

شناسایی مشاهدات دور افتاده یک میدان تصادفی

محسن محمدزاده^۱ یدالله واقعی^۲

چکیده

داده‌های فضایی مشاهداتی همبسته از یک میدان تصادفی^۳ هستند که از موقعیت‌های مختلف در فضای d بعدی جمع‌آوری می‌شوند. ساختار همبستگی این داده‌ها تحت شرط مانایی^۴ بوسیله تابع تغییرنگار بر حسب فاصله بین موقعیت داده‌ها، برآورد و مدل‌سازی می‌شود. وجود داده‌های دور افتاده برآورد تغییرنگار و پیشگویی‌های مبتنی بر آن را تحت تاثیر قرار می‌دهد، لذا شناسایی و برخورد مناسب با آنها در تجزیه و تحلیل داده‌های فضایی بسیار مهم می‌باشد. در این مقاله چند روش گرافیکی برای شناسایی داده‌های دور افتاده فضایی معرفی و مورد بررسی قرار گرفته‌اند و در نتیجه استفاده از نمودار ساقه و برگ یا نمودار پراکنش سه بعدی باقیمانده‌های پیش‌گویی، که بر پایه اعتبار سنجی متقابل تغییرنگار داده‌ها قرار دارد، به عنوان یک روش کارا تر معرفی شده است. بر اساس این روش اگر مقدار واقعی یک داده با مقدار پیش‌گویی شده برای آن که به کمک سایر داده‌ها و به روش کریگیدن بدست آمده اختلاف چشمگیری داشته باشد، دور افتاده محسوب می‌شود.

واژه‌های کلیدی: داده‌های دور افتاده فضایی، تغییرنگار، کریگین، اعتبار سنجی متقابل

۱. مقدمه

همیشه وجود مشاهدات غیرمعقول، غیرعادی^۵ یا دور افتاده^۶ در مجموعه داده‌ها مورد توجه آمارشناسان و محققان بوده است. این‌ها مشاهداتی هستند که اختلافی غیرعادی با سایر داده‌ها دارند به گونه‌ای که با بحث‌های نظری قابل توجیه نیستند. به‌عنوان مثال ممکن است مشاهده (مشاهداتی) از مجموعه داده‌های مورد مطالعه از مفروضات، مدل یا توزیع مسئله مورد نظر پیروی نکنند و موجب بروز خطا یا مسأله‌ای غیرعادی در تجزیه و تحلیل آنها گردند. وجود این‌گونه مشاهدات سبب کاهش اطلاعات موجود در منابع داده‌ها یا خراب کردن آنها شده و در نتیجه داده‌ها را با نوعی آلودگی مواجه می‌سازد. طبیعی است که قبل از هرگونه اقدامی برای تجزیه و تحلیل داده‌ها به دنبال روش‌ها و ابزارهای شناسایی داده‌های دور افتاده باشیم تا بتوان در خصوص نحوه برخورد با آنها تصمیمی مناسب اتخاذ نمود، یا حداقل روش‌هایی را برای تجزیه و تحلیل داده‌ها بیابیم که تأثیر آلودگی‌ها را در نتایج کاهش دهند.

وقتی مشاهدات مستقل از یکدیگر هستند، روش‌های آماری متنوعی برای شناسایی و برخورد با داده‌های دور افتاده ارائه شده است. به عنوان مثال در مسأله رگرسیون برآورد پارامترها، بویژه در روش ماکزیمم درستنمایی، به شدت تحت تأثیر داده‌های دور افتاده قرار دارند، به همین علت روش‌های استوار^۷ برای برآورد پارامترهای رگرسیونی ارائه شده‌اند که تا حد ممکن اثرات داده‌های دور افتاده را در برآورد پارامترهای رگرسیونی کم می‌کنند (بارنت و لوییس، [۳]). داده‌هایی که نوعاً بر حسب موقعیت و مکان قرار گرفتن آنها در فضای مورد مطالعه وابسته هستند، داده‌های فضایی نامیده شده و در آمار فضایی برای مدل بندی آنها از میدان تصادفی $\{Z(t); t \in D \subseteq R^d\}$ ، $d \geq 1$ استفاده می‌شود، که در آن D یک مجموعه اندیس گذار شامل تمام موقعیت‌های ممکن است. مفهوم و نحوه تشخیص داده‌های دور افتاده فضایی با روش‌های معمول اندکی متفاوت است. در داده‌های فضایی مشاهداتی که به طور قابل توجهی با مشاهدات موجود در همسایگی یا اطراف خود اختلاف قابل توجهی داشته باشند به عنوان داده دور افتاده شناخته می‌شوند. برای شناسایی داده‌های دور افتاده فضایی، گاهی، از روش‌های معمولی

۱- گروه آمار دانشگاه تربیت مدرس

۲- گروه آمار دانشگاه بیرجند

۳- Random Field

۴- Stationary

۵- Atypical

۶- Outlier

۷- Robust

ساده‌ترین روش‌های گرافیکی هستند که در شناسایی داده‌های دور افتاده فضایی مؤثر هستند. در این مقاله روش‌های گرافیکی شناسایی داده‌های دور افتاده فضایی معرفی می‌شوند و هر یک از آنها برای مجموعه داده‌های فضایی میزان بروز سل ریوی سال ۱۳۷۸ مربوط به ۲۶۲ شهرستان کشور (واقعی، [۱]) بکار گرفته و نقاط دور افتاده شناسایی و مورد بررسی قرار می‌گیرند. محاسبات و تهیه نمودارها با برنامه‌نویسی در محیط نرم‌افزار R که با استفاده از S-Plus تهیه شده‌اند، انجام شده است.

۲. نمودار پراکنش h

از رسم مقادیر $Z(t)$ در برابر $Z(t+h)$ به ازای یک فاصله h خاص یک نمودار پراکنش دوبعدی بدست می‌آید، که نمودار پراکنش h نامیده می‌شود. این نمودار کلیه زوج داده‌هایی را که در همسایگی h از یکدیگر قرار دارند در برابر یکدیگر نشان می‌دهد. اگر نقاط متناظر با زوج داده‌ها حول خط نیمساز ربع اول پراکنده باشند، می‌توان نتیجه گرفت داده دورافتاده وجود ندارد و چنانچه بعضی از نقاط به طور چشمگیری از خط نیمساز دور باشند، نشان‌دهنده وجود داده‌های دورافتاده هستند.

نمودار پراکنش h را به دو صورت سوار^۷ و بی‌سو یا همسانگرد می‌توان رسم کرد. هنگامی که موقعیت داده‌ها در یک شبکه منظم، با فواصل a واحد از یکدیگر، باشند به راحتی می‌توان نمودار پراکنش h را با استفاده از زوج داده‌هایی که در سوی غربی- شرقی یا شمالی- جنوبی به فواصل منظم $a, 2a, 3a, \dots$ قرار دارند، به طور مجزا، رسم کرد. این نمودار را به صورت بی‌سو نیز می‌توان رسم نمود، مشروط بر آنکه شرط همسان‌گردی تا حدودی برقرار باشد.

اگرچه ایده‌ی اصلی نمودار پراکنش h برای داده‌های منظم (مشبکه) است، برای داده‌های نامنظم نیز با در نظر گرفتن یک ناحیه تحمل می‌توان آن را رسم کرد. در این صورت مثلاً کلیه زوج داده‌هایی که فاصله بین آنها \circ تا \circ واحد و زاویه آنها بین 25° تا 25° درجه باشد به عنوان نمودار پراکنش h در سوی \circ درجه (غربی - شرقی) رسم می‌شوند. با توجه به اینکه دورافتادگی در داده‌های فضایی نسبت به داده‌های مجاور یا همسایه سنجیده می‌شود، بهتر است نمودار پراکنش h در فواصل کوچک که متناظر با فاصله‌های نخست برآورد تغییرنگار هستند مورد بررسی قرار گیرد. نمودار پراکنش h علاوه بر شناسایی داده‌های دورافتاده در بررسی فرض‌های

استفاده می‌شود، اما از آنجا که این روش‌ها موقعیت و همبستگی داده‌ها را نادیده می‌گیرند، امکان دارد در تشخیص داده‌های دورافتاده فضایی دچار خطا شوند. به عنوان مثال امکان دارد یک داده بدون در نظر گرفتن موقعیت فضایی دورافتاده باشد، ولی نسبت به مشاهدات اطراف خود دورافتاده نباشد. برعکس امکان دارد یک داده نسبت به داده‌های اطراف خود دورافتاده باشد، در حالی که در روش‌های معمول به عنوان داده دورافتاده شناخته نمی‌شود. کِرسی [۵] روشی را برای شناسایی داده‌های دورافتاده فضایی هنگامی که موقعیت داده‌ها در یک مشبکه منظم در فضای دوبعدی قرار دارند ارائه کرده است. بر پایه این روش چنانچه میانگین و میانه یک سطر (یا ستون) از مشبکه اختلاف زیادی داشته باشند می‌توان به آن سطر (یا ستون) برای دارا بودن داده‌های دورافتاده مشکوک شد. هاینینگ [۹] نیز روشی مشابه برای داده‌های مشبکه‌ای منظم ارائه داده است. میانه پرداخت^۱ روش دیگری برای شناسایی داده‌های دورافتاده برای داده‌های منظم مشبکه است (کِرسی، [۵]). روش هاوکینز [۱۰] بر اساس مقایسه یک داده مشکوک به دورافتادگی با همه‌ی مقادیر همسایگی آن طراحی شده است. استفاده از فاصله ماله‌لانوبیس که از روش‌های شناسایی داده‌های دورافتاده چندمتغیره است، برای داده‌های فضایی نیز سابقه دارد [۱۳]. بارنت و لوپس [۳] روشی را با استفاده از ماتریس عملگر برای شناسایی داده‌های دورافتاده معرفی کردند. باردوسی، کوندزویکس [۲] روش جک‌نایف^۲ را براساس فاصله اطمینان $Z(t_0)$ که با استفاده از توزیع گاوسی، داده‌ها را برای دور افتادگی آزمون می‌کند ارائه نمودند. هادی [۸] روش جستجوی پیشرو^۳ را برای کشف داده‌های دورافتاده در داده‌های چندمتغیره ارائه داد که به دنبال آن ریانی و سریالی [۴] آن را به داده‌های فضایی تعمیم داده و یک الگوریتم جستجوی پیشرو ارائه دادند که داده‌ها را از هماهنگ‌ترین تا ناهماهنگ‌ترین داده با یک مدل همبستگی معین مرتب می‌کند و منجر به شناسایی داده‌های دورافتاده چندگانه فضایی می‌گردد.

روش‌های گرافیکی در تجزیه و تحلیل اکتشافی داده‌ها و شناسایی داده‌های دور افتاده سهم بسیار زیادی دارند (سری‌آلی و ریانی، [۴])، و وایز و همکاران، [۱۴]). حتی اگر هیچ داده دورافتاده‌ای وجود نداشته باشد، نمودارهای گرافیکی درباره شناخت مدل و فرضیات آن اطلاعات مفیدی را در اختیار تحلیل‌گر قرار می‌دهند. نمودارهای ساقه و برگ^۴، جعبه‌ای^۵ و پراکنش سه‌بعدی^۶ از جمله

ابر تغییرنگار (یا شبه تغییرنگار) را می‌توان در جهات مختلف نیز رسم کرد و علاوه بر تشخیص داده‌های دورافتاده، برای بررسی همسانگردی نیز از آن استفاده نمود.

مانایی و گوسی بودن توزیع داده‌های فضایی کاربرد دارد (پاناتیر، [۱۱]).

۳. ابر تغییرنگار

در نمودار پراکنش h وجود اختلاف زیاد بین $Z(t_i)$ و $Z(t_j)$ ، که در فاصله $h \equiv t_i - t_j$ از یکدیگر قرار دارند، سبب می‌شود که نقطه متناظر با آنها دور از خط نیمساز قرار گیرد. هرچه دو مشاهده اختلاف کمتری داشته باشند نقطه متناظر با آنها به خط نیمساز نزدیک‌تر است و اگر اختلاف، صفر باشد، نقطه روی خط نیمساز قرار می‌گیرد، بنابراین اختلاف $Z(t_i)$ و $Z(t_j)$ اهمیت دارد. بر مبنای این اختلاف می‌توان دوام تفاضل‌های $Z(t_i)$ و $Z(t_j)$ یعنی $[Z(t_i) - Z(t_j)]^2$ را برای مقادیر مختلف محاسبه و در برابر $d_{ij} = \|t_i - t_j\|$ رسم کرد. در اینصورت توده نسبتاً متراکمی از نقاط به وجود می‌آید که ابر توان دوم تفاضل‌ها یا ابر تغییرنگار^۱ نامیده می‌شود. دلیل این نام‌گذاری آن است که از میانگین مقادیر توان دوم تفاضل‌ها در فاصله مشخص h (معمولاً با در نظر گرفتن یک ناحیه تحمل یا فاصله) برآورد کلاسیک تغییرنگار در آن فاصله بدست می‌آید گاهی به جای توان دوم تفاضل‌های فوق نصف آنها یعنی $[Z(t_i) - Z(t_j)]^2 / 2$ در برابر فواصل $d_{ij} = \|t_i - t_j\|$ رسم می‌شود و به آن ابر شبه تغییرنگار می‌گویند (اکبر و گل‌فاند، [۷]).

۴. نمودار جعبه‌ای ابر تغییرنگار

با توجه به تراکم بیش از حد نقاط در ابر تغییرنگار، که شناسایی داده‌های دورافتاده را مشکل می‌سازد، بهتر است نمودار جعبه‌ای مقادیر «توان دوم تفاضل‌های» $[Z(t_i) - Z(t_j)]^2$ در فاصله‌های مختلف رسم شده و مورد بررسی و کنکاش قرار گیرد. گاهی نیز نمودار جعبه‌ای ابر شبه تغییرنگار را با استفاده از مقادیر «نصف توان دوم تفاضل‌ها» رسم می‌نمایند. شکل ۲ نمودار جعبه‌ای شبه تغییرنگار داده‌های سل ریوی را در ۱۸ فاصله متناظر با شکل ۱ نشان می‌دهد. با توجه به این شکل، مشخص است که اولاً مقادیر «نصف توان دوم تفاضل‌ها»، که در برآورد کلاسیک شبه تغییرنگار مورد استفاده قرار می‌گیرند، شدیداً نامتقارن و چوله به راست هستند و ثانیاً مقادیر بسیار بزرگی مشاهده می‌شوند که از داده‌های دورافتاده بدست آمده‌اند. بزرگی این مقادیر به حدی است که چارک‌های اول، دوم و سوم بر روی جعبه‌ها قابل تشخیص نیستند.

به منظور رؤیت این چارک‌ها، در شکل ۳ نمودارهای جعبه‌ای شبه تغییرنگار با صرف نظر کردن از مقادیر بالای ۱۰۰۰ مجدد رسم شده است. این نمودار تراکم نقاط بالای چارک سوم را بهتر از نمودار قبل نشان می‌دهد. در این شکل نمودار برآورد کلاسیک شبه تغییرنگار نیز نمایش داده شده است. همانطور که ملاحظه می‌شود در بیشتر فاصله‌ها نمودار برآورد تغییرنگار بالای چارک سوم قرار دارد که این ناشی از چوله به راست بودن مقادیر «توان دوم تفاضل‌ها» می‌باشد.

اگرچه به دلیل کاهش تعداد نقاط تفسیر و استفاده از این نمودار ساده‌تر از ابر شبه تغییرنگار است، ولی باز هم تشخیص اینکه کدام داده‌ها (ها) سبب بوجود آمدن مقادیر بسیار بزرگ در نمودار جعبه‌ای شده‌اند کار بسیار مشکلی است. از آنجا که طبق تعریف، داده‌هایی که با مشاهدات «نزدیک» یا «همسایه» خود اختلاف زیادی داشته باشند، دورافتاده تلقی می‌شوند، در نمودار ابر تغییرنگار یا نمودار جعبه‌ای آن، به فاصله‌های نخست می‌بایست توجه بیشتری نمود.

شکل ۱، ابر شبه تغییرنگار داده‌های سل ریوی در $n=262$ شهرستان کشور را نشان می‌دهد، که با $n(n-1)/2 = 3419$ زوج از داده‌ها بوجود آمده است. علاوه بر آن، در این شکل، حدود فاصله‌ها (لگ‌ها) با خطوط عمودی نمایش داده شده‌اند. از میانگین عرض نقاط واقع در هر فاصله، که در شکل مذکور با دو خط عمودی محصور شده‌اند، برآورد کلاسیک شبه تغییرنگار در آن فاصله بدست می‌آیند. همان‌طور که در شکل ۱ ملاحظه می‌شود، ابر شبه تغییرنگار در نزدیکی محور افق (مقادیر بین ۰ تا ۱۲۰۰) شامل توده ابر مانندی از نقاط است، تعداد زیادی از نقاط نیز به صورت انفرادی یا جمعی در بالای توده ابر قرار دارند که ناشی از مقادیر بزرگ در داده‌ها می‌باشند. بعنوان مثال، نقاطی که بالای ۲۰۰۰ قرار دارند زوج‌هایی را مشخص می‌کنند که از یک یا چند داده دور افتاده بوجود آمده‌اند. این نقاط، با توجه به مقادیر بسیار بزرگشان، شدیداً برآورد تغییرنگار را تحت تأثیر قرار می‌دهند.

۵. استفاده از باقیمانده‌های پیشگویی

اعتبارسنجی متقابل^۲ به عنوان ابزاری برای سنجش دقت و

برای همه‌ی فواصل h منعکس می‌کند. به عبارت دیگر در ابر تغییرنگار محور افقی فاصله را نشان می‌دهد، لذا هر ابر تغییرنگار متناظر با چند نمودار پراکنش h است. تشخیص دورافتادگی در ابر تغییرنگار از نمودار پراکنش h ساده‌تر است، زیرا تشخیص نزدیکی توان دوم تفاضل‌ها به صفر (خط افق) در ابر تغییرنگار از تشخیص نزدیکی نقاط به خط نیمساز در نمودار پراکنش h ، ساده‌تر است. تفسیر و استفاده از نمودار جعبه‌ای ابر تغییرنگار بدلیل کاهش تعداد نقاط ساده‌تر از ابر تغییرنگار است، اما باز هم تشخیص اینکه کدام داده‌ها (ها) سبب بوجود آمدن مقادیر بزرگ در نمودارهای جعبه‌ای می‌شوند تا حدودی دشوار است.

به نظر می‌رسد از بین روش‌های مختلف ارائه شده در این مقاله روش استفاده از باقیمانده‌های پیشگویی کاراتر از سایر روش‌ها باشد، زیرا این روش اولاً با انجام کریگیدن موقعیت و شکل وابستگی داده‌ها را مورد استفاده قرار می‌دهد، ثانیاً تصمیم‌گیری برای دورافتادگی به کمک آن ساده‌تر از سایر روش‌هاست. منطق این روش نیز جالب است: اگر مقدار واقعی یک داده با مقدار پیشگویی شده برای آن که به کمک سایر داده‌ها بدست آمده اختلاف چشمگیری داشته باشد، دورافتاده محسوب می‌شود. پس از یافتن مقادیر باقیمانده نسبتاً بزرگ می‌توان با توجه به اندیس یا نمودار پراکنش سه بعدی آنها موقعیت یا نام متناظر این داده‌های دور افتاده را پیدا نمود. سپس برای کاهش اثر آنها بر دقت تجزیه و تحلیل داده‌ها باید آنها را کنار گذاشت یا روش‌های استوار را مورد استفاده قرار داد.

نکته قابل ذکر در پایان این است که به منظور جذابیت بیشتر و نیز کاربردی نمودن مقاله از مثالی با داده‌های واقعی استفاده شد. این مثال صرفاً یک مطالعه موردی است و به تنهایی نمی‌تواند نشان دهنده برتری یک روش نسبت به روش‌های دیگر باشد. مقایسه بین روش‌ها در اصل به یک مطالعه شبیه سازی جامع نیاز دارد که موضوع این مقاله نیست.

مقایسه‌ی مدل‌های تغییرنگار مورد توجه می‌باشد. ایده‌ی این روش براساس مقایسه مقادیر داده‌های واقعی با مقادیر پیشگویی شده برای آنها استوار است. از این ایده می‌توان برای شناسایی داده‌های دورافتاده فضایی نیز کمک گرفت. فرض کنید $\hat{Z}_{-j}(t_j)$ نشان دهنده‌ی مقدار پیشگویی شده برای $Z(t_j)$ ، پس از حذف $Z(t_j)$ از داده‌ها و $\hat{\sigma}_{-j}(t_j)$ انحراف معیار آن باشد. چنانچه مقدار پیشگویی $\hat{Z}_{-j}(t_j)$ با مقدار واقعی $Z(t_j)$ اختلاف بسیار زیادی داشته باشد، می‌تواند دلیلی بر دورافتاده بودن مشاهده $Z(t_j)$ باشد. بنابراین می‌توان به روش اعتبارسنجی متقابل، داده‌ها را یکی‌یکی حذف نموده و پس از محاسبه پیشگویی آنها بر اساس مابقی داده‌ها، باقیمانده‌های پیشگویی استاندارد شده را بصورت

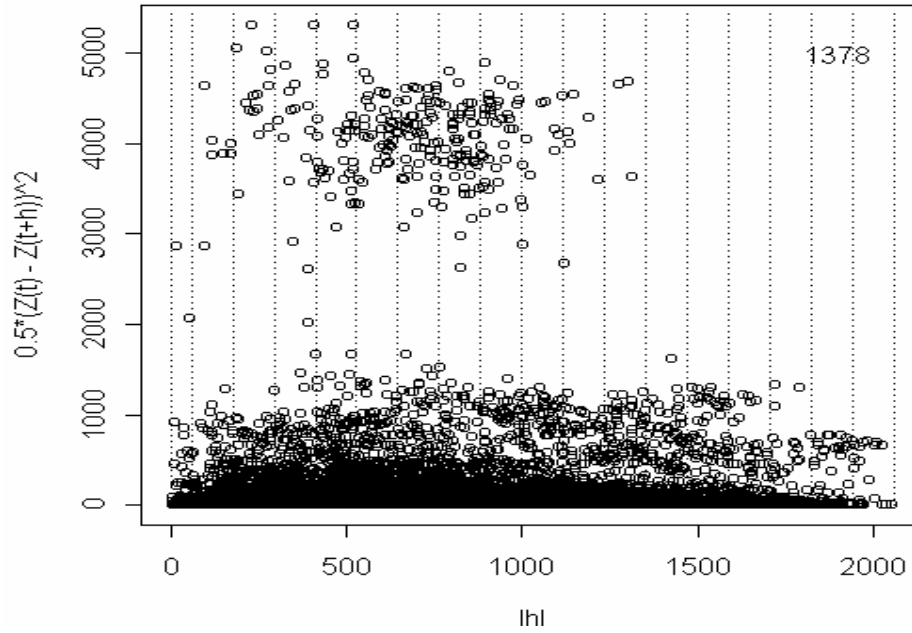
$$e_j = \frac{Z(t_j) - \hat{Z}(t_j)}{\sigma_{-j}(t_j)}, \quad j = 1, \dots, n$$

محاسبه نمود. مقادیر بزرگ باقیمانده‌های استاندارد نشان دهنده داده دور افتاده هستند. با رسم نمودار ساقه و برگ، نمودار پراکنش سه بعدی یا نقشه e_j ها می‌توان تغییرات آنها را مورد بررسی قرار داد.

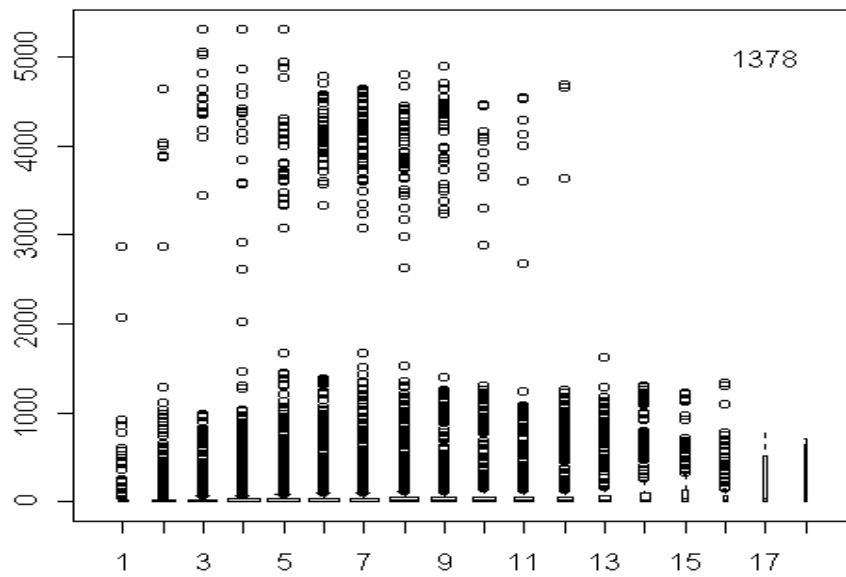
به عنوان مثال شکل ۴، نمودار ساقه و برگ باقیمانده‌های پیشگویی استاندارد شده داده‌های سل ریوی، که در آن $\hat{Z}_{-j}(t_j)$ به روش کریگیدن محاسبه شده است (واقعی، [۱])، را نشان می‌دهد. همان‌طور که در شکل به وضوح ملاحظه می‌شود یک مقدار خیلی بزرگ در دم راست توزیع وجود دارد، که مربوط به شهرستان مینودشت (با مقدار باقیمانده ۱۳/۳) است. علاوه بر آن مقادیر باقیمانده نسبتاً بزرگ ۴/۵، ۵/۶، و ۶/۴ نشان می‌دهند که داده‌های مربوط به شهرستان‌های چابهار، شمیرانات و زابل را نیز می‌توانند دورافتاده تلقی کرد. با توجه به تاثیر سوء این‌گونه داده‌ها بر برآورد کلاسیک تغییرنگار، در تجزیه و تحلیل‌های نهایی ضرورت دارد که این داده‌ها از برآورد تغییرنگار حذف شوند یا از برآوردهای استوار تغییرنگار استفاده شود [۶].

۶. بحث و نتیجه گیری

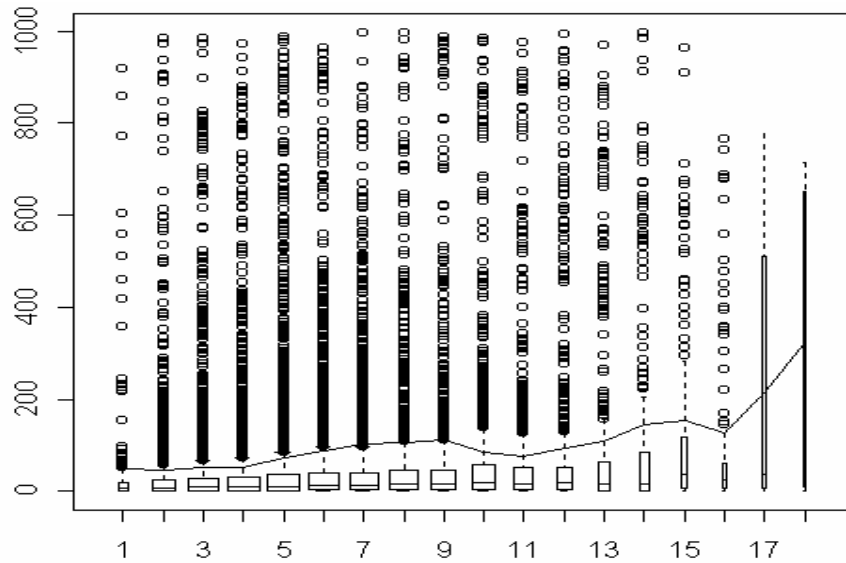
در این مقاله روش‌های مختلف شناسایی داده‌های دورافتاده فضایی معرفی شدند که بیشتر آنها بر پایه نمودارهای گرافیکی هستند. نمودار پراکنش h و ابر تغییرنگار تا حدودی مشابه هستند. یک نمودار پراکنش h فقط به ازای یک بردار فاصله h خاص (و حداکثر در محدوده ناحیه تحمل حول h) وضعیت $Z(t_i)$ در برابر $Z(t_j)$ را نشان می‌دهد. حال آنکه ابر تغییرنگار وضعیت را

شکل ۱: نمودار ابر شبه تغییرنگار داده‌های سل بر حسب فاصله $|h|$ 

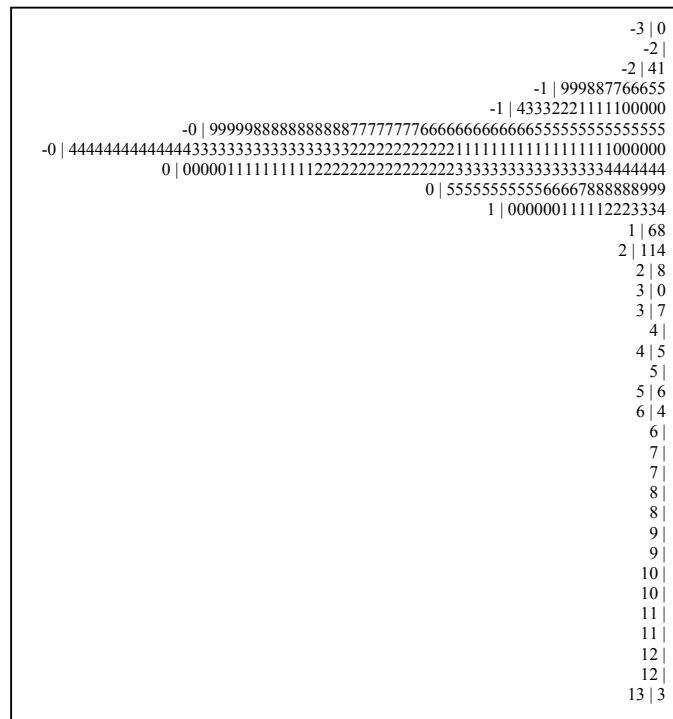
شکل ۲: نمودار جعبه‌ای ابر شبه تغییرنگار داده‌های سل در فواصل مختلف



شکل ۳: نمودار جعبه‌ای ابر شبه تغییرنگار داده‌های سل در فاصله ۰ تا ۱۰۰۰



شکل ۴: نمودار ساقه و برگ باقیمانده‌های پیشگویی استاندارد شده



مراجع

- [۱] واقعی، یدالله ۱۳۸۱، تجزیه و تحلیل داده‌های فضایی نامانا و کاربرد آن در همه‌گیری شناسی جغرافیایی بیماریها، پایان‌نامه دکترای آمارزیستی، دانشگاه تربیت مدرس.
- [2] Bardossy, A. and Kundzewics, Z. W. 1989, Geostatistics methods for detection of outliers in groundwater quality spatial fields, *Journal of Hydrology*, 115, 343-359.
- [3] Barnett, V. and Lewis, T. 1994, *Outliers in Statistical Data*, 3rd Edition, John Wiley. New York.
- [4] Cerioli, A. and Riani, M. 1999, The ordering of spatial data and the detection of multiple outliers, *Journal of Computational and Graphical Statistics*, 8, 239-258.
- [5] Cressie, N. 1993, *Statistics for Spatial Data*, John Wiley, New York.
- [6] Cressie, N. and Hawkins, D. M. 1980, Robust estimation of the variogram, *Journal of the International Association for Mathematical Geology*, 12, 115-125.
- [7] Ecker, M. D. and Gelfand, A. E. 1997, Spatial Modeling and Prediction under Range Anisotropy, *Technical Report*, 97-26, Department of Statistics, University of Connecticut, <http://citeseer.nj.nec.com/ecker99spatial.html>.
- [8] Hadi, A. S. 1992, Identifying multiple outliers in multivariate data, *Journal of the Statistical Society, B*, 54, 761-771.
- [9] Haining, R. P., 1990, *Spatial Data Analysis in Social and Environmental Sciences*, University Press, Cambridge.
- [10] Hawkins, D. M., 1980, *Identification of Outliers*, Chapman and Hall, London.
- [11] Pannatier, Y., 1996, *VARIOWIN: Software for Spatial Data Analysis in 2D*, Springer-Verlag, New York.
- [12] Riani, M. and Cerioli, A., 1999, Graphical tools for the detection of multiple outliers in spatial statistics models, *Proceeding of the 22nd Annual GFKI Conference Dresden*.
- [13] Rousseeuw, P. J. and Leory, A. M., 1987, *Robust Regression and Outlier Detection*, John Wiley, New York.
- [14] Wise, S., Haining R. P. and Signoretta, 1999, Scientific visualization and the exploratory analysis of area data, *Environment and Planning, A*, 31, 1825-1838.