محاسبه ضرایب همبستگی رتبه‌ای اسپیرمن در جدول‌های (۴-۵) و (۴-۶) آورده شده است.

از مقایسه ضرایب همبستگی داده‌های بسته و باز شده می‌توان دریافت که یک افزایش در ضریب همبستگی مقادیر داده‌های بسته وجود دارد که غیرواقعی و کاذب است. در این حالت مقدار جزئی که بیشترین تغییرپذیری را در حالت بسته دارد، تغییرات معکوسی نسبت به سایر متغیرها دارد، یعنی با افزایش مقدار این جزء، اجزای دیگر باید الزاماً کاهش پیدا کنند تا مقدار مجموع کل به صد برسد. چون کانی‌سازی از نوع سولفید توده‌ای است، انتظار می‌رود که Cu ، Zn ، Pb و S همبستگی بالایی نشان دهند.

جدول (۴-۵): ضرایب همبستگی اسپیرمن برای داده‌های بسته

Ni	۰/۶۴۲	-۰/۴۹۴	۰/۸۷۷	۰/۷۳۹	۰/۶۰۱	-۰/۸۵۱	۰/۹۱۶	۰/۶۶۶	۰/۶۷۸	۰/۷۲۲	۰/۶۶۶	-۰/۰۰۳	۱
As	-۰/۰۵	۰/۲۵۳	۰/۰۶۶	-۰/۱۰۲	-۰/۰۷۴	۰/۱۰۷	-۰/۰۳۰	۰/۰۰۴	-۰/۰۱۶	-۰/۱۹۸	۰/۰۰۲	۱	
S	۰/۱۷۲	-۰/۴۷۹	۰/۴۴۶	۰/۹۰۲	۰/۸۳۵	-۰/۸۴۶	۰/۵۹۵	۰/۹۴۳	۰/۸۴۸	۰/۸۳۱	۱		
Pb	۰/۱۷۲	-۰/۳۴۸	۰/۵۲۲	۰/۷۵۳	۰/۷۴۰	-۰/۷۶۴	۰/۶	۰/۸۴۶	۰/۸۵۰	۱			
Zn	۰/۳۳۲	-۰/۴۲۵	۰/۵۲۳	۰/۸۷۳	۰/۷۵۰	-۰/۸۷۶	۰/۶۱۰	۰/۸۷۲	۱				
Cu	۰/۱۷۶	-۰/۳۶۷	۰/۴۴۹	۰/۸۴۹	۰/۸۲۰	-۰/۸۱۷	۰/۵۶۰	۱					
K ₂ O	۰/۰۸۶	-۰/۵۰۲	۰/۹۲۸	۰/۶۴۸	۰/۴۹۰	-۰/۷۹۳	۱						
Na ₂ O	۰/۷۳۰	۰/۵۴۶	-۰/۷۰۳	-۰/۹۳۵	-۰/۸۰۹	۱							
CaO	-۰/۰۴	-۰/۲۹۶	۰/۴۰۸	۰/۹	۱								
MgO	۰/۱۳۲	-۰/۵۱۵	۰/۵۵۰	۱									
Al ₂ O ₃	۰/۲۹۶	-۰/۳۶۹	۱										
TiO ₂	-۰/۳۶۵	۱											
SiO ₂	۱												
نقره													

جدول (۴-۶): ضرایب اسپیرمن برای داده‌های باز شده به روش clr

Ni	۰/۵۶۷	۰/۴۶۰	۰/۶۶۷	۰/۰۰۷	۰/۱۶۰	۰/۴۲۷	۰/۶۴۷	۰/۴۸۷	۰/۱۱۳	۰/۰۵۳	۰/۵۱۳	۰/۰۷۳	۱
As	۰/۱۴۷	۰/۲	۰/۰۷۳	۰/۴۸۰	۰/۴۲۰	۰/۳۴۷	۰/۰۵۳	۰/۳۳۳	۰/۳۶۰	۰/۳۲۷	۰/۳۴۷	۱	
S	۰/۳۳۳	۰/۷۷۳	۰/۷۱۳	۰/۳۰۷	۰/۰۰۷	۰/۷۶۷	۰/۶۸۰	۰/۶۹۳	۰/۳۶۰	۰/۱۲۷	۱		
Pb	۰/۰۰۷	۰/۰۰۷	۰/۱۳	۱/۰۰۷۰	۰/۰۵۳	۰/۰۲۷	۰/۰۳۳	۰/۳	۰/۲۸۷	۱			
Zn	۰/۳۲۰	۰/۳۶۰	۰/۲۴۷	۰/۳۷۳	۰/۱۹۳	۰/۴۲۰	۰/۲۶۷	۰/۵۰۷	۱				
Cu	۰/۵۶۰	۰/۵۳۳	-۰/۱۵	۰/۲۴۰	۰/۱	۰/۵۶۷	۰/۵۰۷	۱					
K ₂ O	۰/۸۸۰	۰/۷۶۰	۰/۸۷۳	۰/۱۲۰	۰/۱۱۳	۰/۷۱۳	۱						
Na ₂ O	۰/۸۳۳	۰/۸۴۷	۰/۶۹۳	۰/۳۶۷	۰/۰۱۳	۱							
CaO	۰/۰۸۷	۰/۱۲۷	۰/۱۳۳	۰/۴۶۰	۱								
MgO	۰/۲۲۷	۰/۲۸۰	۰/۱۲۷	۱									
Al ₂ O ₃	۰/۸۳۳	۰/۷۶۷	۱										
TiO ₂	۰/۸۴۰	۱											
SiO ₂	۱												
شماره	SiO ₂	TiO ₂	Al ₂ O ₃	MgO	CaO	Na ₂ O	K ₂ O	Cu	Zn	Pb	S	As	Ni

همان‌گونه که در جداول (۴-۵) و (۴-۶) مشاهده می‌شود، این عناصر همبستگی بالایی دارند. برای

داده‌های بسته میزان ضریب همبستگی بزرگتر از میزان این ضریب در داده‌های باز شده است. این امر

ممکن است به دلیل از بین رفتن قید از روی داده‌های بسته باشد، زیرا با برداشته شدن این قید رابطه بین عناصر تحمیلی نیست و اعداد می‌توانند آزادانه تغییر نمایند.

۳-۴ تحلیل مولفه‌های اصلی به روش معمولی و مقاوم

روش تجزیه و تحلیل مولفه‌های اصلی (PCA^۱) تکنیکی برای پیدا کردن ترکیبات خطی از متغیرهای اولیه همبسته که تشکیل یک دستگاه محور مختصات جدید را می‌دهند است. این ترکیبات خطی، مولفه‌های اصلی نامیده می‌شوند و دارای خواص زیر است:

۱- بخش اعظمی از تغییرپذیری می‌تواند به وسیله تعداد محدودی از مولفه و متغیرهای جدید (k) توجیه شود. در این صورت گفته می‌شود که تعداد بعدها از p به k کاهش یافته است.

۲- مولفه‌های جدید، که محصول ترکیب خطی متغیرهای اولیه هستند، بین خود همبستگی نشان نمی‌دهند. این امر آزمون روش مورد نظر را آسان می‌کند.

در استفاده از ابزار PCA به دنبال دستیابی به دستگاه محورهای مختصاتی هستیم که بتواند یک فضای p بعدی را با تعداد محدودی بعد (متغیر جدید) ارزیابی نمود به گونه‌ای که بخش عمده تغییرپذیری را توجیه کند. از خصوصیات فوق می‌توان نتیجه گرفت: متغیرهای اولیه بایستی همبسته باشند تا بتوان روش PCA را به کار برد (جدول ۴-۵ و ۴-۶).

هدف از روش تجزیه و تحلیل مولفه‌های اصلی، آن است که بخش اعظمی از تغییرپذیری بین مشاهدات را که به صورت n نقطه در فضای P بعدی ظاهر می‌شوند، در فضایی با بعد کمتر توجیه کرد. فرآیند فوق معادل آن است که محورهای اولیه (متغیرهای اولیه) را در فضای P بعدی در جهت‌های مختلف طوری بچرخانیم که نقاط (مشاهدات) در امتداد آنها بیشترین گسترش و کشیدگی را داشته باشند [۳۱][۳۲].

¹-Principal Component Analysis

۴-۳-۱ روش مقاوم در برابر روش‌های کلاسیک تحلیل مولفه‌های اصلی

همانگونه که قبلاً مطرح شد، به کار بردن روش‌های مقاوم مزایای بسیاری دارد. استفاده از روش‌های مقاوم با استفاده از داده‌های باز شده به وسیله تبدیل مرکزی منجر به مشکلات عددی می‌شود. بنابراین می‌توان از روش تبدیل ایزومتريک به عنوان جایگزین استفاده کرد، تا مشکلات قبلی حل شود هنگام باز کردن داده‌ها با استفاده از تبدیل ایزومتريک، متغیرهای مصنوعی ایجاد می‌شود. بنابراین رابطه با متغیرهای اصلی از بین می‌رود. رسم کردن مقدار PC و بار به دست آمده نموداری به دست می‌دهد که قابل تفسیر نیست. بنابراین باید نتایج را دوباره با استفاده از تبدیل معکوس به فضای سیمپلکس برد و سپس با استفاده از تبدیل clr باز کرد. سپس می‌توان نتایج را به صورت یک نمودار دوتایی قابل تفسیر رسم کرد.

۴-۳-۲ بررسی داده‌ها برای انجام PCA

آماره‌هایی وجود دارد که از طریق آن پژوهشگر قادر به تشخیص مناسب بودن داده‌ها برای تحلیل مولفه‌های اصلی است. از جمله این روش‌ها استفاده از ضریب KMO^1 است که مقدار آن همواره بین صفر و یک در نوسان است و از رابطه (۲-۴) به دست می‌آید [۳۳]:

$$KMO = \frac{\sum \sum r_{ij}^2}{\sum \sum r_{ij}^2 + \sum \sum a_{ij}^2} \quad (2-4)$$

که در آن r_{ij}^2 ضریب همبستگی ساده بین متغیرهای i و j و a_{ij}^2 ضریب همبستگی جزئی بین آنهاست. اگر مجموع ضرایب همبستگی جزئی بین همه زوج متغیرها در مقایسه با مجموع مجذورات ضرایب همبستگی کوچک باشد، اندازه KMO نزدیک به یک خواهد بود.

در صورتیکه مقدار KMO کمتر از ۰/۵ باشد، داده‌ها برای تحلیل مولفه‌های اصلی مناسب نخواهد بود و اگر مقدار آن بین ۰/۵ و ۰/۶۹ باشد می‌توان با احتیاط بیشتر از این روش استفاده کرد. اما در

¹-Keiser Meyer Olkin coefficient

صورتیکه مقدار آن بزرگتر از ۰/۷ باشد، همبستگی‌های موجود در بین داده‌ها برای تحلیل عاملی مناسب خواهد بود. مقادیر کوچک KMO بیانگر آن است که همبستگی بین زوج متغیرها نمی‌تواند توسط متغیرهای دیگر تبیین شود، بنابراین کاربرد تحلیل مولفه‌های اصلی متغیرها ممکن است قابل توجیه نباشد.

جدول (۴-۷): جدول قضاوت درباره ضریب KMO [۳۳]

مقدار KMO	تناسب داده‌ها برای تحلیل مولفه‌های اصلی
بزرگتر یا مساوی ۰/۹	عالی
۰/۸-۰/۹	خیلی خوب
۰/۷-۰/۷۹	خوب
۰/۶-۰/۶۹	متوسط
۰/۵-۰/۵۹	ضعیف
کمتر از ۰/۵	غیرقابل پذیرش

برای اطمینان از مناسب بودن داده‌ها برای تحلیل عاملی افزون بر این که ماتریس همبستگی‌هایی که پایه تحلیل قرار می‌گیرند در جامعه برابر صفر نیست، باید از آزمون کرویت بارتلت^۱ براساس معادله (۴-۲) استفاده کرد [۳۳]:

$$\chi^2 = -\left(n - 1 - \frac{2p+5}{6}\right) \ln|R| \quad (۴-۲)$$

که در آن n معرف تعداد آزمودنی‌ها، p تعداد متغیرها، |R| قدر مطلق دترمینان همبستگی است. این آماره که دارای توزیع مربع کای دو با درجه آزادی 0.5p(p-1) است. مقدار اطلاعات موجود در |R| را با بررسی رابطه بین تعداد مشاهدات و تعداد متغیرها ارزشیابی می‌کند و احتمال خطا را برای رد کردن فرضیه عدم وجود تفاوت از ماتریس همانی می‌آزماید.

آزمون بارتلت این فرضیه را که ماتریس همبستگی‌های مشاهده شده متعلق به جامعه‌ای با متغیرهای ناهمبسته است را می‌آزماید. برای آن که مدل تحلیل مولفه‌های اصلی مفید و دارای معنا

¹ - Bartlett's test of sphericity

باشد، لازم است که متغیرها همبسته باشند، در غیر این صورت دلیلی برای تبیین مدل PCA وجود ندارد. اگر فرضیه متغیرها با هم رابطه ندارند، رد نشود، کاربرد تحلیل مولفه‌های اصلی زیر سوال خواهد رفت [۳۳].

با استفاده از نرم افزار SPSS مقدار این دو آماره محاسبه شد. مقدار آماره KMO برابر ۰/۷۱ به دست آمد که براساس جدول (۷-۴) برای انجام تحلیل مولفه‌های اصلی مناسب است. مقدار آماره آزمون کرویت بارتلت برابر صفر به دست آمد. مقادیر کم این آماره نشان دهنده مناسب بودن داده‌ها برای تحلیل PCA است.

۴-۴ نتایج PCA و RPCA

برای انجام دادن^۱ RPCA بر روی داده‌ها از نرم افزار آماری R و بسته راب کومپوزیشن^۲ استفاده شد. برای این که بتوان RPCA را بر روی داده‌ها اعمال کرد، ابتدا داده‌ها با استفاده از تبدیل نسبت لگاریتمی ایزومتریک باز شده سپس RPCA بر روی داده‌ها اعمال شد. از آنجائیکه در نتیجه باز کردن داده‌ها به روش ایزومتریک رابطه بین متغیرهای اولیه و داده‌های باز شده از بین می‌رود و تفسیر نتایج مشکل است. نتایج دوباره با استفاده از تبدیل نسبت لگاریتمی مرکزی به درون فضای clr آورده می‌شود. در مورد داده‌های بسته نیز با استفاده از نرم‌افزار R و با استفاده از دستور prcomp() انجام شد. امتیازهای PC برای PCA و RPCA در پیوست آورده شده است. مقادیر ویژه برای تحلیل مولفه‌های اصلی معمولی و مقاوم در جدول (۸-۴) آورده شده است. با توجه به جدول (۸-۴)، می‌توان دریافت که میزان λ_1 چند برابر مقادیر ویژه دیگر در نتایج مربوط به PCA است. در حقیقت این موضوع که یکی از مقادیر ویژه چندین برابر مقدار مقادیر ویژه است، بیانگر آن است که یکی از عناصر تغییرپذیری بسیار بیشتری نسبت به عناصر دیگر دارد.

1 - Robust Principal Component Analysis

2 - robCompositions

جدول (۴-۸): مقادیر ویژه برای PCA و RPCA

مقادیر ویژه برای RPCA	مقادیر ویژه برای PCA	ردیف
۱/۶۸	۸/۰۷۹	λ_1
۰/۶۱۶	۱/۷۸۱	λ_2
۰/۰۵۹	۱/۳۰۴	λ_3
۰/۰۲۵	۰/۸۴۸	λ_4
۰/۰۱۲	۰/۴۵۹	λ_5
۰/۰۰۸	۰/۲۴۴	λ_6
۰/۰۰۴۳	۰/۱۱۷	λ_7
۰/۰۰۱۸	۰/۰۶۸	λ_8
۰/۰۰۰۹	۰/۰۵۲۲	λ_9
۰/۰۰۰۶	۰/۰۱۹	λ_{10}
۰/۰۰۰۵	۰/۰۰۹۷	λ_{11}
۰/۰۰۰۴	۰/۰۰۸۵	λ_{12}
-	۰/۰۰۵۴	λ_{13}

با مراجعه به جدول (۴-۱۰)، می‌توان مشاهده کرد که عنصر SiO_2 تغییرپذیری بسیار بالایی دارد. تغییرپذیری زیاد این عنصر باعث می‌شود که در بررسی‌ها تغییرپذیری عناصر دیگر در سایه قرار گیرد. این موضوع به دلیل تفاوت مقدار عنصرها در آنالیزهای شیمیایی روی می‌دهد.

در حالتی که داده‌ها باز شده‌اند، با این که λ_1 هنوز نسبت به دیگر مقادیر ویژه بزرگتر است، می‌توان مشاهده کرد که بزرگی آن به طور قابل ملاحظه‌ای کاهش پیدا کرده است. بنابراین با اعمال تبدیل‌های لگاریتمی می‌توان از کم تاثیر شدن عنصرهایی با مقدار کوچک‌تر جلوگیری کرد.

۴-۴-۴ نمودار دوتایی^۱

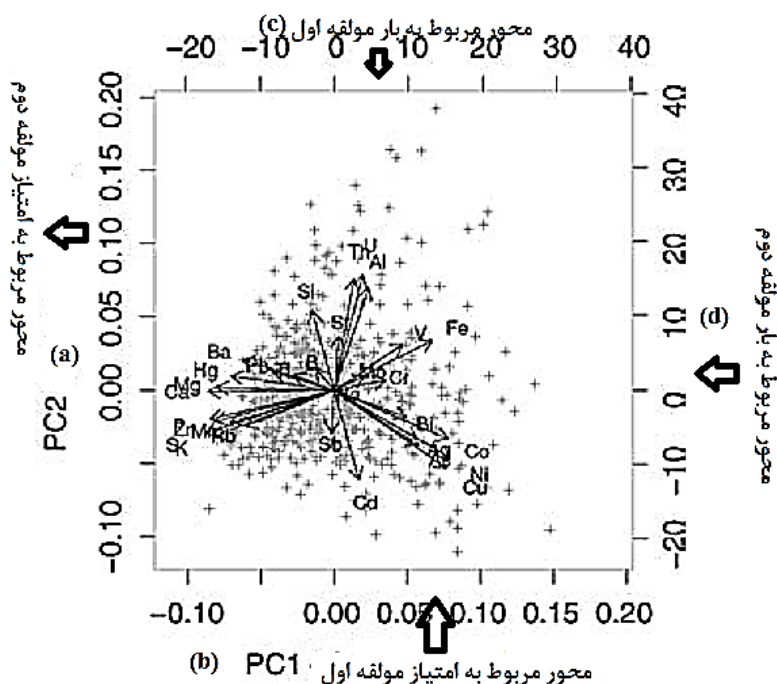
با استفاده از نمودار دوتایی می‌توان مقادیر PC و بار را به صورت همزمان به تصویر کشید. دلیل آنکه به این نمودارها را نمودار دوتایی می‌گویند آن است که هم در مورد مقادیر PC (نقاطی که نشان‌دهنده نقاط مربوط به نمونه‌برداری هستند) و هم در مورد بار (پیکان‌ها) اطلاعات به دست می‌دهد و نه به این

^۱ - Biplot

دلیل که دو تا از مولفه‌های اصلی را در برابر یکدیگر بررسی می‌کند. بنابراین با استفاده از این نمودار می‌توان رابطه بین متغیرها (از طریق بار) و مشاهدات (از طریق مقادیر PC) را به تصویر کشید. بارها (پیکان‌ها) نشان‌دهنده عناصر هستند. طول این پیکان‌ها در این نمودار به طور مستقیم در ارتباط با میزان تغییرپذیری توضیح داده شده توسط دو مولفه نمایش داده شده در نمودار است و زاویه بین هر دو پیکان نشان‌دهنده میزان همبستگی آن دو عنصر است. برای مثال یک پیکان بسیار کوتاه نشان‌دهنده آن است که در دو مولفه اصلی مورد استفاده در رسم نمودار تقریباً هیچ اطلاعاتی در مورد این عنصر وجود ندارد.

چون نمودار پراکندگی تنها دارای دو بعد است، فقط می‌توان دو مولفه را در برابر یکدیگر به تصویر کشید. در شکل (۴-۵) مثالی از نمودار دوتایی آورده شده است. در این شکل محور (a) مربوط به امتیازات مولفه دوم و محور (b) مربوط به امتیازات مولفه اول است. محورهای مربوط به بار مولفه اول و دوم هم به ترتیب در شکل با (d) و (c) مشخص شده است [۳۲].

تفسیر نمودارهای دوتایی برای داده‌های ترکیبی نسبت به نمودارهای دوتایی معمولی تا حدودی متفاوت است. در نمودارهای معمولی اندازه شعاع را می‌توان مستقیماً به تغییرپذیری عنصر مربوطه مرتبط کرد. ولی در نمودارهای دوتایی برای داده‌های ترکیبی اندازه شعاع متناسب با واریانس نسبت لگاریتمی عنصر مربوطه به میانگین هندسی ترکیبی مجموعه داده است (معادله (۳-۳۴)). بنابراین نمی‌توان اندازه شعاع را به تغییرپذیری خود عنصر به تنهایی تفسیر کرد. مهم‌ترین جزء در نمودار دوتایی برای داده‌های ترکیبی لینک است. لینک خط راستی است که نوک پیکان دو شعاع را به هم وصل می‌کند. هر چه اندازه لینک بزرگتر باشد، واریانس نسبت لگاریتمی (واریانس clr) بین این دو عنصر بیشتر است. در نتیجه نسبت لگاریتمی بین این دو سهم بیشتری در واریانس clr کلی دارد. در شکل (۴-۶) نمودار دوتایی را برای نتایج PCA و RPCA نشان می‌دهد. در قسمت ب شکل (۴-۶) مشاهده



شکل (۴-۵): نمودار دوتایی و توضیحات مربوط به محورهای آن [۳۲]

می‌شود که بزرگترین شعاع مربوط به دو عنصر S و As می‌باشد. شعاع‌های مربوط به این دو عنصر در جهت‌های متفاوت از یکدیگر قرار دارد. بنابراین بزرگترین لینک‌ها متعلق به این دو عنصر می‌باشد. بنابراین بیشترین میزان واریانس Clr توجیه شده توسط مولفه اول و دوم تحلیل مولفه‌های مقاوم به این دو عنصر تعلق دارد. به عبارت دیگر بیشترین میزان اطلاعات در مولفه اول و دوم مربوط به این دو عنصر است.

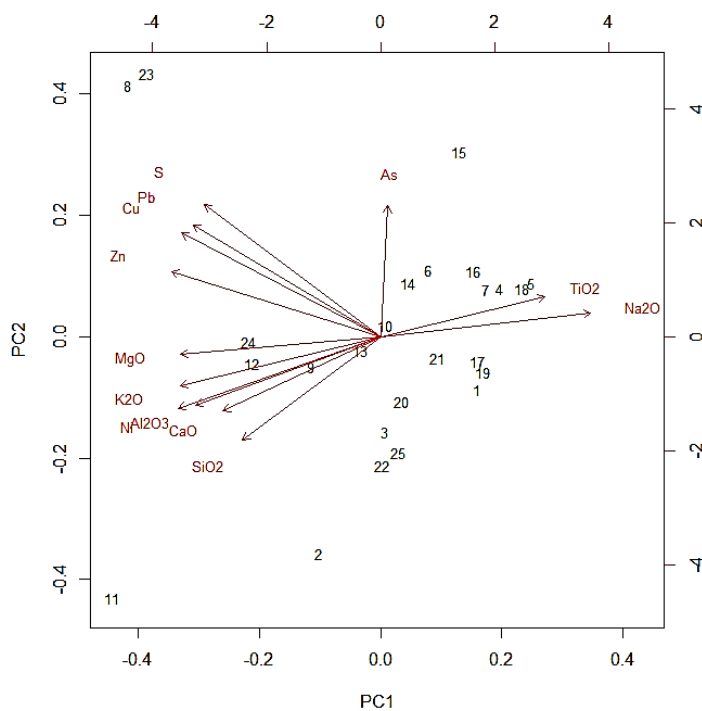
در نمودار دوتایی معمولی زاویه بین دو شعاع نشان‌دهنده تقریبی از همبستگی بین دو عنصر است. در نمودارهای دوتایی داده‌های ترکیبی، کسینوس زاویه بین دو لینک متقاطع نشان‌دهنده تقریبی از همبستگی بین نسبت‌های لگاریتمی دو جفت عنصر است (معادله ۳-۳۵).

اگر نوک پیکان شعاع‌ها برای دو یا چند عنصر در یک نقطه قرار گیرد و یا بسیار نزدیک به هم باشد، نشان‌دهنده آن است که نسبت لگاریتمی بین این عناصر صفر و یا نزدیک به صفر است. در نتیجه کم‌ترین اطلاعات در مورد این عناصر در دو مولفه در حال بررسی وجود دارد. همچنین ثابت ماندن

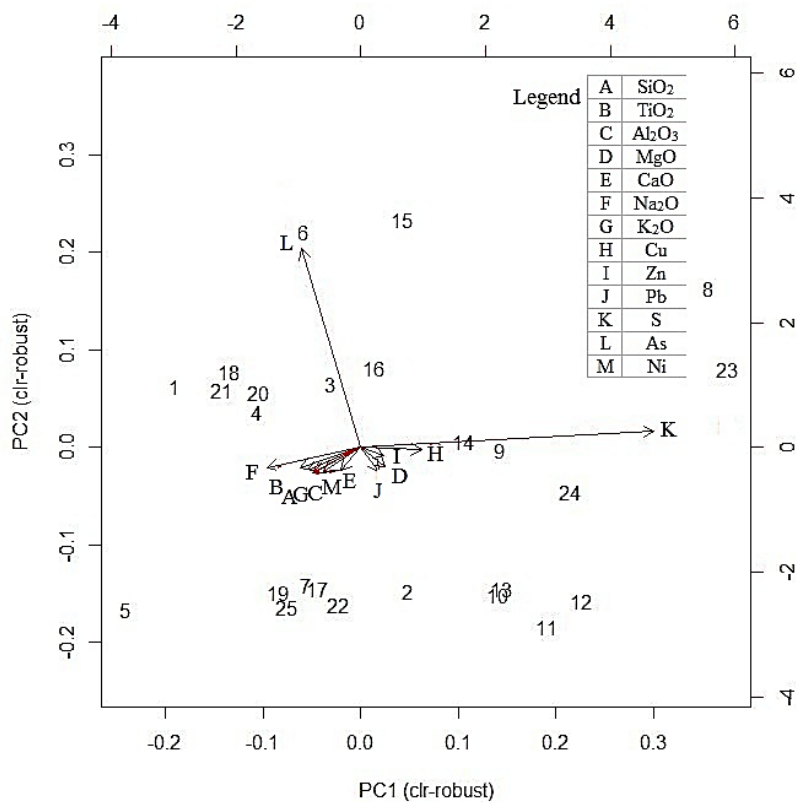
نسبت لگاریتمی بین این عناصر نشان‌دهنده همبستگی بالای بین آنها نیز هست. زیرا نسبت بین این دو عنصر ثابت باقی مانده و با افزایش یک عنصر مقدار عنصر دیگر نیز افزایش پیدا کرده است و بالعکس. در شکل (۷-۴) قسمت ب، عناصر Cu، Pb، Zn، MgO و CaO در کنار یکدیگر قرار گرفته‌اند. در حالی که عناصر SiO_2 ، TiO_2 ، Al_2O_3 و K_2O در کنار یکدیگر قرار گرفته‌اند. با توجه به این موضوع می‌توان دریافت که این عناصر سهم قابل توجهی در واریانس کل دارند و این دو گروه از عناصر همبستگی خوبی دارند.

در نمودارهای دوتایی معمولی، اندازه شعاع و زاویه بین دو شعاع به ترتیب نشان‌دهنده میزان سهم عنصر در واریانس کلی و همبستگی بین دو عنصر است. با توجه به شکل (۷-۴) قسمت الف، اندازه شعاع برای تمام عناصر تقریباً یکسان است بنابراین میزان سهم تمام عناصر در واریانس کلی به یک اندازه است. عناصر Cu، Pb، Zn و S دارای همبستگی مثبت هستند. عناصر TiO_2 و Na_2O با یکدیگر و عناصر SiO_2 ، Al_2O_3 ، CaO، MgO با یکدیگر دارای همبستگی مثبت هستند.

با مقایسه شکل (۷-۴) قسمت الف و ب می‌توان متوجه شد که هر دوی این نمودارهای دوتایی همبستگی بین عناصر را تقریباً مشابه یکدیگر نشان می‌دهند. عناصر کانی‌سازی Cu، Pb، Zn در هر دو شکل دارای همبستگی مثبت هستند. ولی در شکل الف، همبستگی بین عناصر کانی‌سازی با عنصر S نیز قابل مشاهده است. در شکل ب، به دلیل پیچیده‌تر بودن محاسبه همبستگی بین عناصر همبستگی بین عنصر S و عناصر کانی‌سازی قابل مشاهده نیست. لازم به ذکر است میزان گوگرد در نمونه‌های ماسه‌بادی از دو فرآیند در منطقه تاثیر پذیرفته است. یکی از این فرآیندها کانی‌سازی و دیگری تبخیر بالا در سطح و تشکیل ژئپس است. در مورد توجیه واریانس باید توجه داشت که برای محاسبه واریانس کل در داده ترکیبی باید از معادله‌های خاصی که پیشنهاد شده است، استفاده کرد.



(الف)



(ب)

شکل (۴-۶): نمودار دوتایی برای الف) تحلیل مولفه‌های اصلی معمولی و ب) مقاوم

داده‌های ترکیبی هندسه متفاوتی (سیمپلکس) نسبت به اعداد حقیقی دارند و برای بررسی واریانس و دیگر آماره‌های آن باید از معادله‌های مخصوص به سیمپلکس استفاده کرد. برای بررسی واریانس درون سیمپلکس از یک ماتریس با عنوان ماتریس واریانس استفاده می‌شود. این ماتریس به صورت مربعی و با همان ابعاد داده‌ها تعریف می‌شود. درآیه‌های این ماتریس را می‌توان از معادله (۴-۴) محاسبه کرد [۲۸]:

$$t_{ij} = var \ln\left(\frac{x_i}{x_j}\right) \quad (4-4)$$

که در آن t_{ij} درایه i ام ماتریس واریانس و x_i و x_j عناصر i و j هستند. در واقع با استفاده از این رابطه نسبت تغییرات دو عنصر نسبت به یکدیگر سنجیده می‌شود. در جدول (۴-۹) ماتریس واریانس برای داده‌ها نمایش داده شده است. در این شکل هر چه رنگ گرم‌تر شود، میزان تغییرپذیری نسبت لگاریتمی آن عنصر در داده‌ها بیشتر است. در ستون آخر واریانس clr داده‌ها نمایش داده شده است. مشاهده می‌شود که بیشترین میزان تغییرپذیری برای عناصر S و As است.

جدول (۴-۹): ماتریس واریانس داده‌ها

Variation array:

	Variance ln(Xi/Xj)													
$X_i \backslash X_j$	SiO2	TiO2	Al2O3	MgO	CaO	Na2O	K2O	Cu	Zn	Pb	S	As	Ni	clr variances
SiO2		0.0051	0.0021	0.0715	0.0635	0.0432	0.0011	0.1474	0.0915	0.0681	1.2278	0.3202	0.0110	0.0345
TiO2	-3.8480		0.0106	0.0983	0.0873	0.0272	0.0093	0.1805	0.1233	0.0912	1.3165	0.3038	0.0256	0.0518
Al2O3	-1.4598	2.3882		0.0594	0.0554	0.0607	0.0006	0.1302	0.0769	0.0568	1.1742	0.3214	0.0050	0.0268
MgO	-3.4344	0.4136	-1.9746		0.0216	0.2155	0.0603	0.0499	0.0265	0.0444	0.8087	0.4189	0.0395	0.0238
CaO	-3.7154	0.1326	-2.2556	-0.2810		0.1910	0.0574	0.0868	0.0485	0.0689	0.9713	0.4418	0.0390	0.0406
Na2O	-4.2414	-0.3935	-2.7816	-0.8070	-0.5260		0.0558	0.3140	0.2398	0.1847	1.0114	0.3369	0.0932	0.1366
K2O	-3.2902	0.5578	-1.8304	0.1443	0.4252	0.9513		0.1300	0.0777	0.0566	1.1729	0.3242	0.0061	0.0267
Cu	-9.7744	-5.9264	-8.3146	-6.3400	-6.0590	-5.5330	-6.4842		0.0196	0.0273	0.5671	0.4336	0.1063	0.0452
Zn	-8.8059	-4.9579	-7.3461	-5.3715	-5.0905	-4.5644	-5.5157	0.9685		0.0199	0.7261	0.4056	0.0571	0.0237
Pb	-10.1457	-6.2977	-8.6859	-6.7113	-6.4303	-5.9043	-6.8555	-0.3713	-1.3398		0.7819	0.3747	0.0466	0.0166
S	-8.2281	-4.3801	-6.7683	-4.7937	-4.5127	-3.9866	-4.9379	1.5463	0.5778	1.9176		1.4049	1.0873	0.8656
As	-11.6608	-7.8128	-10.2010	-8.2264	-7.9454	-7.4194	-8.3706	-1.8864	-2.8549	-1.5151	-3.4327		0.3366	0.2937
Ni	-10.0912	-6.2432	-8.6314	-6.6568	-6.3758	-5.8498	-6.8011	-0.3168	-1.2853	0.0545	-1.8632	1.5696		0.0191
	Mean ln(Xi/Xj)													1.6047 Total Variance

با بررسی واریانس داده‌های بسته که در جدول (۴-۱۰) آورده شده است، مشاهده می‌شود که بیشترین واریانس متعلق به عنصر SiO_2 است. با مراجعه به شکل (۴-۴) انتظار می‌رفت که بزرگترین

شعاع متعلق به عنصر SiO_2 باشد، که این گونه نیست. اما در قسمت ب شکل (۴-۴) بزرگترین لینکها متعلق به As و S است که این موضوع با نتایج جدول (۴-۸) مطابقت دارد. بنابراین نمودار دوتایی برای داده‌های ترکیبی در توجیه تغییرپذیری بهتر عمل کرده است.

جدول (۴-۱۰): واریانس داده‌های بسته

عنصر	واریانس	عنصر	واریانس
SiO_2	۱/۴۳	K_2O	۰/۰۱۲
TiO_2	۰/۰۰۶	Cu	۰
Al_2O_3	۰/۷۸۳	Zn	۰
MgO	۰/۳۶۳	Pb	۰
CaO	۰/۲۸۷	S	۰/۰۲
Na_2O	۰/۰۲۸	As	۰
Ni	۰		

بارهای فاکتوری محاسبه شده برای تحلیل مولفه‌های اصلی معمولی و مقاوم به ترتیب در جداول (۴-۱۰) و (۴-۱۱) آورده شده است.

با توجه به جدول (۴-۱۱) و (۴-۱۲)، برای انتخاب PC مناسب برای انجام مرحله بعدی و تعیین حد زمینه و آنومال بیشترین مقدار بار فاکتوری برای عناصر مس، سرب و روی در جدول انتخاب شد. برای مس در PCA، PC شماره دو و در مقاوم PC شماره نه انتخاب شد. برای سرب و روی هم به ترتیب در تحلیل مولفه‌های اصلی معمولی و مقاوم PC های یازده و هشت و سه انتخاب شد.

۴-۵ نمودار احتمال

مدل‌سازی نمودار احتمال می‌تواند کاربردهای مختلفی داشته باشد. یکی از کاربردهای عمده نمودار احتمال در بررسی چگونگی رفتار داده‌های ژئوشیمیایی است [۳۴]. با استفاده از نمودار احتمال می‌توان یک جامعه مرکب از دو یا چند زیرجامعه را به اجزاء آن (ناهنجاری و زمینه) تفکیک کرد [۳۱].

جدول (۴-۱۱): بار فاکتوری RPCA

عنصر	SiO ₂	TiO ₂	Al ₂ O ₃	MgO	CaO	Na ₂ O	K ₂ O	Cu	Zn	Pb	S	As	Ni
PC1	-۰/۱۳	-۰/۱۶	-۰/۱۲	۰/۱۱۴	۰/۰۶۶	-۰/۲۷	-۰/۱۱	۰/۱۸۳	۰/۱۳۴	۰/۰۳۵	۰/۷۷۳	-۰/۴۱	-۰/۰۷
PC2	-۰/۱۴	-۰/۱۱	-۰/۱۴	-۰/۰۲	-۰/۰۷	-۰/۱۸	-۰/۱۴	-۰/۰۱	-۰/۰۶	-۰/۱۴	۰/۳۲۳	۰/۵۵۹	-۰/۱۰
PC3	-۰/۰۶	-۰/۲۲	۰/۰۹۱	۰/۴۳۱	۰/۲۳۲	-۰/۵۷	۰/۰۱۱	-۰/۰۶	۰/۴۰۵	-۰/۱۶	-۰/۳۳	۰/۰۴۰	۰/۳۲۳
PC4	۰/۰۸	۰/۲۰	۰/۱۵۳	۰/۵۱۲	-۰/۲۰	-۰/۰۸	۰/۱۱۴	-۰/۵۰	-۰/۴۹	-۰/۱۲	۰/۲۱۲	-۰/۰۵	۰/۱۷۹
PC5	۰/۰۸۹	-۰/۲۴	-۰/۱۶	۰/۰۷۰	۰/۵۳۷	۰/۴۳۵	-۰/۰۴	-۰/۵۱	۰/۱۹۱	-۰/۲۳	۰/۱۰۶	•	-۰/۲۲
PC6	۰/۱۶۴	۰/۰۳۶	-۰/۰۸	۰/۴۱۷	-۰/۵۴	۰/۱۹۹	۰/۰۵۹	۰/۱۲۷	۰/۳۸۰	-۰/۲۱	-۰/۰۴	•	-۰/۴۸
PC7	۰/۲۲۶	-۰/۱۹	۰/۲۸۸	-۰/۳۳	-۰/۱۱	۰/۰۲۴	۰/۳۰۵	۰/۱۲۱	۰/۰۲۹	-۰/۷۱	۰/۱۵۵	-۰/۰۵	۰/۲۴۷
PC8	۰/۰۰۴	-۰/۶۰	-۰/۳۲	۰/۱۴۸	-۰/۳۰	-۰/۳۱۳	۰/۰۲۳	۰/۰۱۶	•	۰/۲۱	-۰/۰۱	۰/۰۱۶	۰/۵۲۰
PC9	۰/۰۸۷	-۰/۱۸	۰/۰۴۱	۰/۳۵۸	۰/۲۰	۰/۱۴۷	-۰/۲۸	۰/۵۴۵	-۰/۴۹	-۰/۲۳	-۰/۱۳	-۰/۰۱	-۰/۱۲
PC10	-۰/۰۹	۰/۵۴۰	-۰/۵۶	۰/۰۳۷	۰/۳۹۱	۰/۰۹۵	-۰/۲۲	۰/۰۹۸	۰/۱۲۸	-۰/۳۶	-۰/۰۲	-۰/۰۴	۰/۳۹۰
PC11	۰/۰۴۹۱	-۰/۰۲	-۰/۵۵	-۰۰۰/۰۴	۰/۱۴۲	-۰/۳۰	۰/۵۱۴	۰/۰۶۶	-۰/۱۹	۰/۰۸	-۰/۰۳	۰/۰۱۱	-۰/۱۶
PC12	۰/۷۲۵	•	۰/۰۷۴	-۰/۱۲	-۰/۰۹	-۰/۱۴	-۰/۶۱	-۰/۱۲	۰/۰۸۸	۰/۰۹	۰/۰۳	۰/۰۰۳	۰/۰۹۲

جدول (۴-۱۲): بار فاکتوری PCA

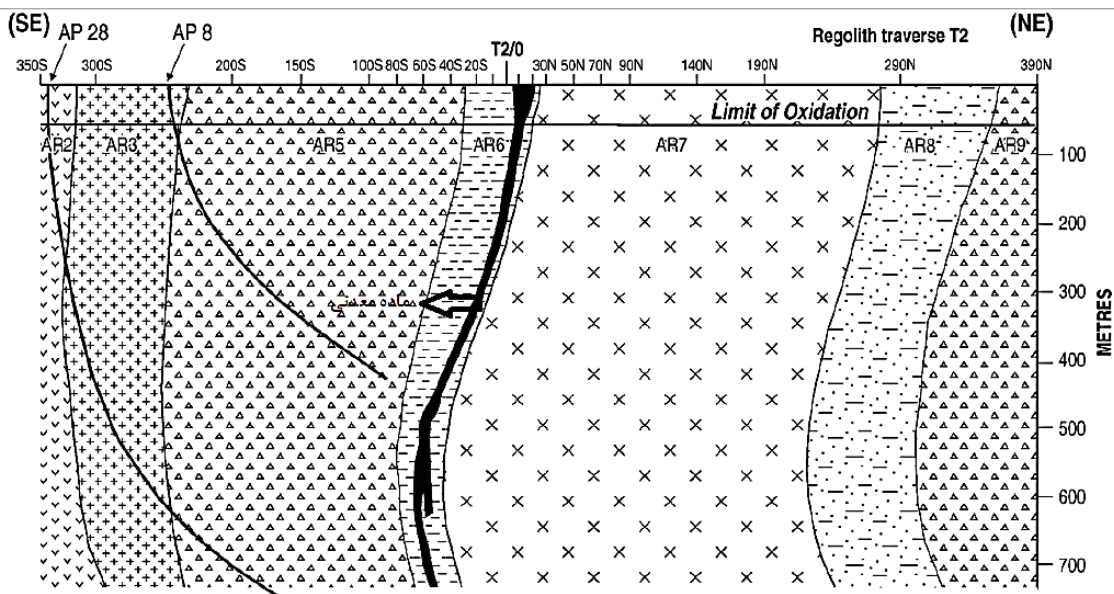
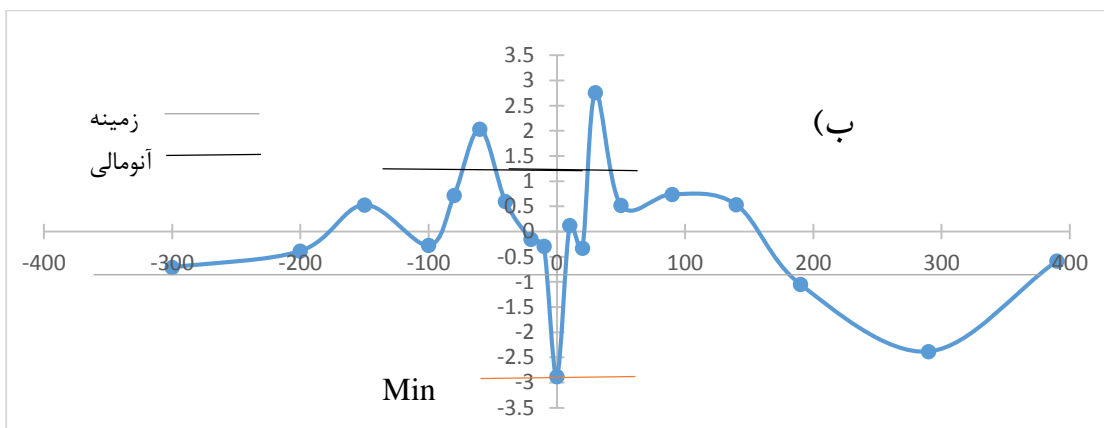
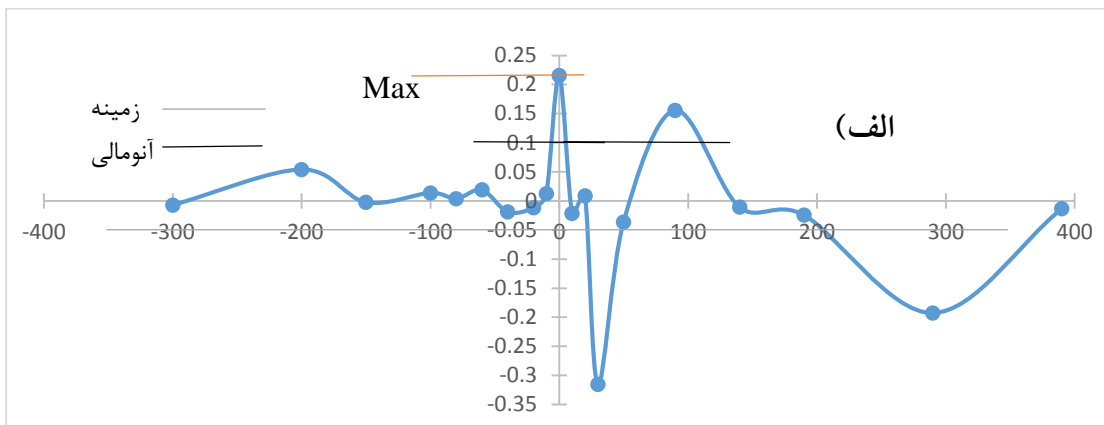
عنصر	SiO ₂	TiO ₂	Al ₂ O ₃	MgO	CaO	Na ₂ O	K ₂ O	Cu	Zn	Pb	S	As	Ni
PC1	-۰/۲۱	۰/۲۵۳	-۰/۲۸	-۰/۳۰	-۰/۲۴	۰/۳۲۳	-۰/۳۰	-۰/۳۰	-۰/۳۲	-۰/۲۸	-۰/۲۷	۰/۰۱۰	-۰/۳۱
PC2	-۰/۳۳	۰/۱۳۰	-۰/۲۲	-۰/۰۵	-۰/۲۴	۰/۰۷۷	-۰/۱۶	۰/۳۴۰	۰/۲۱۴	۰/۳۶۷	۰/۴۳۳	۰/۴۶۹	-۰/۲۳
PC3	۰/۴۸	۰/۲۱۳	۰/۳۸۳	-۰/۲۹	-۰/۴۳	۰/۱۴۶	۰/۳۰۷	-۰/۰۸	-۰/۱۵	۰/۰۵۲	۰/۰۹۶	۰/۳۶۸	۰/۰۹۱
PC4	-۰/۱۰	۰/۳۱۴	۰/۰۶۳	۰/۲۶۹	۰/۳۳۱	-۰/۲۳	-۰/۱۲	-۰/۰۸	-۰/۰۴	-۰/۲۳	-۰/۲۳	۰/۶۹۵	۰/۱۸۴
PC5	-۰/۲۵	۰/۷۸۸	۰/۲۲۵	۰/۰۶۳	-۰/۰۵	-۰/۱۸	۰/۰۸۳	۰/۰۷۳	-۰/۰۱	۰/۱۵۲	۰/۰۱۲	-۰/۱۴	۰/۰۸۰
PC6	۰/۵۶۶	۰/۲۸۲	-۰/۰۱	۰/۰۷۹	۰/۴۶۵	۰/۳۵۳	-۰/۳۳	۰/۰۹۴	۰/۰۶۸	۰/۱۸۹	۰/۰۶۶	-۰/۰۵	-۰/۲۸
PC7	-۰/۲۹	-۰/۰۵	۰/۱۸۱	-۰/۵۸	۰/۴۰۹	۰/۳۳۱	-۰/۰۷	۰/۰۳۲	۰	۰/۱۶۰	۰	۰/۰۳۲	۰/۴۷۱
PC8	۰/۱۰	۰/۰۹۹	-۰/۵۸	۰/۲۷۲	-۰/۰۳	۰/۲۰۶	۰/۱۸۹	۰	-۰/۴۴	۰/۳۳۷	۰/۰۸	۰	۰/۴۷۷
PC9	-۰/۱۸	-۰/۱۶	۰/۳۰۳	۰/۰۵۹	۰/۲۵۷	-۰/۱۱	۰/۰۸۱	-۰/۰۳	-۰/۷۲	۰/۲۰۱	۰/۲۴۱	۰/۰۴۳	-۰/۳۶
PC10	۰/۰۴	-۰/۱۰	۰/۲۰۲	۰/۱۶۶	-۰/۲۴	-۰/۱۲	-۰/۶۴	-۰/۴۸	۰	۰/۲۷۰	۰/۱۹۵	-۰/۰۴	۰/۲۸۹
PC11	-۰/۰۷	۰/۰۲۷	-۰/۱۵	-۰/۰۴	۰/۱۵۱	-۰/۰۳	۰/۳۷۲	-۰/۱۶۰	۰/۲۶۴	۰/۵۱۰	-۰/۲۲	۰/۰۹۲	-۰/۲۲
PC12	-۰/۲۰	۰/۰۵۲	۰/۱۴۰	۰/۳۲۵	۰/۰۸۲	۰/۴۷۹	۰/۲۲۴	-۰/۳۳	۰/۱۷۸	-۰/۳۸	۰/۵۰۳	-۰/۰۱	۰
PC13	۰/۱۹۲	۰/۱۳۲	-۰/۳۴	-۰/۴	۰/۲۰۱	-۰/۴۸	۰/۰۱۸	-۰/۲۱	۰	-۰/۲۵	۰/۵۱۴	-۰/۰۳	۰/۰۱۰

در جدول (۴-۱۳) نتایج مدل‌سازی نمودار احتمال بر روی PC های انتخاب شده برای عناصر مس، سرب و روی آورده شده است. برای اعمال نمودار احتمال باید تمام داده‌ها مثبت باشند. به همین دلیل داده‌ها با یک عدد جمع شد تا داده‌ها دارای مقدار مثبت شوند. پس از انجام نمودار احتمال دوباره داده‌ها از آن مقدار کم شد و نتایج بر روی نمودار ترسیم شد.

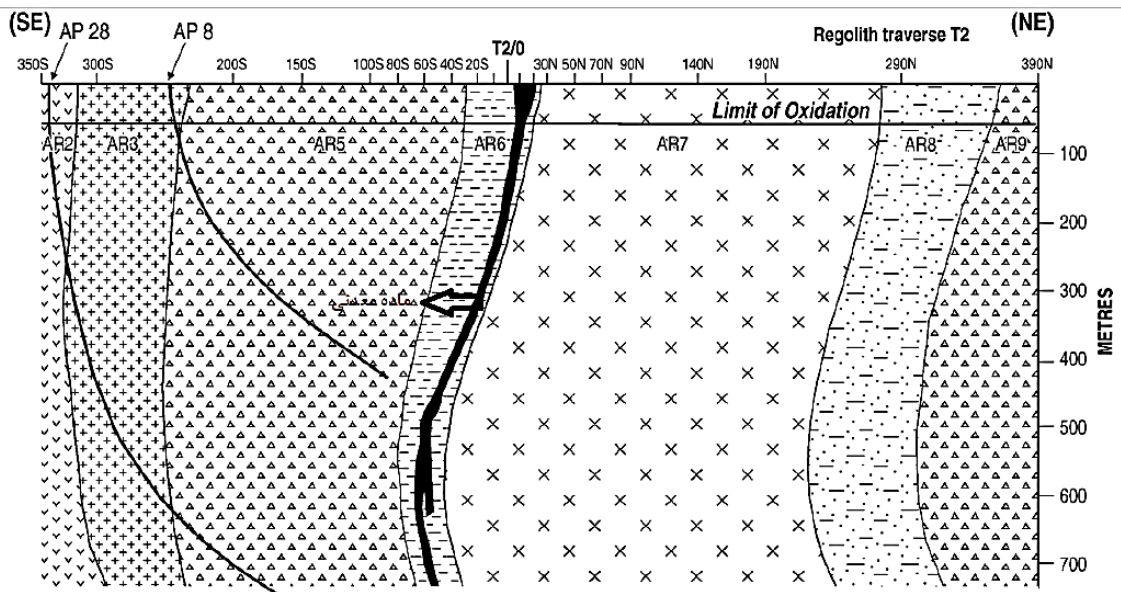
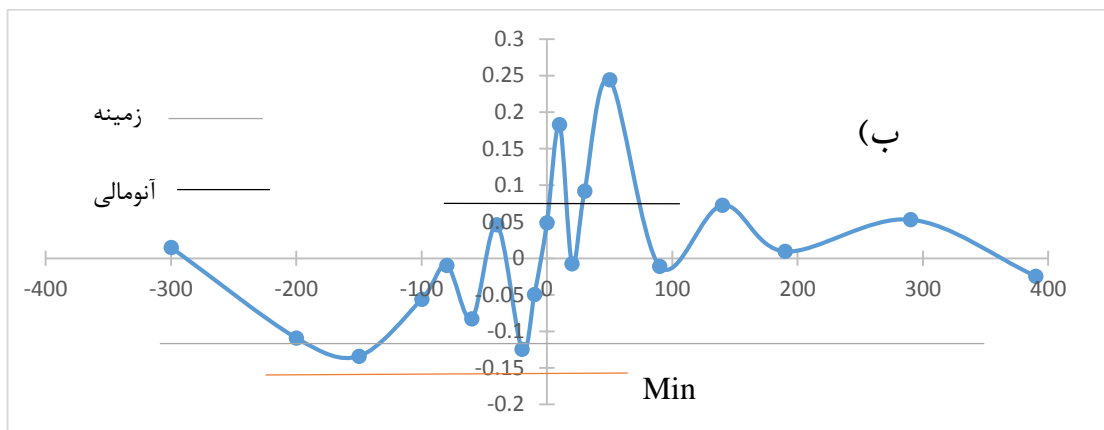
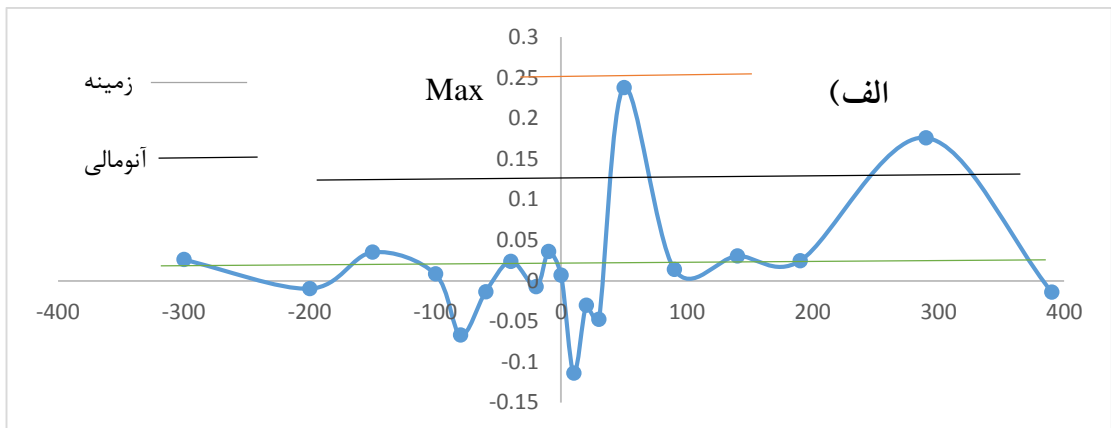
جدول (۴-۱۳): نتایج نمودار احتمال

آنومالی	حد گذر	زمینه	
$0/1 <$	$-0/05 - 0/1$	min-- $0/05$	Cu RPCA
$1/2 <$	$-1/1 - 1/2$	min-- $1/1$	Cu PCA
$0/121 <$	$-0/07 - 0/207$	min-- $0/08$	Pb RPCA
$0/1 <$	$-0/09 - 0/1$	min-- $0/09$	Pb PCA
$-0/5 <$	min-- $0/5$	min-- $0/5$	Zn RPCA
$0/1 <$	$-0/09 - 0/1$	min-- $0/09$	Zn PCA

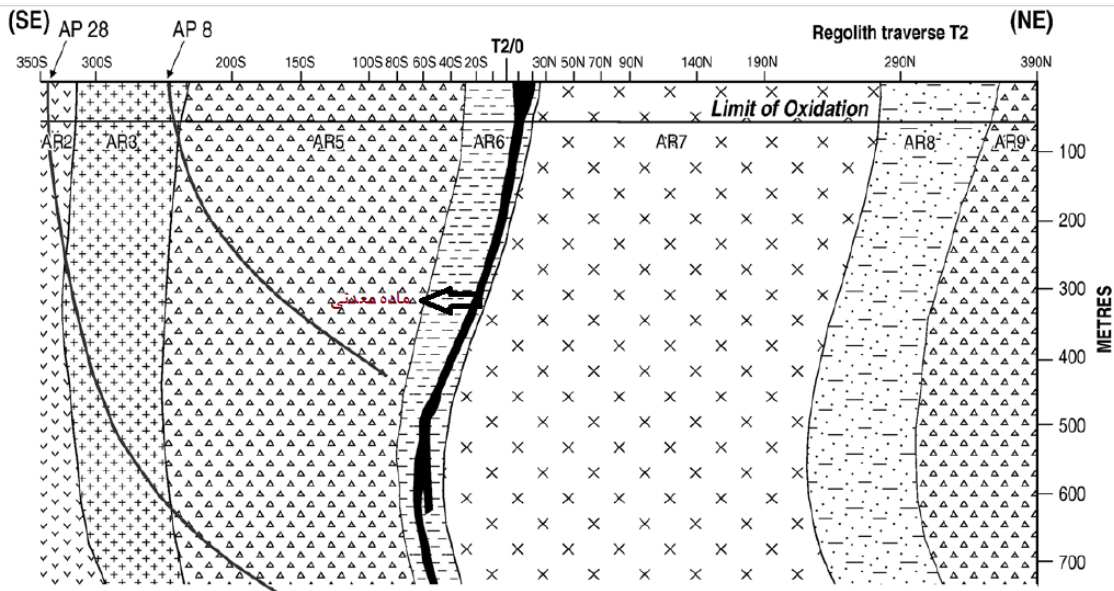
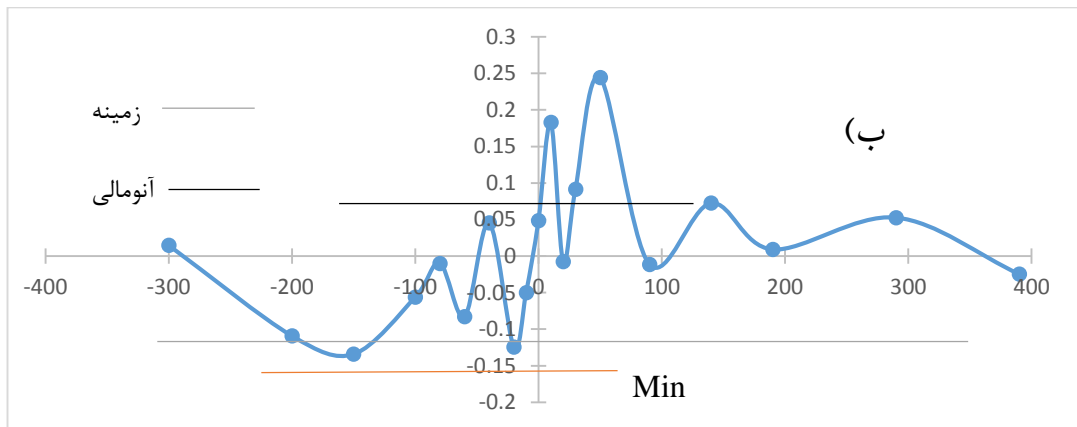
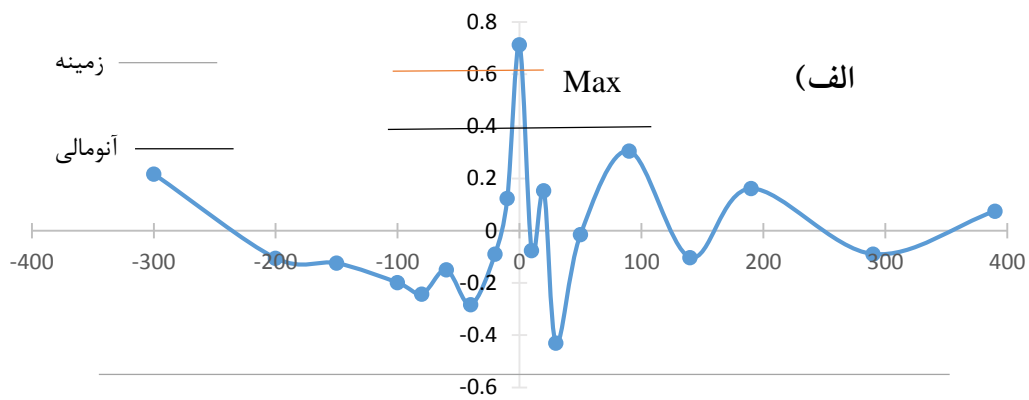
PC های منتخب از PCA و RPCA برای عناصر مربوط به کانی‌سازی در منطقه (Cu,Pb,Zn) به صورت جداگانه بر روی نمودار ترسیم شد. سپس حد زمینه و آنومالی به دست آمده از روش نمودار احتمال روی این نمودارها مشخص شد. هر یک از این نمودارها با مقطع کانی‌سازی در زیر پروفیل نمونه‌برداری مقایسه شد تا مشخص شود که کدام یک عملکرد بهتری برای تشخیص محل کانی‌سازی داشته‌اند. شکل های (۴-۷) تا (۴-۹) نشان‌دهنده این نمودارها به همراه مقطع زیر پروفیل نمونه‌برداری است.



شکل (۴-۷): نمودار منطقه آنومال برای عنصر مس (الف) RPCA (ب) PCA (ج) مقطع کانی‌سازی در منطقه



شکل (۴-۸): نمودار منطقه آنومال برای عنصر سرب الف (RPCA ب) PCA ج) مقطع کانی سازی در منطقه



شکل (۴-۹): نمودار منطقه آنومال برای عنصر روی الف) RPCA ب) PCA ج) مقطع کانی‌سازی در منطقه

برای تشکیل مولفه‌های اصلی در روش‌های PCA و RPCA، تمام عناصر مورد آنالیز با توجه به بار فاکتوری متناظر آنها، دخالت دارند. اگرچه عنصری که بیشترین بار فاکتوری را دارد، بیشترین تاثیر را بر آن مولفه دارد ولی عناصر دیگر نیز با توجه به بار فاکتوری خود در این مولفه تاثیرگذارند. پس مولفه اصلی به دست آمده تجمیعی از هاله‌های ژئوشیمیایی عناصر ایجاد کننده مولفه هستند. پس برای بررسی نمودارهای مولفه‌های اصلی انتخاب شده برای عناصر کانی‌سازی در روش‌های PCA و RPCA باید هم به نقاط پیک موجود در نمودار و هم به روند ایجاد شده توسط این نقاط پیک توجه کرد. (شکل ج ۴-۷)

در شکل (۴-۷) برای عنصر مس، با توجه به مدل‌سازی نمودار احتمال، مناطق آنومال برای روش PCA و RPCA مشخص شده است. در قسمت الف شکل (۴-۷) نمودار مربوط به RPCA و قسمت ب، نمودار مربوط به PCA برای عنصر مس را نشان می‌دهد. در سمت راست این نمودارها روند خاصی مشاهده نمی‌شود. تغییرات ایجاد شده در این قسمت از نمودار از روند خاصی پیروی نمی‌کند و نامنظم است. این موضوع نشان‌دهنده آن است که در این قسمت پروفیل نمونه‌برداری کانی‌سازی وجود ندارد. در قسمت چپ نمودار و از سمت مرکز محور مختصات به سمت راست در نمودار مربوط به PCA تغییرات به صورت منظم‌تر و با روند کاهشی انجام شده است. این موضوع نشان‌دهنده آن است که در این قسمت پروفیل کانی‌سازی احتمالی وجود دارد. در حالی که همین قسمت از نمودار برای RPCA برخلاف نمودار PCA عمل کرده است و در مقابل هر پیک در نمودار PCA در نمودار RPCA یک کاهش مشاهده می‌شود. جدایش موجود در موقعیت مکانی پیک‌ها در روی نمودارهای PCA و RPCA نشان‌دهنده شیب‌دار بودن کانی‌سازی است. برای بهینه کردن نقطه حفاری پیشنهادی نیز می‌توان از نمودار مربوط به هاله‌های تحت کانساری استفاده کرد. در نقطه پایان روند کاهشی در روی نمودار PCA با یک پیک در نمودار RPCA همراه است. این موضوع نشان‌دهنده پایان کانی‌سازی بر روی پروفیل نمونه‌برداری است. در قسمت ج شکل (۴-۷) مقطع کانی‌سازی در زیر پروفیل نمونه‌برداری نشان داده شده است که نشان‌دهنده وجود یک کانی‌سازی شیب‌دار در مرکز مختصات پروفیل نمونه‌برداری است.

با مراجعه به شکل‌های مربوط به عناصر سرب و روی نیز می‌توان همین روندها را برای نمودارهای مربوط به PCA و RPCA مشاهده کرد. در قسمت راست نمودارها، روند تغییرات نامنظم است. در قسمت چپ نمودار PCA روند تغییرات کاهشی است. در قسمت آخر پروفیل در سمت چپ در نمودار PCA با یک نقطه عطف مواجه می‌شویم که نشان دهنده پایان کانی‌سازی است. مشابه نمودار مربوط به عنصر مس، نمودارهای PCA و RPCA مربوط به عناصر سرب و روی نیز تمام تفاسیر مربوط به نمودار مس را برای نمودارهای سرب و روی به کار برد. این موضوع به دلیل نوع کانی‌سازی (VHMS) در این منطقه است.

با بررسی نمودارهای مربوط به PCA و RPCA و تطبیق آن با مقطع پروفیل نمونه‌برداری می‌توان متوجه شد که روش RPCA در شناسایی زون کانی‌سازی بهتر از روش PCA عمل کرده است در حالی که روش PCA با چند افت پایایی به خوبی محدوده سنگ بستر را نشان می‌دهد.

فصل پنجم:

نتیجه گیری و پیشنهاد

۵-۱ نتیجه گیری

یکی از مشکلات آماری داده‌های ژئوشیمیایی ترکیبی مسئله همبستگی‌های غیرواقعی است. این مشکل در تمام داده‌های ترکیبی وجود دارد و باعث اشتباه در تفسیر داده‌ها می‌شود. یکی از روش‌های غلبه بر این مشکل، استفاده از تبدیل‌های نسبت لگاریتمی برای باز کردن آنها است. در این پایان‌نامه از روش تبدیل نسبت لگاریتمی ایزومتريك برای باز کردن داده‌ها استفاده شد. با استفاده از این تبدیل می‌توان داده‌ها را در یک محور مختصات متعامد تصویر کرد. سپس از روش‌های تحلیل مولفه‌های اصلی به روش معمول برای پردازش چند متغیره داده‌های بسته و روش تحلیل مولفه‌های اصلی مقاوم برای پردازش چندمتغیره داده‌های باز شده به روش ایزومتريك استفاده شد. در روش PCA مقدار ویژه مولفه اول بسیار بزرگتر از مقدار ویژه مولفه اول روش RPCA است. این امر ممکن است در نتیجه حضور یک متغیر با تغییرپذیری بسیار بالا باشد که باعث کم‌اثر شدن متغیرهای دیگر در مولفه‌های اول و دوم می‌شود. در روش RPCA مقدار بردار ویژه مولفه اول به طور قابل ملاحظه‌ای کم‌تر شده است. در توجیه مقدار تغییرپذیری نیز با بررسی واریانس برای داده‌های بسته و ماتریس واریانس برای داده‌های باز شده، می‌توان مشاهده کرد که روش RPCA بهتر عمل کرده است.

روش نمودار احتمال یک روش غیرساختاری برای مشخص کردن جوامع مختلف در یک مجموعه داده است. با استفاده از این روش حد زمینه و آنومال برای PC های منتخب محاسبه و روی نمودار تصویر شد. با مقایسه نتایج این دو روش می‌توان به نتایج زیر رسید:

۱- به طور کلی می‌توان گفت هنگامی که روش‌های معمول آنالیز شیمیایی به کار گرفته می‌شود، مشکلات و تفاسیر نادرست ناشی از واریانس داده‌ها است. به طور اخص، این مشکلات هنگامی ظاهر می‌شوند که یک جزء از مجموعه داده‌ها واریانس بزرگتری داشته باشد. ممکن است که واریانس اجزای دیگر زیر سایه واریانس بزرگ این جزء قرار گیرد. در هنگام استفاده از روش تحلیل مولفه‌های اصلی،

این موضوع با مقدار ویژه بسیار بزرگ برای مولفه اول خود را نشان می‌دهد. با استفاده از روش تبدیل نسبت لگاریتمی ایزومتریک می‌توان این مشکل تا حدود زیادی برطرف گردیده است.

۲- فضای سیمپلکس که فضای نمونه‌ای داده‌های ترکیبی است، ساختار برداری متفاوتی نسبت به R (مجموعه داده‌های حقیقی) دارد. استفاده از روش‌های آماری معمول برای بررسی داده‌های ترکیبی به صورت بسته و دورن سیمپلکس ممکن است دارای خطا باشد. مقایسه نتایج نمودار دوتایی داده‌های باز و بسته نشان می‌دهد که با باز کردن داده‌ها بهتر می‌توان ساختار داده‌ها را بررسی کرد. نتایج مربوط به ماتریس واریانس با نتایج نمودار دوتایی داده‌های باز با هم مطابقت کامل دارد. در حالی که واریانس تحت تاثیر مقدار بزرگ SiO_2 قرار دارد، ماتریس واریانس به خوبی نشان می‌دهد که بیشترین سهم در واریانس کلی به S و As متعلق است. نمودار دوتایی مربوط به داده‌های باز نیز به خوبی این موضوع را با بلند بودن پیکان مربوط به این دو عنصر نشان می‌دهد.

۳- با مقایسه نتایج به دست آمده از نمودار احتمال برای داده‌های باز و بسته که بر روی پروفیل نمونه‌برداری تصویر شده است می‌توان نتیجه گرفت که روش RPCA به خوبی زون کانی‌سازی را شناسایی کرده در حالی که روش PCA محدوده سنگ در برگیرنده کانی‌سازی را به خوبی مشخص می‌کند. برای انتخاب نقطه حفاری نیز می‌توان از روش PCA استفاده کرد.

۵-۲ پیشنهاد

روش‌های تبدیل نسبت لگاریتمی نتایج قابل قبولی برای پردازش و تفکیک هاله‌های ژئوشیمیایی از خود نشان داد. علاوه بر این روش‌های معمول نیز برای این منظور بسیار مفید و کارآمد هستند. پیشنهاد می‌شود که در کنار روش‌های معمول برای پردازش داده‌های ژئوشیمیایی از روش‌های جدید نیز استفاده شود.

پوست

جدول پ-۱: داده‌های مورد استفاده (واحد اندازه‌گیری به درصد است).

Ni	As	S	Pb	Zn	Cu	K ₂ O	Na ₂ O	CaO	MgO	Al ₂ O ₃	TiO ₂	SiO ₂	شماره
23	8	40	18	65	21	2.26	1.13	1.21	1.41	14.29	1.37	63.12	1
30	3	185	23	83	29	2.52	0.88	1.17	1.88	15.31	1.21	64.11	2
27	8	124	19	79	26	2.36	0.88	1.45	2.1	14.66	1.29	62.41	3
24	7	71	22	67	25	2.19	1.13	1.34	1.5	13.53	1.39	60.43	4
22	3	20	20	70	24	2.12	1.1	1.27	1.48	12.85	1.36	59.62	5
24	16	109	19	85	29	2.24	0.89	1.36	2.16	14.38	1.36	61.71	6
21	3	78	21	86	29	2.2	1.01	1.22	1.57	13	1.25	59.85	7
29	12	1895	53	174	86	2.5	0.76	1.58	2.67	15.75	1.23	64.04	8
27	6	369	26	119	45	2.34	0.73	1.92	2.82	14.81	1.34	62.07	9
25	3	337	27	103	38	2.26	0.85	1.78	2.56	13.76	1.38	61.14	10
34	3	403	31	143	56	2.47	0.49	3.67	3.65	16.1	1.14	63.7	11
28	3	569	29	154	57	2.35	0.74	2.31	2.85	14.54	1.19	61.84	12
26	3	332	26	99	43	2.32	0.79	1.46	2.44	14.26	1.31	61.23	13
25	6	304	28	90	43	2.29	0.94	1.54	2.02	14.27	1.45	61.24	14
25	16	228	22	84	39	2.19	0.93	1.4	1.99	13.23	1.41	59.95	15
23	8	172	21	73	31	2.21	1.02	1.41	1.7	13.56	1.34	60.34	16
22	3	93	21	71	29	2.24	1.06	1.29	1.49	14.13	1.4	61.76	17
22	8	59	18	65	24	2.15	1.16	1.2	1.3	13.12	1.32	61.13	18
22	3	69	21	66	25	2.19	1.11	1.23	1.63	13.55	1.3	62.04	19
27	8	69	21	79	29	2.36	0.89	1.36	1.73	15.26	1.39	61.87	20
26	8	53	22	93	25	2.27	0.99	1.22	1.4	14.69	1.39	62.06	21
27	3	106	22	81	28	2.35	0.88	1.45	2.11	14.85	1.33	62.42	22
29	8	1885	49	175	86	2.48	0.63	1.55	2.68	15.49	1.22	61.5	23
29	5	576	30	152	59	2.35	0.73	2.3	2.83	14.53	1.22	61.97	24
27	3	71	22	79	29	2.34	0.9	1.38	1.74	15.24	1.38	62.11	25

جدول پ-۲: PC های به دست آمده توسط RPCA

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7	Comp.8	Comp.9	Comp.10	Comp.11	Comp.12
1	-1.39131	-0.02208	0.075495	-0.04412	0.057849	-0.00191	-0.005	-0.01386	-0.01337	-0.01164	-0.00019	0.006251
2	0.381655	-0.42708	-0.08906	0.23991	-0.09956	-0.02801	0.203298	0.176099	-0.1926	-0.06922	-0.02999	0.045
3	-0.33024	0.323347	0.161649	0.195807	0.079945	-0.03069	0.073782	0.024759	-0.02436	-0.00505	0.001185	-0.0024
4	-0.82351	0.017828	-0.10214	-0.03739	0.041665	-0.09168	-0.0893	0.030598	-0.0107	0.002032	-0.00352	-0.00236
5	-1.43031	-1.07237	0.305619	-0.2766	-0.04257	0.066905	-0.21379	-0.00064	0.155778	0.091108	0.036771	-0.0448
6	-0.68394	0.890693	0.206877	0.044993	0.008659	0.117605	-0.03185	-0.02211	0.014637	-0.00965	-0.01067	0.000657
7	-0.27476	-0.63155	-0.0148	-0.16585	0.005645	0.145078	-0.00504	0.014034	-0.0362	0.002533	0.00369	-0.00665
8	2.028553	1.320999	-0.43011	-0.23405	-0.35611	0.062136	-0.10751	0.238044	-0.3153	-0.302	-0.04582	0.100689
9	0.899071	0.348761	0.15337	0.039172	-0.00024	0.010127	-0.02909	-0.04759	0.008698	-0.00537	0.011219	0.001669
10	1.027504	-0.26728	-0.07578	0.140542	0.088765	0.006719	-0.07442	-0.03005	-0.02124	0.026116	0.000954	0.001681
11	1.499972	-0.26924	0.713366	-0.06744	0.061793	-0.37336	-0.10733	-0.11355	0.21587	-0.05858	0.157268	-0.01408
12	1.634178	-0.14716	0.123631	-0.1468	0.082169	-0.0155	0.035461	0.006804	0.012781	0.001338	-0.00847	-0.00373
13	1.031343	-0.24232	-0.08881	0.124956	-0.12095	0.067819	0.038649	0.036084	-0.01183	-0.00947	-0.00069	0.006647
14	0.584755	0.28281	-0.28334	-0.00235	-0.09553	-0.03614	-0.03778	-0.00728	-0.01886	-0.02575	-0.00528	0.000964
15	-0.06654	1.107843	-0.14951	-0.01546	-0.07961	0.025688	0.005096	0.024047	0.018975	0.039574	0.002423	-0.00331
16	-0.09139	0.436333	-0.24195	0.045156	0.059655	-0.04165	0.05264	-0.01359	0.004023	-0.01746	0.001841	-0.00638
17	-0.21849	-0.61092	-0.19878	-0.02568	-0.01413	0.001619	0.055677	-0.06661	0.013859	0.004181	-0.002	-0.00149
18	-1.05317	0.133619	-0.124	-0.10876	0.063396	-0.00831	0.049279	0.008383	-0.00228	0.016868	0.001936	0.009522
19	-0.47289	-0.6908	-0.10487	0.050522	0.043863	0.042157	-0.03833	0.035095	0.054051	-0.0182	0.004787	2.4E-05
20	-0.80579	0.112493	0.217738	-0.06316	-0.13075	-0.05532	-0.00777	-0.00959	-0.00694	-0.01034	0.006112	-0.0043
21	-1.06294	0.019596	0.183867	-0.24206	-0.07165	0.002942	-0.04104	0.026112	-0.2216	-0.0011	-0.05188	0.063453
22	-0.02494	-0.59947	0.151433	0.158273	-0.00465	-0.02623	0.00045	0.020251	0.006338	0.010392	-0.00154	-0.00267
23	2.252999	1.02931	-0.32102	-0.19392	-0.42443	0.052594	-0.04999	0.172427	-0.32388	-0.26127	-0.01884	0.095479
24	1.431742	0.287258	0.13291	-0.17755	0.028413	-0.04654	-0.00263	0.022385	0.004407	-0.00817	0.003292	0.004208
25	-0.37164	-0.7294	0.160867	-0.01229	-0.11417	-0.06284	0.010192	-0.01177	0.008008	0.018069	-0.00509	0.001653

جدول پ-۳: PC های به دست آمده توسط روش PCA

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13
1	2.270206	-0.58571	1.545027	-0.11213	-0.56558	0.911782	-0.12306	-0.1069	-0.0309	-0.13619	-0.02456	0.109132	0.086536
2	-1.44427	-2.38174	1.949724	-1.24327	-0.70703	-0.84822	-0.06723	0.494095	-0.18049	-0.13689	0.052775	-0.03767	0.003552
3	0.102804	-1.04911	0.795158	0.489015	-0.56726	-0.52056	-0.23027	0.131051	0.104009	0.014471	0.009233	0.158394	0.065983
4	2.80751	0.531823	-0.22102	0.07636	0.157248	0.102516	0.647272	0.337218	0.036093	0.081649	0.072503	0.173573	-0.16564
5	3.531905	0.584369	-1.54516	-0.70812	0.174852	-0.09288	0.30119	0.112534	-0.0192	0.200878	-0.01121	0.020441	0.030676
6	1.130451	0.739179	0.919096	2.132248	-0.82894	-0.10542	-0.57361	-0.38201	0.137854	0.124631	0.044201	0.000816	-0.09187
7	2.474595	0.521604	-1.64321	-1.25583	-0.73122	-0.65833	-0.12666	-0.34864	0.156342	-0.14271	0.244232	-0.11411	-0.00944
8	-5.91879	2.758723	1.95478	-0.75501	-0.4583	0.777288	-0.01831	0.173992	0.024607	0.04921	0.091589	-0.0669	-0.06994
9	-1.61542	-0.3333	-0.42666	0.773559	0.61383	0.13763	-0.61344	-0.12067	-0.11755	0.014929	-0.00743	0.064171	-0.0082
10	0.112576	0.122804	-1.26877	0.035921	0.984929	0.375053	-0.5198	0.359357	-0.09822	0.118571	0.182912	0.080051	0.121717
11	-6.30749	-2.87778	-1.6047	1.238418	-0.44662	0.541247	0.509272	0.054581	0.43968	0.044612	0.048603	-0.08629	0.046097
12	-3.01123	-0.29389	-1.73615	-0.33169	-0.478	0.021646	-0.01732	-0.29502	-0.40935	-0.12153	-0.04969	0.128963	-0.07797
13	-0.47471	-0.15415	-0.75214	-0.37259	0.574921	-0.53644	-0.53712	0.126241	0.056185	0.058981	-0.12444	-0.13799	-0.05817
14	0.653696	0.595598	0.114631	0.306989	1.397123	0.447718	0.065887	0.235374	0.006929	-0.28606	0.045676	-0.07313	-0.0715
15	1.830608	2.035866	-0.18174	2.223125	-0.25545	-0.45132	0.122512	0.434945	-0.13743	-0.10114	-0.08296	-0.13105	0.076157
16	2.177302	0.719963	-0.54652	0.221104	-0.28423	-0.20954	0.188496	-0.00696	0.326206	-0.11816	-0.00985	0.068212	-0.01508
17	2.266185	-0.27503	0.249161	-0.78738	0.630636	0.664283	-0.10613	-0.2969	0.160392	-0.23514	-0.05612	-0.02844	0.072502
18	3.312021	0.528757	-0.15564	-0.27828	-1.08859	0.302222	0.239013	0.070874	-0.0185	0.020367	-0.13405	-0.01219	0.080176
19	2.40982	-0.38585	-0.23454	-1.19752	-0.64822	0.647841	-0.28934	0.085024	0.138272	0.190353	-0.10911	-0.1013	-0.0744
20	0.481336	-0.71282	1.35182	0.691683	0.690767	-0.49131	0.313978	-0.28955	0.113428	-0.10464	0.014884	0.023318	-0.04542
21	1.323347	-0.23408	1.220232	0.281956	0.269822	0.047172	0.484299	-0.38208	-0.5012	0.230538	0.138887	-0.11214	0.074465
22	0.030624	-1.41177	0.477937	-0.30499	0.465887	-0.19354	-0.20511	0.111738	0.038755	0.162002	-0.02996	0.042091	-0.02685
23	-5.48037	2.890303	0.234667	-1.00984	0.399381	-0.71952	0.170915	-0.1957	0.250921	0.064667	-0.08484	0.09179	0.10752
24	-3.08893	-0.0578	-1.37401	0.152891	-0.36737	0.064066	0.11194	-0.05683	-0.52455	-0.10157	-0.09585	-0.00946	-0.02786
25	0.426219	-1.27596	0.878023	-0.26661	1.067415	-0.21338	0.272653	-0.24576	0.047719	0.108177	-0.12542	-0.05028	-0.02306

منابع و مأخذ

- [1] Aitchison.J. (1983) “ The Statistical Analysis of Compositional Data” Journal Of the Royal Statistich Society .,Series B , 44, pp 139-177.
- [2] Aitchison. J. (1986) “The Statistical Analysis of Compositional Data”. London. Chapman Hall.
- [3] Egozcue.J.J, Pawlowsky-Glahn.V., Mateu-Figueras and Bacelo-Vidal.C.(2003) “Isometric Logratio Transformations for Compositional Data Analysis” Mathmatical Geology., 35,3,pp279-300.
- [4] Antonella Buccianti, (2012), “Is compositional data analysis a way to see beyond illusion?” Journal of Computer & Geosciences, 50, pp165-173.
- [5] Tolosana-Degado.R., Otero.N., Pawlowsky-Glahn.V., (2005) “ Some Basic concepts of Compositional Geometry” Mathmatical Geology., 37,7,pp 673-680.
- [6] Flizmore.P., Horn.K., Reimann.C., (2009) “ Principal component analysis for compositional data” Enviromentics, 20, pp. 621-632.
- [7] Ghavami-Riabi.R., Theart.H.F.J.,De Jager.C.(2008) “Detaection of concealed Cu-Zn massive sulfide mineralization below eolian sand and a accrete cover in the eastern part of the Namaqua Metamorphic Province, South Africa” Journal of Geochemical Exploration, 97,pp 83-111.
- [8] Rossouw.R. (2003) Ms.c. thesis “A tectonical risk evaluation of the Katienpan volcanic hosted massive sulphide (VHMS) deposit and its financial viability” Faculty of Natural & Agricultral Science, University of Pretoria.
- [9] Jacobs, J., Weber, K., (1993) “ Accretion and indention tectonics at southern edge of the Kaapvaal craton during the Kibaran (Grenville) orogeny” Geology, 21, pp. 203-206.
- [10] Middelton, R., C., (1976) “ The geology of the Prieska copper mines (Pty) Ltd” Economic Geology, 71, pp. 328-350.
- [11] Geriner, G., J., Humpherys, H., C., Scheepers, D., J. (1994) “Lithostratigraphy, protolithology, and tectonic setting of the Areachap group along the eastern margin of the Namaqua mobile belt, South Africa” South African Journal of Geology, 97, pp. 78-100.

[12] Ghavami-Riabi.R.(2006) “Geochemical exploration for base metal sulphide deposits in an arid environment (eastern Namaqua Metamorphic Province), South Africa” Ph.D thesis, Faculty of Natural & Agricultural Science, University of Pretoria.

[13] Voet, H., W., King, B.,H. (1986) .The Areachap copper-zinc deposit, Gordonia district, pp. 1529, In: “ Mineral Deposits of southern Africa” S.Maaske, Geological society of South Africa.

[14] Theart, H.,F., J., Cornell, D., H., Schade,J., (1989) “ Geochemistry and metamorphism of Prieska Zn-Cu deposit, South Africa” Economic Geology, 84, pp. 34-49.

[15] Rodgers.A.W, Du Toit.A.L, (1908) “ Report on the geology of parts of Prieska, Hay,Britstown, Carnarvon and Victoria West. A. Rep.Geol.Dcom. Cape Hope, Vol 13, pp 8-127.

[16] Pat Shanks.W.C.,Thurston.R. (2012)“ Volcanogenic Massive Sulfide Occurance” Us Geological Survey, Reston, Virginia, pp. 37-53.

[۱۷] محمد حسن کریمپور، سعید سعادت، (۱۳۸۴)، زمین شناسی اقتصادی کاربردی، چاپ دوم، انتشارات ارسلان، مشهد، ص ۲۵۵-۲۷۸.

[18] Zuo.R., Xia.Q., Wang.H. (2012) “ Compositional data analysis in the study of integrated geochemical anomalies with mineralization” Applied Geochemistry, 28,pp.202-211.

[19] Davis.J.C., (2002) “Statistics and Data Analysis in Geology” John Waley & Sons, New York, New York.pp. 11-50.

[20] Chiasera.B., (2010) “A new treatment of compositional data, The Stromboli,Italy case” Geology Depart., State University of Newyork.

[21] Tennant, C.,B., White, M.,L.,(1959) “ Study of distribution of some geochemical data” Econ.Geol., 54, pp.1281-1290.

[22] Sinclair.A.J., (1974) “SELECTION OF THRESHHOL VALUES IN GEOCHEMICAL DATA USING PROBABILITY GRAPHS” J.Geochem.Explor, 3, pp. 129-149.

- [23] Pearson, K., (1897) "Mathematical contribution to the theory of evolution. On a form of spurious correlation which may arise when indices are used in measurement of organs" Proc.R.Soc.Lond., v.LX, pp.489-502.
- [24] Chayes, F., (1960) "On correlation between variables of constant sum" J.Geophys.Res., 65,12, pp.4185-4193.
- [25] Sarmanov, O.V., Vistelius, A.B., (1959) "On the correlation of percentage values" Dokl.Akad.Nawk.SSSR., 126, pp. 22-25.
- [26] Mosimann, J.E., (1962) "On the compound multinomial distribution, the multivariate β -distribution and correlation proportions" Biometrika, 49,1-2, pp.65-82
- [27] Pawlowsky-Glahn, V., Egozcue, J.J., (2001) "Geometric Approach to statistical analysis on the simplex" Stochastic Environ.Ress.Risk Assess., 15,5, pp.384-398
- [28] Egozcue, J.J., Pawlowsky-Glahn, V., (2006) "Compositional data and their analysis: an introduction" pp. 1-10, In: "**Compositional data Analysis in Geoscience: From theory to practice**" Mateu-Figueras, G., Buccianti, A., Pawlowsky-Glahn, V., Geological Society, Special Publications, London.
- [29] Pawlowsky-Glahn, V., Egozcue, J.J., Tolosana-Delgado, R., (2015) "**Modeling and Analysis of Compositional Data**" John Wiley & Sons, United Kingdom, pp.23-30, 32-59, 67-68.
- [30] Gabriel, K.R., (1971) "The biplot-graphical display of matrices with application to principal component analysis" Biometrika, 58,3, pp.453-467.

[۳۱] حسنی پاک ع، شرف‌الدین م، ۱۳۸۰، تحلیل داده‌های اکتشافی (جدایش زمین‌شناسی از آنومالی، آمار و احتمال مهندسی، تخمین ذخیره) انتشارات دانشگاه تهران، ص ۳۸۸، ۲۲۲-۲۲۳، ۳۱۷-۳۲۵، ۴۱۸.

- [32] Flizmore, P., Horn, K., Reimann, C., Garret, G.R., Dutter, R. (2008) "**Statistical Data Analysis Explained: Applied Environmental Statistics with R**" John Wiley & Sons, United Kingdom, pp. 181-191, 219-220.

[۳۳] زارع چاهوکی.م.(۱۳۸۹) روش‌های تحلیل چندمتغیره در نرم‌افزار *SPSS* ، انتشارات دانشگاه

تهران، تهران.

[34] Sinclair, A.J. (1981), Application of Probability Graphs in Mineral Exploration, The Association of Exploratin Geochemists, P. 95.

Abstract

The closed nature of geochemical data has been referred in many studies. In order to analysis the compositional data and use standard statistical methods, firstly the data should be opened. The data might be opened by various transformations provided. One of the most popular of applied transformations is the logratio transform.

To investigate the halos and concealed mineralization, data of wind blown sands related to VHMS deposit was analyzed in the form of closed and opened data. Data used in the current research was collected from the wind blown sands from the surface at defunct Areachap Cu-Zn mine from South Africa. Principal component analysis, PCA, is very popular for multivariate analysis of geochemical data. But PCA is very sensitive to outliers. Using robust methods, RPCA, could resolve this problem. For this purpose, the closed data was evaluated by PCA method and the opened data was analysed by RPCA. Later, the anomaly was separated from background on mineralized factor of PCA and RPCA by using probability-plot modelling. The results were shown that the mineralized factor of RPCA could identify the mineralized zone and PCA identify the host rock. The PCA can optimizes location of drillhole for the followup stage.

Key words: Wind blown sand, Compositional data, Robust PCA, PCA, Opening data



Faculty of mining, Petroleum and Geophysics Engineering

MSc Thesis Mineral exploration

**Investigation of impact of data closure on multivariate analysis of
geochemical data**

By:

Jamal kalte

Supervisor:

Dr. Seyyed Reza Ghavami Riabi

Advisor:

Hamid Moeni

June 2017