بررسی آلودگی چاه های آب اصفهان با استفاده از آمار فضایی

نصراله ایران پناه، محبوبه شهریاری، پریسا گودرزیان
گروه آمار، دانشگاه اصفهان

دربسیاری از موارد با مشاهداتی سروکار داریم که مشاهدات مستقل از هم نبوده و برحسب مکان و موقعیت قرار گرفتنشان در فضای مورد مطالعه به یکدیگر وابسته هستند. این گونه مشاهدات داده های فضایی نامیده می‌شوند و چون شرط اساسی استقلال داده ها را ندارند از این رو برای تحلیل آنها از آمار فضایی استفاده می‌کنیم. در این مقاله ابتدا مفاهیم اصلی آمار فضایی معرفی می‌شود. سپس با تحلیل فضایی اکتشافی داده های آلودگی چاه های آب اصفهان از جمله هدایت الکتریکی، فرضیات مدل مانند مانایی و همسانگردی بررسی می‌گردد. در نهایت با تعیین ساختار همبستگی فضایی، پیشگوی فضایی کریگینگ و واریانس پیشگو در این منطقه تعیین می‌گردد.

واژه‌های کلیدی: تغییرنگار، همسانگردی، کریگینگ، هدایت الکتریکی.

۱ مقدمه

در آمار معمولاً فرض می‌شود داده‌های به دست آمده از جامعه، مستقل از یکدیگر هستند. اما در مطالعات محیطی مواردی وجود دارند که مشاهدات مستقل از هم نبوده و بر حسب مکان و موقعیت قرار گرفتنشان در فضای مورد مطالعه به یکدیگر وابسته هستند. اگر این وابستگی تابعی از فاصله بین موقعیت مشاهدات باشد، به گونه ای که مشاهدات نزدیک به هم وابسته تر و مشاهدات دورتر از هم وابستگی کمتری داشته باشند، این گونه مشاهدات داده های فضایی نامیده می شوند. بدیهی است که تحلیل آماری داده های فضایی با روش های آماری معمول مقدور نیست، زیرا شرط اساسی استقلال داده ها محقق نمی‌باشد. لذا شاخه آمار فضایی برای تحلیل این گونه مشاهدات شکل گرفته است. درواقع آمارفضایی شاخه ای از علم آمار است که در آن به بررسی متغیرهایی پرداخته می شود که ازخود ساختار همبستگی فضایی نشان می دهند و تلاش می شود این ساختار، که همان ارتباط بین مقادیر متغیر و فاصله و جهت قرارگیری آنهاست، تعیین و برای افزایش دقت در تحلیل آماری آنها مورد

Archive of SID
www.SID.ir

استفاده قرار گیرد.

اولین اظهارات آماری داده های فضایی توسط هالی (۱۶۸۶) جغرافیدان و ستاره شناس انگلیسی صورت گرفته است. وی برای بررسی وزش بادهای موسمی، به نقشه زمین های کشاورزی در نزدیکی نواحی گرمسیر استوایی، مسیر حرکت بادهای موسمی را اضافه کرد و تلاش نمود علل طبیعی آنها را تعیین کند. فیشر (۱۹۳۵) با توجه به آگاهی از وجود همبستگی فضایی در آزمایش های کشاورزی کرت های آزمایشی را در فاصله های بسیار دور از یکدیگر برمی‌گزید. وی قواعد تصادفی سازی، بلوک بندی و تکرار را بنا نهاد. گام تکمیلی بعدی برای توسعه روش های آماری در علوم زمین توسط ماترون (۱۹۶۲) پژوهشگر فرانسوی با انتشار مقاله ای بنیادین برداشته شد. وی با انتشار این مقاله توانست شاخه جدیدی از علم آمار تحت عنوان زمین آمار را پایه گذاری کند. زمین آمار به عنوان زیر بخشی از آمار فضایی و ابزاری کارا برای مطالعه داده های فضایی، که لزوما در گستره علوم زمین قرار ندارند، مطرح شده است. جورنل و هیوجبرگ (۱۹۸۷) به طور مفصل روشهای زمین آماری را با اصطلاحات خاص شرح دادند. ریپلی (۱۹۸۱) در کتاب آمار فضایی پیشگویی فرآیند های تصادفی را مطرح نمود. با ادامه فعالیت ها در زمینه آمار فضایی، کرسی (۱۹۹۳) با جمع‌آوری کلیه اطلاعات پیشین و انتشار کتاب آمار برای داده های فضایی به طور مبسوط به بررسی مباحث مختلف آمار فضایی پرداخته و در تکامل این شاخه از علم آمار قدم بسیار بزرگی برداشته است.

در سال های اخیر نیز توجه ویژه ای به مباحث نظری و کاربردی آمار فضایی در زمینه های مختلف شده و نتایج حاصل از مطالعات آنان در منابع قابل دسترس می‌باشند. برای آشنایی بیشتر با آمار فضایی در یک مرجع فارسی به محمدزاده (۱۳۸۵) مراجعه کنید. آمار فضایی در علوم مختلف مانند زمین شناسی، جغرافیا، کشاورزی، معدن، پردازش تصاویر ماهواره‌ای کاربرد دارد. ایران‌پناه و همکاران (۲۰۰۷) تحلیل فضایی را برای داده های واتنش در زمین شناسی به کار بردند. در این مقاله، کاربرد آمار فضایی را برای تحلیل داده های آلودگی چاه های آب اصفهان به کار می‌بریم. برای این منظور در بخش ۲ مدل آمار فضایی به همراه تعاریف و فرضیات مقدماتی ارائه می‌گردد. در بخش ۳ ساختار همبستگی فضایی ارائه می گردد. پیشگوی فضایی کریگینگ در بخش ۴ ارائه شده است. در بخش ۵ در یک مثال کاربردی تحلیل فضایی داده های آلودگی چاه های آب اصفهان ارائه شده است.

Archive of SID

www.SID.ir

۲ مدل آمار فضایی

برای تجزیه و تحلیل داده‌های فضایی، لازم است یک مدل آماری در نظر گرفته شود. معمولاً یک میدان تصادفی به عنوان مدل آماری برای داده‌های فضایی در نظر گرفته می‌شود. میدان تصادفی مجموعه‌ای از متغیرهای تصادفی مانند $\{Z(s); s \in D \subseteq R^d; d \geq ۱\}$ است، که در آن D یک مجموعه اندیس گذار است. هر میدان تصادفی را می‌توان به صورت $Z(s) = \mu(s) + \delta(s)$ تجزیه کرد، که درآن $\mu(s)$ تغییرات مقیاس بزرگ یا روند و $\delta(s)$ تغییرات مقیاس کوچک یا فرآیند خطای میدان تصادفی است. اگر توزیع توأم هر تعداد متناهی از متغیرهای تصادفی یک میدان دارای توزیع نرمال باشند، میدان تصادفی گوسی خواهد بود. اگر میانگین میدان تصادفی ثابت و به موقعیت s بستگی نداشته باشد، یعنی $E[Z(s)] = \mu$ و واریانس عبارت $[Z(s) - Z(s+h)]$ فقط تابعی از فاصله موقعیتها باشد، یعنی $Var[Z(s) - Z(s+h)] = ۲\gamma(h)$، آنگاه میدان تصادفی را مانای ذاتی می‌نامند. اگر علاوه بر مانایی در میانگین، کوواریانس بین $Z(s)$ و $Z(s+h)$ فقط تابعی از فاصله موقعیتها باشد، یعنی $Cov[Z(s), Z(s+h)] = \sigma(h)$ آنگاه میدان تصادفی را مانای مرتبه دوم می‌نامند. توابع $۲\gamma(h)$ و $\sigma(h)$ که ساختار همبستگی فضایی را مشخص می‌کنند، به ترتیب تغییرنگار[1] و همتغییرنگار[2] نامیده می‌شوند و رابطه $\gamma(h) = \sigma(۰) - \sigma(h)$ بین آنها برقرار است. توابع تغییرنگار و همتغییرنگار به ترتیب همیشه منفی شرطی و همیشه مثبت هستند. تغییرنگار و همتغییرنگار علاوه بر مقدارh، به جهت آن نیز بستگی دارند. اگر مقادیر $۲\gamma(h)$ و $\sigma(h)$ در جهتهای مختلف یکسان و به جهت بستگی نداشته باشند، به عبارت دیگر فقط تابعی از اندازه فاصله h یعنی $| h | = | s_i - s_j |$ باشند، آنها (یا میدان تصادفی) را همسانگرد[3] می‌نامند.

۳ برآورد ساختار همبستگی

در آمار فضایی میزان همبستگی داده‌ها با تغییرنگار و همتغییرنگار اندازه‌گیری می‌شود. افزایش تغییرنگار بیانگر کاهش همبستگی داده‌ها است. با توجه به اینکه معمولاً بعد از فاصله‌ای همبستگی داده‌های فضایی ناچیز و حتی مستقل از یکدیگر می‌شوند، تغییرنگار از این فاصله

[1] Variogram

[2] Covariogram

[3] Isotropic

Archive of SID

www.SID.ir

به بعد تغییر چندانی ندارد و به صورت تخت درمی‌آید، این فاصله، دامنه[۴] تغییرنگار (a) نامیده می‌شود. تغییرنگار بعد از دامنه به مقدار ثابتی گرایش می‌یابد که آستانه[۵] $(c_\circ + c)$ نام دارد. پارامتر دیگر تغییرنگار که معمولاً به دلیل خطاهای اندازه‌گیری ظاهر می‌شود، اثر قطعه‌ای[۶] $c_\circ = \lim_{h\to\circ}\gamma(h)$ است.

در عمل تغییرنگار نامعلوم است و باید بر اساس داده‌های فضایی برآورد شود. یک برآورد تجربی برای تغییرنگار براساس مشاهدات $\{Z(s_1),\ldots,Z(s_n)\}$ به صورت

$$۲\hat{\gamma}(h) = \frac{۱}{N_h}\sum_{i=۱}^{N_h}[Z(s_i+h)-Z(s_i)]^۲,$$

تعریف می‌شود، که در آن N_h تعداد زوج نقاط متمایزی است که به فاصله h از یکدیگر قرار دارند. $۲\hat{\gamma}(h)$ یک برآوردگر نااریب است، ولی مستقیماً نمی‌توان آن را در روشهای مختلف تجزیه و تحلیل داده‌های فضایی از جمله پیش‌بینی به کار برد. زیرا این بر آوردگر لزوماً خاصیت همیشه منفی بودن شرطی تغییرنگار را ندارد. برای تحلیل داده‌های فضایی از جمله پیش‌بینی، لازم است یک مدل به تغییرنگار تجربی داده‌ها برازش داده شود. برای این منظور ابتدا تغییرنگار تجربی مشاهدات در فواصل مختلف محاسبه و نمودار آن رسم می‌شود، سپس با استفاده از روشهای برازش یک مدل پارامتری مناسب برای تغییرنگار انتخاب و پارامترهای آن برآورد می‌گردد. مدلهای مختلفی از جمله مدلهای کروی، نمایی، خطی، گوسی، توانی و موجی مورد استفاده قرار می‌گیرد.

۴ پیشگویی فضایی کریگینگ

کریگینگ یک روش پیش‌بینی کننده در آمار فضایی به ویژه در زمین آمار می‌باشد و معادل بهترین پیش‌بینی کننده نااریب خطی $(BLUP)$ می‌باشد. مسئله پیش‌بینی $Z(s_\circ)$ بر اساس مشاهدات $Z=(Z(s_1),\ldots,Z(s_n))$ را در نظر بگیرید. فرض کنید $P(Z;s_\circ)$ پیش‌بینی کننده بر اساس تابع زیان درجه دوم باشد، در این صورت بهترین پیش‌بینی کننده خطی نااریب و واریانس آن به صورت

$$P(Z;s_\circ) = C'\Sigma^{-۱}(Z-\mu)+\mu(s_\circ),$$

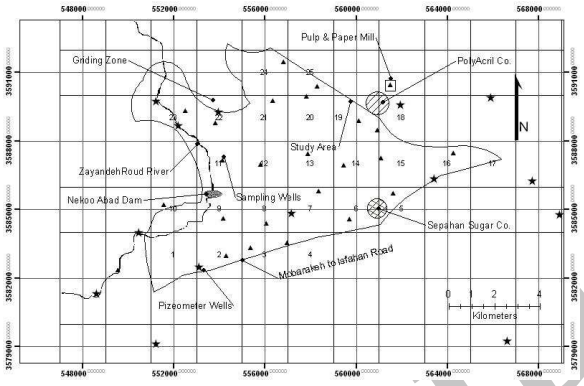
[4] Range

[5] Sill

[6] Nugget Effect

Archive of SID

www.SID.ir



شکلِ ۱: نقشه جغرافیایی منطقه مورد مطالعه.

$$\sigma^{۲}(s_{۰}) = \sigma(s_{۰}, s_{۰}) - C'\Sigma^{-۱}C,$$

به دست می‌آید، که در آن $\mu = (\mu(s_{۱}), \ldots, \mu(s_n))$ ساختار میانگین، Σ یک ماتریس $n \times n$ با عناصر $\sigma(s_i, s_j)$ و $C \equiv (\sigma(s_{۰}, s_{۱}), \ldots, \sigma(s_{۰}, s_n))$ است. با فرض معلوم بودن μ بهترین پیش‌بینی کننده خطی فضایی را کریگینگ ساده، μ ثابت ولی نامعلوم کریگینگ معمولی و چنانچه μ ثابت نباشد، کریگینگ عمومی نامیده می‌شود.

۵ مثال کاربردی

دراین بخش به تحلیل فضایی داده‌های آلودگی چاه‌های آب منطقه لنجان سفلی اصفهان می پردازیم. این داده‌ها برگرفته از مطالعه انجام شده توسط آقای ابراهیمی دانشجوی دکترای بهداشت محیط دانشگاه علوم پزشکی اصفهان است. این داده‌ها در خصوص بررسی آسیب پذیری آبهای زیرزمینی ناشی از اثر مواد شیمیایی کشاورزی شامل آفت کش‌ها، هدایت الکتریکی، تعداد ذرات معلق در آب، کربن آلی، نیترات، جامدات محلول و برخی آنیون‌ها و کاتیون‌ها می‌باشد. این داده ها از ۲۳ چاه نمونه در منطقه‌ای به وسعت ۸۰ کیلومتر مربع در چهار فصل سال ۱۳۸۷ جمع آوری گردیده است. از میان شاخص های آلودگی ذکر شده فقط به بررسی متغیر هدایت الکتریکی[۷] آب چاه در فصل بهار می‌پردازیم و تحلیل فضایی

[7] Electerical Conductivity

Archive of SID
www.SID.ir

سایر متغیرها به طور مشابه انجام پذیرفته است. در نمونه آب چاه‌ها هر چه غلظت یونهای محلول بیشتر باشد شاخص هدایت الکتریکی آب بالاتر خواهد بود. در شکل ۱ موقعیت جغرافیایی منطقه مورد مطالعه نشان داده شده است.

در شکل ۲(a) مقادیر متغیر هدایت الکتریکی بر حسب موقعیت چاه‌ها در منطقه مورد مطالعه نشان داده شده است. با رسم نمودار پراکنش مقادیر متغیر هدایت اکتریکی در مقابل طول و عرض جغرافیایی فرض ایستایی در میانگین مورد پذیرش قرار گرفته است. با رسم تغییرنگار تجربی داده‌ها در چهار جهت صفر، ۴۵، ۹۰ و ۱۳۵ درجه و همچنین رویه تغییر نگار فرض همسانگردی تغییرنگار مورد بررسی قرار می‌گیرد. برای تعیین ساختار همبستگی داده‌ها، ابتدا نیمه‌تغییرنگار تجربی را برآورد نموده و یک مدل تغییرنگار پارامتری کروی

$$۲\gamma(h;\theta)=\begin{cases}۰ & |h|=۰\\ c_۰+c_۱[(\frac{۳}{۲}\frac{|h|}{a})-(\frac{۱}{۲})(\frac{|h|}{a})^۳] & ۰<|h|\le a\\ c_۰+c_۱ & |h|\ge a\end{cases}$$

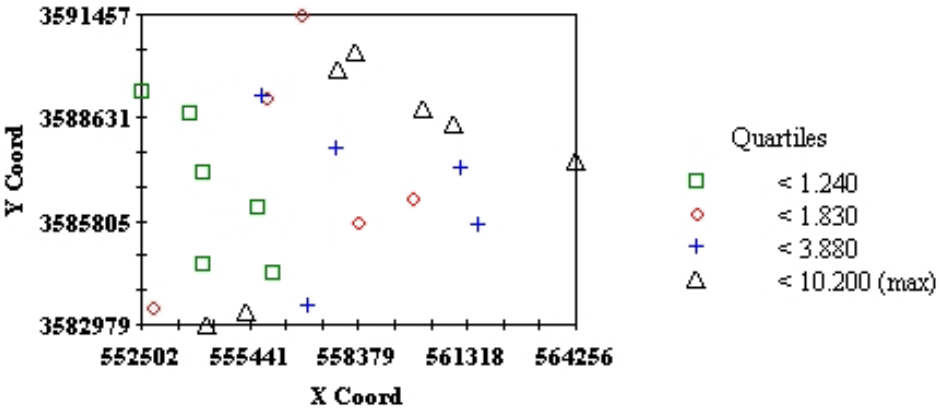
به آنها برازش داده شده است. پارامترهای اثر قطعه‌ای، آستانه جزیی و دامنه با استفاده از نرم افزار $GS+$ به روش حداقل مربعات وزنی به ترتیب به صورت $\hat{\theta}=(\hat{c}_۰,\ \hat{c}_۱,\ \hat{a})=(۲/۵۸, ۵/۵۸, ۷۲۵۰)$ برآورد شده‌اند. مدل تغییرنگار پارامتری کروی از بین مدلهای مختلف نمایی، گوسی، خطی و توانی بر اساس ملاکهای $R^۲$ و RSS انتخاب گردیده است. نمودار نیمه‌تغییرنگار تجربی و برازش نیمه‌تغییرنگار کروی در شکل ۲(b) نشان داده شده است. برای گزارش مقادیر پیشگویی در مکان های جدید یک روش شهودی و کارا ترسیم پهنه بندی پیشگویی است. شکل ۲(c) رویه دو بعدی مقادیر پیشگویی شده متغیر را توسط کریکینگ معمولی نشان می دهد. این شکل نشان می‌دهد در شمال شرق منطقه لنجان آلودگی هدایت الکتریکی بیشتر و در غرب این منطقه این آلودگی کمتر است. وجود کارخانجات صنعتی در این منطقه از جمله عوامل تاثیر گذار بر وجود این نوع آلودگی است. شکل ۲(d) رویه دو بعدی انحراف معیار مقادیر پیشگویی شده را نشان می دهد که بیانگر آن است که پیشگویی در کدام منطقه از دقت بالاتری برخوردار است.

مراجع

[۱] محمد زاده، م. (۱۳۸۵)، آشنایی با آمار فضایی، نشریهٔ دانشجویی آمار (ندا)، شماره ۲، سال ۴، ۱۲ـ۱.

Archive of SID

www.SID.ir



شکلِ ۲: a) موقعیت داده‌ها، b) نیمه‌تغییرنگار تجربی و برازش مدل کروی، c) رویه کریگینگ، d) رویه انحراف معیار کریگینگ.

www.SID.ir

[2] Cressie, N. (1993), *Statistics for Spatial Data*, John Wiley, New York.

[3] Fisher, R. A. (1935), *The Design of Experiments*, Oliver and Boyd, Edinburgh.

[4] Halley E. (1686), *An Historical of the Trade Winds and Monsoons, Observable in the Seas between and near the Tropicks; With an attempt to assign the phisical cause of Said Winds*, *Ohilosophical Thransaction*, **183**, 153-168.

[5] Iranpanah, N., Mohammadzadeh, M., Vahidi Asl, M. G. and Yassaghi A. (2007), Spatial Data Analysis of Finite Strain Data Across a Thrust Sheet Using R Package, *Computer and Geosciences* , **35**, 626-634.

[6] Journel, A. G. and Huijbregts, C. J. (1978), *Mining Geostatistics*, Academic Press, London.

[7] Matheron, G. (1963), *Principles of Geostatistics*, *Economic Geology*, **58**, 1246-1266.

[8] Ripley, B. D. (1981), *Spatial Statistics*, John Wiley, New York.

Archive of SID

www.SID.ir