



دانشکده علوم ریاضی
گروه آمار

پایان نامه

برای دریافت درجه کارشناسی ارشد در رشته
آمار، گرایش آمار ریاضی

عنوان

تشخیص داده‌های پرت فضایی

اساتید راهنما

دکتر داود شاهسونی - دکتر محمد کنشلو

استاد مشاور

دکتر حسین باغیشنی

دانشجو

مریم مقدم

بهمن ۱۳۹۲

به نام آن که جان را فکرت آموخت

پروردگارا...!

نه می توانم مویشان را که در راه عزت من سفید شد، سیاه کنم و نه برای دست های
پینه بسته شان که شمره تلاش برای افتخار من است، مرهمی دارم. پس توفیقم ده که
هر سحر گزاشان باشم و ثانیه های عمرم را در عصبای دست بودنشان بگذارم.

اگر تنها ترین تنها شوم، باز خدا هست

او جانشین همه ندانستن هاست...



مدیریت تحصیلات تکمیلی
فرم شماره (6)

باسمه تعالی

شماره:
تاریخ:
ویرایش:

فرم صورت جلسه دفاع از پایان نامه تحصیلی دوره کارشناسی ارشد

با تأییدات خداوند متعال و با استعانت از حضرت ولی عصر (عج) نتیجه ارزیابی جلسه دفاع از پایان نامه کارشناسی ارشد خانم مریم مقدم- رشته آمار گرایش آمار ریاضی تحت عنوان تشخیص داده های پرت فضایی که در تاریخ ۱۳۹۲/۱۱/۲۷ با حضور هیأت محترم داوران در دانشگاه صنعتی شاهرود برگزار گردید به شرح ذیل اعلام می گردد:

قبول (با درجه : عالی) امتیاز (۱۹.۴۸) دفاع مجدد مردود

1- عالی (20 - 19) 2- بسیار خوب (18/99 - 18)

3- خوب (17/99 - 16) 4- قابل قبول (15/99 - 14)

5- نمره کمتر از 14 غیر قابل قبول

عضو هیأت داوران	نام و نام خانوادگی	مرتبه علمی	امضاء
1- استاد راهنما	دکتر داود شاهسونی	استادیار	
2- استاد راهنما	دکتر محمد کنشلو	استادیار	
3- استاد مشاور	دکتر حسین باغیشنی	استادیار	
4- نماینده شورای تحصیلات تکمیلی	دکتر احمد زیره	دانشیار	
5- استاد ممتحن	دکتر محمد آرشی	دانشیار	
6- استاد ممتحن	دکتر بهزاد تخم چی	دانشیار	

رئیس دانشکده: دکتر احمد زیره
امضاء
دانشگاه صنعتی شاهرود

تقدیم بہ خانوادہ عزیزم

سپاس‌گزاری

سپاس خدای را که سخنوران، در ستودن او بمانند و شمارندگان، شمردن نعمت‌های او ندانند و کوشندگان، حق او را گزاردن نتوانند، و سلام و دورد بر محمد و خاندان پاک او، طاهران معصوم، هم آنان که وجودمان وامدار وجودشان است. به نشانه سپاس از الطافش، دستان مهربان‌ترین بندگانش، پدر و مادر عزیزم را بوسه می‌زنم.

اینک که این مرحله را پشت سر می‌گذارم، از تلاش‌ها، زحمات و راهنمایی‌های ظریف و ارزشمند اساتید فرزانه و گرانمایه‌ام، جنابان دکتر داود شاهسونی، دکتر محمد کنشلو و دکتر حسین باغیشنی کمال تشکر را دارم. باشد که این خردترین، بخشی از زحمات آنان را سپاس گوید. در پایان از جناب دکتر مهدی حسین‌پوری و سرکار خانم زهرا قیومی که در این مدت کمک‌های زیادی به من کردند تشکر می‌کنم.

مریم مقدم
بهمن ۱۳۹۲

تعمدنامه

اینجانب مریم مقدم دانشجوی کارشناسی ارشد رشته آمار دانشکده علوم ریاضی دانشگاه شاهرود، نویسنده پایان نامه با عنوان تشخیص داده‌های پرت فضایی، تحت راهنمایی دکتر داود شاهسونی و دکتر محمد کنشلو متعهد می‌شوم:

- تحقیقات در این پایان نامه توسط اینجانب انجام شده است و از صحت و اصالت برخوردار است.
- در استفاده از نتایج پژوهش‌های دیگر پژوهش‌گران، به مرجع مورد استفاده استناد شده است.
- مطالب این پایان نامه، تا کنون توسط خود، یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی در هیچ جا ارایه نشده است.
- حقوق معنوی این اثر، به دانشگاه شاهرود متعلق دارد، و مقالات مستخرج با نام “ دانشگاه شاهرود “ یا “ Shahrood University “ به چاپ خواهد رسید.
- حقوق معنوی تمام افرادی که در به دست آوردن نتایج اصلی پایان نامه تاثیرگذار بوده‌اند، در مقالات مستخرج از پایان نامه رعایت می‌گردد.
- در تمام مراحل انجام این پایان نامه، در مواردی که از موجود زنده (یا بافت‌های آنها) استفاده شده است، ضوابط و اصول اخلاقی رعایت شده است.
- در تمام مراحل انجام این پایان نامه، در مواردی که به حوزه اطلاعات شخصی افراد دسترسی یافته (یا استفاده) شده است، اصل رازداری و اصول اخلاق انسانی رعایت شده است.

مریم مقدم
بهمن ۱۳۹۲

مالکیت نتایج و حق نشر

- تمام حقوق معنوی این اثر و محصولات آن (مقالات مستخرج، کتاب، برنامه‌های رایانه‌ای، نرم افزارها و تجهیزات ساخته شده) متعلق به دانشگاه شاهرود می‌باشد. این مطلب باید به نحو مقتضی، در تولیدات علمی مربوطه ذکر شود.
- استفاده از اطلاعات و نتایج موجود در این پایان نامه بدون ذکر منبع مجاز نمی‌باشد.

چکیده

در ذخایر معدنی، تشخیص مقادیری که در دم راست توزیع واقع شده‌اند، از نظر اقتصادی اهمیت ویژه‌ای دارند؛ در نتیجه تشخیص مرز این‌گونه مشاهدات در صنعت معدن اهمیت بسزایی در اکتشاف و استخراج دارد. روش‌های موجود برای تشخیص داده پرت فضایی، مقدار این داده‌ها را ارائه می‌کنند و معیاری جهت شناسایی مرز مقادیر بالا ارائه نمی‌دهند. از این رو در این پایان‌نامه برای شناسایی مرز مقادیر بالا، روش مرز-بالا پیشنهاد می‌شود، که بر مبنای عدم وجود اثر مرزی استوار است. در صورت عدم وجود اثر مرزی، مقادیر بالای مرز، هیچ‌گونه تاثیر پذیری از مقادیر کمتر از مرز ندارد. در این روش، نسبت تغییرنگارهای شاخص متقابل به تغییرنگار شاخص برای مرزهای پیشنهادی محاسبه شده و در صورتی که تابعی از فاصله موقعیت‌ها نباشد، عدم وجود اثر مرزی را نشان می‌دهد. همچنین اگر تغییرنگار باقی‌مانده شاخص مرز پیشنهادی، اثر قطعه‌ای تام را نشان دهد، آنگاه مرز پیشنهادی به عنوان مرز مقادیر بالا انتخاب می‌گردد. ارزیابی عملکرد این روش و نیز برخی روش‌های تشخیص داده پرت، در یک مطالعه شبیه سازی انجام شده است. همچنین به عنوان یک مورد مطالعاتی واقعی، از این روش برای تشخیص مرز مقادیر بالای عیار داده‌های طلای ساری گونای در استان کردستان، استفاده شده و در نتیجه مرز $4/63 \text{ ppm}$ ، به عنوان مرز عیار بالای طلای ساری گونای انتخاب گردید.

کلمات کلیدی: داده پرت فضایی، مقادیر بالا، اثر مرزی، تغییرنگار شاخص، تغییرنگار شاخص متقابل، روش مرز-بالا

مقاله مستخرج از پایان نامه

۱. مقدم، م.، کنشلو، م. و شاهسونی، د. (۱۳۹۲)، "کاربرد روش *Top-Cut* در تعیین مرز مقادیر بالای عیار طلا"، اولین کنفرانس ملی مهندسی اکتشاف منابع زیرزمینی، ص ۲۱، شاهرود.

پیش‌گفتار

داده‌های فضایی مشاهده‌اتی هستند که بر حسب موقعیت قرار گرفتن‌شان در فضای مورد مطالعه به یکدیگر وابسته‌اند. برآورد تغییرنگار و پیش‌گویی فضایی از مباحث مهم در آمار فضایی هستند. وجود داده‌های پرت فضایی یکی از مهم‌ترین عوامل بروز انحرافات، در این مباحث می‌باشد. بنابراین روش‌های تشخیص داده پرت فضایی، حائز اهمیت هستند. در این پایان‌نامه به شرح برخی از روش‌های تشخیص داده پرت فضایی می‌پردازیم. در ذخایر معدنی، منظور از داده پرت، مشاهده‌اتی هستند که دارای مقادیر بالایی نسبت به سایر داده‌ها می‌باشند. مقادیر بالا (عیار بالا) در ذخایر معدنی اهمیت بسزایی در استخراج دارند. این‌گونه مقادیر در داده‌های طلا به صورت مرز شناخته می‌شوند. روش‌های موجود برای تشخیص داده‌های پرت فضایی، معیاری جهت تشخیص مرز مقادیر بالا ارائه نمی‌دهند. لذا در این پایان‌نامه به منظور تشخیص مرز مقادیر بالا، به شرح روشی به نام مرز-بالا می‌پردازیم که برگرفته از مدل مرز-بالا است.

مطالب فوق را در ۵ فصل بیان می‌کنیم:

- در فصل اول، پس از بیان مفاهیم و تعاریف آمار فضایی، به منظور شناخت اولیه زمین‌شناسی، به بیان تعاریف و مفاهیم آن پرداخته‌ایم و در آخر، اطلاعاتی در مورد طلا از جمله عیار، انواع کانسارهای طلا و ... ارائه نموده‌ایم.
- در فصل دوم، ابتدا به معرفی انواع داده پرت فضایی پرداخته و برخی روش‌های تشخیص آن‌ها را بیان نموده‌ایم. سپس مسایل مربوط به داده پرت در زمین‌آمار را بیان نموده و با مثال کوتاهی به اهمیت تشخیص مرز مقادیر بالا، در ذخایر معدنی پرداخته‌ایم.
- در فصل سوم، برای تشخیص مرز مقادیر بالا، روش مرز-بالا معرفی شده است. از آنجایی که این روش بر مبنای عدم وجود اثر مرزی استوار است قبل از معرفی آن، مفاهیم اثر مرزی و عدم وجود اثر مرزی را بیان نموده‌ایم.
- در فصل چهارم، در یک مطالعه شبیه‌سازی به ارزیابی عملکرد روش‌های تشخیص داده پرت فضایی و تشخیص مرز مقادیر بالا پرداخته‌ایم و همچنین به منظور تشخیص مرز عیار بالای یکی از معادن طلای ایران (ساری‌گونای)، از روش مرز-بالا استفاده نموده‌ایم. بدین منظور در ابتدا به تحلیل اکتشافی داده‌های معدن ساری‌گونای پرداخته‌ایم.

فهرست مطالب

ز	فهرست تصاویر
۱	آشنایی با مفاهیم مقدماتی آمار فضایی ۱
۱	۱.۱ مقدمه ۱
۲	۲.۱ تعاریف و مفاهیم آمار فضایی ۲
۶	۱.۲.۱ مانایی ۶
۷	۲.۲.۱ تغییرنگار ۷
۸	۳.۲.۱ هم‌تغییرنگار ۸
۹	۴.۲.۱ برآورد تغییرنگار ۹
۱۰	۵.۲.۱ تغییرنگار متقابل ۱۰
۱۱	۶.۲.۱ همبستگی‌نگار ۱۱
۱۱	۳.۱ تعاریف و مفاهیم زمین‌شناسی ۱۱
۱۳	۴.۱ مقدمه‌ای بر مفاهیم مرتبط با عنصر طلا ۱۳
۱۳	۱.۴.۱ مقدمه ۱۳
۱۳	۲.۴.۱ تاریخچه ۱۳
۱۳	۳.۴.۱ عیار طلا ۱۳
۱۴	۴.۴.۱ ذخایر جهانی طلا ۱۴
۱۵	۵.۱ مدل‌سازی ذخایر طلا ۱۵
۱۵	۱.۵.۱ انواع کانسارهای طلا ۱۵
۱۸	۲.۵.۱ تولید طلا ۱۸
۱۸	۶.۱ نتایج ۱۸
۱۹	۲ روش‌های تشخیص داده‌های پرت ۱۹
۱۹	۱.۲ داده پرت فضایی ۱۹
۲۱	۲.۲ تقسیم‌بندی روش‌های تشخیص داده‌های پرت ۲۱

۲۳	تشخیص داده پرت فضایی با یک صفت کیفی ساده	۳.۲
۲۳	نمودار پراکنش موران	۱.۳.۲
۲۳	نمودار پراکنش	۲.۳.۲
۲۵	الگوریتم Z	۳.۳.۲
۲۵	الگوریتم میانه	۴.۳.۲
۲۶	روش میانه-میانگین	۵.۳.۲
۲۶	ابر تغییرنگار	۴.۲
۲۸	نمودار جعبه‌ای	۵.۲
۲۹	داده پرت در زمین‌آمار	۶.۲
۳۰	برش زدن داده پرت	۱.۶.۲
۳۳	نتیجه‌گیری	۷.۲
۳۵	روش مرز-بالا برای تشخیص مرز مقادیر بالا	۳
۳۵	مقدمه	۱.۳
۳۶	اثر مرزی	۲.۳
۳۶	عدم وجود اثر مرزی	۳.۳
۳۸	روش مرز-بالا	۴.۳
۳۸	مقدار بریده‌شده	۱.۴.۳
۳۹	مدل مرز-بالا	۲.۴.۳
۳۹	فرضیه‌های مدل مرز-بالا	۳.۴.۳
۴۲	انتخاب مقدار مرز-بالا	۴.۴.۳
۴۵	شبیه‌سازی و تحلیل اکتشافی داده‌های پروژه زرکوه	۴
۴۵	مطالعه شبیه‌سازی	۱.۴
۴۷	ارزیابی روش‌های تشخیص مرز مقادیر بالا	۲.۴
۴۸	ارزیابی عملکرد روش‌های تشخیص مقدار داده پرت فضایی	۳.۴
۵۰	تحلیل اکتشافی داده‌های زرکوه	۴.۴
۵۰	ساری‌گونای	۵.۴
۵۰	تحلیل اکتشافی داده مورد مطالعه	۶.۴
۵۱	بررسی وجود همبستگی فضایی در داده‌ها	۱.۶.۴
۵۴	نمودار جعبه‌ای	۷.۴
۵۴	ابر تغییرنگار	۸.۴
۵۵	روش مرز-بالا	۹.۴

۵۹	آ	
۵۹	۱.آ	ساری گونای
۶۱	۱.۱.آ	زمین شناسی ساری گونای
۶۱	۲.۱.آ	نحوه کانه‌زایی در کانسار
۶۲	۳.۱.آ	زمین شناسی ساختاری ساری گونای
۶۳		ب کدهای مربوط به فصل‌های چهارم
۶۳	۱.ب	شبیه سازی
۶۷	۲.ب	روش مرز-بالا
۶۹	۳.ب	تحلیل اکتشافی
۷۰		مراجع
۷۴		واژه‌نامه فارسی به انگلیسی
۷۶		واژه‌نامه انگلیسی به فارسی

فهرست تصاویر

۴	موقعیت داده خاکستر زغال سنگ (نمونه‌ای از داده زمین‌آماری)	۱.۱
	(حق‌شناس تجن گوکه، ۱۳۸۹) نقشه متوسط اجاره‌بهای پرداختی ماهانه در شهر	۲.۱
۵	تهران به تفکیک ۲۲ منطقه شهرداری	۵
۵	نمونه‌ای از الگوی نقطه‌ای	۳.۱
۹	نمودار تغییرنگار و هم‌تغییرنگار (محمدزاده، ۱۳۹۱)	۴.۱
۲۱	تقسیم‌بندی روش‌های تشخیص داده پرت	۱.۲
۲۴	نمودار پراکنش موران برای یک مجموعه داده فضایی	۲.۲
۲۴	نمودار پراکنش برای یک مجموعه داده فضایی	۳.۲
۲۷	نمودار پراکنش مقادیر خاکستر زغال سنگ در موقعیت‌های فضایی	۴.۲
۲۸	نمودار ابر تغییرنگار برای داده خاکستر زغال سنگ	۵.۲
۲۹	نمودار جعبه‌ای	۶.۲
۲۹	نمودار جعبه‌ای برای داده خاکستر زغال سنگ	۷.۲
۳۱	موقعیت داده پرت در بلوک	۸.۲
۳۷	نمونه‌ای از اثر مرزی	۱.۳
۴۴	نسبت واریوگرام‌های شاخص برای مرزهای ۵، ۱۰ و ۲۰	۲.۳
۴۴	تغییرنگارهای باقی‌مانده‌های شاخص برای $10g/t - 5$ و $20g/t - 10$	۳.۳
۴۶	شبکه منظم داده‌های شبیه‌سازی شده	۱.۴
	الف: بافت‌نگار نمونه‌های شبیه‌سازی شده ب: بافت‌نگار نمونه‌های شبیه‌سازی شده	۲.۴
۴۶	آلوده به داده‌های پرت	۴۶
	معیار ارزیابی $TPrate$ برای روش‌های نمودار جعبه‌ای و مرز-بالا جهت تشخیص	۳.۴
۴۷	مرز مقادیر بالا	۴۷
۴۸	نمودارهای معیارهای سنجش برای الگوریتم میانه در $k = 2$	۴.۴
۴۸	نمودارهای معیارهای سنجش برای الگوریتم میانه در $k = 8$	۵.۴

۴۹	$\alpha/2 = 0/001$	نمودارهای معیارهای سنجش برای الگوریتم میانه در	۶.۴
۴۹	$\alpha/2 = 0/005$	نمودارهای معیارهای سنجش برای الگوریتم میانه در	۷.۴
۴۹	$\alpha/2 = 0/025$	نمودارهای معیارهای سنجش برای الگوریتم میانه در	۸.۴
۴۹		نمودارهای معیارهای سنجش برای الگوریتم میانه و روش میانه میانگین	۹.۴
۵۰		موقعیت کانسار طلای ساری گونای در ایران	۱۰.۴
۵۲		پراکندگی داده‌ها نسبت به محورهای مختصات	۱۱.۴
			الف: نمودار چندک چندک داده‌ها، ب: نمودار چندک چندک داده‌ها همراه با کاهش	۱۲.۴
۵۳		چولگی	
			الف: نمودار جعبه‌ای داده‌های عیار طلا، ب: نمودار جعبه‌ای داده‌های عیار طلا همراه با کاهش	۱۳.۴
۵۵		چولگی	
۵۵		نمودارهای ابر تغییرنگار داده‌های عیار طلای زرکوه	۱۴.۴
۵۷		الف: نمودار $\frac{\gamma I(2/6, 4/63)(h)}{\gamma I(2/6)(h)}$ ب: نمودار $\frac{\gamma I(4/63, 6/59)(h)}{\gamma I(4/63)(h)}$ ج: نمودار $\frac{\gamma I(6/59, 8/412)(h)}{\gamma I(6/59)(h)}$	۱۵.۴
۵۷			تغییرنگارهای باقی‌مانده شاخص برای $2/6-4/63$ ، $4/63-6/59$ و $6/59-8/412$	۱۶.۴
۶۱		نمای منطقه پروژه ساری گونای از جنوب غربی	۱.آ
۶۲		زون‌های پرعیار و کم عیار ساری گونای	۲.آ

فصل ۱

آشنایی با مفاهیم مقدماتی آمار فضایی

۱.۱ مقدمه

در روش‌های معمول آماری، اغلب فرض می‌شود که مشاهدات حاصل از نمونه استخراج شده از جامعه مستقل از یکدیگرند. اما در بسیاری از موارد، مشاهدات مستقل نیستند و برحسب موقعیت قرار گرفتن‌شان در فضای مورد مطالعه، به یکدیگر وابسته هستند. اگر این وابستگی تابعی از فاصله بین موقعیت‌های مشاهدات باشد، به‌گونه‌ای که مشاهدات نزدیک به هم، وابسته‌تر و مشاهداتی که دور از یکدیگر هستند، وابستگی کمتری داشته باشند، آنگاه این‌گونه مشاهدات را داده‌های فضایی^۱ می‌نامند (محمدزاده، ۱۳۹۱).

مبانی آمار فضایی و گسترش روش‌های آن، به دنبال پژوهش‌های پروفسور ماترون^۲ بنا نهاده شد. این پژوهشگر فرانسوی با انتشار مقاله بنیادین خود توانست شاخه جدیدی از علم آمار با عنوان ”زمین‌آمار”^۳ را پایه‌گذاری کند. همچنین ماترون، روش کریگینگ^۴ را، که منتسب به دی.جی. کریگ^۵، مهندس معدن آفریقای جنوبی است، مطرح کرد. دیوید^۶ (۱۹۷۷) نیز در مورد پیش‌بینی‌های فضایی ذخایر معدنی، مطالب بسیاری را ارائه کرده است. کرسی^۷ (۱۹۹۳) با انتشار کتاب ”آمار برای داده‌های فضایی” در تکامل این شاخه از علوم قدم بزرگی را برداشت.

نخستین تجربه‌ها برای به‌کارگیری روش‌های آماری در پیش‌گویی ذخایر معادن، با شناسایی مقدماتی الگوی توزیع طلا در معادن آفریقای جنوبی آغاز شد. در آن زمان نکته جالب توجه محققان این بود

^۱Spatial Data

^۲Matheron

^۳Geostatistics

^۴Kriging

^۵Daniel Gerhardus Krige

^۶David

^۷Cressie

که در یک منطقه معدنی، بخش‌های پرعیار و کم‌عیار به‌طور معمول در کنار یکدیگر قرار دارند و باید نوعی رابطه میان این بخش‌ها وجود داشته باشد (محمدزاده، ۱۳۹۱). بخش‌های پرعیار (مقادیر بالا) در ذخایر معدنی، داده پرت^۱ نامیده می‌شود. در راستای تشخیص داده پرت فضایی، تحقیقات زیادی انجام شده است. از جمله کو^۲ و همکاران (۲۰۰۷)، الگوریتمی را بر مبنای k -نزدیکترین همسایگی، برای تشخیص نقاط یا ناحیه‌های معمولی که خیلی شبیه به داده‌هایی هستند که در همسایگی‌شان قرار دارند، در نظر گرفتند. الگوریتم پیشنهادی آن‌ها، به علت دقت در تشخیص داده‌های پرت فضایی و توانایی تشخیص داده‌های پرت ناحیه‌ای، نسبت به روش‌های دیگر تشخیص داده پرت فضایی برتری دارد. همچنین الگوریتم k نزدیکترین همسایگی مبنای^۳ $(BKNN)$ ، توسط کاو^۴ (۲۰۱۳a) و همکاران که بر مبنای اصول نمودار KNN است، به منظور رفع این مشکل که مشاهدات فضایی ممکن است توسط همسایگی‌های غیرمعمول‌شان، تاثیرپذیر باشند ارائه گردید. کاو و همکاران، (۲۰۱۳b) نیز نادیده گرفتن تاثیر مشخصه‌های فضایی بر روی درجه پرت بودن داده‌های پرت فضایی را مشکل روش‌های رایج تشخیص داده پرت فضایی می‌داند و برای این منظور ایده‌ای بر مبنای داده پرت وزن‌دار بیان می‌کند. این مطالعه، تاثیرات مشخصه‌های فضایی را بر روی محاسبه درجه پرتی فضایی تشخیص می‌دهد و این تاثیرات را با مشخصه‌های غیرفضایی ترکیب می‌کند و دو الگوریتم Z -مقدار توسعه یافته^۵ و الگوریتم تفاضل وزن‌دار^۶ را پیشنهاد می‌کند.

از آنجاکه بخش‌های پرعیار و کم‌عیار معدن در کنار یکدیگر قرار دارند، لذا تمرکز ما در این پایان‌نامه، مطالعه روش تشخیص مرز مقادیر بالا برای ذخایر معدنی است که از این روش، برای تشخیص مرز بالای عیار طلا در یکی از معادن طلای ایران استفاده شده است.

۲.۱ تعاریف و مفاهیم آمار فضایی

برای مدل‌سازی و تجزیه و تحلیل داده‌های فضایی، نیاز به تعریف میدان تصادفی است.^۷

تعریف ۱.۲.۱. میدان تصادفی: فرض کنید $s \in \mathbb{R}^d$ ، موقعیت داده در فضای اقلیدسی d بعدی ($d \geq 1$) و $Z(s)$ یک کمیت تصادفی در موقعیت s باشد. حال اگر s روی مجموعه اندیس‌گذار $D \subset \mathbb{R}^d$ تغییرکند، موجب تولید میدان تصادفی

$$\{Z(s); s \in D\}, \quad (1.1)$$

^۱Outlier

^۲Kou

^۳Based K-Nearest Neighbor

^۴Cao

^۵Improved Z-value

^۶Weighted difference algorithm

^۷مفاهیم این بخش، از کتاب دکتر محمدزاده (۱۳۹۱) اقتباس گردیده است.

می‌شود که $\{z(s); s \in D\}$ یک تحقق بیان شده از میدان تصادفی (۱.۱) است. توابع میانگین، کوواریانس و واریانس میدان تصادفی $Z(\cdot)$ به ترتیب به صورت زیر بیان می‌شوند:

$$\mu(s) = E[Z(s)], \quad s \in D,$$

$$\begin{aligned} C(s_i, s_j) &= \text{Cov}[Z(s_i), Z(s_j)] \\ &= E[(Z(s_i) - \mu(s_i))(Z(s_j) - \mu(s_j))], \quad s_i, s_j \in D, \end{aligned}$$

$$\text{Var}[Z(s)] = E[Z(s) - \mu(s)]^2 = C(s, s).$$

هر میدان تصادفی (۱.۱) را می‌توان به صورت

$$Z(s) = \mu(s) + \delta(s), \quad s \in D,$$

تجزیه کرد که در آن $\mu(\cdot)$ ، تغییرات بزرگ مقیاس^۱ یا روند^۲ میدان تصادفی است، که ممکن است ناشی از تغییرپذیری بین موقعیت‌های مشاهده شده باشد، و $\delta(\cdot)$ ، فرآیند خطا یا تغییرات کوچک مقیاس^۳ میدان تصادفی است، که ممکن است ناشی از خطای اندازه‌گیری یا تغییرپذیری در درون موقعیت مشاهده شده باشد.

مثال ۲.۲.۱. فرض کنید برای یک میدان تصادفی در پنج موقعیت

$$D = \{s_1 = (1, 1), s_2 = (2, 2), s_3 = (3, 4), s_4 = (1, 4), s_5 = (3, 1)\}$$

مقادیر زیر مشاهده شده‌اند.

$$z(s_1) = 1, z(s_2) = 2, z(s_3) = 20, z(s_4) = 4, z(s_5) = 3$$

در این صورت $\{z(s_1), z(s_2), z(s_3), z(s_4), z(s_5)\}$ ، تحقق از میدان تصادفی $\{Z(s); s \in D\}$ است.

در آمار فضایی، متغیرهای مورد اندازه‌گیری ممکن است گسته یا پیوسته باشند. فضای موقعیت یا مکان مشاهدات نیز ممکن است پیوسته یا گسسته، نقطه‌ای یا ناحیه‌ای، منظم یا نامنظم باشد. با توجه به انواع موقعیت‌ها، مشاهدات فضایی به سه گروه عمده، داده‌های زمین‌آماری^۴، داده‌های شبکه‌ای^۵ و الگوهای نقطه‌ای^۶ تقسیم می‌شوند (کرسی، ۱۹۹۳).

داده‌های زمین‌آماری: این نوع داده‌ها در موقعیت‌های ثابت و مشخص، در ناحیه‌ای پیوسته مشاهده می‌شوند. مکان و متغیر مورد مطالعه، ممکن است گسسته یا پیوسته باشد. به عنوان مثال، برای حالت پیوسته می‌توان غلظت مواد معدنی در داخل یک کانسار^۷، مقدار ریزش باران ثبت شده در ایستگاه‌های

^۱Large scale variation

^۲Trend

^۳Small scale variation

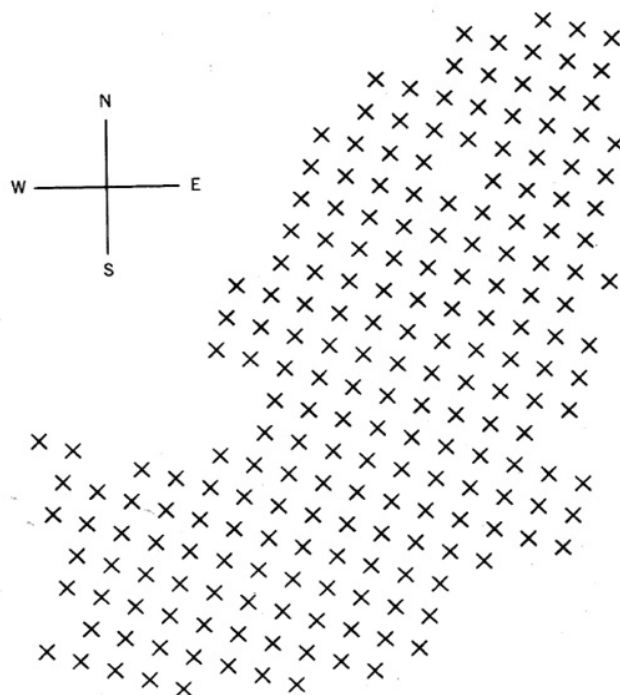
^۴Geostatistical data

^۵Lattice data

^۶Point pattern

^۷Mineral deposit

هواشناسی، و برای حالت گسسته می‌توان تعداد نوعی جانور دریایی در مکان‌های نمونه‌گیری شده در طول یک ساحل را نام برد. داده‌های مربوط به موقعیت خاکستر زغال سنگ^۱ نمونه‌ای از داده‌های زمین‌آمار، در شکل (۱.۱) نمایش داده شده است.



شکل ۱.۱: موقعیت داده خاکستر زغال سنگ (نمونه‌ای از داده زمین‌آمار)

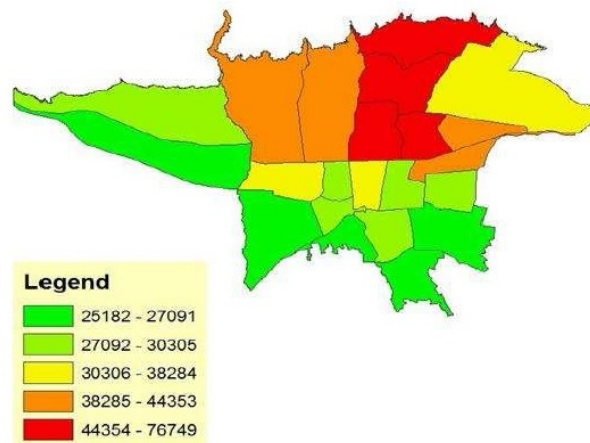
داده‌های شبکه‌ای: این نوع داده‌ها مربوط به مکان‌های ناحیه‌ای هستند، که این مکان‌ها ممکن است منظم یا نامنظم باشند. تصاویر ماهواره‌ای از سطح زمین که در آن‌ها سطح زمین به تعدادی عنصر تصویر^۲ کوچک به صورت شبکه منظم در R^2 تقسیم شده، مثالی برای داده‌های شبکه‌ای منظم است. تعداد افراد مبتلا به سرطان در مناطق مختلف خدمات درمانی کشور، مثالی برای داده‌های شبکه‌ای نامنظم است. در این حالت در هر شهرستان، کل افراد مبتلا به سرطان، در مرکز خدمات درمانی آن شهرستان، در نظر گرفته می‌شوند. به طور معمول، هدف از تحلیل داده‌های فضایی شبکه‌ای، مدل‌بندی احتمالاتی مشاهدات است، در صورتی که در زمین‌آمار، پیش‌گویی مقدار متغیر در یک موقعیت جدید مورد نظر است. برای مثال نقشه متوسط اجاره‌بهای پرداختی ماهانه در شهر تهران به تفکیک ۲۲ منطقه شهرداری، در شکل (۲.۱) نمایش داده شده است.

الگوی نقطه‌ای: در این گروه از داده‌ها، مکان یا موقعیت مشاهده‌شده، متغیر تصادفی مورد نظر است. الگوی نقطه‌ای، شامل تعدادی متناهی از مکان‌ها در یک ناحیه است که در آن‌ها یک صفت خاص اندازه‌گیری می‌شود. به طور معمول، الگوی‌های نقطه‌ای به سه دسته به طور کامل تصادفی فضایی^۳

^۱Coal Ash Data

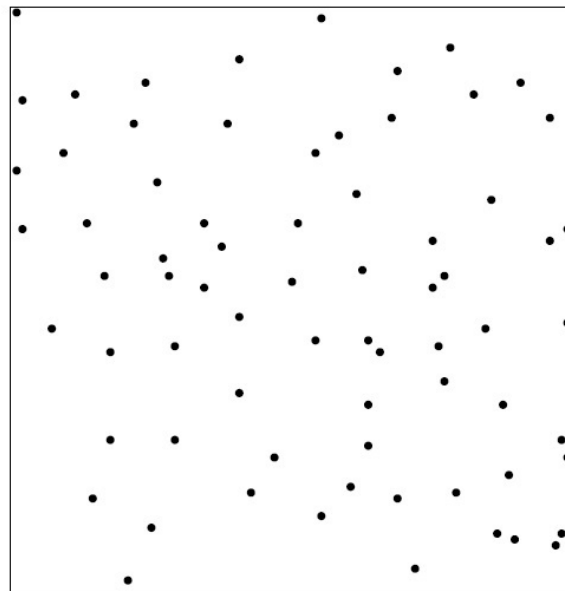
^۲Pixel

^۳Complete spatial randomness



شکل ۲.۱: (حق شناس تجن گوکه، ۱۳۸۹) نقشه متوسط اجاره‌بهای پرداختی ماهانه در شهر تهران به تفکیک ۲۲ منطقه شهرداری

(CSR)، منظم^۱ و خوشه‌ای^۲ تقسیم گردیده و به مدل‌بندی آن‌ها اقدام می‌شود. یک مثال برای این نوع مشاهدات، موقعیت گونه‌ای از درختان در یک ناحیه جنگلی یا موقعیت مراکز زلزله است. نمونه‌ای از الگوی نقطه‌ای که موقعیت آن‌ها به صورت تصادفی قرار گرفته‌اند، در شکل (۳.۱) نمایش داده شده است.



شکل ۳.۱: نمونه‌ای از الگوی نقطه‌ای

در آمار کلاسیک برای نشان دادن میزان پراکنندگی و وابستگی خطی داده‌ها، به ترتیب از واریانس

^۱Regularity

^۲Clustering

و کوواریانس استفاده می‌شود. در آمار فضایی بر حسب این‌که مشاهدات فضایی، مربوط به کدامیک از دسته‌های فوق باشند، بررسی و تحلیل آن‌ها نیز متفاوت است. به‌عنوان مثال، محاسبه وابستگی فضایی برای داده‌هایی که از نوع زمین‌آماری باشند، بر حسب تغییرنگار^۱ و برای داده‌ها مشبکه‌ای، بر اساس ماتریس همسایگی محاسبه می‌شود.

۱.۲.۱ مانایی

تعریف ۳.۲.۱. مانایی قوی^۲: $Z(s)$ مانا است، اگر برای فاصله h ، و همه موقعیت‌های s_1, \dots, s_n که n متناهی است، توزیع توام $Z(s_1), \dots, Z(s_n)$ همانند توزیع توام $Z(s_1 + h), \dots, Z(s_n + h)$ باشد. یعنی به بیان دیگر، در صورت انتقال موقعیت‌های s_1, \dots, s_n در راستای $h \in \mathbb{R}$ توزیع توام $Z(s_1), \dots, Z(s_n)$ تغییر نمی‌کند (کاکس و میلر^۳، ۱۹۶۵).

تعریف ۴.۲.۱. مانایی مرتبه دوم^۴: به دو دلیل، یکی مشکل بودن آزمودن مانایی قوی و دیگری این‌که در این مانایی (مانایی قوی) وجود گشتاورها الزامی نیست، اغلب نوع ضعیف‌تری از آن که با عنوان مانایی مرتبه دوم شناخته شده است، به‌کار گرفته می‌شود.

$Z(s)$ مانای مرتبه دوم است، اگر $\text{Cov}(Z(s+h), Z(s))$ وجود داشته باشد و تنها به h بستگی داشته باشد یعنی میزان تغییرپذیری یا همبستگی متغیر در تاخیر h ، تابعی از فاصله بین موقعیت‌های آن‌ها می‌باشد. این شرط، وجود $\text{Var}(Z(s))$ و عدم وابستگی آن به S را الزامی می‌دارد یعنی واریانس میدان تصادفی به موقعیت فضایی بستگی ندارد و تغییرپذیری میدان همه‌جا یکسان است. به همین ترتیب $E(Z(s))$ نیز وجود دارد و به S وابسته نمی‌باشد (کاکس و میلر، ۱۹۶۵). بدیهی است، مانایی به معنای برقراری مانایی مرتبه دوم نمی‌باشد بدین معنا که یک تابع تصادفی مانا ممکن است به علت عدم وجود گشتاورهای مراتب اول دوم، مانای مرتبه دوم نباشد (میرس^۵، ۱۹۸۹).

تعریف ۵.۲.۱. میدان تصادفی $Z(\cdot)$ مانای ذاتی^۶ نامیده می‌شود، هرگاه

$$1. \text{ میانگین میدان تصادفی ثابت یا مستقل از } s \text{ باشد، یعنی } E[Z(s)] = \mu$$

۲. واریانس عبارت $(Z(s) - Z(s+h))$ فقط تابعی از فاصله موقعیت‌های s و $s+h$ باشد، یعنی

$$\text{Var}[Z(s+h) - Z(s)] = 2\gamma(h).$$

(محمدزاده، ۱۳۹۱).

^۱Variogram

^۲Strong stationary

^۳Cox and Miller

^۴Second order stationary

^۵Myers

^۶Intrinsic stationary

تعریف ۶.۲.۱. میدان تصادفی $Z(s)$ ، دارای خاصیت مانایی ضعیف با انحراف^۱ است، اگر

$$Z(s) = Y(s) + m(s)$$

که در آن $Y(s)$ متغیری است که در شروط مانای ذاتی صدق می‌کند و $E[Z(s)] = m(s)$ ، مقدار $m(s)$ به صورت ترکیب خطی از تابع‌های مستقل از هم، قابل بیان است (میرس، ۱۹۸۹).

۲.۲.۱ تغییرنگار

میدان تصادفی $Z(\cdot)$ را برای دو موقعیت s و $s+h$ به صورت دو متغیر تصادفی $Z(s)$ و $Z(s+h)$ در نظر بگیرید، که h تفاضل این دو موقعیت، "تاخیر"^۲ نامیده می‌شود. همچنین $(Z(s+h) - Z(s))$ را نمو متغیر گوئیم. چون $Z(s)$ و $Z(s+h)$ هر دو متغیر تصادفی هستند، تفاضل آن‌ها نیز متغیری تصادفی است. اگر میزان تغییرات (واریانس) این تفاضل زیاد باشد به معنای آن است که نقاط s و $s+h$ در موقعیت‌های دور از هم واقع‌اند و اگر میزان تغییرات کم باشد، این دو موقعیت به هم نزدیک هستند. لذا $\text{Var}(Z(s+h) - Z(s))$ ، ساختار وابستگی را نشان می‌دهد یعنی هر چه مقدار این تابع بزرگتر باشد، وابستگی کمتر بوده و هرچه مقدار این تابع کوچکتر باشد، وابستگی بیشتر است.

وابستگی فضایی را با تابع تغییرنگار که به صورت زیر بیان می‌شود، نشان می‌دهند:

$$\gamma(h) = \text{Var}(Z(s+h) - Z(s)). \quad (2.1)$$

اگر میانگین میدان تصادفی، ثابت و تغییرنگار آن فقط تابعی از فاصله موقعیت‌ها باشد، این میدان تصادفی، مانای ذاتی است. در این صورت می‌توان تغییرنگار را به صورت

$$\gamma(h) = E(Z(s+h) - Z(s))^2$$

نوشت (کرسی، ۱۹۹۳). ماترون (۱۹۶۲)، $\gamma(h)$ را نیم‌تغییرنگار^۳ نامید. مقدار تغییرنگار معمولاً به اندازه و جهت h بستگی دارد. اگر در میدان تصادفی، مقدار تغییرنگار فقط تابعی از فاصله بین مشاهدات باشد و به جهت آن بستگی نداشته باشد، تغییرنگار را همسان‌گرد^۴ می‌نامند.

پارامترهای تغییرنگار به صورت زیر تعریف می‌گردند:

- دامنه^۵: فاصله‌ای که تغییرنگار به حد فوقانی خود می‌رسد و در خارج از آن بازه، تغییرنگار ثابت می‌ماند، دامنه نامیده می‌شود. به عبارتی دیگر، در خارج از دامنه، همبستگی فضایی بین نقاط از بین می‌رود و مشاهدات را می‌توان مستقل از هم در نظر گرفت.

^۱Weakly Stationary with Drift

^۲Lag

^۳Semivariogram

^۴Isotropic

^۵Range

- ازاره^۱: به طور معمول، تغییرنگار تابعی افزایشی از تاخیر $\|h\|$ است و ممکن است به کران بالایی منتهی شود. به این مقدار ثابت تغییرنگار بعد از دامنه، ازاره یا آستانه گویند و عبارت است از

$$c = \lim_{\|h\| \rightarrow \infty} \gamma(h).$$

- اثر قطعه‌ای^۲: مقدار تغییرنگار در نزدیکی مبدأ مختصات ($h \rightarrow 0$) را اثر قطعه‌ای نامیده و به صورت

$$c_0 = \lim_{\|h\| \rightarrow 0} \gamma(h),$$

تعریف می‌شود. اثر قطعه‌ای، مجموع دو نوع خطای اندازه‌گیری^۳ و خطای ریزمقیاس^۴ است، یعنی

$$c_0 = c_{ME} + c_{MC},$$

به عبارت دیگر، اثر قطعه‌ای بیان‌گر این است که در فواصل بسیار کوچک، رفتار متغیر بسیار نامنظم است. متغیرهای متعددی در زمین‌شناسی وجود دارند که چنین رفتاری از خود نشان می‌دهند. عیار فلزات گران‌بها نمونه‌ای از آن‌ها هستند. دلیل نام‌گذاری اثر قطعه‌ای، این است که اولین بار این پدیده در یک معدن طلا در آفریقای جنوبی و در بررسی متغیر عیار طلا دیده شد. به طوری که میزان عیار به یکباره از صفر در خارج از قطعه، به مقادیر بسیار بالایی در داخل قطعه رسید. به همین دلیل است که از اثر قطعه‌ای برای بیان این قبیل تغییرات شدید در فواصل بسیار کم استفاده می‌شود (آرمسترانگ^۵، ۱۹۹۷).

۳.۲.۱ هم‌تغییرنگار

کوواریانس دو متغیر $Z(s)$ و $Z(s+h)$ به صورت

$$C(h) = Cov(Z(s), Z(s+h)) = E\{[Z(s) - \mu(s)][Z(s+h) - \mu(s+h)]\} \quad (3.1)$$

تعریف شده و آن را هم‌تغییرنگار^۶ می‌نامند. هم‌تغییرنگار، $C(h)$ ، میزان شباهت تغییرپذیری دو متغیر $Z(s)$ و $Z(s+h)$ را نمایش می‌دهد.

در شکل (۴.۱) مفاهیم ذکر شده در قالب یک نمودار نشان داده شده است.

^۱Sill

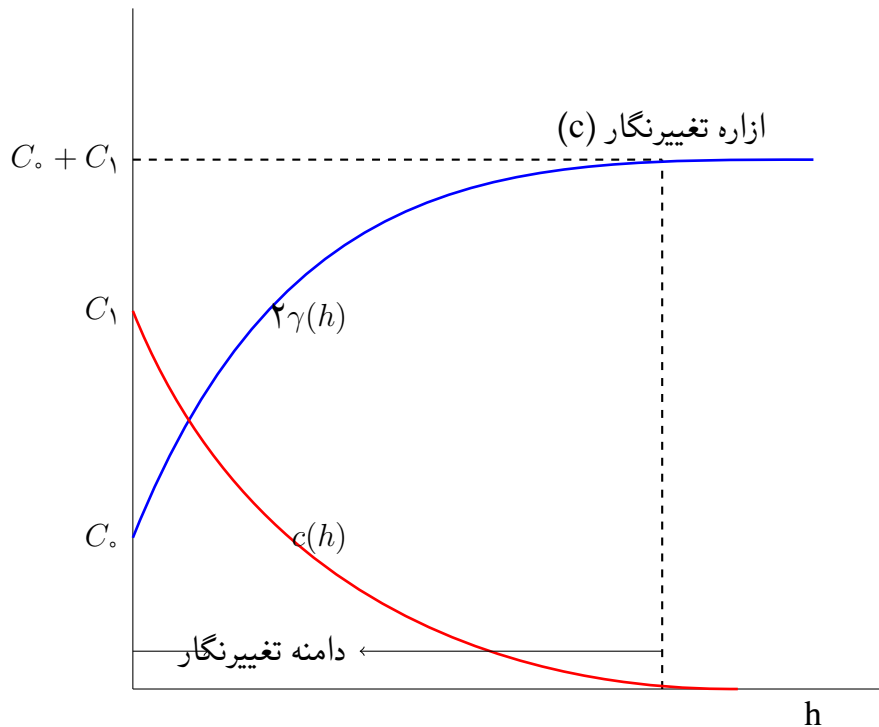
^۲Nugget effect

^۳Measurement error

^۴Micro scale error

^۵Armstrong

^۶Covariogram



شکل ۴.۱: نمودار تغییرنگار و هم‌تغییرنگار (محمدزاده، ۱۳۹۱)

برخی ویژگی‌های تغییرنگار

اگر میدان تصادفی، مانای ذاتی باشد، آنگاه

- تغییرنگار موجود است (یک تابع متناهی و قابل تعیین است).
- اگر تغییرنگار ثابت باشد، برای هر $s_i \neq s_j$ و $Z(s_i)$ و $Z(s_j)$ ناهمبسته‌اند.
- تغییرنگار تابع معین منفی شرطی است^۱، یعنی برای هر دنباله اعداد ثابت $\{a_i\}_{i=1}^m$ که $\sum_{i=1}^m a_i = 0$ ،
 $\sum_{i=1}^m \sum_{j=1}^m a_i a_j \gamma(s_i - s_j) \leq 0$ ، (کرسی، ۱۹۹۳).

۴.۲.۱ برآورد تغییرنگار

برای برآورد تابع تغییرنگار روش‌های مختلفی از قبیل روش‌های گشتاوری، حداقل مربعات و درست‌نمایی ماکسیمم ارائه شده‌اند (کرسی، ۱۹۹۳). در این پایان‌نامه، از روش گشتاوری در برآورد تغییرنگار استفاده شده است. به همین منظور، به‌طور خلاصه این روش را توضیح می‌دهیم. برای جزئیات بیشتر می‌توان به کتاب کرسی (۱۹۹۳)، مراجعه کرد.

^۱Conditional negative definite

روش گشتاوری

با فرض میانگین ثابت (برای میدان تصادفی مانای ذاتی $Z(\cdot)$)، یک برآورد تجربی تغییرنگار بر مبنای روش گشتاوری، توسط ماترون (۱۹۶۲) به صورت

$$\hat{\gamma}(h) = \frac{1}{|N(h)|} \sum (Z(s_i) - Z(s_j))^2, \quad h \in \mathbb{R}^d,$$

ارائه شد، که در آن

$$N(h) = \{(s_i, s_j) : s_i - s_j = h \quad i, j = 1, \dots, n\}$$

و $|N(h)|$ ، تعداد اعضای مجموعه $N(h)$ است. قابل ذکر است که، اگرچه $\hat{\gamma}(-h) = \hat{\gamma}(h)$ ، اما $N(-h) \neq N(h)$ (کرسی، ۱۹۹۳).

تغییرنگار تجربی، نسبت به داده‌های پرت حساس است. به طوری که، اگر داده پرت وجود داشته باشد، توان دوم اختلاف‌های کمیت‌های تصادفی میدان، مقدار بزرگی را نشان می‌دهد. در نتیجه تغییرنگار تجربی مقدار بالایی را می‌گیرد. در این صورت، نسبت به زمانی که داده پرتی در مشاهدات فضایی وجود نداشته باشد، ازاره تغییرنگار بیشتر و دامنه تغییرنگار کمتر است.

۵.۲.۱ تغییرنگار متقابل

فرض کنید برای $s \in D$ ، $Z(s) = (Z_1(s), Z_2(s), \dots, Z_k(s))$ یک میدان تصادفی چندمتغیره باشد. با فرض این که میدان مولفه‌ای آن دارای تغییرنگار $\gamma_{ii}(h) = \text{Var}(Z_i(s+h) - Z_i(s))$ باشد، آنگاه برای محاسبه وابستگی متقابل بین $Z_i(\cdot)$ و $Z_j(\cdot)$ دو راه وجود دارد. معمول‌ترین آن تغییرنگار متقابل^۱ است که برابر

$$\gamma_{ij}(h) = \text{Cov}(Z_i(s+h) - Z_i(s), Z_j(s+h) - Z_j(s)), \quad h \in \mathbb{R}^d,$$

می‌باشد، صورت دیگر آن را می‌توان به صورت زیر نوشت

$$\hat{\gamma}_{ij}(h) = \text{Cov}[(Z_i(s+h) - Z_i(s)), (Z_j(s+h) - Z_j(s))], \quad h \in \mathbb{R}^d,$$

(کرسی، ۱۹۹۳). مقادیر تغییرنگار متقابل نمونه‌ای نیز با استفاده از رابطه زیر برآورد می‌شود:

$$\hat{\gamma}_{ij}(h) = \frac{1}{\hat{\gamma}_n(h)} \sum_{\alpha}^n \sum_{\beta}^m \{Z_i(s_{\alpha}) - Z_i(s_{\beta})\} \{Z_j(s_{\alpha}) - Z_j(s_{\beta})\}$$

$n(h)$ ، تعداد جفت مشاهدات از متغیر اول Z_i و متغیر دوم Z_j است، که در موقعیت‌های s_{α} و s_{β} به فاصله h از هم قرار دارند (پبسم^۲، ۱۹۹۶).

^۱Cross Variogram

^۲Pebesma

۶.۲.۱ همبستگی‌نگار

ضریب همبستگی بین $Z(s)$ و $Z(s+h)$ ، همبستگی‌نگار^۱ نام دارد. با توجه به تعریف هم‌تغییرنگار (رابطه (۳.۱))، اگر $C(0) > 0$ آنگاه می‌توان آن را به صورت

$$\rho(s, s+h) = \rho(h) = \frac{C(h)}{C(0)} = 1 - \frac{\gamma(h)}{C(0)}$$

تعریف نمود. همان‌طور که دیده می‌شود، همبستگی‌نگار نیز تابعی از h است، که با افزایش فاصله بین موقعیت‌ها، مقدار آن کاهش و با کاهش فاصله، مقدار آن افزایش پیدا می‌کند.

۳.۱ تعاریف و مفاهیم زمین‌شناسی

بعضی از واژه‌های زیر از یک کشور به کشور دیگر و نیز از یک سیستم اقتصادی به سیستم اقتصادی دیگر دارای تعاریف متفاوتی هستند^۲.

- **کانه^۳**: کانی‌ای است که به‌عنوان منبع اقتصادی یک فلز مفید مورد استفاده قرار می‌گیرد.
- **باطله^۴**: بیشتر کانه‌ها همراه با مواد بی‌ارزشی می‌باشند به نام باطله. باطله ممکن است سولفور، سیلیکات، کربنات یا فلورید باشد. یک کانی ممکن است در یک معدن باطله و در معدن دیگر کانی با ارزش محسوب شود. به عنوان مثال باریتی که به مقدار کم در معادن سرب وجود دارد، باطله محسوب می‌شود اما در معادنی که باریت به‌صورت رگه‌های قطور حادث شده است، به‌عنوان ماده‌ای با ارزش استخراج می‌شود.
- **کانسنگ یا سنگ معدن^۵**: عبارت از سنگ‌ها یا کانی‌هایی می‌باشند که استخراج و ارائه آن‌ها به بازار یا بخش صنعتی سودآور باشد.
- **کانسار^۶**: به محل پیدایش طبیعی کانی یا کانی‌هایی که معمولاً دارای ارزش اقتصادی می‌باشند، کانسار اطلاق می‌گردد.
- **معدن^۷**: به کانسار یا کانساری فلزی اطلاق می‌شود که در حال بهره‌برداری یا آن‌که قبلاً از آن بهره‌برداری شده است.

^۱ Correlogram

^۲ مفاهیم این بخش، از کتاب دکتر جمشید شهاب‌پور اقتباس گردیده است

^۳ Ore mineral

^۴ Gangue

^۵ Ore

^۶ Mineral deposit

^۷ Mine

- **توده معدنی^۱**: توده‌ای است پیوسته با شکل مشخص که مقدار ماده معدنی آن به حدی است که استخراج آن از نقطه نظر اقتصادی امکان‌پذیر است.
 - **کلارک^۲**: میانگین عیار یک عنصر در پوسته زمین به کلارک آن عنصر موسوم است.
 - **عیار^۳**: مقدار فلز موجود در یک سنگ معدن عیار آن نامیده می‌شود که به صورت درصد یا پی‌پی‌ام^۴ بیان می‌شود.
 - **عیار حد^۵**: عبارت است از پایین‌ترین عیار قابل استخراج در یک توده معدنی.
 - **ضریب غنی‌شدگی^۶**: برای تشکیل یک توده معدنی، عنصر یا عناصر مورد نظر می‌بایست نسبت به مقدار آن‌ها در پوسته معمولی زمین تا حد بسیار زیادی غنی‌شده باشند. مقدار این غنی‌شدگی، ضریب غنی‌شدگی نامیده می‌شود.
 - **نسبت تمرکز^۷**: نسبت غلظت میانگین یک عنصر در مواد مورد مطالعه یا در پوسته زمین، نسبت تمرکز نامیده می‌شود.
 - **ذخایر**: این واژه به توده‌های معدنی در حال استخراج، توده‌های معدنی‌ای که موجودیت آن‌ها به کمک حفاری یا سایر روش‌ها مشخص گردیده است و توده‌های معدنی‌ای که وجودشان در مناطق خاصی با اطمینان حدس زده شده است، اطلاق می‌شود.
 - **منابع**: این واژه به ذخایر، مواد کشف‌شده‌ای که در حال حاضر استخراج آن‌ها غیراقتصادی می‌باشد و توده‌های معدنی هنوز کشف‌نشده، اطلاق می‌گردد.
- به علت مطالعه‌ای که بر روی داده‌های طلای رخساره ساری گونای انجام شده‌است، در بخش بعدی به‌طور مختصر اطلاعاتی در مورد فلز طلا ارائه می‌دهیم^۸.

^۱Orebody

^۲Clarke

^۳Grade

^۴ppm (parts per milion)

^۵Cutoff grade

^۶Enrichment factor

^۷Concentration ratio

^۸مفاهیم این بخش، از پایگاه ملی داده‌های علوم زمین کشور اقتباس گردیده است.

۴.۱ مقدمه‌ای بر مفاهیم مرتبط با عنصر طلا

۱.۴.۱ مقدمه

طلا از گذشته‌های بسیار دور، به دلیل جلای زیبا، شکل‌پذیری خوب و کمیابی، مقاومت بالا در مقابل اکسیداسیون و دیگر عوامل شیمیایی، در طول تاریخ مورد توجه بشر بوده و دارای اهمیت ویژه‌ای می‌باشد. طلا به عنوان مهمترین استاندارد پولی جهان مطرح بوده و بیشترین مورد مصرف آن در ساخت سکه و شمش به عنوان ذخایر پولی بین‌المللی است. این فلز به علت زیبایی و مقاومت، به صورت زیورآلات و کارهای هنری نیز استفاده می‌شود. همچنین در ساخت لوازم الکترونیکی دقیق مورد استفاده قرار می‌گیرد به طوری که در آینده پیش بینی می‌شود رده اول مصرف طلا از آن این نوع لوازم باشد. فلز طلا به عنوان یک سرمایه ملی و پشتوانه اقتصادی کشورها مطرح می‌باشد. بنابراین اطلاع‌رسانی در مورد آمار قیمت، تولید، ذخیره و سایر ویژگی‌های این فلز گران‌بها در ایران در مقایسه با دیگر کشورها برای برنامه‌ریزی بهتر در جهت استخراج و استفاده از آن مفید و حتی ضروری است.

کمترین مقدار طلایی که از نظر تجاری ارزش استخراج دارد (بسته به اندازه و وسعت کانسار و جنبه اقتصادی بهره‌برداری از آن) ۱ تا ۱۰ گرم در هر تن سنگ معدن است که معادن ۰/۰۰۰۱ تا ۰/۰۰۱ درصد می‌باشد (ضرغام معزز لسکو، ۱۳۸۱).

۲.۴.۱ تاریخچه

طلا از واژه گوتیک (Gult) اقتباس شده است. این فلز از قدیمی‌ترین فلزات شناخته شده است و به سادگی شکل‌پذیر می‌باشد. هندی‌ها اولین بار در کتاب مقدس‌شان (۴۰۰۰ سال قبل از میلاد مسیح) به طلا اشاره کرده‌اند. قدیمی‌ترین معدن طلا (۴۰۰۰ سال قبل از میلاد مسیح) در بین‌النهرین، در خاورمیانه و مربوط به سومریان بوده است. یافته‌ها نشان می‌دهد که بین سال‌های ۲۶۵۴ تا ۲۶۸۵ قبل از میلاد در آیدوس^۱ و نگادا^۲ مصر، زرگری مرسوم بوده است. این زمان، به پیش از تاریخ استفاده از نقره برمی‌گردد. در این زمان استفاده از طلا در جواهرسازی و تزئینات، به جز برای فرعونیان، خانواده آن‌ها، کاهنین و مقامات بلندپایه ممنوع بود. براساس اطلاعات محلی در مصر، وجود طلا برای اولین بار در رود نیل و معادن خاص گزارش شده است که به ۲۲۰۰ سال قبل از میلاد برمی‌گردد.

۳.۴.۱ عیار طلا

عیار، در حقیقت مقدار طلای خالص در ۲۴ قسمت از کل فلز است. یعنی این‌که عیار طلای ۱۰۰ درصد خالص، برابر ۲۴ است. بهترین طلای تجارتي که در ساخت زیورآلات به کار می‌رود دارای عیار ۱۵ تا

^۱Abydos

^۲Nagada

۱۸ است. عیار سکه‌های طلا عموماً ۶/۲۱ می‌باشد. یعنی در آن‌ها ۹۰ درصد طلا وجود دارد. برای تعیین عیار طلا در ایران مقادیر معینی را به عنوان مبنا قرار داده‌اند. مثلاً طلای با عیار ۹/۹۹۹. طلای ۲۲ را با عدد ۹۱۶ و عیار ۱۸ را با عدد ۷۵۰ و عیار ۱۴ را با عدد ۵۸۵ نشان می‌دهند. وقتی می‌گوییم که طلایی عیارش ۱۸ است منظور آن است که ۷۵۰ قسمت آن از طلای خالص و بقیه آن از فلزاتی مانند نقره، مس و نیکل است. برای تهیه عیارهای رنگی طلا می‌توانیم به سه مورد زیر اشاره کنیم:

۱. برای تهیه طلای سفید با عیار ۱۸ می‌بایست ۷۵۰ قسمت آن را طلا، ۵۰ قسمت را نقره و ۲۰۰ قسمت را پالادیم، یا ۷۵۰ قسمت آن را طلا، ۱۳۵ قسمت آن را نیکل، ۸۵ قسمت را مس و ۳۰ قسمت را روی در نظر بگیریم.

۲. برای تهیه طلای به رنگ ارغوانی باید ۴۹/۷۸ درصد طلا و باقی آن فلز آلومینیوم باشد.

۳. برای تهیه طلای آبی روشن، نیاز به ۱۷/۴۶ درصد طلا و ۸۳/۵۳ درصد ایندیم است.

عیار طلا به روش‌های مختلفی قابل تعیین است، از جمله روش سنتی با استفاده از سنگ محک و تیزاب سلطانی، و همچنین توسط دستگاه اسپکتروفتومتر و روش کوپلاسیون.

۴.۴.۱ ذخایر جهانی طلا

کل ذخایر جهانی طلا در دنیا تقریباً ۳/۱ بیلیون اونس بوده که از آن مقدار، تقریباً ۳/۲ آن در آفریقای جنوبی قرار دارد. لازم به ذکر است که منظور از ذخایر، ذخایر قطعی و ثابت می‌باشند و ذخایر پایه شامل مجموع منابعی است که در حال حاضر، اقتصادی، کمی اقتصادی و نیمه اقتصادی است. میزان تولید سالانه جهانی طلا در سال ۱۸۴۵ حدود ۶۰ تن بود. اولین افزایش چشمگیر در تولید جهانی طلا به سال ۱۸۴۹ برمی‌گردد که در آن سال در غرب ایالات متحده، کانسارهای طلای زیادی کشف شد و براساس این اکتشاف، سهم ایالات متحده در تولید جهانی طلا از چند درصد به حدود ۵۰ درصد افزایش یافت. تولید طلا تا سال ۱۹۱۰، روند افزایشی نشان می‌دهد ولی از سال ۱۹۱۰ تا کمی بعد از جنگ جهانی اول، تولید شدیداً کاهش یافت و تا سال ۱۹۲۵ این روند ادامه داشت. از این سال تا شروع جنگ جهانی دوم افزایش چشمگیری در تولید جهانی طلا مشاهده می‌شود که با شروع جنگ جهانی دوم کاهش شدیدی در تولید طلا رخ داد. نیاز به سرمایه برای بازسازی جوامع بعد از جنگ و توسعه آن‌ها، دلیل افزایش تولید در سال‌های بعد از جنگ بود. این افزایش تولید برای بیشتر از دو دهه ادامه داشت به طوری که تولید طلا از ۷۵۰ تن به حدود ۱۵۰۰ تن در سال رسید و علت عمده این افزایش، کشف کانسارهای طلای نوع کارلین بود.

۵.۱ مدل‌سازی ذخایر طلا

مانند هر ماده معدنی دیگر، پیشینیان بیشتر تمایل به استخراج ذخایر پرعیار سطحی طلا داشته‌اند، زیرا از یک طرف خطر سرمایه‌گذاری اکتشافی آن‌ها ناچیز بوده و از طرف دیگر معدن‌کاری این نوع کانسارها با سودهای کلان همراه بوده است. امروزه احتمال پیدایش چنین کانسارهایی در کشورهای توسعه‌یافته به دلیل شناخت قبلی این کانسارها و بهره‌برداری در سطح وسیع از آن‌ها (به منظور سرعت بخشیدن به برنامه‌های توسعه‌ای) اندک می‌باشد. ولی در کشورهای در حال توسعه، احتمال وجود چنین کانسارهایی به دلیل ناشناخته ماندن آن‌ها هنوز بالا است.

مدل‌سازی رخدادهای طبیعی و روابط علت و معلولی مرتبط با آن‌ها، همواره یکی از مهم‌ترین موضوعات مورد پژوهش بوده است. مدل‌سازی را می‌توان روش ساده‌سازی، کلی‌نگری و سهولت‌بخشی برای شناخت رخدادهایی دانست که دارای ویژگی‌های مشترک می‌باشند (حسنی پاک، ۱۳۷۸). انواع مدل‌ها عبارت‌اند از:

- **مدل توصیفی:** مدل‌سازی را می‌توان روش ساده‌سازی، کلی‌نگری و سهولت‌بخشی برای شناخت رخدادهایی دانست که دارای ویژگی‌های مشترک می‌باشند. بنا به تعریف، اگر داده‌های معرف عمومی‌ترین ویژگی‌های توصیفی یک نوع کانسار طلای خاص به‌طور سیستماتیک و استاندارد مرتب‌گردند، مدل توصیفی آن نوع کانسار طلا به دست می‌آید.
- **مدل عیار-ذخیره:** هر مدل عیار-ذخیره می‌تواند از دو نمودار تشکیل شود (کاکس و سینگر، ۱۹۸۶). روی محور افقی این نمودارها میزان عیار یا ذخیره کانسار طلا و روی محور عمودی آن فراوانی تجمعی نسبی کانسارهای طلا می‌توانند قرار گیرند. مدل‌های عیار-ذخیره به مهندسين اکتشاف در مورد تخمین عیار و ذخیره کانسار احتمالی دید اولیه‌ای می‌دهد. با توجه به این مدل‌ها، طراح پژوه اکتشافی، می‌تواند ارزیابی اولیه‌ای از حدود تغییرات عیار و ذخیره کانسار مورد اکتشاف داشته باشد و سپس با توجه به عوامل اقتصادی، نسبت به عملیات اکتشافی تصمیم‌گیری کند.
- **مدل عددی**

۱.۵.۱ انواع کانسارهای طلا

تاکنون تقسیمات متعددی توسط زمین‌شناسان مختلف برای کانسارهای طلا ارائه شده‌اند. دانشمندی همچون بیتمن و ایتانتون هر یک طلا را در گروه‌های متفاوتی قرار داده‌اند. تقسیم‌بندی کانسارهای طلا براساس نظر کریم‌پور (سن سنگ در برگیرنده، منشاء و پتانسیل اقتصادی) عبارتند از:

۱. گروه پرکامبرین

• آرکتن:

- الف) کانسارهای طلا همراه رسوبات شیمیایی آهن دار
- ب) کانسارهای طلای نوع رگه‌ای در شیست‌های سبز آرکتن
- ج) کانسارهای نوع ماسیو سولفاید Cu-Zn
- د) کانسارهای نوع پلاسر

• پروتروزوئیک:

- الف) کانسارهای طلای نوع پلاسر
- ب) کانسارهای ماسیو سولفاید Cu, Zn

۲. گروه فانروزوئیک

• کانسارهای طلای اپی ترمال:

- الف) نوع پراکنده در سنگ‌های کربناته
- ب) نوع پراکنده در سنگ‌های آتشفشانی
- ج) نوع رگه‌ای

• کانسارهای نوع پلاسر

• کانسارهای ماسیو سولفید:

الف) نوع کوروکو

ب) نوع قبرس

• کانسارهای مس پورفیری:

الف) مدل مونزونیتی

ب) مدل دیوریتی

جدول (۱۰۱) نشان‌دهنده مشخصات سنگ‌های در برگیرنده این کانسارها، کانی‌های مهم و ذخیره و عیار طلا در کشورهای است که از این نوع کانسارها در آنها دیده شده‌اند.

جدول ۱.۱: کانسارهای طلا

عیار	ذخیره	نوع کانسار
ppm A	۱۰ میلیون در تن	طلا در رسوبات شیمیایی آهن‌دار
ppm ۱۰۰ تا ۱۵	کمتر از یک میلیون تن	کانسارهای نوع رگه‌ای
ppm ۵ تا ۱۰.۵		ماسیو سولفاید Cu-Zn
متغیر	ناچیز	کانسارهای نوع پلاسیر
ppm ۱۱ تا ۱۰.۵	۱۰ میلیون تن	کانسارهای طلای اپی‌ترمال (نوع پراکنده در سنگ‌های کربناته)
ppm ۵ تا ۱۰.۵	۱ تا ۲۰۰ میلیون تن	کانسارهای طلای اپی‌ترمال (نوع پراکنده در سنگ‌های آتشفشانی)
ppm ۵	کمتر از یک میلیون تن	کانسارهای طلای اپی‌ترمال (نوع رگه‌ای)
ppm ۲.۳ تا ۰.۱	۰.۱ تا ۵ میلیون تن	کانسارهای ماسوسولفاید (نوع کروکو)
ppm ۱۰.۵ تا ۰.۱	۰.۱ تا ۲۵ تن	کانسارهای ماسوسولفاید (نوع قبرس)
ppm ۵	کمتر از یک میلیون تن	کانسارهای طلای اپی‌ترمال (نوع رگه‌ای)
ppm ۰.۵		کانسارهای مس پورفیری

۲.۵.۱ تولید طلا

شتاب چشم‌گیر در تولید طلا از یک طرف در نتیجه توسعه تکنیک‌های اکتشاف، استخراج و بهبود فن‌آوری استحصال این عنصر بوده و از طرف دیگر معرف نیاز جوامع و ملل مختلف به سرمایه برای توسعه می‌باشد. در این میان درک بهتر مهندسين اکتشاف از محیط تشکیل کانسارهای طلا و مدل‌سازی آن‌ها و همچنین تقاضای روزافزون ملل مختلف برای طلا، که خود موجب افزایش قیمت آن در دهه‌های اخیر شده است، بسیار با اهمیت می‌باشد. در این تولید نیز عواملی نظیر قیمت طلا و نیاز به سرمایه برای رهایی از رکود اقتصادی موثر بوده است. به‌طور خلاصه، اگر چه ذخایر طلا در سراسر جهان مورد اکتشاف، استخراج و بهره‌برداری قرار گرفته‌اند و بسیاری از کشورها در چرخه تولید آن سهیم می‌باشند، ولی تولید طلا در جهان به وسیله شش کشور آفریقای جنوبی، آمریکا، استرالیا، چین، کانادا و روسیه که هر یک سهم قابل ملاحظه‌ای در تولید طلا دارند، کنترل می‌شود.

۶.۱ نتایج

با توجه به این‌که نوسانات قیمت طلا بالا است، نوسانات اقتصادی شدیدی در بحث معدن‌کاری طلا ایجاد می‌کند. به‌همین دلیل سرمایه‌گذاران ترجیح می‌دهند که عدم قطعیت ناشی از سایر مسایل را به حداقل برسانند. یکی از این موارد عدم قطعیت‌های ناشی از برآمد و تخمین ذخیره است، که در این میان داده پرت نقش مهمی دارد. ما در این پایان‌نامه قصد داریم با استفاده از روش‌هایی که در فصل‌های بعدی مطرح می‌شوند، به تشخیص این مقادیر بپردازیم.

فصل ۲

روش‌های تشخیص داده‌های پرت

داده‌های پرت، به‌طور معمول حاوی اطلاعاتی درباره فرآیند مورد مطالعه هستند، و در کاربردهای متفاوت دارای اسامی مختلفی دارد همچون آنومالی^۱، استثنا^۲، خطا^۳ و بی‌نظمی^۴ است. همچنین در مجموعه داده‌های فضایی، شناسایی این نوع داده‌ها، نقش حائز اهمیتی را در بسیاری از مسایل، از جمله تخمین ذخایر معدنی، پیش‌بینی آب و هوا، مدیریت ترافیک و غیره دارد (چن^۵ و همکاران، ۲۰۰۸). از این‌رو، در این فصل، ابتدا به تعریف داده‌های پرت فضایی پرداخته و سپس انواع روش‌های تشخیص آن‌ها را معرفی نموده‌ایم.

۱.۲ داده پرت فضایی

برای تحلیل آماری داده‌ها، به‌طور معمول لازم است تا یک بررسی اولیه به منظور شناخت خصوصیات داده‌ها، از جمله شناسایی فرضیه‌ها اولیه و داده‌های پرت انجام شود. داده پرت^۶ به مشاهده‌ای گفته می‌شود که در یک مجموعه داده، انحراف شدیدی از مابقی مشاهدات داشته و همچنین ممکن است توزیع آن، متفاوت از توزیع سایر مشاهدات باشد (والفیش^۷، ۲۰۰۶). بارنت و لوئیس^۸ (۱۹۸۴) داده پرت را مشاهده‌ای تعریف می‌کنند که به نظر می‌آید با مشاهدات دیگر در مجموعه داده، ناسازگار است. این تعریف دو جنبه از مسئله داده پرت را بیان می‌کند، یکی این‌که تصمیم قبلی در مورد گروه‌بندی داده‌ها ممکن است درست نباشد، به‌طوری‌که تشخیص داده پرت، وجود زیرجامعه دیگری را که قبلاً

^۱Anomalie

^۲Exception

^۳Fault

^۴Irregularitie

^۵Chen

^۶Outlier data

^۷Walfish

^۸Barnett and Lewis

سازماندهی نشده است را فراهم کند. دیگری در مورد رفتاری است که داده‌ها باید داشته باشند. امکان دارد فرضیه‌هایی به اشتباه در نظر گرفته شده باشند. هرگاه مشاهده‌ای به عنوان داده پرت انتخاب می‌شود، نیاز است ابتدا بررسی شود که این مقدار ناشی از ثبت نادرست یا اندازه‌گیری غلط نباشد و اگر پس از بررسی به این نتیجه برسیم که این داده از فرآیند واقعی پشتیبان داده‌ها آمده است و تصمیم قبلی‌مان در مورد جامعه مناسب بوده است، نباید به عنوان یک مشاهده نادرست کنار گذاشته شود.

داده‌های پرت با تعریف فوق، تنها یک دسته خاص از داده‌های پرت فضایی هستند. دسته دیگر مشاهداتی‌اند که نسبت به مقادیر مجاور خود غیرعادی می‌باشند. بنابراین در یک مجموعه داده فضایی، ممکن است دو نوع داده پرت وجود داشته باشد (محمدزاده، ۱۳۹۱).

- دسته اول: داده‌هایی هستند که رفتار آن‌ها بدون در نظر گرفتن ساختار فضایی و فقط براساس مقدار مشاهده‌شده، نسبت به عمده داده‌ها، غیرمعمول تشخیص داده می‌شود. این مساله معمولاً با استفاده از نمودار ساقه و برگ^۱ و نمودار جعبه‌ای^۲ بررسی می‌شود و همچنین این‌گونه داده‌های پرت در مشاهدات فضایی با روش هاینینگ^۳ نیز مورد شناسایی قرار می‌گیرد.
- دسته دوم: بیان‌گر موقعیت‌هایی هستند که مقادیر آن‌ها تفاوت معناداری نسبت به مقادیر موقعیت‌های همسایه‌شان دارند، هرچند ممکن است با عمده داده‌ها تفاوت معناداری نداشته باشند.

تفاوت‌هایی در تشخیص این دو دسته داده پرت وجود دارند که عبارتند از:

- الف. داده پرت نوع اول، مشاهده‌ای است که با کل مجموعه داده مقایسه می‌شود، در حالی که داده پرت نوع دوم به اختلاف موضعی در بین همسایگی‌های فضایی می‌پردازد.
- ب. روش‌های تشخیص داده پرت نوع اول، به‌طور عمده برای مقادیر، ویژگی‌ها و خوشه‌ها به‌کار گرفته می‌شوند، درحالی‌که روش‌های تشخیص داده پرت نوع دوم، داده‌های فضایی پیچیده‌تری مانند نقاط، خطوط، چندضلعی‌ها و اشیا سه‌بعدی را پردازش می‌نمایند.
- ج. در تشخیص داده‌های پرت فضایی نوع دوم، نیاز است وابستگی فضایی در نظر گرفته شود. به همین علت در آمار فضایی بیشتر مورد توجه قرار می‌گیرد.

در تعریف داده فضایی، صفت کیفی^۴ به دو قسمت اساسی، صفت کیفی فضایی و صفت کیفی غیرفضایی تقسیم می‌شود. صفت‌های کیفی فضایی، اطلاعات مربوط به موقعیت‌ها، مرزها، جهت‌ها، اندازه‌ها و حجم‌ها را ثبت می‌کنند، که همبستگی فضایی بین همسایه‌ها را تعیین کنند. مقدار صفت‌های کیفی

^۱ Steam and leaf

^۲ Box plot

^۳ Haining

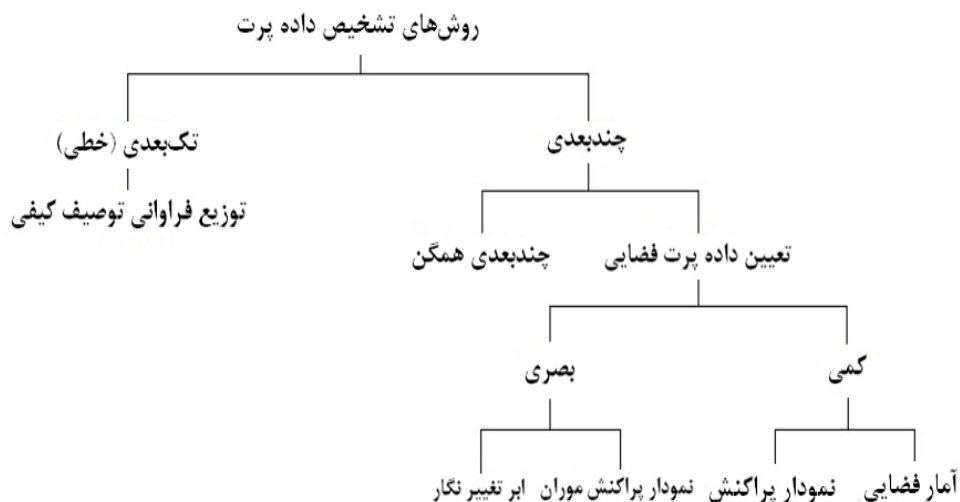
^۴ Attribute

غیرفضایی می‌توانند بر مبنای همسایگی، مشاهدات غیرعادی را تشخیص دهند، مانند سن، حس‌گر ترافیکی و عیار طلا.

همسایگی فضایی در یک مجموعه داده فضایی، زیرمجموعه‌ای از داده‌های فضایی بر مبنای بعد و موقعیت فضایی است. همسایگی‌های فضایی ممکن است بر اساس صفت‌های کیفی فضایی همچون مسافت و مجاورت، که در روابط فضایی استفاده شده‌اند، تعریف شوند. بنابراین یک مشاهده پرت فضایی، داده‌ای است که مقادیر صفت کیفی غیرفضایی آن از دیگر داده‌های فضایی در همسایگی فضایی، به طور معناداری متفاوت هستند، هرچند ممکن است نسبت به کل جامعه متفاوت نباشند. برای مثال، یک خانه جدید در یک محله قدیمی از یک منطقه رو به رشد وابسته به پایتخت، یک داده پرت بر مبنای صفت کیفی غیرفضایی عمر خانه می‌باشد (شخار^۱ و همکاران، ۲۰۰۳).

۲.۲ تقسیم‌بندی روش‌های تشخیص داده‌های پرت

روش‌های متعددی برای تشخیص داده پرت وجود دارند، که به دو گروه اصلی روش‌های تک‌بعدی^۲ و چندبعدی^۳ تقسیم بندی می‌شوند. این دو گروه را همانند آنچه در شکل (۱.۲) آمده است، می‌توان به زیر گروه‌های دیگر تقسیم نمود، که در جدول (۱.۲) نیز مقایسه مشخصه‌های آن‌ها آمده است (شخار و همکاران، ۲۰۰۳).



شکل ۱.۲: تقسیم‌بندی روش‌های تشخیص داده پرت

^۱Shekhar

^۲One-dimensional outlier detection

^۳Multi-dimensional outlier detection

جدول ۱۰۲: مقایسه روش‌های تشخیص داده پرت

نوع روش	تعریف همسایگی	تشخیص داده پرت
تک‌بعدي		بر اساس توزیع جامعه
چندبعدي	بر اساس موقعیت و مقدار مشخصه بر اساس موقعیت	بر اساس موقعیت و مقدار مشخصه بر اساس موقعیت و مقدار مشخصه

در روش‌های تشخیص داده پرت تک‌بعدي، موقعیت فضایی هر داده و همبستگی فضایی بین مشاهدات نادیده گرفته می‌شود و مدل توزیعی بر مقادیر مشخصه غیرفضایی، برازش داده می‌شود.

روش‌های تشخیص داده پرت چندبعدي، خود نیز به دو گروه روش‌های چندبعدي همگن^۱ (*HMD*) و روش‌های فضایی (روش‌های چندبعدي دو بخشی^۲) (*BPMD*) تقسیم می‌شوند.

روش‌های چندبعدي همگن، مجموعه داده‌ها را همچون مجموعه‌ای از نقاط، در یک فضای ایزومتریک چندبعدي، مدل‌بندی می‌کنند و آزمون‌ها را بر مبنای مفاهیمی همچون مسافت و چگالی، فراهم می‌کنند. این روش‌ها هیچ تمایزی بین بعدهاى صفت کیفی و بعدهاى زمین-فضایی قائل نمی‌شوند و برای تعریف همسایگی، همانند مقایسه، از همه بعدها استفاده می‌کنند. روش‌های چندبعدي همگن چند محدودیت را به همراه دارند:

- داده پرت نوع اول را بهتر از داده پرت فضایی تشخیص می‌دهد.
- داده‌ها را در فضای متریک، هم‌واحد فرض می‌کنند و بین صفت‌های کیفی غیرفضایی و فضایی تمایزی قائل نمی‌شوند.
- اطلاعات پیشین در مورد توزیع آماری را از داده خاصی استخراج نمی‌کنند.
- به‌ندرت برای داده‌های پرت تشخیص داده‌شده، یک اندازه اطمینان فراهم می‌کنند.

روش‌های *BPMD*، برای تعیین داده‌های پرت فضایی طراحی شده‌اند، که صفت‌های کیفی فضایی را از صفت‌های کیفی غیرفضایی تفکیک می‌کنند. این نوع روش‌ها به دو دسته روش تقسیم می‌شوند: روش‌های بصری مانند تغییرنگار و نمودار پراکنش موران^۳، و روش‌های کمی، مانند نمودار پراکنش، که روش‌های دقیقی برای تشخیص داده پرت هستند (شخار و همکاران، ۲۰۰۳).

^۱ Homogeneous multi-dimensional

^۲ Bi-partite multi-dimensional

^۳ Moran scatterplot

۳.۲ تشخیص داده پرت فضایی با یک صفت کیفی ساده

در این بخش، قصد داریم که برخی روش‌های معمول موجود برای تشخیص داده پرت، در داده‌های فضایی که دارای یک صفت کیفی ساده (یک‌متغیره) هستند را معرفی کنیم.

فرض کنید مجموعه‌ای از موقعیت‌های فضایی $S = \{s_1, s_2, \dots, s_n\}$ ، در فضایی با بعد $d \geq 1$ موجود باشد و تابع مشخصه^۱ f طوری تعریف شده باشد که نگاشتی از S به \mathbb{R} بوده و $f(s_i)$ بیانگر مقدار صفت کیفی در موقعیت‌های فضایی s_i باشد. همچنین فرض کنید $NN_k(s_i)$ ، مجموعه k نزدیک‌ترین همسایگی‌های نقطه s_i را نشان دهد و تابع همسایگی g ، یک نگاشت از S به \mathbb{R} باشد به طوری که برای هر s_i ، $g(s_i)$ ، یک آماره از مقادیر صفت کیفی همه موقعیت‌های فضایی موجود در مجموعه $NN_k(s_i)$ است. برای مثال، $g(s_i)$ می‌تواند میانگین مقدار صفت کیفی k نزدیک‌ترین همسایگی‌های s_i باشد. برای تشخیص داده پرت، مقدار صفت کیفی موقعیت s_i را با مقادیر صفت کیفی، نقاط داخل مجموعه $NN_k(s_i)$ ، مقایسه می‌کنیم. برای مقایسه، از تابع مقایسه h که تابعی از f و g است، استفاده می‌شود. برای مثال، این تابع می‌تواند اختلاف یا نسبت f و g باشد. اگر h_i یک مقدار فرین^۲ در مجموعه $\{h_1, h_2, \dots, h_n\}$ باشد، آنگاه موقعیت s_i نشان‌دهنده موقعیت یک داده پرت فضایی است (چن و همکاران، ۲۰۰۸). حال بنا به تعاریف بیان‌شده، به معرفی برخی روش‌های تشخیص داده پرت فضایی می‌پردازیم.

۱.۳.۲ نمودار پراکنش موران

نموداری است که میانگین همسایگی مقادیر صفت کیفی استانداردشده $Z[f(i)] = \frac{f(i) - \mu_f}{\sigma_f}$ را در مقابل مقادیر صفت کیفی استانداردشده، نشان می‌دهد. $f(i)$ مقدار صفت کیفی و μ_f و σ_f به ترتیب میانگین و واریانس مقادیر صفت کیفی ($f(i)$ ها) هستند. بر اساس این روش، نقطه‌ای به عنوان داده پرت فضایی انتخاب می‌شود که موقعیت آن نقطه، در ربع دوم یا چهارم این نمودار باشد. در شکل (۲.۲)، بنا بر روش نمودار پراکنش موران^۳، نقاطی که به صورت مثلث نشان داده شده‌اند، به عنوان داده پرت در نظر گرفته می‌شوند (شخار و همکاران، ۲۰۰۳).

۲.۳.۲ نمودار پراکنش

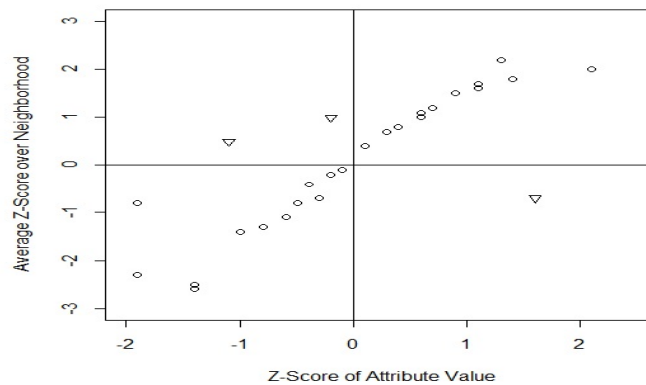
نمودار پراکنش^۴، نموداری است که مقادیر صفت کیفی روی محور افقی و میانگین مقادیر صفت کیفی نقاطی که در همسایگی یک نقطه قرار دارند، روی محور عمودی نمایش داده می‌شود. در این نمودار،

^۱ Attribute function

^۲ Extreme value

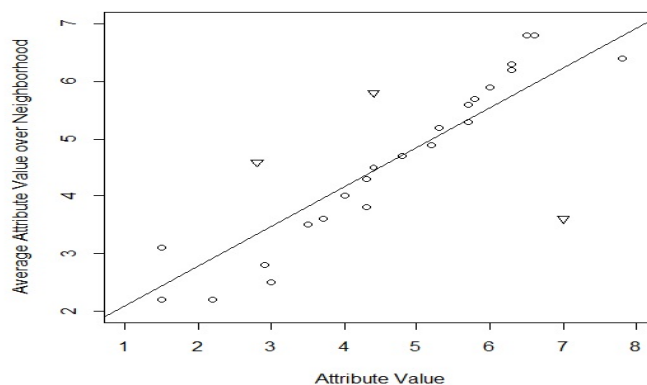
^۳ Moran scatterplot

^۴ Scatterplot



شکل ۲.۲: نمودار پراکنش موران برای یک مجموعه داده فضایی

شیب پراکنش رو به بالا به سمت راست، خودهمبستگی مثبت را نشان می‌دهد در حالی که شیب پراکنش رو به بالا به سمت چپ حاکی از خودهمبستگی منفی است. برای تعیین داده‌های پرت فضایی، از خط رگرسیونی کمترین توان‌های دوم استفاده می‌شود. در این روش، با فرض این‌که خطاها دارای توزیع نرمال باشند، آنگاه $\theta > \left| \frac{\epsilon - \mu_\epsilon}{\sigma_\epsilon} \right|$ ، آزمون برای تشخیص داده پرت فضایی است، که در آن θ بستگی به انتخاب سطح اطمینان دارد. برای مثال، برای سطح اطمینان ۹۵ درصد، θ برابر ۱/۹۶ است. بنابراین در نمودار پراکنش، داده پرت نقطه‌ای است که دارای خطای باقی‌مانده استاندارد شده معنادار، نسبت به خط رگرسیونی کمترین توان‌های دوم است. به این معنا که بیشترین فاصله را نسبت به نقاط دیگر از خط رگرسیونی کمترین توان‌های دوم داشته باشد. در شکل (۳.۲)، بنا بر روش نمودار پراکنش، نقاطی که به صورت مثلث نمایش داده شده‌اند، به‌عنوان داده پرت تعیین می‌شوند (شخار و همکاران، ۲۰۰۳).



شکل ۳.۲: نمودار پراکنش برای یک مجموعه داده فضایی

۳.۳.۲ الگوریتم Z

در روش الگوریتم Z، فرض می‌شود که در یک مجموعه داده فضایی، همه $NN_k(s_i)$ ها (k نزدیکترین همسایگی‌های s_i)، برابر مقدار ثابت k هستند و تابع همسایگی $g(\cdot)$ ، برای موقعیت فضایی s، برابر با میانگین مقدار مشخصه همه k نزدیکترین همسایه‌های s است. با استفاده از تابع همسایگی $g(s)$ و صفت کیفی $f(s)$ ، تابع مقایسه‌ای $h(s)$ را $f(s) - g(s)$ انتخاب می‌کنیم و سپس h_i های موجود در مجموعه $\{h_1, h_2, \dots, h_n\}$ را با $z_i = \frac{h_i - \mu}{\sigma}$ استاندارد می‌کنیم، که در آن μ_h و σ_h به ترتیب میانگین و انحراف معیار مجموعه $\{h_1, h_2, \dots, h_n\}$ است. در این صورت، s_i یک موقعیت داده پرت است اگر $|z_i| > \theta$. انتخاب θ به فاصله اطمینان مشخص شده بستگی دارد.

یکی از عیب‌های این الگوریتم آن است که اگر تعداد داده‌های پرت فضایی مورد انتظار محدود شده باشد، آنگاه ممکن است داده پرت واقعی، به علت اشتباه در انتخاب مشاهده منظم به عنوان داده پرت، نادیده گرفته شود. همچنین ممکن است به علت حضور نقاط همسایگی با مقادیر صفت کیفی خیلی زیاد (یا خیلی کم)، نقاط منظم فضایی، به اشتباه به عنوان داده پرت تشخیص داده شوند. راه حلی که برای این مشکل می‌توان پیشنهاد کرد این است که تابع همسایگی مناسبی، مانند میانه، را انتخاب کنیم. استفاده از میانه به عنوان تابع همسایگی، اثر ایجاد شده توسط نقاط همسایگی کرانگین را کاهش می‌دهد. عیب دیگری که می‌تواند وجود داشته باشد آن است که در این الگوریتم، مجموعه داده $\{h_1, h_2, \dots, h_n\}$ به وسیله میانگین و انحراف معیار نمونه، استاندارد شده است. هنگامی که داده پرت چندگانه^۱ در داده‌ها موجود باشد، این کمیت‌ها اغلب برآوردهای اریبی برای میانگین و انحراف معیار جامعه هستند. در نتیجه برخی داده‌های پرت واقعی، ممکن است انتخاب نشوند و داده‌های منظم، به اشتباه به عنوان داده پرت انتخاب شوند. برای حل این مشکل نیاز است از برآوردهای استواری برای میانگین و انحراف معیار، مورد استفاده قرار گیرند (چن و همکاران، ۲۰۰۸).

۴.۳.۲ الگوریتم میانه

در الگوریتم میانه^۲، فرض کنید $S = \{s_1, s_2, \dots, s_n\}$ مجموعه موقعیت‌های فضایی، f تابع مشخصه، $\alpha \in (0, 1)$ و k مقداری صحیح و مثبت باشد، آنگاه برای هر موقعیت فضایی s_i :

الف) $NN_k(s_i)$ ، مجموعه k نزدیک‌ترین همسایگی را مشخص می‌کنیم.

ب) تابع همسایگی $g(s_i)$ را برابر میانگین مجموعه داده $\{f(s) : s \in NN_k(s_i)\}$ قرار می‌دهیم.

ج) تابع مقایسه‌ای به صورت $h_i = h(s_i) = f(s_i) - g(s_i)$ را محاسبه می‌کنیم.

^۱Multiple outlier

^۲Median algorithm

با فرض این که σ_h^* و μ_h^* به ترتیب برآوردگرهای استوار میانگین و انحراف معیار مجموعه داده $\{h_1, h_2, \dots, h_n\}$ باشند، $y_i = \left| \frac{h_i - \mu_h^*}{\sigma_h^*} \right|$ ، $i = 1, 2, \dots, n$ ، را محاسبه می‌کنیم. آن‌گاه s_i داده پرت است، اگر $y_i \geq z_{\alpha/2}$ ، که در آن $z_{\alpha/2}$ صدک $\alpha/2$ بالایی توزیع نرمال استاندارد است. لازم به ذکر است که در این الگوریتم، μ_h^* می‌تواند میانه، σ_h^* نیز می‌تواند انحراف مطلق میانه^۱ (MAD) باشد که به صورت زیر تعریف می‌شود:

$$\text{median}\{|h_1 - \text{median}(H)|, |h_2 - \text{median}(H)|, \dots, |h_n - \text{median}(H)|\},$$

به طوری که $\text{median}(H)$ ، بیان‌گر میانه مجموعه $\{h_1, h_2, \dots, h_n\}$ است.

۵.۳.۲ روش میانه-میانگین

روش میانه-میانگین، روشی بر مبنای میانه و میانگین داده‌ها است. میانه نسبت به داده‌های غیرعادی استوار است، مقایسه آن با میانگین که یک آماره استوار نسبت به نقاط دورافتاده محسوب نمی‌شود، می‌تواند تردید در مورد وجود داده‌های پرت را قوی یا ضعیف کند. در واقع وسیله دیگر برای شناسایی داده پرت، اختلاف بین میانگین و میانه است. با فرض گاوسی بودن میدان تصادفی، محاسبه این اختلاف مستلزم محاسبه کمیت $u = n^{1/2} \frac{(\bar{f} - \bar{f})}{\sigma / \sqrt{5555}}$ (کرسی، ۱۹۹۳) است، که در آن \bar{f} و \bar{f} به ترتیب میانگین و میانه داده‌ها و $\hat{\sigma}$ برآورد انحراف معیار داده‌ها است که می‌توان آن را از روی چارک اول و سوم به صورت تقریبی $\hat{\sigma} = \frac{q_{0.75} - q_{0.25}}{1.349}$ محاسبه نمود. هر سطر یا ستونی که در آن‌ها مقدار $|u|$ نزدیک یا بیشتر از ۳ باشد، مشکوک به دارا بودن داده پرت خواهد بود. منظور از سطر و ستون این است که، در نمودار مربوط به موقعیت جغرافیایی داده‌ها، مشاهداتی که دارای عرض جغرافیایی یکسان هستند تشکیل سطر داده و مشاهداتی با طول جغرافیایی یکسان، تشکیل ستون می‌دهند. بنابراین نمودار آن‌ها تشکیل شبکه منظمی می‌دهد که دارای سطر و ستون مجزا است.

برای مثال، مقادیر محاسبه شده $|u|$ برای سطرها و ستون‌های داده‌های خاکستر زغال سنگ^۲ که مربوط به معدن روبنا در محله گرین در پنسیلوانیا می‌باشد (کرسی، ۱۹۹۳) و در بسته *gstat* در نرم افزار *R* موجود است، در جدول (۲.۲) آورده شده‌اند. ستون‌های ۵ و ۱۲ و سطر ۳ مشکوک به وجود داده پرت هستند زیرا مقادیر $|u|$ نزدیک به عدد ۳ است (محمدزاده، ۱۳۹۱). در شکل (۴.۲) که مربوط به مقادیر و موقعیت‌های داده‌های خاکستر زغال سنگ است، در ستون پنجم، داده پرت ۱۷/۶ و ستون دوازدهم، داده پرت ۱۱/۹ و همچنین در سطر سوم داده پرت ۱۲/۷ یافت شدند.

۴.۲ ابر تغییرنگار

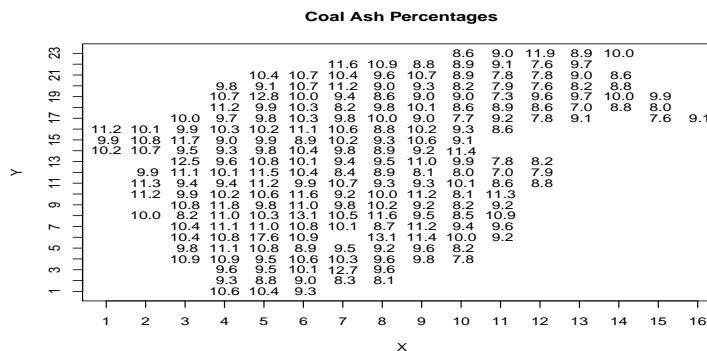
برای نشان دادن اندازه تشابه و وابستگی متقابل مقادیر میدان تصادفی در دو موقعیت s و $s+h$ ، علاوه بر میانگین توان دوم اختلاف‌ها برای هر یک از جفت موقعیت به فاصله h ، می‌توان از دو معیار

^۱Median absolute deviation

^۲coalash

جدول ۲.۲: مقادیر |u| برای داده‌های خاکستر زغال سنگ (کرسی، ۱۹۹۳)

ستون	۱	۲	۳	۴	۵	۶
u	۱/۱۱	۰/۷۶	۰/۷۸	۰/۳۵	۲/۸۷	۰/۰۲
ستون	۷	۸	۹	۱۰	۱۱	۱۲
u	۰/۲۲	۱/۲۹	۱/۲۳	۱/۰۳	۰/۵۸	۳/۱۷
ستون	۱۳	۱۴	۱۵			
u	۱/۲۴	۱/۳۹	۱/۴۸			
سطر	۱	۲	۳	۴	۵	۶
u	۱/۵۴	۰/۴۰	۶/۱۲	۰/۴۵	۰/۳۵	۲/۰۱
سطر	۷	۸	۹	۱۰	۱۱	۱۲
u	۰/۵۶	۰/۰۷	۰/۶۳	۰/۱۸	۲/۱۲	۰/۸۰
سطر	۱۳	۱۴	۱۵	۱۶	۱۷	۱۸
u	۰/۴۶	۰/۷۸	۰/۱۰	۱/۰۵	۰/۶۰	۱/۰۵
سطر	۱۹	۲۰	۲۱	۲۲	۲۳	
u	۰/۱۸	۰/۳۵	۰/۲۵	۱/۳۳	۲/۴۷	



شکل ۴.۲: نمودار پراکنش مقادیر خاکستر زغال سنگ در موقعیت‌های فضایی

اندازه‌گیری زیر استفاده کرد:

$$\gamma^*(h) = \frac{1}{p} [Z(s+h) - Z(s)]^2,$$

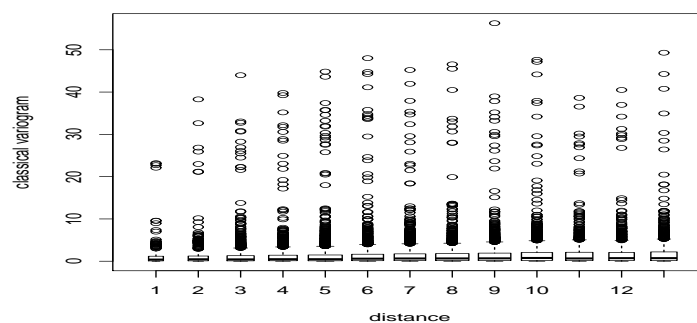
$$\gamma^{**}(h) = \frac{1}{p} \sqrt{|Z(s+h) - Z(s)|}.$$

با رسم $\gamma^*(h)$ و $\gamma^{**}(h)$ در برابر فاصله h ، برای تمام جفت نقاط موجود در مجموعه داده‌ها، نموداری به وجود می‌آید که آن را ابر تغییرنگار می‌نامند (محمدزاده، ۱۳۹۱). این روش برای اولین بار توسط شوو^۱ در سال ۱۹۸۲ پیشنهاد شد، که ابر پراکندگی نقاط را ایجاد می‌کند. این نمودار در واقع نمایشی

^۱Chauvet

از توزیع واریانس بین همه جفت نقاط در تمام فواصل ممکن h است، به همین دلیل آن را ابر تغییرنگار می‌نامند. در واقع ابر تغییرنگار، رابطه نقاط داده‌ها را با همسایگی‌شان نشان می‌دهد. در این پایان‌نامه ابر تغییرنگار را با رسم $\gamma^*(h)$ در مقابل h به دست آورده‌ایم.

بنا به خاصیت داده‌های فضایی، با افزایش فاصله بین موقعیت‌ها، تغییرات در اختلاف‌های مقادیر میدان تصادفی نیز، افزایش می‌یابد. اما اگر برای موقعیت‌های نزدیک، این اختلاف زیاد باشد، ممکن است ناشی از وجود داده پرت باشد. به همین دلیل از ابر تغییرنگار می‌توان به عنوان روشی برای تشخیص داده پرت فضایی استفاده نمود. برای مثال در شکل (۵.۲)، در تاخیر اول ابر تغییرنگار داده‌های خاکستر زغال‌سنگ، جفت نقاطی وجود دارند که اختلاف مقادیر میدان تصادفی‌شان نسبت به جفت نقاط دیگر همان تاخیر بیشتر است.

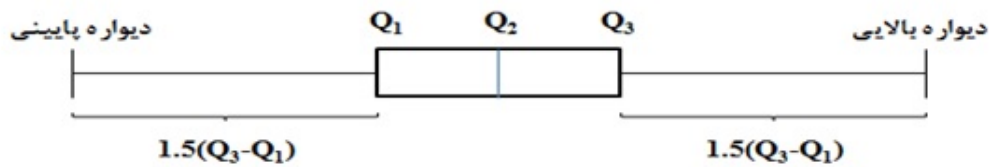


شکل ۵.۲: نمودار ابر تغییرنگار برای داده خاکستر زغال‌سنگ

۵.۲ نمودار جعبه‌ای

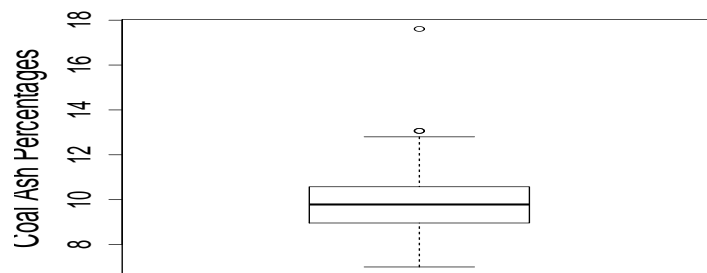
نمودار جعبه‌ای به کمک معیارهای مرکزی و پراکندگی، موقعیت مجموعه داده‌ها را به شکلی بسیار گویا و مفید ارائه می‌دهد. این نمودار ابتدا توسط توکی^۱ ارائه گردید که چارک اول، میانه، چارک سوم، کمترین و بیشترین مقدار مشاهدات را نشان می‌دهد. اختلاف بین چارک اول و چارک سوم ($Q_3 - Q_1$)، دامنه میان چارکی است. اغلب، دیوار بالایی و پایینی این نمودار در یک فاصله ثابتی از دامنه میان چارکی ($1/5(Q_3 - Q_1)$) قرار گرفته است. توسط این نمودار می‌توان تمرکز، پراکندگی و چولگی داده‌ها را نیز تفسیر کرد. از این نمودار برای تشخیص داده پرت دسته اول نیز می‌توان استفاده نمود. مشاهده‌ای داده پرت تشخیص داده می‌شود که خارج از $1/5$ برابر دامنه میان چارکی دیواره‌های بالایی و پایینی قرار گرفته باشد. شکل (۶.۲) نموداری از یک نمودار جعبه‌ای است برای مثال، نمودار جعبه‌ای داده‌های خاکستر زغال‌سنگ در شکل (۷.۲) را در نظر بگیرید. با توجه به این نمودار، داده‌های پرت نوع اول

^۱Tukey



شکل ۶.۲: نمودار جعبه‌ای

قابل شناسایی هستند. در بخش بعدی به مسایل مربوط به داده پرت در زمین آمار می‌پردازیم.^۱



شکل ۷.۲: نمودار جعبه‌ای برای داده خاکستر زغال سنگ

۶.۲ داده پرت در زمین آمار

در ذخایر یک معدن، منظور از داده‌های پرت، مشاهداتی هستند که دارای مقادیر بالایی نسبت به سایر داده‌ها می‌باشند. وجود این‌گونه از مشاهدات در مجموعه داده، موجب بروز برخی مشکلات می‌شوند:

- وجود داده‌های پرت، انحراف اساسی را در برآورد پارامترهای مختلف آماری از جمله میانگین، واریانس و کوواریانس ایجاد می‌کند. همچنین بر برآورد معیارهای خودهمبستگی همچون تغییرنگار تاثیر می‌گذارد.
- در تعیین عیار بلوک ذخایر معدنی، اگر مشاهده پرتی در مرکز بلوک باشد، می‌تواند مقدار بالایی از میانگین را به آن بلوک اختصاص دهد، که در نتیجه منجر به بیش برآورد عیار بلوک خواهد شد.

^۱ مطالب این بخش از کتاب سینکلر (۲۰۰۴) اقتباس گردیده است.

- در پیش‌گویی بلوکی، مثلاً به روش کریگینگ، اگر یک مقدار پرت با وزن منفی وجود داشته باشد، پیش‌گویی حاصل می‌تواند خطای زیادی داشته باشد، حتی ممکن است به مقداری منفی منجر شود.

در مطالعات زمین‌آماري، در هنگام برآورد تغییرنگار، داده پرت حذف می‌شود اما در کریگینگ عادی از داده پرت استفاده می‌شود. همچنین در برخی موارد، داده‌های پرت موجود در مجموعه داده، بیان‌گر جامعه زمین‌شناسی جداگانه هستند، که امکان دارد دارای قلمرو زمین‌شناسی شناخته‌شده‌ای نیز باشند. به منظور برآورد، می‌توانند از قلمرو اصلی مستقل فرض شوند.

به‌طور کلی در ارزیابی ذخایر معدنی، مقادیر بالای غیر عادی دیده می‌شوند. این مقادیر بالا ممکن است به علت اشتباه یا با بازتاب زیرمحیط‌ها و قلمروهای زمین‌شناسی جداگانه‌ای، ایجاد شوند. برای آزمودن این مقادیر بالا و زمینه زمین‌شناسی آن‌ها باید تلاش نمود تا پس از شناسایی، امکان آن باشد که به ترتیب، خطاها را از مقادیر حقیقی تشخیص داده و مشخصات مقادیر حقیقی را بررسی نمود.

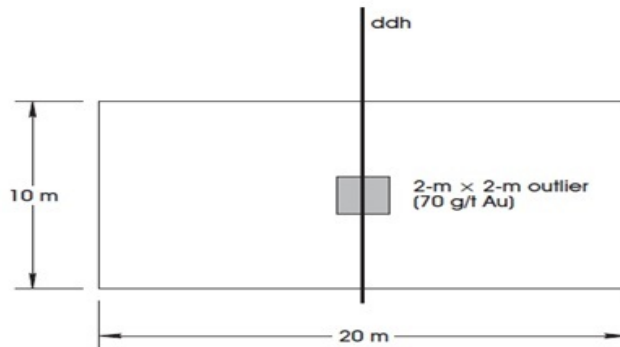
برخورد با داده‌های پرت واقعی در برآورد منابع و ذخایر، مشکلی است که در حال حاضر، راه حل قابل قبول و کلی برای آن وجود ندارد. به خصوص نگرانی این است که مقادیر خیلی بالا، نباید با وزن خیلی بالایی در نظر گرفته شوند یا این‌که در تناژی آشکار، از ماده معدنی با عیار بالا، شرکت کنند که در واقعیت وجود نداشته باشد. تجربه نشان داده است، اگر با مقادیر عیار خیلی بالا همانند مقادیر عیار پایین، در برآورد منابع و ذخایر، رفتار شود، نسبت کوچکی (۱ درصد یا کمتر) از عیار طلا، که مقدار بالایی دارد (شاید یک یا دو برابر بیشتر از میانگین مابقی نمونه‌ها)، می‌تواند منجر به بیش‌برآورد جدی از عیار میانگین بالای عیار حد شود. به‌طور معمول داده‌های پرت، از منظر زمین‌شناسی مشخص هستند و پیوستگی زمین‌شناسی محدودشده‌ای نسبت به مقادیر عیار پایین دارند. بنابراین در سنگ همسایه، می‌توانند به همان فاصله گسترش یابند که نمونه‌های عیار پایین گسترش یافته‌اند، که منجر به بیش‌برآوردی معنادار از منابع و ذخایر می‌شوند.

۱.۶.۲ برش زدن داده پرت

برش زدن مقادیر پرت، در صنعت معدن روش گسترده‌ای است. به‌طور خلاصه در برش زدن، عیارهای بالا را به یک آستانه بحرانی کاهش می‌دهند و همانند عیارهای پایین بلوک، برای برآورد بلوک از آن‌ها استفاده می‌شود. آستانه برش، توسط آزمون و خطا به‌دست می‌آید. کاربرد روش برش زدن را می‌توان با یک مثال ساده دوبعدی نشان داد. در این مثال می‌خواهیم عیار بلوکی را که شامل پنج نمونه محوری در داخل آن است، را برآورد کنیم. چهار نمونه از این نمونه‌ها، متوسط مقدار عیارشان 2 g/t و نمونه دیگر، مقدار عیارش 70 g/t است. میانگین پنج نمونه برابر است با

$$g_1 = [(4 \times 2) + (1 \times 70)] / 5 = 15/6.$$

این مقدار بیانگر یک برآورد ممکن از بلوک است. فرض کنید بنا بر مطالعه تفصیلی زمین‌شناسی این ذخیره، مقادیر پرت طلای بیشتر از 30 g/t ، یک شکل زمین‌شناسی شناخته شده با یک پیوستگی فیزیکی متوسط از ۲ متر در دو جهت اصلی دارند. با توجه به اندازه‌گیری‌های مختلف برای بلوک دوجانبی در



شکل ۸.۲: موقعیت داده پرت در بلوک

شکل (۸.۲)، امکان آن وجود دارد که برآورد بلوکی دیگری را ایجاد کرد، یعنی وزن‌دهی مقادیر عیار، طبق احتمال مناطق باشد. در این حالت وزن برای عیار 70 g/t ، برابر، $(2 \times 2) / (20 \times 10) = 0.02$ و برای عیار متوسط 2 g/t برابر، $[(20 \times 10) - 2 \times 2] / (20 \times 10) = 0.98$ است. این وزن‌ها، عیار بلوکی g_2 را تولید می‌کنند:

$$g_2 = (0.02 \times 70) + (0.98 \times 2) = 3/36.$$

رویکردی دیگر برای برآورد عیار بلوک، برش زدن عیارهای پرت به عیار آستانه از پیش تعیین شده است. برای مثال، فرض کنید عیار آستانه 35 g/t است. در این صورت برآورد عیار بلوک به روش چندضلعی^۱ برابر است با

$$g_3 = (1 \times 35) + (4 \times 2) = 8/6.$$

برش زدن داده پرت به عیار آستانه 35 g/t برخی مشکلات بیش‌برآورد را کاهش داده است، اما با این وجود، آستانه برش منجر به بیش‌برآورد شده است. فرض کنید عیار بلوک مشخص باشد. در این صورت، امکان پیدا کردن مقدار مناسبی برای آستانه که بتوان برآورد مناسبی برای عیار بلوک داشت، وجود دارد. برای مثال، فرض کنید g_2 بهترین برآورد برای بلوک باشد. رابطه زیر شاخص برش مناسب (F) را تعیین می‌کند:

$$3/36 = (1 \times F) + (4 \times 2).$$

بنابراین

$$F = 8/8 \text{ g/t}.$$

^۱Polynognal estimate

به عبارت دیگر، تمامی داده‌های پرت را به مقدار ۸/۸ برش می‌زنیم. در عمل، عیار واقعی بلوک معلوم نیست. در این حالت، میانگین تولید با برآوردهایی که بر اساس شاخص‌های برش مختلف به دست آمده‌اند، مقایسه شده تا این‌که مقداری از F به دست آید که مقایسه مناسبی از برآوردها با تولید باشد. مشکل اصلی برش زدن، انتخاب مقدار برش است. مشکل دیگر این است که مقدار برش مطلوب تعیین شده برای یک بخش محدود یک ذخیره، ممکن است برای بخش دیگری از ذخیره مناسب نباشد. با وجود این محدودیت‌ها، حتی زمانی که تولید امکان‌پذیر نیست، برش زدن عیار بالا به علت اهمیت بیش‌برآورد بالقوه، روش رایجی است. در اصل، برش زدن عیارها نامناسب است. زیرا دلخواه بودن مقدار برش، منجر به عدم قطعیت در برآوردهای تناژ و مقدار عیار می‌شود.

داده‌های پرت و وزن‌های منفی

برخی از روش‌های زمین‌آماری برای برآورد منابع و ذخایر، می‌توانند وزن منفی، به یک یا بیشتر از یک گروه از داده‌های انتخاب‌شده، به منظور برآورد عیار بلوک اختصاص دهند. فرض کنید پنج عیار برای برآورد بلوک در نظر گرفته شود که چهار عیار به‌طور متوسط 2g/t و عیار پنجم 100g/t باشد. فرض کنید حجم بلوک $20 \times 20 \times 10\text{m} (4000\text{m}^3)$ و حجم داده پرت در بلوک $2 \times 2 \times 1\text{m} (4\text{m}^3)$ هستند. بنابراین، عیار بلوک برابر $2/0.98\text{g/t} = 100 \times 4/4000 + (4 - 4000)/4000$ است. برای محاسبه عیار بلوک، شاخص‌های وزنی متفاوتی را برای داده پرت به‌صورت زیر فرض کنید.

$$\text{وزن داده پرت} = 0/01$$

$$42\text{ درصد بیش برآورد دارد} \Rightarrow 2/98 = (0/99 \times 2) + (0/01 \times 100)$$

$$\text{وزن داده پرت} = -0/01$$

$$42\text{ درصد کم برآورد دارد} \Rightarrow 1/02 = (1/01 \times 2) - (0/01 \times 100)$$

$$\text{وزن داده پرت} = -0/03$$

$$\text{این مقدار برای عیار ممکن نیست} \Rightarrow -0/94 = (1/03 \times 2) - (0/03 \times 100)$$

نتیجه‌ای که می‌توان از این محاسبات ساده گرفت این است که با اختصاص وزن بسیار کوچکی به داده‌های پرت، می‌تواند منجر به آریبی معناداری در برآورد عیار بلوک شود. اغلب وزن‌های مثبت منجر به بیش‌برآورد و وزن‌های منفی منجر به کم‌برآورد می‌شوند. در برآورد ذخایر یا منابع، نمی‌توان با داده‌های پرت مانند بیشتر مقادیر دامنه برخورد کرد.

۷.۲ نتیجه‌گیری

با توجه به مطالبی که در این فصل بیان شدند، می‌توان به اهمیت تشخیص مقدار برش، برای برآورد مقدار تناژ و عیار یک ذخیره معدنی، پی برد. از این‌رو، اهمیت روشی که بتواند مرز مقادیر بالا را برای مشاهدات فضایی، تشخیص دهد، قابل توجه است. زیرا به استناد مطالب بیان‌شده، تاکنون روش مناسبی برای تشخیص مرز مناسب مقادیر بالا در داده‌های فضایی ارائه نشده است، در این پایان‌نامه قصد داریم با معرفی روش مرز-بالا^۱، به تشخیص این مرزها بپردازیم. در فصل بعدی، شرح این روش آورده شده است.

^۱Top-Cut

فصل ۳

روش مرز-بالا برای تشخیص مرز مقادیر بالا

۱.۳ مقدمه

برخی ذخایر معدنی، دارای توزیعی با دم راست سنگین^۱ می‌باشند، به طوری که وجود این‌گونه مقادیر بالا موجب برآوردهای ناستوار در گشتاورهای اول و دوم می‌شوند. از آن‌جا که مقادیر بالا در داده‌های طلا، به عنوان مرز شناخته می‌شوند، شناسایی این مرزها کمک بسزایی در استخراج طلا با عیار بالا می‌کند. مقادیر بالا، به طور معمول، توسط روش‌های شناسایی داده‌های پرت بدست می‌آیند. اما این روش‌ها تنها مقدار داده پرت را تشخیص می‌دهند و معیاری جهت شناسایی مرز مقادیر بالا در مشاهدات فضایی ارایه نمی‌دهند. در این پایان‌نامه، قصد داریم روشی به نام روش مرز-بالا معرفی کنیم که قادر به تشخیص مرز مقادیر بالا برای مشاهدات فضایی می‌باشد. همچنین به علت این‌که در زمین‌آمار و مهندسی معدن، داده‌های فرین^۲ نیز نوعی داده پرت محسوب می‌شوند، تمایزی بین مقادیر فرین و داده‌های پرت قائل نشده‌ایم. وجود مقادیر فرین در منابع طبیعی (فلزات گران‌بها، ماهی و غیره) یا آلاینده‌ها رایج می‌باشند. این داده‌های پرت، مقادیری هستند که به طور غیرمنتظره، بزرگ و یا کوچک بوده اما به عنوان مقادیر نادرست، از مجموعه مشاهدات کنار گذاشته نمی‌شوند. در این پایان‌نامه منظور ما از مقادیر فرین، مقادیر بالای دم راست یک توزیع چوله به راست است. در کاربردهای زیست‌محیطی، چنین مقادیر بالایی ممکن است مانند مقادیر بیش از آستانه، موجب برآوردهای ناستوار در گشتاورهای مرتبه اول و دوم (میانگین، واریانس، تغییرنگار) شوند. مدل‌هایی که مقادیر فرین را نیز در برمی‌گیرند، مختلف هستند. به عنوان مثال، در آمار، نظریه مقادیر فرین برلنت^۳ (۲۰۰۴)، مدل مناسبی از دم توزیع‌ها و پیش‌بینی اتفاق‌های نادر همچون سیل را فراهم نموده است. چنین مدل‌هایی در زمین‌آمار برای برآورد

^۱Heavy right tail

^۲Extreme Value

^۳Beirlant

مفید می‌باشند: مانند برآورد عیار الماس (کارس و رومبوتوس^۱، ۱۹۹۶). مدل دیگری به نام مرز-بالا وجود دارد که ما از آن ایده‌ای برای تشخیص مرز مقادیر بالا استفاده می‌کنیم (ریویرارد^۲، ۲۰۱۳). قبل از بیان روش مرز-بالا، به توضیح اثر مرزی^۳ و عدم وجود اثر مرزی^۴ می‌پردازیم.

۲.۳ اثر مرزی

فرض کنید تابع تصادفی $Z(s)$ بیان‌گر مقدار متغیر پاسخ در موقعیت s باشد، که به عنوان مثال مقادیر $0, 1, 2$ و 3 را اختیار کند. مجموعه‌ای مانند A_1 در شکل (۱.۳) را در نظر بگیرید، که شامل نقاطی است که مقدار آن‌ها بزرگتر یا مساوی یک است. لذا می‌توان تابع شاخص را برای این مجموعه به صورت زیر نوشت:

$$I_{s \in A_1} = I_{Z(s) \geq 1} = I$$

به طور معمول ممکن است نقاطی در مجموعه A_1 موجود باشند، که مقدار آن‌ها بزرگتر یا مساوی 2 است. اگر موقعیت نقطه مورد نظر در مجموعه A_1 ، به مرزهای A_1 نزدیک‌تر باشد، این پیشامد با شانس کمتری رخ می‌دهد. بنابراین اثری به نام اثر مرزی وجود دارد که به واسطه آن نقاط نزدیک به مرز A_1 مقادیر کمتری را نسبت به نقاط موجود در مرکز A_1 دارا هستند. در این وضعیت، مقدار میدان تصادفی $Z(0)$ در موقعیت s در مجموعه A_1 ، به این موضوع وابسته است که آیا موقعیت $s+h$ ، به مجموعه A_1 متعلق است یا خیر. به عبارت دیگر اگر مقدار احتمال شرطی زیر به h وابسته باشد، حاکی از وجود اثر مرزی است. این احتمال شرطی در مثال مذکور به صورت زیر قابل بیان است:

$$P[Z(s) \geq 2 | Z(s) \geq 1, Z(s+h) < 1] = \frac{P[Z(s) \geq 2, Z(s+h) < 1]}{P[Z(s) \geq 1, Z(s+h) < 1]}$$

به طور معمول، مقدار احتمال در فواصل کوتاه با افزایش فاصله، افزایش می‌یابد (ریویرارد، ۱۹۹۴)، یعنی اگر موقعیت در یک نقطه خارج از مجموعه A_1 واقع باشد و از فاصله کوتاهی شروع کنیم، با افزایش فاصله، احتمال این‌که موقعیت در مجموعه $Z(s) \geq 2$ واقع شود، بیشتر می‌شود.

۳.۳ عدم وجود اثر مرزی

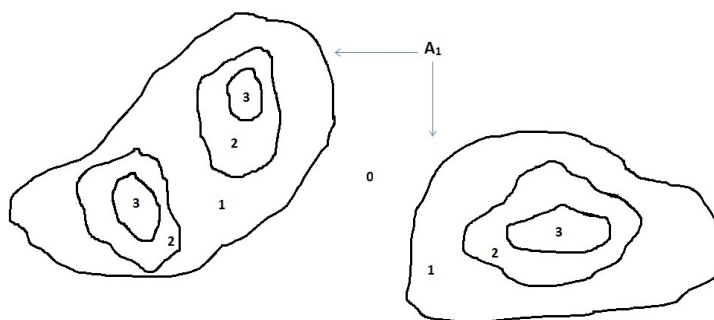
مثال بخش قبل را در نظر بگیرید که در آن تابع تصادفی $Z(s)$ مقادیر $0, 1, 2$ و 3 را در موقعیت s اختیار می‌کند. فرض کنید بر روی مجموعه‌های A_1 (نقاط شامل 1) و A_2 (نقاط شامل 2)، $Z(s) \geq 2$ ، اثر مرزی وجود نداشته باشد. در این صورت، هر موقعیت s در مجموعه A_1 که مقدار آن برابر یا بزرگتر

^۱Caers and Rombouts

^۲Rivoirard

^۳Edge effect

^۴Absence of edge effect



شکل ۱.۳: نمونه‌ای از اثر مرزی

از ۲ است، مستقل از مقادیر موقعیت‌های واقع در همسایگی $s + h$ ، متعلق به مجموعه A_1 است. به بیان دیگر شاخص $I_{Z(s) \geq 2}$ مستقل از شاخص $I_{Z(s) \geq 1}$ است. بنابراین رابطه زیر را داریم:

$$P[Z(s) \geq 2, Z(s+h) \geq 1 | Z(s) \geq 1] = P[Z(s) \geq 2 | Z(s) \geq 1] P[Z(s+h) \geq 1 | Z(s) \geq 1]$$

$$= \frac{P[Z(s) \geq 2, Z(s) \geq 1]}{P[Z(s) \geq 1]} \frac{P[Z(s+h) \geq 1, Z(s) \geq 1]}{P[Z(s) \geq 1]}$$

عبارت فوق را می‌توان به صورت زیر بازنویسی نمود

$$\underbrace{\frac{P[Z(s) \geq 2, Z(s+h) \geq 1]}{P[Z(s) \geq 1]}}_{T_{21}^h} = \frac{P[Z(s) \geq 2]}{P[Z(s) \geq 1]} \underbrace{\frac{P[Z(s+h) \geq 1, Z(s) \geq 1]}{P[Z(s) \geq 1]}}_{T_{11}^h}$$

نتیجه فوق را به صورت زیر نمادگذاری می‌کنیم

$$T_{21}^h = \frac{T_2}{T_1} T_{11}^h$$

در حالت کلی به ازای هر $j > i$ ، توزیع توام را می‌توان به صورت زیر بیان کرد

$$T_{ji}^h = \frac{T_j}{T_i} T_{ii}^h$$

نتیجه دیگری که می‌توان در صورت عدم وجود اثر مرزی گرفت، رابطه رگرسیونی بین توابع شاخص در موقعیت s است که به صورت زیر نوشته می‌شود:

$$E[I_{Z(s) \geq 2} | I_{Z(s) \geq 1}] = \frac{T_2}{T_1} I_{Z(s) \geq 1} \quad (۱.۳)$$

برای به دست آوردن رابطه (۱.۳) دو حالت وجود دارد:

۱. $I_{Z(s) \geq 1} = 0$ که در این صورت $I_{Z(s) \geq 2} = 0$ است.

۲. $I_{Z(s) \geq 1} = 1$ که در این صورت با احتمال $\frac{T_1}{T_1} = P[Z(s) \geq 2 | Z(s) \geq 1]$ ، پیشامد $I_{Z(s) \geq 2} = 1$ رخ می‌دهد، در غیر این صورت $I_{Z(s) \geq 2}$ برابر صفر است. بنابراین در این حالت میانگین برابر $\frac{T_2}{T_1}$ است.

در نتیجه امید ریاضی شرطی (۱.۳) در حالتی که $I_{Z(s) \geq 1} = 1$ است، برابر $\frac{T_2}{T_1}$ است.

۴.۳ روش مرز-بالا

در معادن طلا، بخش استخراج شده به میانگین عیار بالا بستگی دارد که این میانگین مربوط به دم توزیع مقادیر است. در این روش، متغیر پاسخ Z به سه بخش مقدار بریده شده^۱، شاخص وزنی بالای مقدار مرز-بالا و باقی مانده، تقسیم می‌شود. در این چارچوب، باقی مانده ساختار ضعیفی دارد و در بیشتر موارد، تغییرنگار آن اثر قطعه‌ای تام را نشان می‌دهد. همچنین باقی مانده هیچ همبستگی فضایی با مقدار بریده شده و شاخص بالای مقدار مرز-بالا ندارد. این تجزیه، مطالعه واریوگرافی استوارتری را ایجاد می‌کند، زیرا متغیرها (شاخص و مقدار بریده شده) مقادیر بالا را در بر ندارند.

۱.۴.۳ مقدار بریده شده

با فرض این که $Z(s)$ بیانگر مقدار متغیر پاسخ در موقعیت s و z_e مقدار آستانه باشند، مقدار بریده شده را با $Z(s) \wedge z_e$ نشان می‌دهند، به طوری که

$$Z(s) \wedge z_e = \begin{cases} Z(s), & Z(s) < z_e, \\ z_e, & Z(s) \geq z_e. \end{cases} \quad (2.3)$$

همچنین مقدار مازاد^۲ (بخش حذف شده)، با نماد $B_{z_e}(s)$ نشان داده و به صورت زیر تعریف می‌شود:

$$B_{z_e}(s) = [Z(s) - z_e] I_{Z(s) \geq z_e}.$$

یعنی اگر $Z(s)$ کمتر از z_e باشد، برابر صفر است، در غیر این صورت برابر $Z(s) - z_e$ می‌باشد. مفاهیم به کار گرفته شده در روش مرز-بالا، ریشه در اقتصاد معدن دارند. مقدار z_e ، عیار حد اقتصادی^۳ نامیده می‌شود. عیار حد، عیاری است که در آن درآمدها (ارزش محصول) با هزینه‌های معدن کاری و فرآوری آن برابر می‌شود. در نتیجه، این عیار مازاد است که سود اقتصادی را به همراه دارد. به همین دلیل، سود مرسوم در زمین آمار به ازای عیار حد z_e در موقعیت s برابر خواهد بود با

$$b(z_e) = E[B_{z_e}(s)] = E\{[Z(s) - z_e] I_{Z(s) \geq z_e}\} = [m(z_e) - z_e] T(z_e),$$

که در آن $T(z_e) = P\{Z(s) \geq z_e\}$ و $m(z_e) = E[Z(s) | Z(s) \geq z_e]$. بنابراین، اگر عیار کانسار کوچکتر از عیار حد باشد، این بخش استخراج نمی‌شود و سودی از آن حاصل نمی‌گردد.

^۱Truncated grade

^۲Excess

^۳Cutoff

بنابر تعاریف فوق می‌توان متغیر اصلی $Z(s)$ را به دو قسمت مقدار بریده‌شده و مقدار مازاد تقسیم‌بندی نمود:

$$Z(s) = Z(s) \wedge z_e + [Z(s) - z_e]I_{Z(s) \geq z_e}. \quad (۳.۳)$$

بین این دو قسمت همبستگی وجود دارد.

۲.۴.۳ مدل مرز-بالا

هنگامی که مقدار بریده‌شده کمتر از z_e است، مقدار مازاد برابر صفر و اگر مقدار آن برابر با z_e باشد، مقدار مازاد برابر با $Z(s) - z_e$ است. بنابراین رابطه رگرسیونی بین مازاد و مقدار بریده‌شده به صورت زیر بیان می‌شود:

$$\begin{aligned} E[B_{z_e}(s)|Z(s) \wedge z_e] &= E[Z(s) - z_e|Z(s) \wedge z_e = z_e]I\{Z(s) \wedge z_e = z_e\}, \\ &= E[Z(s) - z_e|Z(s) \geq z_e]I_{Z(s) \geq z_e}, \\ &= [m(z_e) - z_e]I_{Z(s) \geq z_e}. \end{aligned}$$

در این رگرسیون، باقی‌مانده عبارت است از

$$\begin{aligned} R_{z_e}(s) &= B_{z_e} - [m(z_e) - z_e]I_{Z(s) \geq z_e}, \\ &= [Z(s) - m(z_e)]I_{Z(s) \geq z_e}. \end{aligned} \quad (۴.۳)$$

همانند تعریف متداول مدل‌های رگرسیونی، می‌توان متغیر پاسخ را به صورت مجموع مقدار بریده‌شده، شاخص وزنی و باقی‌مانده تقسیم نمود.

$$Z(s) = [Z(s) \wedge z_e] + [m(z_e) - z_e]I_{Z(s) \geq z_e} + R_{z_e}(s). \quad (۵.۳)$$

وزن $[m(z_e) - z_e] = E[Z(s) - z_e|Z(s) \geq z_e]$ بیان‌گر میانگین مقدار مازاد است، هنگامی که مقدار متغیر بیشتر از مقدار آستانه باشد. بین مقدار بریده‌شده و شاخص وزنی نیز همبستگی وجود دارد، اما باقی‌مانده هیچ همبستگی با شاخص و مقدار بریده‌شده در موقعیت s ندارد. تجزیه (۵.۳) تا زمانی جالب است که باقی‌مانده، هیچ همبستگی فضایی با شاخص یا مقدار بریده‌شده نداشته باشد. این همان مدل مرز-بالا است.

۳.۴.۳ فرضیه‌های مدل مرز-بالا

فرض کنید $A(z_e)$ مجموعه موقعیت‌هایی از $s \in D$ باشد که در آن $Z(s) \geq z_e$. بنابراین $A(z_e)$ شامل مقادیر بالایی است. اگر رفتار مقادیر بالا در مجموعه خود، از شکل هندسی این مجموعه و مقادیر پایین $Z(s)$ تاثیر نپذیرد، باقی‌مانده‌ها، مستقل از شاخص و متغیر بریده‌شده خواهند بود. در این صورت، رفتار مقادیر بالا همچون یک فرآیند مستقل در مجموعه $A(z_e)$ دیده شده و هیچ اثر مرزی را نشان نخواهد داد.

عدم وجود هرگونه همبستگی فضایی بین باقی مانده‌ها از یک طرف و شاخص و مقدار بریده‌شده از طرف دیگر، از شرط‌های موجود در مدل مرز-بالا است. در ابتدا به عدم وجود همبستگی فضایی بین باقی مانده‌ها و شاخص می‌پردازیم. برای هر بردار h داریم

$$\text{Cov}[R_{z_e}(s)I_{Z(s+h) \geq z_e}] = 0.$$

از آن جایی که $E[R_{z_e}(s)] = 0$ ، داریم

$$E[R_{z_e}(s)I_{Z(s+h) \geq z_e}] = 0.$$

در نتیجه

$$\begin{aligned} & E[(Z(s) - m(z_e))I_{Z(s) \geq z_e}I_{Z(s+h) \geq z_e}] \\ &= E[Z(s)|Z(s) \geq z_e, Z(s+h) \geq z_e] - E[E(Z(s)|Z(s) \geq z_e)|Z(s) \geq z_e, Z(s+h) \geq z_e] \\ &= E[Z(s)|Z(s) \geq z_e, Z(s+h) \geq z_e] - E(Z(s)|Z(s) \geq z_e) = 0. \end{aligned}$$

بنابراین

$$E[Z(s)|Z(s) \geq z_e, Z(s+h) \geq z_e] = m(z_e) = E[Z(s)|Z(s) \geq z_e]. \quad (6.3)$$

بررسی رابطه (۶.۳) بر روی داده‌ها به علت وابستگی آن به مقادیر پرت، دشوار است. از این رو، شرط عدم وجود هرگونه همبستگی فضایی بین باقی مانده‌ها و شاخص، ناشی از عدم وجود اثر مرزی در مجموعه $A(z_e)$ است. در این صورت $Z(s)$ ($s \in A(z_e)$) مستقل از $Z(s+h)$ است بدین معنا که، مقدار $Z(s)$ مستقل از این موضوع است، که آیا موقعیت $s+h$ به مجموعه $A(z_e)$ متعلق است یا خیر. بنابراین رابطه (۶.۳) در صورتی صادق است که برای هر $z' > z_e$ داشته باشیم

$$P[Z(s) \geq z'|Z(s) \geq z_e, Z(s+h) \geq z_e] = P[Z(s) \geq z'|Z(s) \geq z_e], \quad (7.3)$$

که هم ارز است با

$$P[Z(s) \geq z'|Z(s) \geq z_e, Z(s+h) < z_e] = P[Z(s) \geq z'|Z(s) \geq z_e]. \quad (8.3)$$

از روابط (۷.۳) و (۸.۳) این نتیجه حاصل می‌شود که در صورت عدم وجود اثر مرزی، احتمال شرطی $P[Z(s) \geq z'|Z(s) \geq z_e, Z(s+h) < z_e]$ ، مستقل از فاصله h است.

حال به عدم وجود همبستگی فضایی بین باقی مانده‌ها و مقدار بریده‌شده در مدل (۵.۳) می‌پردازیم. با توجه به این که $E[R_{z_e}(s)] = 0$ ، کوواریانس بین باقی مانده‌ها و مقدار بریده‌شده برابر است با

$$E[R_{z_e}(s)\{Z(s+h)I_{Z(s+h) < z_e} + z_e I_{Z(s+h) \geq z_e}\}] = 0.$$

بنابراین

$$E[(R_{z_e}(s) \times Z(s+h))I_{Z(s+h) < z_e} + z_e E[R_{z_e}(s)I_{Z(s+h) \geq z_e}]] = 0.$$

چون $E[R_{z_e}(s)] = 0$ ، در نتیجه

$$E[(Z(s) - m(z_e))Z(x+h)|Z(s) \geq z_e, Z(s+h) < z_e] = 0.$$

این رابطه در صورتی صادق است که برای $s \in A(Z_e)$ و $s+h \notin A(Z_e)$ خارج از $A(z_e)$ ، $Z(s)$ مستقل از $Z(s+h)$ باشد. این دلالت بر عدم وجود اثر مرزی برای مقادیر بالای z_e دارد که فرضیه اساسی مدل مرز-بالا است. با توجه به این که در صورت عدم وجود اثر مرزی، احتمال شرطی بیان شده مستقل از h است، هدف آن است که احتمال شرطی $P[Z(s) \geq z'|Z(s) \geq z_e, Z(s+h) < z_e]$ را بتوان بر اساس نسبتی بین تغییرنگار متقابل شاخص^۱ مرزهای z' و z_e ، و تغییرنگار شاخص^۲ مرز z_e بنویسیم. بر اساس تعریف احتمال شرطی، رابطه (۷.۳) را به صورت زیر می‌نویسیم:

$$P[Z(s) \geq z'|Z(s) \geq z_e, Z(s+h) < z_e] = \frac{P[Z(s) \geq z', Z(s+h) < z_e]}{P[Z(s) \geq z_e, Z(s+h) < z_e]}.$$

نسبت فوق را می‌توان به کمک تغییرنگارهای شاخص نشان داد. تغییرنگار شاخص $I_{Z(s) \geq z_e}$ برابر است با

$$\gamma_{I_{z_e}}(h) = \frac{1}{V} E[I_{Z(s) \geq z_e} - I_{Z(s+h) \geq z_e}]^2.$$

امید ریاضی فوق در صورتی برابر با یک است که مقدار فقط یکی از شاخص‌های فوق برابر یک باشد. بنابراین مقدار تغییرنگار شاخص برابر است با

$$\gamma_{I_{z_e}}(h) = \frac{1}{V} \{P[Z(s) \geq z_e, Z(s+h) < z_e] + P[Z(s+h) \geq z_e, Z(s) < z_e]\}.$$

با فرض برابری این احتمال‌ها داریم

$$\gamma_{I_{z_e}}(h) = P[Z(s) \geq z_e, Z(s+h) < z_e].$$

به همین ترتیب نیز تغییرنگار شاخص متقابل مرزهای z' و z_e به صورت زیر بیان می‌شود:

$$\begin{aligned} \gamma_{I_{z', z_e}}(h) &= \frac{1}{V} E[(I_{Z(s) \geq z'} - I_{Z(s+h) \geq z'}) \cdot (I_{Z(s) \geq z_e} - I_{Z(s+h) \geq z_e})] \\ &= P[Z(s) \geq z', Z(s+h) < z_e]. \end{aligned}$$

در نتیجه احتمال شرطی را می‌توان به صورت نسبت زیر نوشت:

$$P[Z(s) \geq z'|Z(s) \geq z_e, Z(s+h) < z_e] = \frac{\gamma_{I_{z', z_e}}(h)}{\gamma_{I_{z_e}}(h)}. \quad (9.3)$$

بنابراین می‌توان به کمک تغییرنگارهای شاخص، وجود اثر مرزی را بررسی نمود. در حالت کلی، مقدار نسبت (۹.۳) به‌ازای مقادیر کوچک h ، با افزایش فاصله، افزایش می‌یابد. به بیان دیگر، برای رسیدن از یک مقدار کوچک ($Z(s) < z_e$) به یک مقدار بزرگ ($Z(s) \geq z'$)، میل به عبور از مقادیر حد واسط (حوالی z_e) وجود دارد (مانند شکل (۱.۳)). نسبت (۹.۳) نشان می‌دهد که چرا نسبت تغییرنگارهای

^۱Indicator cross variogram

^۲Indicator variogram

شاخص، برای تعیین مقادیر پرت فضایی کاربرد دارد. در بررسی مقادیر پرت فضایی به دنبال مقادیری هستیم که در فضا به صورت ناگهانی واقع شده‌اند. به طور مثال، مقادیر بزرگ که در کنار مقادیر کوچک قرار گرفته‌اند، اگرچه مقادیر معمول مجموعه‌اند ولی از لحاظ فضایی پرت هستند، زیرا هنگامی که این نسبت، مستقل از فاصله h است (عدم وجود اثر مرزی)، انتظار داریم که این تغییرات برای فاصله h ، به صورت تدریجی باشد و نه ناگهانی. یعنی میل کلی برای عبور از مقادیر کمتر از z_e به مقادیر بیشتر از z' بر این نیست که از مقادیر حد واسط z_e تا z' عبور کنیم و این تغییرات در فضا به صورت تقریباً ناگهانی انجام می‌شود.

۴.۴.۳ انتخاب مقدار مرز-بالا

تاکنون فرض بر این بوده که مقدار مرزی، از پیش تعیین شده است. اما چون اثر مرزی بالاتر از مقدار مرزی وجود ندارد، انتخاب مرز با اهمیت‌تر از بررسی عدم وجود اثر مرزی است. مرزها از مقادیری انتخاب می‌شوند که بر روی دامنه توزیع $Z(s)$ گسترده شده‌اند. از احتمال شرطی (۹.۳) به منظور تشخیص این که مرز (Z_{min})، در صورت عدم وجود اثر مرزی در مجموعه مربوطه، قابل قبول است یا خیر، استفاده می‌شود. در این صورت، کران پایینی برای انتخاب مقدار مرز-بالا به دست می‌آید. هنگامی که اثر مرزی وجود نداشته باشد، برای هر $z' > z$ ، با استفاده از امید شرطی (۱.۳) رابطه زیر حاصل می‌شود:

$$I\{Z(s) \geq z'\} = \frac{T(z')}{T(z)} I_{Z(s) \geq z} + IR_{z,z'},$$

که در آن $IR_{z,z'}$ باقی‌مانده رگرسیون بین $I_{Z(s) \geq z'}$ و $I_{Z(s) \geq z}$ است. از آن جایی که هیچ همبستگی فضایی بین باقی‌مانده شاخص و $I_{Z(s) \geq z}$ وجود ندارد (ریویارد، ۱۹۹۴)، لذا رابطه فوق را می‌توان برای تغییرنگارهای شاخص به صورت زیر نتیجه گرفت:

$$\gamma_{z'}(h) = \gamma_z(h) \frac{(T(z'))^2}{(T(z))^2} + \gamma_{IR_{z,z'}}(h).$$

بنابراین تغییرنگار باقی‌مانده شاخص^۱ ($\gamma_{IR}(h)$)، بخشی از تغییرات $I_{Z(s) \geq z'}$ را که از $I_{Z(s) \geq z}$ تاثیر نمی‌پذیرند، بلکه بیان‌گر رفتار $I_{Z(s) \geq z'}$ در مجموعه $A(z)$ است، را نشان می‌دهد. باقی‌مانده‌های شاخص (IR_s) و تغییرنگارشان را می‌توان مرتباً از هر مرز به مرز بالاتر با شروع از Z_{min} محاسبه کرد (ماترون، ۱۹۸۲). بالاترین باقی‌مانده شاخص (IR_s) ساختار ضعیفی است و تغییرنگار آن (γ_{IR}) اثر قطعه‌ای تام را نشان می‌دهد. بنابراین در این حالت کران بالای مرز (Z_{max}) تعیین می‌شود.

به طور خلاصه مرز-بالا بین Z_{min} و Z_{max} تعیین می‌شود. بنابراین هیچ اثر مرزی در مجموعه مقادیر بالا وجود ندارد. بیشتر اوقات وقتی اثر مرزی وجود نداشته باشد، تغییرپذیری باقی‌مانده، قطعه‌ای تام است. در این صورت مقدار Z_{min} و Z_{max} برابر بوده و مقدار مرز-بالا را نشان می‌دهد.

برای نشان دادن اهمیت تشخیص مرز مقادیر بالا در ذخایر معدنی، به کاربرد روش مرز-بالا در یکی از ذخایر طلا که توسط ریویارد (۲۰۱۳) انجام شده است می‌پردازیم. مدل مرز-بالا، برای داده‌های چال‌های

^۱Indicator residual variogram

انفجاری، بر روی رگه شمالی - جنوبی در معدن طلای روباز درحال استخراج، اعمال شده است. این مجموعه داده حاوی ۸۷۰۶ داده عیار طلا است، که از ۴۹۳ چال انفجاری، در شبکه $5 \times 2/5$ متر، استخراج شده است. شاخص آماری برای داده‌ها در حجم پایه ۱ متر، مطابق جدول (۱۰۳) است. لازم به ذکر است که چال انفجاری همانند گمانه اکتشافی است اما نمونه بدست آمده، مغزه نیست بلکه به صورت پودر می‌باشد و بر اساس قانون نمونه‌برداری که بستگی به نوع ماده معدنی دارد، متراژ نمونه‌برداری انجام می‌شود (معمولا ۱ متر و آهک ۱۰ متر). اغلب شبکه چال انفجاری دوبعدی در نظر گرفته می‌شود و حجم پایه، به حجم نمونه برداشت شده اطلاق می‌گردد.

جدول ۱۰۳: شاخص‌های آماری عیار طلا

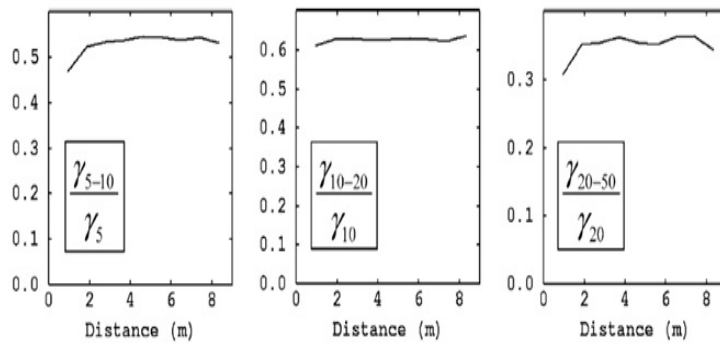
تعداد	کمیته	بیشینه	میانگین	واریانس	ضریب تغییرات
۸/۷۰۶	۰	۴۴۳	۱/۷۶	۱۸۵	۷/۷۴

ضرورت تشخیص مرز مقادیر بالا را می‌توان در جدول (۲۰۳) که مربوط به شاخص‌های آماری مقادیر بالا و کمتر از مرزهای مختلف است، مشاهده کرد. در این جدول، مقادیر بالاتر از 150 g/t فقط $24/0$ درصد از نمونه‌های کانسار را در برمی‌گیرند که از لحاظ تعداد به این معنا است که $\frac{24}{1000}$ از نمونه‌ها بالاتر از 150 g/t هستند که این مقادیر، به علت مشکلی که ممکن است در برآورد ایجاد کنند، کنار گذاشته می‌شوند. شایان ذکر است که اگرچه تعداد این مقادیر کم است، اما 32% محتوی فلز کانسار را تشکیل داده‌اند. همان‌طور که بیان شد در ذخایر معدنی، تشخیص مرز مقادیر بالا اهمیت ویژه‌ای در استخراج دارد.

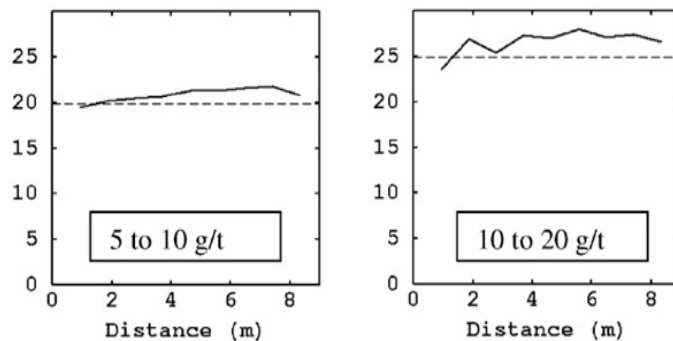
در شکل (۲۰۳) نمودارهای نسبت تغییرنگارهای شاخص برای مرزهای 20 و 10 و همچنین در شکل (۳۰۳)، تغییرنگار باقی‌مانده شاخص‌های 10 - 5 و 20 - 10 نشان داده شده‌اند. در صورتی که احتمال شرطی (نسبت تغییرنگارهای شاخص) نسبت به فاصله ثابت باشد، حاکی از عدم وجود اثر مرزی است. با توجه به شکل (۲۰۳)، نمودار مربوط به آستانه 5 g/t نسبت به فاصله ثابت است و همچنین تغییرنگار باقی‌مانده شاخص 10 - 5 اثر قطعه‌ای تام را نشان می‌دهد. بنابراین مرز مقادیر بالا برای عیار طلا، در این مثال کاربردی برابر 5 g/t است (ریویرارد، ۲۰۱۳).

جدول ۲.۳: شاخص‌های آماری مقادیر بالاتر و کمتر از مرزهای مختلف

مقادیر مرز	۵	۱۰	۲۰	۵۰	۱۰۰	۱۵۰	
مقادیر بالاتر از مرز	تناژ%	۴/۶۷	۲/۴۴	۱/۵۰	۰/۵۲	۰/۲۴	
	فلز%	۷۸/۱۶	۶۹/۳۱	۶۱/۸۸	۴۴/۵۲	۳۵/۶۲	
	میانگین g/t	۲۹/۳۵	۴۹/۹۷	۷۲/۲۱	۱۵۱/۲۳	۲۱۷/۸۱	۲۳۵/۶۷
مقادیر کمتر از مرز	میانگین g/t	۰/۶۲	۰/۷۸	۰/۹۷	۱/۲۳	۱/۴۲	
	واریانس	۱/۴۹	۳/۶۵	۸/۸۰	۲۵/۱۹	۵۱/۶۱	
	ضریب تغییرات	۱/۹۸	۲/۴۴	۳/۰۶	۴/۰۷	۵/۰۷	



شکل ۲.۳: نسبت واریوگرام‌های شاخص برای مرزهای ۵، ۱۰ و ۲۰



شکل ۳.۳: تغییرنگارهای باقی‌مانده‌های شاخص برای ۱۰ - ۲۰ g/t و ۵ - ۱۰ g/t

فصل ۴

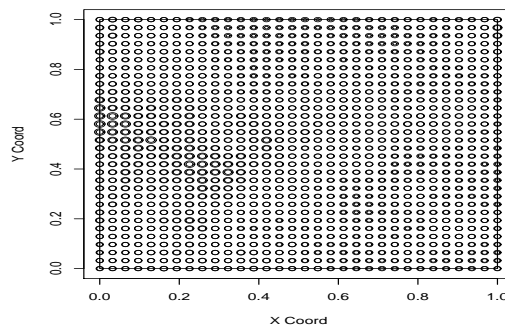
شبیه‌سازی و تحلیل اکتشافی داده‌های پروژه زرکوه

در این فصل با اجرای یک مطالعه شبیه‌سازی، عملکرد برخی روش‌های تشخیص داده پرت، که در فصل دوم و همچنین روش مرز-بالا، که در فصل سوم بیان شده است، را مورد بررسی قرار داده‌ایم. در آخر از روش مرز-بالا، به منظور تشخیص مرز عیار بالای معدن ساری گونای پرداخته‌ایم.

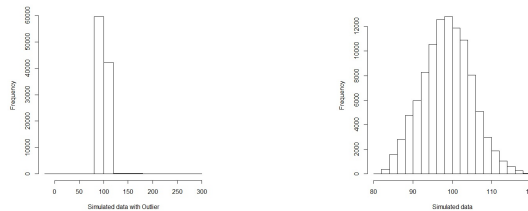
۱.۴ مطالعه شبیه‌سازی

برای مطالعه شبیه‌سازی، ابتدا یک میدان تصادفی گوسی را در یک شبکه منظم تولید کرده‌ایم. این کار، با استفاده از تابع grf در بسته $geoR$ در محیط $R - 2.15.1$ انجام شده است. تعداد تکرارهای شبیه‌سازی را که در موقعیت‌های یکسان انجام شده، برابر ۱۰۰ اختیار کرده‌ایم. در هر تکرار ۱۰۰۰ نمونه با واریانس ۴۰ و میانگین ۱۰۰ تولید شده است. شبکه منظم داده‌های شبیه‌سازی شده، در شکل (۱.۴)، و بافت‌نگار آن، در قسمت الف شکل (۲.۴) نشان داده شده است. در هر تکرار به تعداد ۱٪ کل داده‌ها، داده پرت تولید شد. تولید داده‌های پرت به گونه‌ای صورت پذیرفت که تا حد امکان مقادیر بالا تولید شوند و همچنین توزیع داده‌ها بعد از ایجاد داده‌های پرت، به توزیعی با چولگی مثبت تبدیل شود، که بافت‌نگار آن در قسمت شکل (۲.۴) آورده شده است. به همین منظور، به‌طور تصادفی، ۱۰ عدد از نمونه‌های دهک هشتم به بالا را با توزیع نرمالی با میانگین ۱۳۰ و انحراف معیار ۴۰ جابه‌جا نموده‌ایم.

برای بررسی عملکرد روش‌های تشخیص داده پرت، از برخی معیارهای رده‌بندی استفاده نموده‌ایم. در این مطالعه، تعداد رده‌ها ۲ است، به‌طوری که رده یک، مربوط به مشاهداتی است که داده پرت هستند (P) و رده صفر، مربوط به مشاهداتی است که داده پرت نیستند (N). نتیجه رده‌بندی با هر روش ممکن، می‌تواند توسط جدول (۱.۴) بیان شود که در آن TP بیان‌گر تعداد داده‌های پرتی است که



شکل ۱.۴: شبکه منظم داده‌های شبیه‌سازی شده



(ب)

(الف)

شکل ۲.۴: الف: بافت‌نگار نمونه‌های شبیه‌سازی شده ب: بافت‌نگار نمونه‌های شبیه‌سازی شده آلوده به داده‌های پرت

به‌درستی به‌عنوان داده پرت انتخاب شده‌اند؛ FP تعداد داده‌هایی است که به‌عنوان داده پرت انتخاب شده، در صورتی‌که داده پرت واقعی نبوده‌اند. FN تعداد داده پرتی است که داده پرت، تشخیص داده نشده‌اند. TN تعداد مشاهداتی است که داده پرت نبوده‌اند و به‌درستی به‌عنوان داده پرت نیز تشخیص داده نشده‌اند.

جدول ۱.۴: حالات مختلف ایجاد شده برای تشخیص نقاط پرت

مشاهده معمولی	مشاهده پرت	
FP	TP	تشخیص پرت بودن با روش
TN	FN	تشخیص پرت نبودن با روش

با تعریف این کمیت‌ها، برخی از معیارهای ارزیابی عملکرد به صورت زیر قابل تعریف هستند:

$$Recall = TPrate = \frac{TP}{TP + FN},$$

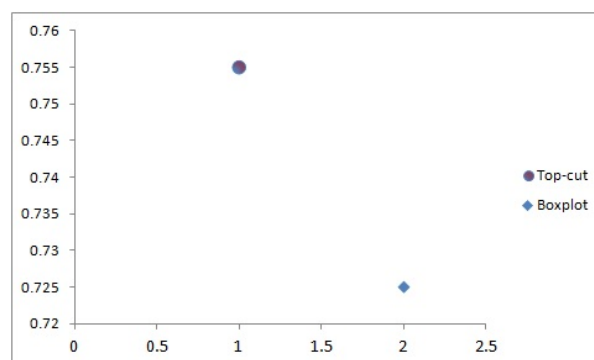
$$Precision = \frac{TR}{TP + FP},$$

$$FNrate = \frac{FN}{TP + FN}.$$

مقادیر بزرگ معیارهای $TPrate$ و $Precision$ و مقادیر کوچک معیار $FNrate$ مبین مناسب‌تر بودن روش رده‌بندی است.

۲.۴ ارزیابی روش‌های تشخیص مرز مقادیر بالا

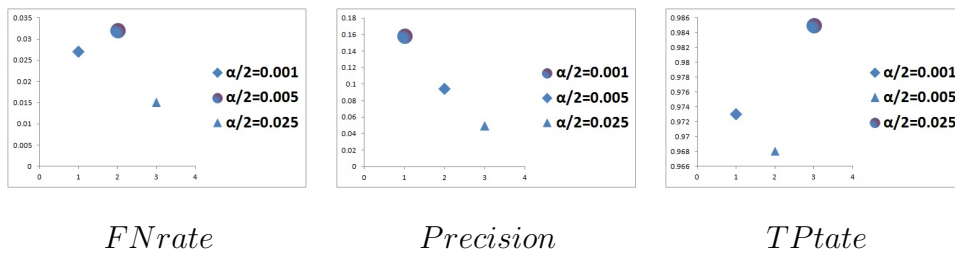
در روش نمودار جعبه‌ای، مرز مقادیر بالا (حدود آستانه‌ای برای پرت بودن یک داده) برابر با $1/5$ برابر دامنه میان چارکی بالای نمودار جعبه‌ای، انتخاب می‌شود. این روش در پروژه تخمین ذخیره طلای ساری گونی، که توسط شرکت خدمات اکتشافی کشور-خدمات سسکو، برای تشخیص حدود آستانه‌ای پرت بودن استفاده شده است. به همین منظور در مطالعه شبیه‌سازی انجام‌شده، برای هر تکرار، تعداد نمونه‌هایی که بیشتر از مقدار $1/5$ برابر دامنه میان چارکی بالای نمودار جعبه‌ای، و در اصل همان داده‌های پرت ایجادشده در شبیه‌سازی است، را محاسبه نمود ($TP_{Boxplot}$) و همچنین با اجرای روش مرز-بالا نیز تعداد نمونه‌هایی که بیشتر از مرز انتخاب شده‌اند و در عین حال به‌عنوان داده پرت واقعی به شمار می‌آیند را محاسبه کرده‌ایم ($TP_{Top-Cut}$). با انتخاب معیار $TPrate$ برای بررسی ارزیابی عملکرد این دو روش؛ مقادیر آن در نمودار (۳.۴) نشان داده شده است. با توجه به شکل، روش مرز-بالا مناسب‌تر است.



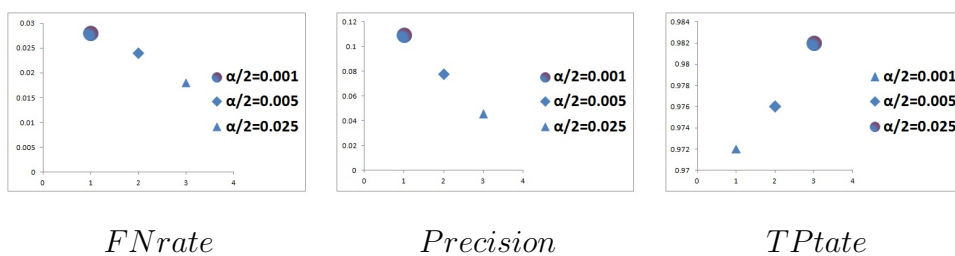
شکل ۳.۴: معیار ارزیابی $TPrate$ برای روش‌های نمودار جعبه‌ای و مرز-بالا جهت تشخیص مرز مقادیر بالا

۳.۴ ارزیابی عملکرد روش‌های تشخیص مقدار داده پرت فضایی

در این بخش در ابتدا به بررسی عملکرد الگوریتم میانه، برای تشخیص داده پرت فضایی، در سطح اطمینان و k نزدیکترین همسایگی‌های مختلف می‌پردازیم و سپس در حالات مختلف مناسب‌ترین را انتخاب نموده و با روش میانه-میانگین مقایسه می‌کنیم. برای این منظور نتایج الگوریتم میانه به ازای $\alpha = 0.001, 0.005, 0.025$ و $k = 2, 8$ مورد بررسی قرار داده‌ایم. نتایج در شکل‌های (۴.۴) تا (۸.۴) نشان داده شده‌اند. بنا بر نتایج حاصل، با ثابت نگه داشتن k ، بر اساس معیار $Precision$ ، با کاهش α الگوریتم میانه برای تشخیص داده پرت مناسب‌تر است و هنگامی که α ثابت است بر اساس معیار $Precision$ با کاهش k این الگوریتم مناسب‌تر است. سرانجام برای هر معیار مناسب‌ترین، الگوریتم میانه را انتخاب نموده‌ایم و با روش میانه-میانگین (شکل (۹.۴)) مقایسه کرده‌ایم. بنا بر معیارهای $FNrate$ و $TPrate$ ، الگوریتم میانه روش مناسب برای تشخیص داده پرت است و بنا بر معیار $Precision$ روش میانه-میانگین روش مناسب می‌باشد.

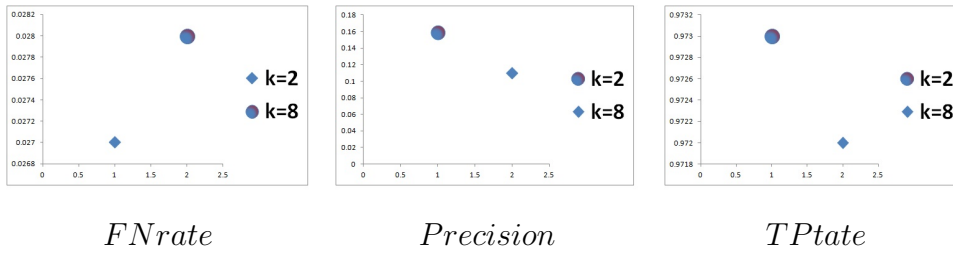


شکل ۴.۴: نمودارهای معیارهای سنجش برای الگوریتم میانه در $k = 2$

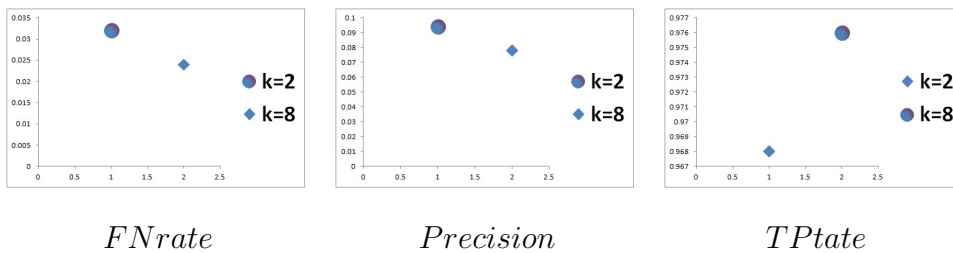


شکل ۵.۴: نمودارهای معیارهای سنجش برای الگوریتم میانه در $k = 8$

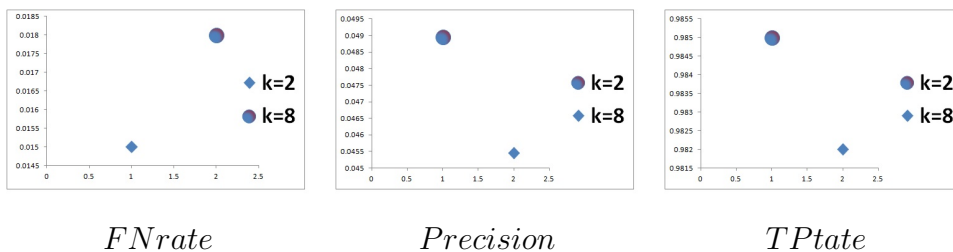
لازم به ذکر است بزرگ بودن مقدار $Precision$ ، بیانگر این موضوع است که تعداد مشاهدات کمتری، به اشتباه به عنوان داده پرت انتخاب شده است. بنا به شکل (۹.۴)، اختلاف زیاد معیار $Precision$ ، بین روش میانه-میانگین و الگوریتم میانه، مبین این است که الگوریتم میانه نسبت به روش میانه-میانگین، مشاهدات زیادی را که داده پرت نیستند به عنوان داده پرت انتخاب نموده است. همچنین با توجه به



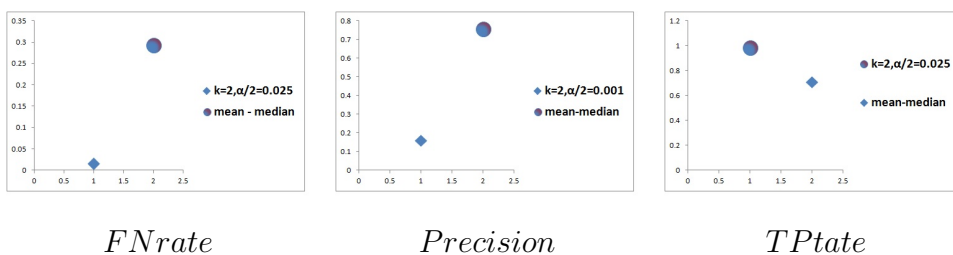
شکل ۶.۴: نمودارهای معیارهای سنجش برای الگوریتم میانه در $\alpha/2 = 0.01$



شکل ۷.۴: نمودارهای معیارهای سنجش برای الگوریتم میانه در $\alpha/2 = 0.05$



شکل ۸.۴: نمودارهای معیارهای سنجش برای الگوریتم میانه در $\alpha/2 = 0.25$



شکل ۹.۴: نمودارهای معیارهای سنجش برای الگوریتم میانه و روش میانه میانگین

اختلاف کم‌تر این دو روش در معیارهای $FNrate$ و $TPrate$ ، نسبت به معیار $Precision$ ، می‌توان نتیجه گرفت روش میانه-میانگین، با این‌که به خوبی الگوریتم میانه داده پرت واقعی را انتخاب نکرده است (بنا به معیارهای $FNrate$ و $TPrate$) ولی روش مناسب‌تری نسبت به الگوریتم میانه است.

۴.۴ تحلیل اکتشافی داده‌های زرکوه

در این بخش قصد داریم، به تحلیل اکتشافی داده‌های پروژه زرکوه بپردازیم. منظور از تحلیل اکتشافی این است، که با توجه به مفروضاتی که داریم، به بررسی ویژگی‌ها و خصوصیات مجموعه داده‌ها بپردازیم و سپس به تناسب این ویژگی‌ها و هدفی که داریم، روش آماری را برای تحلیل داده‌های مورد استفاده قرار دهیم. قبل از پرداخت به این موضوع ابتدا به پیشینه کانسار طلای زرکوه می‌پردازیم.

۵.۴ ساری‌گونای

ساری‌گونای، معدن طلای مرتبط با توده‌های نیمه نفوذی طلای اپی ترمال تشکیل شده، در میوسن میانی به همراه کمپلکس‌های نفوذی پتاسیم بالا در مرکز شمال غربی ایران در ناحیه کردستان و ۶۰ کیلومتری شمال غرب شهر همدان با مختصات جغرافیایی ۱۲ ۳۵° شمالی و ۵ ۴۸° شرقی واقع شده است (برای مطالعه بیشتر به پیوست ۱۰.۴ مراجعه شود).



شکل ۱۰.۴: موقعیت کانسار طلای ساری‌گونای در ایران

۶.۴ تحلیل اکتشافی داده مورد مطالعه

داده مورد مطالعه، مربوط به پروژه زرکوه، شامل ۸۳ گمانه و تعداد نمونه‌ها در مجموع ۳۴۵۵ می‌باشد که برای هر نمونه طول و عرض جغرافیایی و همچنین ارتفاع آن ثبت شده است. داده‌ها در طول ۵ متر کامپوزیت شده است. منظور از کامپوزیت نمودن، تقسیم بندی گمانه‌ها به بخش‌هایی با بعد یکسان و عیاری میانگین است. عیار هر قطعه کامپوزیت، میانگین وزن دار عیارهای نمونه‌های موجود در آن قطعه در طول گمانه است. روش‌های مختلفی جهت کامپوزیت نمودن گمانه‌های اکتشافی وجود دارد. یکی از این روش‌ها، کامپوزیت بر مبنای ارتفاع پله است. کامپوزیت بر مبنای ارتفاع پله این مزیت را دارد که

در نهایت، تخصیص عیار به هر بلوک و یا پله استخراجی، و در نتیجه طراحی و برنامه‌ریزی معدن ساده‌تر خواهد بود. اما روشی که در کامپوزیت نمودن گمانه‌های داده‌های مزبور، به کار گرفته شده است، روش کامپوزیت در طول گمانه است. این روش به گونه‌ای است که در هر طول مشخص (به‌عنوان مثال ۵ متر) در طول گمانه، میانگین نمونه‌هایی که در داخل این بازه قرار می‌گیرند محاسبه می‌شود. این میانگین‌گیری وزن‌دار بوده و وزن هر عیار، طول مشخص شده برای آن عیار است. استفاده از این روش به دلیل آن است که بیشتر گمانه‌های حفر شده در ذخیره ساری گونای مایل بوده و استفاده از روش کامپوزیت کردن قائم سبب بیش از حد یکنواخت شدن تغییرات عیاری می‌شود. از آنجایی که طول نمونه‌ها به طور عمده ۱ متر است، انتخاب ارتفاع ۱ متر برای کامپوزیت سازی مناسب نیست. بنابراین گزینه‌های بعدی یعنی طول‌های ۵ و ۱۰ متر مورد بررسی قرار گرفت. انتخاب طول ۱۰ متر باعث خواهد شد تا در هر کامپوزیت، جهت میانگین‌گیری عیار از ۱۰ نمونه ۱ متری استفاده شود و به دلیل اینکه عیار طلا در ابعاد کوچک نوسانات بالایی دارد این انتخاب باعث هموارسازی بیش از حد نمونه‌ها و در نتیجه تخمین خواهد شد. انتخاب طول ۲.۵ متر نیز از لحاظ شاخص‌های آماری تفاوت چندانی با طول ۵ متر ندارد و به دلیل اینکه انتخاب طول بزرگ‌تر نسبت به ابعاد کوچک‌تر از حیث کمتر نمودن پیچیدگی و حجم محاسبات مزیت دارد، طول ۵ متر مبنای کامپوزیت سازی انتخاب می‌شود.

۱.۶.۴ بررسی وجود همبستگی فضایی در داده‌ها

برای تحلیل اکتشافی داده‌ها، ابتدا به بررسی وجود همبستگی فضایی در داده‌ها می‌پردازیم. برای این کار از آزمون I -موران استفاده می‌کنیم، در این آزمون، فرض H_0 ، نبود همبستگی فضایی می‌باشد. این آزمون را با استفاده از تابع $Moran.I$ در بسته ape انجام داده‌ایم. خروجی این آزمون در جدول (۲.۴) آورده شده است. با توجه به مقدار p -value که برابر ۰، می‌توان نتیجه گرفت که همبستگی فضایی در داده‌ها موجود است. همچنین برای تشخیص نوع همبستگی فضایی داده‌ها، می‌توان از این آزمون استفاده نمود، با توجه به این که مقدار مشاهده موران (I -موران) بیشتر از مقدار انتظار می‌باشد، می‌توان نتیجه گرفت همبستگی فضایی مثبت بین داده‌ها موجود است. منظور از همبستگی فضایی مثبت این می‌باشد که مناطق نزدیک به هم شباهت بیشتری با یکدیگر دارند.

جدول ۲.۴: خروجی آزمون I -موران

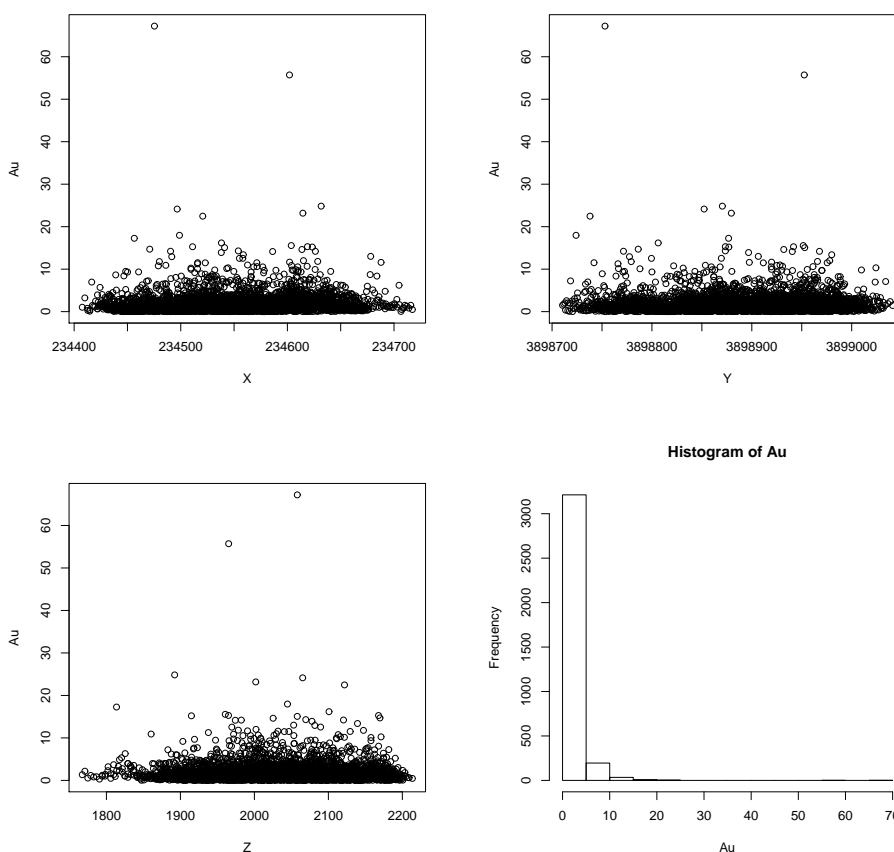
I -موران	مقدار انتظار I -موران تحت فرض H_0	انحراف معیار I -موران تحت فرض H_0	p -value
۰/۰۰۷۹۱۶۹۵۸	-۰/۰۰۰۲۸۹۵۱۹۴	۰/۰۰۰۳۷۷۶۱۶۴	۰

در جدول (۲.۴) شاخص‌های آماری عیار طلا آورده شده است. برای به‌دست آوردن شناخت اولیه از پراکندگی داده‌ها نسبت به موقعیت‌های فضایی، نمودارهای مقادیر اندازه‌گیری شده عیار طلا در مقابل هریک از محورهای x و y و z و همچنین بافت‌نگار این داده‌ها، بدون در نظر گرفتن موقعیت فضایی آن‌ها، در شکل (۱۱.۴) رسم شده است. یکی از کاربردهای نمودار

جدول ۳.۴: شاخص‌های آماری عیار طلای پروژه زرکوه

کمینه	چارک اول	میانه	میانگین	چارک سوم
۰	۰/۶۲۷	۱/۳۲۸	۱/۹۹۹	۱/۵۱۱
بیشینه	مد	واریانس	چولگی	کشیدگی
۶۷/۲۱۰	۰/۰۳۲	۶/۹۵۷۱۵۶	۸/۷۷۵۹۹۴	۱۶۸/۵۴۳۳

پراکندگی داده‌ها نسبت به محورهای مختصات، تشخیص وجود روند در مشاهدات می‌باشد. در شکل (۱۱.۴)، به دلیل تراکم مشاهدات، تشخیص وجود روند از نمودارهای مربوط به پراکندگی داده‌ها مشکل است. در بخش ۱.۶.۴ به رفع این مشکل پرداخته می‌شود.

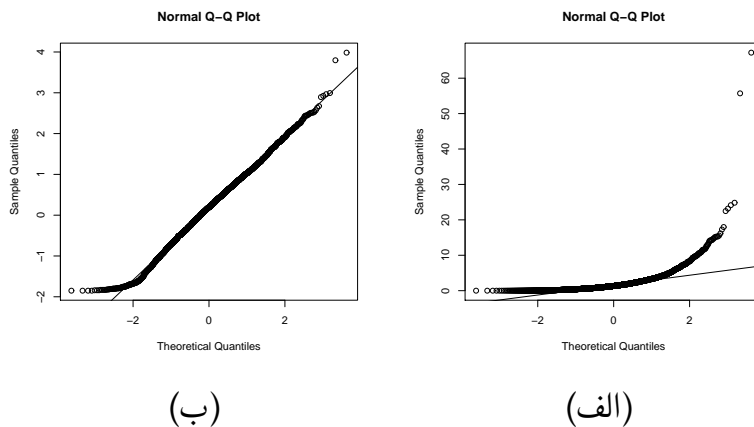


شکل ۱۱.۴: پراکندگی داده‌ها نسبت به محورهای مختصات

بررسی نرمال بودن داده‌ها

اگر میدان تصادفی گاوسی باشد، پیش‌گوی کریجیدن در رده همه پیش‌گوهای فضایی بهینه است، اما اگر میدان تصادفی غیرگاوسی باشد، این پیش‌گو فقط در رده پیش‌گوهای خطی بهینه خواهد بود. در برخی

مسایل کاربردی، با داده‌های مثبتی مواجه هستیم که از توزیع گاوسی پیروی نمی‌کنند، اما تبدیل لگاریتمی آن‌ها گاوسی یا به‌طور تقریبی گاوسی است. با توجه به مقدار چولگی داده‌ها (۸/۷۷۵)، بافت نگار داده‌ها در شکل (۱۱.۴) و همچنین نمودار چندک چندک داده‌ها در قسمت الف، شکل (۱۲.۴) می‌توان نتیجه گرفت داده‌ها دارای چولگی مثبت می‌باشند. وجود چولگی در داده‌ها موجب می‌شود، برخی داده‌ها به اشتباه به عنوان داده پرت تشخیص داده شوند. به همین علت برای تشخیص داده‌های پرت واقعی بهتر است تا حد امکان چولگی داده‌ها را برطرف نمود. با توجه به این‌که داده‌ها مربوط به عیار طلا هستند و مقدار مثبتی دارند و از توزیع گاوسی پیروی نمی‌کنند، می‌توان با انتخاب تبدیل لگاریتمی مناسبی بر روی داده‌ها، توزیع آن‌ها را به توزیع گاوسی یا به‌طور تقریبی گاوسی نزدیک نمود. داده‌ها با استفاده از تبدیل لگاریتم دو پارامتری $\log(\alpha + \beta Au)$ به منظور کمینه‌سازی چولگی به داده‌هایی با توزیعی نزدیک به نرمال تبدیل شده‌اند. بهینه‌سازی پارامترهای مذکور توسط روش نیوتن-رافسون انجام شده است، که مقادیر بهینه پارامترها عبارت‌اند از $\alpha = 0/157$ و $\beta = 0/797$ است. شایان ذکر است که انجام این بهینه‌سازی توسط بسته stats در نرم افزار R صورت گرفته است. نمودار چندک چندک در قسمت ب، شکل (۱۲.۴) نمایانگر کاهش چولگی داده‌ها پس از تبدیل انجام شده جهت کاهش چولگی، می‌باشد.



شکل ۱۲.۴: الف: نمودار چندک چندک داده‌ها، ب: نمودار چندک چندک داده‌ها همراه با کاهش چولگی

بررسی روند داده‌ها

در صورت وجود روند در داده‌ها، میدان تصادفی، نامانا و برآورد تغییرنگار، اریب است. به همین دلیل، در تحلیل اکتشافی داده‌ها، بررسی وجود روند داده‌ها مورد مطالعه قرار می‌گیرد. در این پایان‌نامه، برای تشخیص مرز عیار بالا در معدن ساری گونای، از روش مرز-بالا استفاده نموده‌ایم و در این روش، تغییرنگار را براساس روش گشتاوری برآورد کرده‌ایم. به علت داشتن برآوردگر نارایب، نیز به بررسی وجود روند در داده‌ها پرداخته‌ایم. برای شناسایی وجود روند در داده‌ها، لازم است نمودار مشاهدات میدان تصادفی $Z(s)$ در مقابل تمام مولفه‌های s رسم شوند. اگر در تمام این نمودارها،

نقاط به‌صورت تصادفی پراکنده شده باشند و حالت سیستماتیک نداشته باشند، داده‌ها فاقد روند هستند. بررسی نمودارها در شکل (۱۱.۴) نشان می‌دهد در هیچ جهتی و در هیچ یک از نواحی، ساختار منظم روند قابل تشخیص نبوده و احتمالاً وجود ندارد ولی با توجه به مقدار p -value برای مدل‌های مختلفی که در جهات مختلف برای روند داده‌ها استفاده کرده‌ایم. در راستای y و z ، مدل $\mu(s_i) = -8/722 \times 10^4 + 2/237 \times 10^{-2}y + 4/131 \times 10^0z - 1/06 \times 10^{-5}yz$ را به‌عنوان مدل روند، انتخاب می‌کنیم. برای حذف روند داده‌ها، برای هر موقعیت روند $(\mu(s_i))$ را از مشاهدات $(z(s_i))$ کم کرده تا مانده‌ها $(\delta(s_i))$ حاصل شوند، در این صورت مانده‌ها داده‌های روندزودوده هستند. لازم به‌ذکر است تغییرات ناگهانی در عیار طلا، تا فاصله ۵۰ متر بین جفت نقاط مدنظر بوده. برای این منظور تاخیرها، طوری در نظر گرفته شده است که تاخیر اول مربوط به جفت نقاطی است که در فاصله ۵ متر از یکدیگر قرار گرفته‌اند، به همین ترتیب تا تاخیر دهم که مربوط به جفت نقاطی است، که در فاصله ۵۰ متر از یکدیگر قرار گرفته‌اند.

۷.۴ نمودار جعبه‌ای

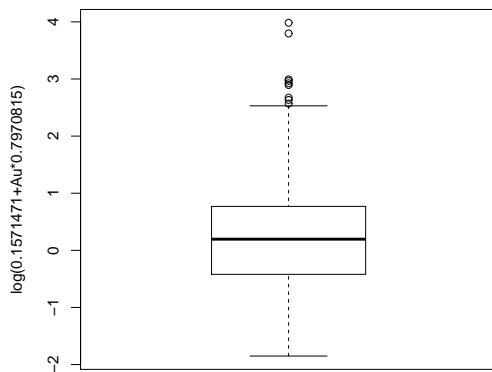
نمودار جعبه‌ای یک روش تشخیص داده پرت نوع اول است. همچنین در پروژه طلای ساری گونای که توسط شرکت خدمات اکتشافی کشور-خدمات فنی سسکو انجام شده است، از این روش برای تشخیص حدود آستانه‌ای مقادیر پرت استفاده شده است. دیواره بالایی نمودار جعبه‌ای داده‌ها، به عنوان حدود آستانه داده پرت انتخاب می‌شود.

با توجه به نمودار جعبه‌ای (۱۳.۴) مربوط به داده‌های عیار طلای پروژه زرکوه، وجود داده پرت نوع اول (بدون در نظر گرفتن موقعیت مشاهدات، فقط بر اساس مقدار مشاهده شده) دیده می‌شود. این نتیجه را می‌توان گرفت، که تعداد داده‌های پرت پس از تبدیل کاهش چولگی کمتر شده و همچنین حد آستانه پرت بودن افزایش یافته و برابر ۱۵/۹۲۸ است، در صورتی که قبل از تبدیل کاهش چولگی، حد آستانه‌ای برابر ۵/۳۳۶ بوده است.

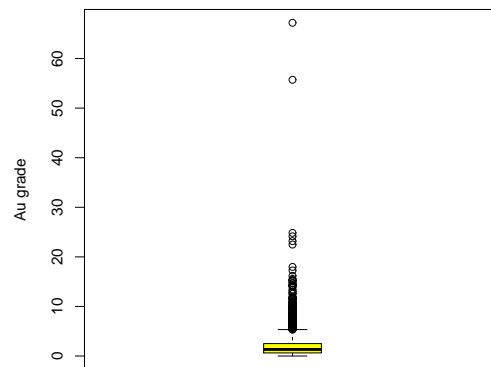
لازم به‌ذکر است که در تشخیص حدود آستانه‌ای پرت بودن توسط این روش، همبستگی فضایی داده‌ها در نظر گرفته نشده است.

۸.۴ ابر تغییرنگار

شکل (۱۴.۴) (الف)، نمودار ابر تغییرنگار مشاهدات عیار طلای زرکوه رسم شده است. این نمودار بیانگر وجود داده‌های پرت در مشاهدات است. با رسم نمودار جعبه‌ای به‌ازای هر تاخیر در نمودار ابر تغییرنگار، نمودار قسمت ب، شکل (۱۴.۴) حاصل می‌شود. با توجه به این‌که، نمودارهای جعبه‌ای در هر تاخیر متقارن نیستند، می‌توان به وجود داده پرت در مشاهدات پی‌برد. همچنین به علت این‌که در

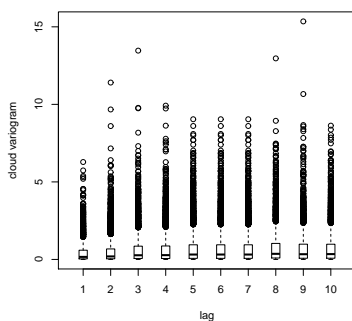


(ب)

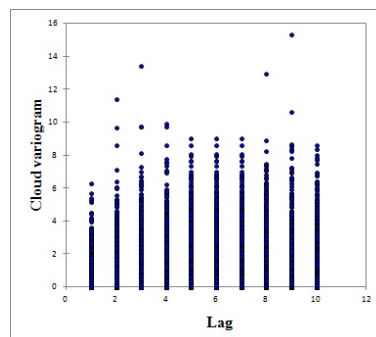


(الف)

شکل ۱۳.۴: الف: نمودار جعبه‌ای داده‌های عیار طلا، ب: نمودار جعبه‌ای داده‌های عیار طلا همراه با کاهش چولگی همه تاخیرها، بازوهای تمام نمودارهای جعبه‌ای به یک سمت است، نشان‌دهنده وجود همبستگی فضایی در مشاهدات است.



(ب)



(الف)

شکل ۱۴.۴: نمودارهای ابر تغییرنگار داده‌های عیار طلای زرکوه

۹.۴ روش مرز-بالا

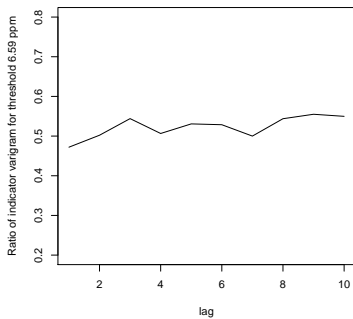
برای تشخیص مرز مقایر بالای داده‌های عیار طلای پروژه زرکوه از روش مرز-بالا استفاده نموده‌ایم. به علت این‌که مشاهدات دارای وابستگی فضایی بوده و در روش مرز-بالا نیز وابستگی فضایی مشاهدات، لحاظ می‌شود، این روش برای تشخیص مرز مقادیر بالا روش مناسبی است.

در این روش برای انتخاب مرزها، ابتدا مرز اول را به‌گونه‌ای انتخاب می‌کنیم که مقدار آن بالاتر از میانگین داده‌ها (۱/۹۹۹)، باشد. به همین منظور در این مجموعه داده‌ها، مرز پیشنهادی اول را ۲/۶، و مرزهای بعدی را با اختلاف‌های نزدیک ۲ انتخاب می‌کنیم، زیرا باید در نظر داشت که در روش مرز-بالا اگر مرزها، خیلی نزدیک به هم انتخاب شوند، ممکن است در تمامی مرزها اثر قطعه‌ای تام دیده شود، در این صورت تشخیص مرز واقعی دشوار است. در جدول (۴.۴)، مرزهای پیشنهادی همراه با اطلاعاتی در مورد میانگین، واریانس و ضریب تغییرات مقادیر کمتر و بالاتر هر مرز و همچنین تعداد مقادیر بالاتر هر مرز، آمده است.

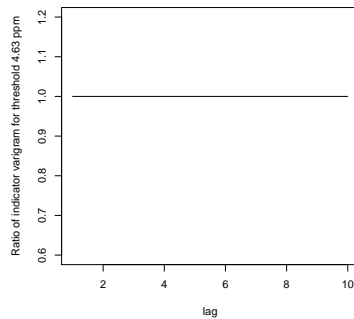
جدول ۴.۴: شاخص‌های آمار مقادیر بالاتر و کمتر از مرزهای مختلف

مقادیر مرز	۲/۶	۴/۶۳	۶/۵۹	۸/۴۱۲
مقادیر بالاتر از مرز	میانگین	۴/۹۹	۸/۰۳	۱۰/۵۸
	واریانس	۱۵/۹۲	۳۱/۵۳	۴۹/۸۷
	ضریب تغییرات	۱/۲۵	۱/۴۳	۱/۵۰
	تعداد	۸۲۳	۲۸۴	۱۴۲
مقادیر کمتر از مرز	میانگین	۱/۰۶	۱/۴۶	۱/۶۳
	واریانس	۰/۴۸	۱/۲۱	۱/۸۴
	ضریب تغییرات	۱/۵۳	۱/۳۲	۱/۲

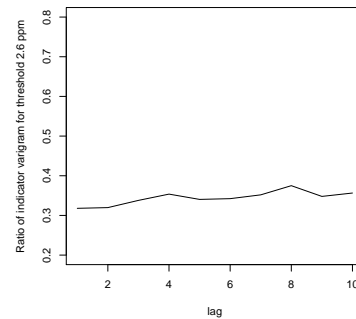
برای مرزهای پیشنهادی، نمودارهای نسبت تغییرنگارهای متقابل شاخص و تغییرنگارهای شاخص را به ازای تاخیرهای متفاوت در شکل (۱۵.۴) نشان داده‌ایم. از آنجایی که در توضیح روش مرز-بالا، در فصل سوم گفته شده است، بالاتر از مقدار مرز-بالا، اثر مرزی دیده نمی‌شود. به عبارتی دیگر احتمال شرطی (نسبت تغییرنگارهای شاخص) مربوط به مرز مقادیر بالا نباید تغییراتی را نسبت به فاصله نشان دهد. قسمت ب، شکل (۱۵.۴) مربوط به مرز پیشنهادی ppm ۴/۶۳، نسبت به تاخیر، ثابت است که بیانگر عدم وجود اثر مرزی می‌باشد.



(ج)



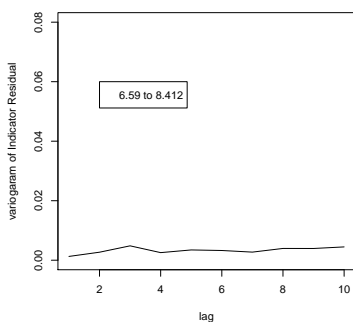
(ب)



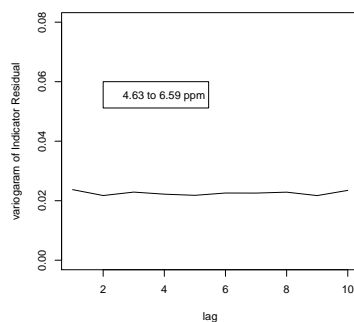
(الف)

شکل ۱۵.۴: الف: نمودار $\frac{\gamma I(2/6, 4/63)(h)}{\gamma I(2/6)(h)}$ ب: نمودار $\frac{\gamma I(4/63, 6/59)(h)}{\gamma I(4/63)(h)}$ ج: نمودار $\frac{\gamma I(6/59, 8/412)(h)}{\gamma I(6/59)(h)}$

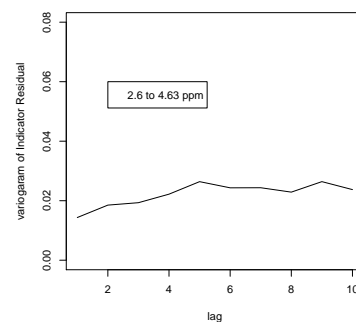
حال باید بررسی شود، آیا تغییرنگار باقی مانده شاخص مرز ۴/۶۳ ppm نسبت به مرزهای دیگر اثر قطعه‌ای تام را نشان می‌دهد یا خیر. برای این منظور نمودارهای مربوط به تغییرنگار باقی مانده شاخص برای مرزهای پیشنهادی، در شکل (۱۶.۴) رسم شده‌اند. بنا به نمودار (ب) که بیانگر اثر قطعه‌ای تام برای تغییرنگار شاخص مرز ۴/۶۳ ppm است، می‌توان مرز ۴/۶۳ ppm را به عنوان مرز مقادیر بالای معدن طلای پروژه زرکوه انتخاب نمود.



(ج)



(ب)



(الف)

شکل ۱۶.۴: تغییرنگارهای باقی مانده شاخص برای ۲/۶-۴/۶۳، ۴/۶۳-۶/۵۹ و ۶/۵۹-۸/۴۱۲

پیشنهادات لازم برای آینده تحقیق

با توجه به مطالب بیان شده و مطالعه شبیه‌سازی در این پایان‌نامه، می‌توان به موارد زیر به عنوان آینده تحقیق اشاره کرد:

۱. از آنجاکه در این پایان‌نامه، برای داده‌های طلای ساری گونای، مرز مناسب برای عیارهای بالا را پیدا کرده‌ایم، می‌توان مدل مرز-بالا را برای برازش این داده‌ها استفاده کرد.
۲. مقادیر بالاتر از مرز انتخابی، که توسط روش مرز بالا شناسایی شده‌اند را می‌توان به عنوان مقادیر فرین در نظر گرفت و برای برازش، از مدل‌هایی که برای مجموعه داده‌های، دارای مقادیر فرین، مناسب هستند، استفاده کرد.
۳. در یک مطالعه شبیه‌سازی، مدل مرز-بالا را با مدل‌هایی که برای داده‌های فرین مناسب هستند، مقایسه نمود.
۴. با توجه به این‌که داده‌های طلای ساری گونای دارای میدان غیر گوسی است، می‌توان برای تعیین ساختار داده‌ها، میدان را چوله گوسی در نظر گرفت.

پیوست آ

۱. آ ساری گونای

در این پیوست، برخی مطالب مربوط به ساری گونای، از پایگاه ملی داده‌های علوم زمین کشور اقتباس گردیده است. ساری گونای تقریباً در ۶۰ کیلومتری شمال غرب شهر همدان در گوشه‌ی جنوب شرقی استان کردستان در شمال غرب ایران واقع است. در مطالعات قبلی، کانسار به عنوان کانه‌زایی کوچک استینیت، اورپیمنت و رگه‌های منظم آنتیموان و آرسنیک استخراج شده است. در فاصله‌ی دورتر در شمال غرب، میزبان‌های رسوبی مرتبط با توده‌های نفوذی طلای تیپ - اپی ترمال زرشوران و آغدره وجود دارند. این توده‌های نفوذی، دورتر از بقیه توده‌های نفوذی ایران، مانند کمر بند مس پورفیری در جنوب شرق (به‌طور مثال سرچشمه و میدوک) و کانسار مس پورفیری سونگون در شمال غربی ایران واقع شده است. منطقه پروژه، از دو تپه‌ی سنگی اصلی ساری گونای و آق داغ تشکیل شده است، که توسط تپه‌های متعددی در شمال، احاطه شده و در سمت غرب و در فاصله‌ی دورتر، دشت وسیعی جای گرفته است. ناحیه‌ی نیمه خشک با بارش محدود، به بهار و طوفان‌ها و برف زمستانی است. به دلیل تاریخچه‌ی طولانی کشاورزی در حد معاش مرتع داری و دامداری در منطقه، پوشش گیاهی بومی، تنها به مناطق پرشیب سنگی محدود شده و متشکل از علف‌های بسیار کوتاه است. درخت‌ها تنها در کنار چشمه‌ها و نهرها رشد می‌کنند. محل پروژه بر روی یک معدن کوچک آرسنیک و آنتیموان رگه‌ای در داخل منطقه‌ای به وسعت ۱۶ کیلومتر مربع از سنگ‌های پیروکلاستیک و نیمه آتشفشانی حد واسط دگرسان شده الیگومیوسن با دگرسانی آرژیلیک و فیلیک یافت شد. این منطقه متشکل از نیمه آتشفشان‌های دگرسان شده‌ی هیدروترمال با رگه‌های سولفید آنتیموان / آرسنیک، دارای پتانسیل کانه‌زایی گسترده‌ی طلای اپی ترمال شناخته شده است. سیلیسی شدن سنگ‌های این منطقه در برخی مکان‌ها بسیار شدید است. تاریخ معدن کاری ساری گونای به هزاران سال باز می‌گردد، زمانی‌که این سنگ‌های پورفیری شدیداً سیلیسی و توف‌ها برای سنگ آسیاب استخراج می‌شدند (به ویژه در آق داغ).

اولین مطالعات شناسایی، در می و آگوست ۱۹۹۹ به کشف یک سیستم هیدروترمال مهم به وسعت ۱۶ کیلومتر مربع همراه با کانه‌زایی طلا، آنتیموان و آرسنیک به سن الیگومیوسن در داخل داسیت و

آندزیت پورفیری و توف‌های آتشفشانی انجامید. از این کارهای مقدماتی دانسته شد که نشانه‌هایی از سیستم طلای اپی ترمال غنی از سولفید دیده می‌شود. اساس تجاری پروژه‌ی ساری گونای، همکاری مشترک بین ریوتینتو بر اساس جواز بهره برداری سسکو است، که برای آنتیموان صادر شده است. سسکو طی توافق به عمل آمده در اکتبر ۲۰۰۰ یک همکاری مشترک با ریوتینتو را امضا کرد، که طبق آن ۷۰ درصد متعلق به ریوتینتو و ۳۰ درصد متعلق به سسکو بود. کارهای اکتشافی سطحی انجام شده توسط ریوتینتو و سسکو در نیمه‌ی اول ۲۰۰۰ به کشف موفقیت آمیز آنومالی خاک طلا دار به وسعت ۱۳۰۰m در ۴۰۰m در ساری گونای منجر شد و به علاوه چندین آنومالی کوچکتر، شامل آنومالی ۴۵° در ۱۵۰ متر در دامنه‌ی شمال غربی کوه آق داغ نیز شناسایی گردید. حفاری‌های بعدی انجام شده در مناطق آنومال از اواخر ۲۰۰۰ تا اواخر ۲۰۰۲، در مجموع ۳۸ چاه، موجب شناسایی ذخیره‌ی قابل توجه طلای ساری گونای و نشانه‌های کانه‌زایی در رگه‌ی ۱۱ و آق داغ گردید. حفاری‌های انجام شده تا سال ۲۰۰۴ به صورت موفقیت آمیزی به کشف یک توده با ۱۲۰۰ متر طول و حداکثر تا ۶۰۰ متر عرض و عمق حداقل ۳۵۰ متر در منطقه‌ی اصلی ساری گونای انجامید. مطالعات متالوژی، پتروگرافیکی و کانی شناسی مختلف برای درک کانه‌زایی انجام شده است.

نام پروژه در اوایل ۲۰۰۴ از داشکسن به ساری گونای تغییر یافت. چرا که نام سابق با توجه به منطقه‌ی پروژه که تمرکز اصلی را در خود جای داده بود، باعث ایجاد سردرگمی بین پروژه و روستای مجاور با نام مشابه شده بود و همچنین نام ساری گونای، برای منطقه‌ی حاوی تمرکز اصلی کانه‌زایی بیشتر کارایی دارد. در ژانویه‌ی ۲۰۰۴ یک شرکت جدید ایرانی با نام زرکوه تشکیل شد و هدف اصلی آن جستجو و توسعه‌ی احتمالی معدن کاری در کانسار طلای ساری گونای بود. ۷۰ درصد زرکوه متعلق به ریوتینتو و ۳۰ درصد متعلق به سسکو است. طی ۲۰۰۳ فاز چهارم حفاری‌ها در ساری گونای هدایت شد. کل حفاری‌های انجام شده در این مرحله در چهار مقطع به فاصله‌ی ۳۳ متر از مقاطع حفاری شده‌ی قبلی (۹۴۰۰ متر شمال، ۹۶۰۰ متر شمال، ۹۸۰۰ متر شمال و ۱۰۰۰۰ متر شمال) صورت گرفت.

این حفاری‌ها با فاصله ۳۳ متر برای تعیین مقدار تغییرات کانه‌زایی طلا، در فاصله کوتاه و همچنین تعیین فاصله بهینه حفاری، برای مطالعه احتمالی آینده در ساری گونای لازم بود. متأسفانه مطالعه تغییرنگارها از حفاری‌های ۲۰۰۳ نشان داد، که بر خلاف حفاری‌ها با فاصله‌ی ۳۳ متری، که شباهت بیشتری را در فاصله بزرگتر نشان می‌دادند، تغییرات بیشتری در عیار، بین فاصله‌های کوتاه تر وجود دارد. به دلیل این مقدار تغییرات عیار، این نتیجه حاصل شد که فاصله‌بندی حفاری‌های موجود برای تعیین تغییرنگار دقیق زمین آماری کانسار کافی نیست. به دلیل شکست ظاهری برنامه‌ی حفاری در ۲۰۰۳، برای تهیه‌ی تغییرنگار منطقی طلا از ساری گونای، خدمات فنی ریوتینتو به گروه زرکوه پیوسته و یک برنامه‌ی حفاری دقیق از شبکه‌بندی با فاصله‌ی ۲۰ متر و ۳۶ چاه طرح ریزی کرد. برنامه‌های اصلی دیگر ۲۰۰۴، شامل حفاری‌های تعیینی فراوان در ساری گونای به جهت کسب اطلاعات بیشتر برای محاسبه‌ی ذخیره در آینده، چندین حفاری اکتشافی در آنومالی‌های اصلی طلا در اطراف ساری گونای، برنامه‌ی حفاری چاه‌های دیده‌بانی آب و آزمایش‌های مداوم متالوژی بوده است.

۱.۱. آ. زمین شناسی ساری گونای

کانسار طلای ساری گونای با میزبانی کمپلس های ولکانیک میوسن میانی که از تپه های کوچک در ناحیه به طول ۱۰ کیلومتر گسترش یافته است. کانسار آنتیموان - آرسنیک - طلای داشکسن در باختر ایران و ۴۲ کیلومتری شمال خاوری شهرستان قروه واقع گردیده است. این کانسار از دیدگاه تقسیمات زمین ساختی، در زون دگرگونی - ماگمایی سنندج - سیرجان قرار می گیرد. بر پایه مشاهدات زمین شناسی، کهن ترین سنگ های منطقه را اسلیت، فیلیت و کوارتزیت های ژوراسیک و جوان ترین آن ها را کنگلومرای ولکانو کلاسیک نئوژن و جریان های بازالتی، گدازه های بلوکی و آگلومرای کواترنری تشکیل می دهد. سنگ های نفوذی منطقه مورد بررسی را یک توده نیمه عمیق میکروگرانیتی - میکروگرانودیوریتی با ماهیت کالکو آلکالین و بافت میکروگرانولار - پورفیری به سن نئوژن تشکیل می دهد، که در بخش های فوقانی به صورت گنبد های داسیتی - ریوداسیتی (آق داق و ساری داق) رخنمون پیدا کرده است (ریچارد و همکاران، ۲۰۰۶).

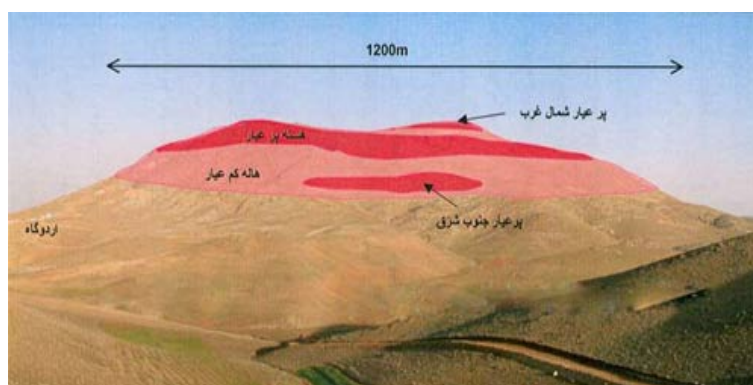


شکل ۱. آ: نمای منطقه پروژه ساری گونای از جنوب غربی

۲.۱. آ. نحوه کانه زایی در کانسار

کانه زایی در کانسار داشکسن، از نوع رگه ای و زون های سیلیسی - سولفیدی طلا دار است، که به وضوح توسط ساختارهای گسلی کنترل می شود. سنگ میزبان رگه ها و زون های کانه دار را سنگ های نفوذی نیمه عمیق و سنگ های داسیتی - ریوداسیتی سازنده گنبد های آق داغ و ساری داغ تشکیل می دهند، که با انواع دگرسانی هیدروترمال به ویژه آرژیلی و سیلیسی شدن همراهی می گردند. کانی شناسی رگه ها شامل: کوارتز، استینیت، پیریت، رآلگار، اریمنت، پیروتیت - کالکوپیریت، بورنیت، گالن، بولانزیت، آروستیبیت، طلا، استیبکونیت، کرمزیت و هیدروکسید های آهن می باشد. عیار طلا در رگه های کانه دار از ۵۵ تا ۲۶۶ پی پی ام و در زون های سیلیسی سولفیدی از ۱۰ تا ۱۸ پی پی ام تغییر می کند. طلا در ابعاد ۱۰ تا ۱۵۰ میکرون بیشتر به همراه کوارتز و در اندازه های کوچک تر در داخل کانی های پیریت

اکسیده مشاهده می‌شود. مطالعات میکروپروب الکترونی (SEM) مقادیری از عناصر آنتیموان، نقره و جیوه را در درون ذرات طلا نشان می‌دهد. نتایج آزمایشگاهی نمونه‌های شش مقطع لیتوژئوشیمیایی عمود بر روند رگه‌های اصلی استیبینیت در نواحی آق داغ و ساری داغ، نشان‌گر آن است که مقادیر بالای طلا محدود به رگه‌های استیبینیت و همچنین زون‌های دگرسانی سیلیسی - پیریتی است. مطالعات میکروترمومتری سیالات درگیر معرف دمای هموژنیزاسیون بین ۲۵۵-۱۸۳ درجه سانتیگراد و درجه شوری از ۸.۳ تا ۱۸۰.۸ معادل درصد وزنی کلرید سدیم برای سیالات درگیر اولیه است. در نهایت مجموعه بررسی‌های صحرائی، میکروسکپی، سیالات درگیر و سایر شواهد و داده‌های موجود نشان از رابطه نزدیک بین کانی‌سازی آنتیموان - آرسنیک - طلا و محلول‌های سیلیسی مربوط به توده ساب ولکانیک در ناحیه داشته و حاکی از آن است که کانسار داشکسن بیشترین شباهت را با کانسارهای تیپ اپی ترمال، به‌ویژه نوع اسید - سولفات دارد (ریچارد و همکاران، ۲۰۰۶).



شکل ۲.۰: زون‌های پرعیار و کم‌عیار ساری گونای

۳.۱.۱ زمین‌شناسی ساختاری ساری گونای

اگرچه گسل‌های قابل برداشت به ندرت در منطقه به دلیل وجود رسوبات کواترنری قابل برداشت هستند و گسل‌های منطقه‌ای توسط توزیع رخنمون‌ها و ساختارهای زمین ریخت‌شناسی، غالب‌ترین ساختارها شمال و شمال غرب و جنوب و جنوب شرق هستند که در ولکانیک میوسن ساری گونای محصور هستند. همچنین توسط ولکانیک‌های پلئوسن تا کواترنری که در مرکز شمال و شمال غرب محدوده قرار دارند، کنترل می‌شوند. گسترش ناحیه‌ای و خطواره‌ها با منشا امتداد لغز با جهت‌گیری کلی شبیه به مقیاس‌های منطقه‌ای و گسل‌های امتداد لغز که منشا انتقالات فشاری کوه‌زایی در مرکز ایران را دارند. در منطقه ساری گونای برش‌هایی توسط مجموعه‌های کوچک شمال شرق و تمایل بلوک‌های گسلی جنوب غرب با جابجایی عمودی وجود دارد. دو دسته از این گسل‌ها نقشه برداری شده‌اند یکی گسل غرب، ساری گونای و دیگری بین ساری گونای و آق داغ قرار دارد (ریچارد و همکاران، ۲۰۰۶).

پیوست ب

کدهای مربوط به فصل‌های چهارم

ب.۱ شبیه سازی

دستورات لازم برای مطالعه شبیه سازی در این بخش بیان شده است، و همچنین به‌طور جداگانه، کد مربوط به روش‌هایی که در شبیه سازی به کار برده‌ایم را گزارش می‌کنیم.

- تولید میدان گوسی مشبکه‌ای منظم با حجم نمونه ۱۰۰۰ (چون شبکه منظم ۱۰۲۴ نمونه تولید می‌کند) و تعداد تکرار ۱۰۰، و نحوه آلوده کردن آن به داده پرت:

```
require(geoR)
simulate<-grf(1000,grid="reg",cov.pars=c(40,3),nsim=100,mean=100)
loc<-simulate[[1]]
points(simulate)
sim<-simulate[[2]]
simreplace<-sim
loc.N<-matrix(NA,nrow=10,ncol=100)
for(j in 1:100){
loc.N[,j]<-as.vector(sample(which(simreplace[,j]>=
quantile(simreplace[,j],prob=0.8)),size=10, replace=FALSE))
}
sample<-as.vector(sample(200:2344,size=100, replace=FALSE))
for(i in 1:100){
rn<-rnorm(10,mean =130, sd =40)
for(j in 1:10){
simreplace[loc.N[j,i],i]<-rn[j]
}
}
}
```

- روش تشخیص مرز مقادیر بالا توسط نمودار جعبه‌ای و شمارش تعداد داده‌های پرتی که توسط این روش در داخل مرز قرار دارند:

```

box.tershold<-vector(length=100)
for(i in 1:100){
  box.tershold[i]<-quantile(simreplace[,i],prob=.75)+
  (1.5*(quantile(simreplace[,i],prob=.75)-quantile(simreplace[,i],prob=.25)))
}
boxresult<-matrix(NA,ncol=100,nrow=10)
b<-list()
for(i in 1:100){
b[[i]]<-which(simreplace[,i]>= box.tershold[i])
lb<-c()
n<-0
for(t in 1:10){
  for(j in 1:length(b[[i]])){
    if(loc.N[t,i]==b[[i]][j]){
      n<-n+1
    }
  }
}
lb<-c(lb,n)
}
boxres[,i]<-lb
}

```

- روش تشخیص داده پرت توسط الگوریتم میانه و شمارش تعداد داده‌های پرتی که به درستی انتخاب شده و داده‌هایی که، به اشتباه به عنوان داده پرت انتخاب شده است:

```

h<-function(x,y){
  h1<-sqrt(sum((x-y)^2))
  return(h1)
}

dist<-matrix(NA,ncol=1024,nrow=1024)
for(i in 1:1024){
  for(j in 1:1024){
    dist[i,j]<-h(loc[i,c(1,2)],loc[j,c(1,2)])
  }
}

```



```

sortdist<-matrix(NA,ncol=1024,nrow=1024)
for(i in 1:1024){
  sortdist[i,<-sort(dist[i,])
}
simk<-matrix(NA,nrow=1024,ncol=8)
for(i in 1:1024){
  for(j in 1:8{ # for k=8
    for(t in 1:1024){
      if(sortdistk[i,j]==dist[i,t]){
        simk[i,j]<-simreplace[t,1] # for first replicate
      }
    }
  }
}
gx<-apply(simk,1,mean)
hx<-vector(length=1024)
for(i in 1:1024){
  hx[i]<-(simreplace[i,1]-gx[i]) # for first replicate
  mu<-median(hx)
  sigma<-median(abs(hx-mu))
  y<-abs((hx-mu)/sigma)
  a<-which(y >3.09) # for alpha/2=0.001
}
##### detect
detect<-length(a)
lt<-c()
n<-0
for(i in 1:10){
  for(j in 1:length(a)){
    if(loc.N[i,1]==a[j]){ # for first replicate
      n<-n+1
    }
  }
}
lt<-c(lt,n)
}
wrong<-detect-n

```

• روش میانه و میانگین در تشخیص داده پرت :

```

count1<-function(x,a){
n<-length(x)
t<-0
for(i in 1:n){
if(x[i]==a) t<-t+1
}
t
}
#####
count2<-function(x){
x<-sort(x)
z<-10^(-5)
t<-a<-x1<-0
for(i in 1:length(x)){
if(z!=x[i]) { t<-t+1
a[t]<-count1(x,x[i])
z<-x[i]
x1[t]<-z}}
return(list(t=t,x1=x1,a=a))
}
#####
data<-cbind(loc,simreplace)
Mean.median<-function(data){
data<-data[sort.list(data[,1]),] # sort x
k1<-count2(data[,1])
c<-mx1<-mx2<-my1<-my2<-ucolumn<-urow<-0
for(i in 1:k1$t){
c[i+1]<-k1$a[i]+c[i]
mx1[i]<-mean(data[(c[i]+1):c[i+1],3])
mx2[i]<-median(data[(c[i]+1):c[i+1],3])
sigmahat<-IQR(data[(c[i]+1):c[i+1],3])/(2*0.6745)
ucolumn[i]<-length(data[(c[i]+1):c[i+1],3])^(0.5)*
(mx1[i]-mx2[i])/(0.7555*sigmahat)
}

data<-data[sort.list(data[,2]),] # sort y
k2<-count2(data[,2])
c1<-0
for(i in 1:k2$t){

```

```

c1[i+1]<-k2$a[i]+c1[i]
my1[i]<-mean(data[(c1[i]+1):c1[i+1],3])
my2[i]<-median(data[(c1[i]+1):c1[i+1],3])
sigmahat<-IQR(data[(c1[i]+1):c1[i+1],3])/(2*0.6745)
urow[i]<-length(data[(c1[i]+1):c1[i+1],3])^(0.5)*
(my1[i]-my2[i])/(0.7555*sigmahat)
}
return(list(nx=k1$t,ny=k2$t,x=k1$x1,y=k2$x1,meanx=mx1,meany=my1,
medianx=mx2,mediany=my2,Ucolumn=ucolumn,Urow=urow))
}
##### result of simulation #####
ucol<-matrix(0,ncol=100,nrow=32)
urow<-matrix(0,ncol=100,nrow=32)
for(i in 1:100){
urow[,i]<- Mean.median(data[,c(1,2,i+2)])$Urow
ucol[,i]<- Mean.median(data[,c(1,2,i+2)])$Ucolumn
}
##### plot
plot(loc[,1], loc[,2],type="n",lab=c(32,32,150),xlab="x",ylab="y")
text(loc[,1], loc[,2],format(round(simreplace[,1],1))) #for first replicate
title("simreplace1",cex=1.5)

```

ب.۲ روش مرز-بالا

برای یک مرز پیشنهادی روش مرز-بالا توسط کد زیر اجرا می‌شود.

• ابتدا شاخص را محاسبه می‌کنیم

```

indicator<-c()
for(i in 1:n){
# number of data
if(data< cut off){
indicator[i]<-0
}else{
indicator[i]<-1
}
}
}

```

• تغییرنگار شاخص را برای مرز پیشنهادی محاسبه می‌کنیم

```

semivariog<-c()
Ncol<-c()
lag<-c()
for(l in 1:L){
    # l=number of lag
    squar<-c()
    n<-0
    for(i in 1:n){
        # number of data
        if(h(loc[i,],loc[j,])>5*(1-1)&&h(loc[i,],loc2[j,])<=5*1){ # loc=matrix of location
            squar<-c(squar,(indicator[i]-indicator[j])^2)
            n<-n+1
        }
    }
}
Ncol<-c(Ncol,n)
semivariog<-c(semivariog,(1/(2*n))*sum(squar))
lag<-c(lag,l )
}

```

• محاسبه تغییرنگار متقابل شاخص بین دو مرز

```

crossvariog<-c()
Ncol<-c()
lag<-c()
for(l in 1:L){
    # L=number of lag
    squar<-c()
    n<-0
    for(i in 1:n){
        # number of data
        if(h(loc[i,],loc[j,])>5*(1-1)&&h(loc[i,],loc2[j,])<=5*1){ # loc=matrix of location
            squar<-c(squar,(indicator1[i]-indicator1[j])*(indicator2[i]-indicator2[j]))
            n<-n+1
        }
    }
}
Ncol<-c(Ncol,n)
crossvariog<-c(crossvariog,(1/(2*n))*sum(squar))
lag<-c(lag,l )
}

```

• محاسبه نسبت تغییرنگارها

```
ratio=crossvariog/semivariog
```

• سپس محاسبه تغییرنگار باقی‌مانده شاخص که کد این برنامه نیز واضح است.

ب.۳ تحلیل اکتشافی

کد مربوط به آزمون I -موران

```
Au.dists <- as.matrix(dist(loc[,c(1,2)]))
Au.dists.inv <- 1/Au.dists
diag(Au.dists.inv) <-0
require(ape)
Moran.I(dataAu, Au.dists.inv)
```

محاسبه ابر تغییرنگار

```
variog.c<-list()
mean<-list()
for(l in 1:L){ # L=number of lag
  A<-c()
  for(i in 1:208){
    for(j in i:208){
      if(h(loc[i,],loc[j,])>5*(1-1)&&h(loc[i,],loc2[j,])<=5*1){ # loc=matrix of location
        A<-rbind(A,cbind(data[i],data[j]))
      }
    }
  }
  vario.c[[a]]<-((A[,2]-A[,1])^2)/2
  mean[[a]]<-mean(vario.c[[a]])
}
boxplot(variog.c,boxwex = 0.3,col="yellow",main="cloud variogram for data",
xlab="lag",ylab="cloud variogram")
```

مراجع

[۱] حسنی پاک، ع، (۱۳۷۸)، "اکتشاف ذخایر طلا"، چاپ اول، موسسه انتشار و چاپ دانشگاه تهران، دانشگاه تهران، تهران، ایران.

[۲] حق شناس تجن گوکه، ف، (۱۳۸۹)، "تحلیل بیزی میدان‌های تصادفی مارکوفی گاوسی با استفاده از پیشین مرجع"، پایان‌نامه کارشناسی ارشد، دانشگاه تربیت مدرس، ایران.

[۳] شهاب پور، ج، (۱۳۹۰)، "زمین‌شناسی اقتصادی"، چاپ ششم، مرکز نشر آثار علمی، دانشگاه شهید باهنر، کرمان، ایران.

[۴] محمدزاده، م، (۱۳۹۱)، "آمار فضایی و کاربردهای آن"، چاپ اول، مرکز نشر آثار علمی، دانشگاه تربیت مدرس، تهران، ایران.

[۵] معزز لسکو، ض، (۱۳۸۱)، "کافی‌شناسی"، چاپ اول، موسسه چاپ دانشگاهی فردوسی، مشهد، ایران.

[6] Armstrong, M. and Jacques. (1997), *Geostatistique Lineaire Application au domaine minier*, Paris.

[7] Barnett, V. and Lewis, T. (1984), *Outliers in Statistical Data*, John Wiley, New York.

[8] Beirlant, J. Goegebeur, Y. Teugels, J. Segers, J. DeWaal, D; and Ferro, C. (2004), *Statistics of extremes, theory and applications*. Wiley, Chichester.

-
- [9] Caers, J. Rombouts, L. (1996), Valuation of primary diamond deposits by extreme value statistics. *Econ Geol* 91(5):841–854.
- [10] Cao, L. Liu, X. Wang, Z, and Zhang, Z. (2013a), The Spatial Outlier Mining Algorithm based on the KNN Graph, *Journal of Software*, 8:3158-3165.
- [11] Cao, L. Liu, X. Wang, Z, and Zhang, Z. (2013b), WSO-based Spatial Outlier Detection Algorithms, *Journal of Networks*, 8:1582-1588.
- [12] Chen, D. Lu, C.-T. Kou, Y, and Chen, F. (2008), On Detecting Spatial Outliers, *Geoinformatica*, 12:455-475.
- [13] Cox, D. R. and Miller, H. D. (1965), *The Theory of Stochastic Processes*, Methuen, London.
- [14] Cressie, N. (1993), *Statistics for Spatial Data*, John Wiley, New York.
- [15] David, M. (1977), *Geostatistical Ore Reserve Estimation*, Elsevier, Amsterdam.
- [16] Kou, Y. Lu, C. (2007), Spatial Outlier Detection: A Graph-based Approach. *IEEE Computer Society*, 139:281-289.
- [17] Matheron, G. (1962), Traite de Geostatistique Appliquee, Tome I. *Memoires du Bureau de Recherches Geologiques et Minières*, 14, Editions: Technip, Paris.
- [18] Matheron, G. (1969), *Le Krigeage Universal*, *Cahiers du Centre de Morphologie Mathematique, No.1. Fontainebleau, France.*, *Jornal of Econometrics*, 37:87-114.

- [19] Matheron, G. (1982), La destructurement des hautes teneurs et le krigeage des indicatrices (The destructuring of high grades and indicator kriging), *Technical Report N-761*, Centre de Géostatistique, Fontainebleau, France.
- [20] Myers. D. E (1989), To Be or Not to Be ... Stationary? That Is the Question *Mathematical Geology*, 21:347-362.
- [21] Pebesma, E.J. (1997), Mapping Groundwater Quality in the Netherlands. *Journal of Hydrology*, 200:364-386.
- [22] Philip, G. M. and Watson, D. F. (1986), *Matheronian Geostatistics: Quo Vadis: Math. Geol.*, 18:93-117
- [23] Richard, J. Wilkinson, D, and Ulirich.T, (2006), SGeology of the Sari Gunay Epithermal Gold Deposit, Northwest Iran, *Economic Geology*, 101:1455-1496.
- [24] Ripley, B. D. (1981), *Spatial Statistics*, John Wiley, New York.
- [25] Rivoirard, J. (1994), *Introduction to disjunctive kriging and nonlinear geostatistics*, Clarendon, Oxford.
- [26] Rivoirard, J. Demange, C. Freulon, X. Lecureuil, A, and Bellot, N. (2013), A Top-Cut Model for Deposits with Heavy-Tailed Grade Distribution, *International Association for Mathematical Geosciences*, 45:967-982.
- [27] Shekhar, S. Lu, C, and Zhang, P. (2003), A Unified Approach to Detecting Spatial Outiers, *GeoInformatica*, 7:139-166.
- [28] Sinclair, A, and Blackwell, G. (2004), *Applied Mineral Inventory Estimation*, Cambridge, United Kingdom.

-
- [29] Walfish, S. (2006), A Review of Statistical Outlier Methods, *Pharmaceutical Technology*, 2:1-5.

واژه‌نامه فارسی به انگلیسی

Spatial statistics	آمار فضایی
Variogram cloud	ابر تغییرنگار
Nugget effect	اثر قطعه‌ای
Edge effect	اثر مرزی
Sill	ازاره
Point pattern	الگوی نقطه‌ای
Estimate	برآورد
Estimator	برآوردگر
Large scale variation	تغییرات بزرگ مقیاس
Small scale variation	تغییرات کوچک مقیاس
Variogram	تغییرنگار
Indicator residual variogram	تغییرنگار باقی‌مانده شاخص
Indicator variogram	تغییرنگار شاخص
Cross variogram	تغییرنگار متقابل
Indicator cross variogram	تغییرنگار متقابل شاخص
Outlier data	داده‌های پرت
Deterrended data	داده‌های روندزوده
Geostatistical data	داده‌های زمین‌آماری
Spatial data	داده‌های فضایی
Coal Ash Data	داده‌های خاکستر زغال سنگ
Lattice data	داده‌های شبکه‌ای
Range	دامنه
Interquartile range	دامنه میان‌چارکی
Outlier	پرت
Trend	روند

Geostatistics.....	زمین آمار.....
Kriging	گریگینگ
Intrinsically stationary.....	مانای ذاتی.....
Weakly stationary with drift.....	مانای ضعیف با انحراف.....
Strong stationary	مانای قوی
Second order stationary.....	مانای مرتبه دوم.....
location	موقعیت
Box plot	نمودار جعبه‌ای.....
Steam and leaf plot.....	نمودار ساقه و برگ.....
Semi-Variogram.....	نیم‌تغییرنگار.....
Spatial dependence.....	وابستگی فضایی.....
Correlogram.....	هم‌بستگی‌نگار.....
Covariogram	هم‌تغییرنگار.....
Isotrop	همسان‌گرد.....

واژه‌نامه انگلیسی به فارسی

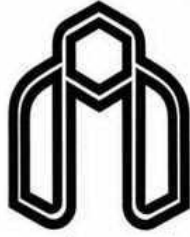
Bias	اریبی
Box plot	نمودار جعبه‌ای
Coal Ash Data	داده‌های خاکستر زغال سنگ
Conditional negative definite	همیشه منفی شرطی
Correlation coefficient	ضریب همبستگی
Correlogram	هم‌بستگی‌نگار
Covariogram	هم‌تغییرنگار
Cross variogram	تغییرنگار متقابل
Deterended data	داده‌های روندزودوده
Edge effect	اثر مرزی
Estimate	برآورد
Estimator	برآوردگر
Gaussian	گاوسی
Geostatistical data	داده‌های زمین‌آمار
Geostatistics	زمین‌آمار
Indicator variogram	تغییرنگار شاخص
Indicator residual variogram	تغییرنگار باقی‌مانده شاخص
Point pattern	الگوی نقطه‌ای
Range	دامنه
Robust	استوار
Second order stationary	مانای مرتبه دوم
Semi-variogram	نیم‌تغییرنگار
Sill	ازاره
Skew Gaussian	چوله گاوسی
Small scale variation	تغییرات کوچک مقیاس

Spatial data	داده‌های فضایی
Spatial dependence	وابستگی فضایی
Spatial statistics	آمار فضایی
Steam and leaf plot	نمودار ساقه و برگ
Strong stationary	مانای قوی
Trend	برآورد
Variogram	تغییرنگار
variogram cloud	ابر تغییرنگار
Weakly stationary with drift	مانای ضعیف با انحراف

Aabstract

In ore deposits, detecting the grades that are in the right tail of the distribution are economically important. As a result, detecting the cutoff of these observations is important in mining industry extraction. Existing methods for detecting spatial outlier indentify the outlier values but do not provide criteria for identifying cutoff for high values. Hence, in this thesis, Top-Cut method is suggested to identify the cutoff for high values which is based on absence of edge effect. In the absence of edge effect, highe values have no influence from lower grades. In this method, ratio of the indicator cross variogram and indicator variogram in each suggested cutoff is calculated and if this ratio is constant, in terms of function of distances, it shows absence of edge effect. Also if the variogram of the residual indicator for suggested cutoff shows pure nugget effect, then the suggested cutoff is chosen for topcut grade. Evaluation of these methods and some outlier detection methods has been carried out in a simulation study. Moreover as a real case study, this method has been used to determine the cutoff of high-grade of Sari-Gunay deposit in Kordestan province, Iran. We found that the value 4.63 ppm as the cutoff for high-grade of Sari-Gunay deposit.

keywords: Spatial Outlier, high-grade, Edge effect, Indicator variogram, Indicator cross variogram



Shahrood University Of Technology

Shahrood University of Technology
Faculty Of Mathematical Sciences

Dissertation Submitted in Partial
Fulfillment of The Requirements For The
Degree of Master of Science in
Statistics

Spatial Outlier Detection

Supervisors

Dr.D. Shahsavani and Dr.M. Koneshloo

Advisor

Dr.H. baghishani

by

Maryam Moghaddam

2014