

بررسی عملکرد دو روش پیش‌بینی داده‌های فضایی

روشنک علی محمدی^{1*}، صدیقه خندان²

1. دانشیار، آمار، دانشگاه الزهرا

2. کارشناسی ارشد، آمار، دانشگاه الزهرا

تاریخ دریافت: 1395/05/01 تاریخ پذیرش: 1395/06/25

Studying of two Spatial Data Prediction Methods Performances

R. AliMohammadi^{*1}, S. Khandan²

1. Associate Professor in Statistics, Alzahra University

2. M.A., Statistics, Alzahra University

Received: 2016/07/22 Accepted: 2016/09/15

Abstract

Correlation structure of spatial data is related to their positions and distances. Then spatial data analysis has various applications in applied areas.

In this research, we study two spatial interpolation methods, that is Ordinary Kriging and Universal Kriging. In this purpose, data sets are simulated and the performance of each of the methods is studied under the considered conditions. Besides, different sample sizes and variograms is considered to assess the effects of various sample sizes and different variogram functions. Furthermore, for every set of simulated data, cross validation criteria is applied to compare the methods. As an application of the results, the methods are applied for data of a mine.

Keywords

Spatial Data, Ordinary Kriging, Universal Kriging, Cross Validation.

چکیده

ساختار وابستگی داده‌های فضایی با موقعیت و فاصله بین آن‌ها مرتبط است؛ لذا تحلیل داده‌های فضایی دارای کاربردهای متعددی در زمینه‌های کاربردی است.

در این پژوهش به مطالعه دو روش درونیابی فضایی یعنی کریجینگ معمولی و عام پرداخته می‌شود. بدین منظور، به شبیه‌سازی مجموعه‌های داده‌ها پرداخته، عملکرد هر یک از این روش‌ها تحت شرایط مورد نظر، ارزیابی می‌شود. همچنین، برای بررسی اثر اندازه نمونه‌های مختلف و توابع تغییرنگار متفاوت، داده‌ها با اندازه‌های نمونه و تغییرنگارهای متفاوت شبیه‌سازی شده‌اند. علاوه بر این، برای هر یک از مجموعه داده‌های شبیه‌سازی شده از معیارهای اعتبارسنجی متقابل برای مقایسه این دو روش استفاده شده است. به عنوان ارائه کاربردی از نتایج حاصل، دقت روش‌های مورد نظر در برآورد مقادیر مربوط به یک معدن استفاده شده است.

واژگان کلیدی

داده‌های فضایی، کریجینگ معمولی، کریجینگ عام، اعتبارسنجی متقابل.

* نویسنده مسئول: روشنک علی محمدی

*Corresponding Author: r_alimohammadi@alzahra.ac.ir

مقدمه

آمار فضایی، شاخه‌ای از آمار است که به تحلیل داده‌هایی که دارای وابستگی فضایی هستند، می‌پردازد. منظور از داده‌های فضایی، داده‌هایی است که از نظر جغرافیایی با یکدیگر ارتباط دارند. یکی از روش‌های پیش‌بینی فضایی کریجینگ است که بر مبنای مطالعات تجربی کریج (1951) شکل گرفته است. ماترون (1962) این روش را از لحاظ نظری گسترش کرد.

روش درون‌یابی کریجینگ، با استفاده از ساختار فضایی مشاهدات به برآورد متغیر در مکان فاقد نمونه (یا مکان هدف) می‌پردازد. کریجینگ نوعی روش میانگین متحرک وزنی است که با استفاده از مشاهداتی که در همسایگی مکان هدف قرار دارند و وزن دهی به این مشاهدات، مقدار متغیر را برآورد می‌کند.

در آمار فضایی میدان تصادفی $Z(\cdot)$ به صورت رابطه (1) تعریف می‌شود:

$$\{Z(s): s \in D\} \quad (1)$$

به طوری که $s \in \mathbb{R}^d$ مکانی متعلق به فضای اقلیدسی d بعدی است و در مجموعه اندیس‌گذار $D \in \mathbb{R}^d$ تغییر می‌کند و $Z(s)$ متغیری تصادفی در مکان s است. میدان تصادفی $Z(\cdot)$ به صورت زیر قابل تجزیه است:

$$Z(s) = \mu(s) + \epsilon(s), \quad s \in D \quad (2)$$

در رابطه (2)، $\mu(\cdot)$ میانگین میدان تصادفی $Z(\cdot)$ در نظر گرفته می‌شود. در کریجینگ معمولی و نشانگر $\mu(\cdot)$ مقدار ثابت و نامعلوم است اما در کریجینگ عام، نامعلوم و تابعی عموماً خطی فرض می‌شود.

علی محمدی (2009) در مقایسه دو روش قطعی اسپالین و روش زمین‌آمار کریجینگ برای یک مجموعه از داده‌های واقعی در داده‌های پزشکی پرداخته است.

متکان و همکاران (2010) در مقایسه دو روش زمین‌آمار کریجینگ معمولی و کوکریجینگ با دو روش رگرسیون کمترین توان‌های دوم معمولی و رگرسیون جغرافیایی وزنی نشان می‌دهد کوکریجینگ نسبت به سه مدل دیگر نتایج بهتری ارائه داده است.

در این نوشتار با مطالعه شبیه‌سازی دو روش کریجینگ معمولی و عام، به بررسی عملکرد این دو روش پرداخته شده است. بدین منظور، از دو شاخص اعتبارسنجی متقابل شامل مجموع توان‌های دوم خطا (SSE) و ریشه میانگین توان

دوم خطا (RMSE) استفاده شده است. هر یک از این دو شاخص، تابعی از میزان خطای هر مدل را محاسبه می‌کند (کرسی، 1993).

کریجینگ معمولی و عام

در این بخش، به مرور مفاهیم دو روش کریجینگ معمولی و عام پرداخته می‌شود.

کریجینگ معمولی

در کریجینگ معمولی مقدار میانگین $(\mu(\cdot))$ در رابطه (2)، نامعلوم اما برابر مقداری ثابت فرض می‌شود. کریجینگ معمولی را ماترون در اوایل دهه شصت میلادی گسترش داد و به دلیل اینکه نسبت به شرایط مانایی منعطف است، پرکاربردترین روش کریجینگ محسوب می‌شود.

در کریجینگ معمولی، برآورد $Z(\cdot)$ در مکان نامعلوم s_0 به صورت رابطه (3) محاسبه می‌شود:

$$\hat{Z}(s_0) = \sum_{i=1}^N I_i Z(s_i), \quad (3)$$

$$\sum I_i = 1$$

به طوری که N تعداد مشاهدات و $Z(s_i)$ مقدار واقعی مشاهده در مکان s_i می‌باشد. I_i وزن‌های به دست آمده با استفاده از تغییرنگار نظری است. شرط $\sum I_i = 1$ نارایب بودن برآوردگر را تضمین می‌کند؛ یعنی تحت این شرط میانگین خطای برآوردگر صفر خواهد بود:

$$E[\hat{Z}(s_0) - Z(s_0)] = 0 \quad (4)$$

نسخه نامانای کریجینگ معمولی، کریجینگ عام نام دارد که در ادامه بررسی خواهد شد.

کریجینگ عام

زمانی که شرط مانایی (حتی مانایی ذاتی) برقرار نباشد از کریجینگ عام استفاده می‌شود. در این روش $\mu(\cdot)$ یا میانگین میدان تصادفی $Z(\cdot)$ ، نامعلوم و تابعی عموماً خطی است و روند نامیده می‌شود. در کریجینگ عام نیز میدان تصادفی $Z(\cdot)$ به صورت رابطه کلی (2) قابل تجزیه است که در آن $\mu(s)$ جزء قطعی و $\epsilon(s)$ جزء تصادفی در نظر گرفته می‌شود. استین (2005) روشی برای برآورد اجزای این مدل ارائه کرد.

همان‌طور که اشاره شد در کریجینگ عام، میانگین ثابت نیست و می‌تواند تابعی از مختصات یا تابعی از یک یا چند

معیارهای اعتبارسنجی متقابل هر یک از روش‌های کریجینگ برای ارزیابی بهترین روش، مقایسه شده است. (لازم به ذکر است که در شبیه‌سازی داده‌ها برای منطقی بودن انتخاب پارامترهای شبیه‌سازی از حدود مجموعه داده‌های واقعی معدن استفاده شده است. همچنین نتایج ارائه شده با میانگین گرفتن از تکرار شبیه‌سازی‌ها در هر مورد به دست آمده است.)

جدول 1. محاسبه شاخص‌های اعتبارسنجی متقابل دو روش کریجینگ (تغییرنگار مدور)

مدل تغییرنگار	انداز ه نمونه	شاخص‌های اعتبارسنجی متقابل	شاخص‌های اعتبارسنجی متقابل	
			SSE	RMSE
مدور	49	کریجینگ معمولی	499.529	3.181
		کریجینگ عام	516.915	3.237
	100	کریجینگ معمولی	629.817	2.502
		کریجینگ عام	631.288	2.504
	225	کریجینگ معمولی	1033.698	2.141
		کریجینگ عام	1039.946	2.148
	400	کریجینگ معمولی	1471.582	1.917
		کریجینگ عام	1479.723	1.922

با توجه به نتایج جدول 1، شاخص‌های اعتبارسنجی متقابل برای روش کریجینگ معمولی (در اندازه‌های متفاوت و برای تغییرنگار مدور) مقادیر کمتری از روش دیگر نشان داده است.

جدول 2. محاسبه دقت روش‌های کریجینگ (تغییرنگار کروی)

مدل تغییرنگار	انداز ه نمونه	شاخص‌های اعتبارسنجی متقابل	شاخص‌های اعتبارسنجی متقابل	
			SSE	RMSE
کروی	49	کریجینگ معمولی	507.089	3.205
		کریجینگ عام	525.248	3.261
	100	کریجینگ معمولی	639.766	2.521
		کریجینگ عام	642.256	2.526
	225	کریجینگ معمولی	1041.527	2.150
		کریجینگ عام	1049.267	2.158
	400	کریجینگ معمولی	1436.198	1.894
		کریجینگ عام	1438.518	1.895

متغیر کمکی باشد. در این تحقیق برای محاسبه کریجینگ عام، از وجود یک متغیر کمکی بهره گرفته شده است به طوری که علاوه بر متغیر اصلی یک متغیر کمکی نیز شبیه‌سازی شده است.

اعتبارسنجی متقابل

اعتبارسنجی متقابل مشهورترین روش در زمینه ارزیابی اعتبار روش درون‌یابی به کار برده شده است. روش کار به این صورت است که هر یک از داده‌ها به نوبت حذف شده، سپس با استفاده از داده‌های مجاور (به روش کریجینگ) مقدار آن برآورد می‌شود، سپس خطای برآورد با استفاده از حاصل تفریق مقدار برآورد از مقدار واقعی محاسبه می‌شود. تکرار این روش برای تک تک مشاهدات، نوعی ارزیابی از تغییرپذیری خطای برآورد به دست می‌دهد. برای انتخاب بهترین روش درون‌یابی از بین چند گزینه موجود، می‌توان از اعتبارسنجی متقابل استفاده کرد.

در این تحقیق از دو آماره تشخیصی اعتبارسنجی متقابل یعنی مجموع توان دوم خطا (SSE) و ریشه میانگین توان دوم خطا (RMSE) استفاده شده است و معیار مقبولیت برای هر دو شاخص نزدیک به صفر بودن آن است. در ادامه معادلات مربوط به هر یک از این دو شاخص آورده شده است:

$$SSE = \sum_{i=1}^N [z(s_i) - \hat{Z}(s_i)]^2 \quad (5)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N [z(s_i) - \hat{Z}(s_i)]^2} \quad (6)$$

به طوری که در روابط (5) و (6)، N تعداد مشاهدات، $z(s_i)$ مقدار واقعی مشاهده در مکان s_i و $\hat{Z}(s_i)$ مقدار برآورد شده با فرض حذف مقدار مشاهده شده و برآورد آن به وسیله سایر مقادیر در مکان s_i است.

شبیه‌سازی و مقایسه روش‌های کریجینگ

در این بخش نتایج هر یک از دو روش کریجینگ معمولی و عام بر روی داده‌های شبیه‌سازی شده آمده است. شبیه‌سازی برای اندازه‌های متفاوت نمونه و مدل‌های متفاوت تغییرنگار - مدور، کروی، نمایی، گاوسی و موجی - تکرار شده است و با استفاده از اعتبارسنجی متقابل،

برای اندازه نمونه‌های بزرگ، هر دو روش برای تغییرنگار گاوسی تقریباً یکسان عمل می‌کنند.

جدول 5. محاسبه دقت دو روش کریگینگ (تغییرنگار موجی)

مدل تغییرنگار	اندازه نمونه	روش درون‌یابی	شاخص‌های اعتبارسنجی متقابل	
			SSE	RMSE
موجی	49	کریگینگ معمولی	486.665	3.090
		کریگینگ عام	505.283	3.133
	100	کریگینگ معمولی	520.640	2.274
		کریگینگ عام	525.321	2.284
	225	کریگینگ معمولی	922.929	2.024
		کریگینگ عام	926.822	2.029
	400	کریگینگ معمولی	1386.048	1.860
		کریگینگ عام	1390.956	1.863

با توجه به جدول 5، شاخص‌های اعتبارسنجی متقابل برای روش کریگینگ معمولی در تمامی اندازه‌های نمونه با تغییرنگار موجی مقادیر کمتری از روش کریگینگ عام دارد. چنان که ملاحظه می‌شود، نتایج حاصل از شبیه‌سازی‌ها در جدول‌های 1 تا 5 نشان‌دهنده عملکرد بهتر کریگینگ معمولی نسبت به کریگینگ عام به منظور درون‌یابی داده‌های فضایی است.

کاربرد

به منظور ارائه کاربردی از نتایج حاصل در این مقاله، از داده‌های مربوط به یک معدن مس استفاده شده است. از لحاظ عملی، مهندسین معدن ترجیح می‌دهند تا با روش‌های فضایی بتوانند درصد و عیار ماده معدنی را در نقاط حفاری نشده برآورد کنند، زیرا حفاری و مشاهده مستقیم می‌تواند باعث آسیب به معدن و افزایش هزینه‌های اکتشاف گردد.

داده‌ها مربوط به عیار فلز مس در نقاط نمونه‌گیری شده از معدن است که طول و عرض جغرافیایی نقاط به عنوان موقعیت آن‌ها به کار می‌رود. هدف، یافتن روشی مناسب برای پیش‌بینی در نقاط بدون داده است. در این مسئله از دو روش کریگینگ معمولی و عام با مدل‌های تغییرنگار گاوسی، کروی، نمایی و موجی استفاده شده و مقدار حاصل

با توجه به جدول 2، شاخص‌های اعتبارسنجی متقابل برای روش کریگینگ معمولی (در اندازه‌های متفاوت و برای تغییرنگار کروی) مقادیر کمتری از روش دیگر کریگینگ دارد.

جدول 3. محاسبه دقت دو روش کریگینگ (تغییرنگار نمایی)

مدل تغییرنگار	اندازه نمونه	روش درون‌یابی	شاخص‌های اعتبارسنجی متقابل	
			SSE	RMSE
نمایی	49	کریگینگ معمولی	547.030	3.327
		کریگینگ عام	564.923	3.381
	100	کریگینگ معمولی	710.282	2.658
		کریگینگ عام	716.687	2.670
	225	کریگینگ معمولی	1129.498	2.239
		کریگینگ عام	1137.716	2.247
	400	کریگینگ معمولی	1414.604	1.879
		کریگینگ عام	1417.98	1.881

جدول 3 نشان می‌دهد که شاخص‌های اعتبارسنجی متقابل برای روش کریگینگ معمولی (در اندازه‌های متفاوت و برای تغییرنگار نمایی) مقادیر کمتری از روش کریگینگ عام دارد. در نتیجه در داده‌های شبیه‌سازی شده با تغییرنگار نمایی نیز روش کریگینگ معمولی برآورد دقیق‌تری از روش دیگر داشته است.

جدول 4. محاسبه دقت روش‌های کریگینگ (تغییرنگار گاوسی)

مدل تغییرنگار	اندازه نمونه	روش درون‌یابی	شاخص‌های اعتبارسنجی	
			SSE	RMSE
گاوسی	49	کریگینگ معمولی	525.632	3.261
		کریگینگ عام	549.166	3.333
	100	کریگینگ معمولی	601.348	2.443
		کریگینگ عام	608.531	2.457
	225	کریگینگ معمولی	999.106	2.105
		کریگینگ عام	999.290	2.105
	400	کریگینگ معمولی	1452.752	1.905
		کریگینگ عام	1448.225	1.901

شاخص‌های اعتبارسنجی متقابل برای روش کریگینگ معمولی در اندازه نمونه‌های کوچک‌تر نشان‌دهنده دارای عملکرد بهتر این روش نسبت به کریگینگ عام است و

مقایسه‌ای بین مقادیر معیارهای اعتبارسنجی متقابل دو روش کریجینگ معمولی و عام با تغییرنگار کروی در جدول‌های 6 و 7 نشان می‌دهد که کریجینگ معمولی دارای دقت بیشتری نسبت به کریجینگ عام در پیش‌بینی نقاط حفاری نشده برای مجموعه داده‌های مورد بررسی است.

نتیجه‌گیری

روش‌های کریجینگ دارای کاربردهای فراوان در درون‌یابی و پیش‌بینی داده‌های فضایی است. در این مقاله به بررسی دقت دو روش کریجینگ معمولی و عام پرداخته شده است. با توجه به نتایج شبیه‌سازی که برای اندازه‌های مختلف و مدل‌های متفاوت تغییرنگارها تکرار شده‌اند، شاخص‌های اعتبارسنجی متقابل (که نشان‌دهنده برای روش کریجینگ معمولی تقریباً در تمام موارد، مقادیر کمتری از کریجینگ عام حاصل شده است. لذا تحت شرایط مورد نظر، روش کریجینگ معمولی عملکرد بهتری از روش دیگر دارد و به کارگیری آن به منظور درون‌یابی داده‌های فضایی توصیه می‌شود.

از معیارهای اعتبارسنجی متقابل در جدول‌های 6 و 7 آمده است.

جدول 6. معیارهای اعتبارسنجی متقابل برای داده‌ها با برازش چهار مدل تغییرنگار در کریجینگ معمولی

	کروی	نمایی	گاوسی	موجی
SSE	837.690	903.869	869.483	1048.369
RMSE	3.459	3.593	3.524	3.869

چنان که در این جدول ملاحظه می‌شود، تغییرنگار کروی بهترین برازش در بین چهار مدل تغییرنگار برای روش کریجینگ معمولی به داده‌ها را دارد.

جدول 7. معیارهای اعتبارسنجی متقابل برای داده‌ها با برازش چهار مدل تغییرنگار در کریجینگ عام

	کروی	نمایی	گاوسی	موجی
SSE	901.395	930.467	935.637	1077.974
RMSE	3.588	3.645	3.655	3.924

نتایج جدول 7 نشان می‌دهد که تغییرنگار کروی نسبت به سایر مدل‌ها دارای خطای کمتری در روش کریجینگ عام است و از این مدل می‌توان در برآورد داده‌ها استفاده کرد.

منابع

- Alimohammadi, R. (2009). Comparison of Spline with Kriging in an Epidemiological Problem. *Ip Journal of Biostatistics* 3 (3), 187 – 193.
- Cressie, N. (1993). *Statistics for Spatial Data*. Wiley. New York.
- Krige, D.G. (1951). A Statistical Approach to Some Basic Mine Valuation Problem on the Witwatersand. *Journal of the Chemical, Metallurgical and Mining Society of South Africa*. 52(6). 119-139.
- Matheron, G. (1963). Principles of Geostatistics. *Economic Geology*, 58(8). 1246-1266.
- Matkan, A., Shakiba, A., Mirbagheri, B. & Tavooosi, H. (2010), A comparison between Kriging, Cokriging and Geographically Weighted Regression models for estimating rainfall over north west of Iran. *EMS Annual Meeting Abstracts*. Sep 2010. Zurich. Switzerland.
- Stein, M.L. (2005). Statistical Methods for Regular Monitoring Data. *Journal of Royal Statistical Society B*, 67, 667-687.