مقاله

مقایسه دو روش درون یابی IDW و KRIGING

جهت تهیه نقشه پهنه بندی املاک (عرصه) ناحیه ۵ شهرداری منطقه یک

هرمز صفری / کارشناس ارشد شرکت پردازش و برنامه ریزی شهری (TGIC)

تقدیر و تشکر:
جا دارد از زحمات آقای مهندس فلاح مدیریت محترم برنامه ریزی و فروش که فرصت این تحقیق را فراهم ساختند و همچنین همکاران گرامی‌آقای مهندس جباری و خانم‌ها پدرامی‌و تکه که در نحوه انتخاب و جانمایی نمونه‌ها کمک کردند، تشکر نمایم.

چکیده

برای تهیه نقشه یکپارچه و پیوسته از مکان مورد نظر نیاز به روش‌های درون یابی به منظور پیش بینی مقادیر نامعلوم می‌باشد. در این خصوص روش‌های مختلف درون یابی و تخمین‌های زمین آماری وجود دارد. عوامل مهمی مانند تعداد، توزیع مکانی نقاط نمونه برداری و توانایی مدل درون یابی نقش موثری در دقت تهیه نقشه پهنه بندی دارند. روش‌های متعددی برای درون یابی یک متغیر ارائه شده‌اند که روشهای چند ضلعی‌های تیسن و وزن دهی بر مبنای عکس فاصله[1] (IDW)از جمله روش‌های مرسوم می‌باشند. به هر حال، این روشها به دلیل در نظر نگرفتن همبستگی بین داده‌ها و بهینه نبودن اندازه و شکل همسایگی مورد استفاده، غالبا از دقت کافی برخوردار نمی‌باشند نظریه زمین آماری شاخه‌ای از آمار کاربردی است که نقش موثری در تشریح و تجزیه وتحلیل مشاهدات زمینی ایفا می‌کند. توسعه تئوری‌ها و تکنیکهای زمین آمار در سالهای اخیر تحولات زیادی در تجزیه و تحلیل داده‌های مکانی در محیط GISایجاد کرده است. استفاده از آنالیزهای مبتنی بر تئوری‌های زمین آمار نواقص و محدودیت‌های مربوط به آمار کلاسیک را که غالبا" با فرضیه توزیع تصادفی متغیرها و فرایندها توسعه یافته‌اند تا حد زیادی کاهش می‌دهد. درونیابی داده‌ها در تجزیه و تحلیل مکانی به کمک GIS از اهمیت ویژه‌ای برخوردار است، چرا که بسیاری از نقشه‌های مورد استفاده در عملیات GISاز طریق درون یابی تولید می‌شوند. در واقع تولید مدلهای هموار و پیوسته از توزیع مکانی و زمانی داده‌های مورد بررسی از طریق درون یابی امکان پذیر می‌باشد.. در تحقیق، دو روش IDW، Kriging برای پهنه بندی قیمت زمین مورد بررسی قرار گرفت و به معایب و مزایای آنها پرداخته شد. ارزیابی و تست خطا دو مدل درون یابی با روش (Willmott,۱۹۸۴)، مورد بررسی قرار گرفت که حاکی از دقت بالاتر روش Krigingنسبت به روش IDWمی‌باشد.

کلید واژه ها : پهنه بندی، IDW، Kriging

مقدمه

GIS مجموعه‌ای است از یک سیستم سخت افزاری و نرم افزاری که قادر است اطلاعات مکانی و توصیفی را بهم مرتبط سازد. GIS دارای اجزای مختلفی از جمله ورودی داده ها، مدیریت داده ها(ذخیره، بازیابی، کنترل و تهیه بانک داده و ...)، تجزیه و تحلیل داده‌ها و خروجی بصورت Soft copy و Hard cooy می‌باشد. پهنه بندی یکی از روشهای تحلیل مکانی که از داده‌های مکانی بصورت نقطه و خطوط هم ارزش، سطح پیوسته‌ای بوجود می‌آورد که این سطح نشان دهنده تغییرات ارزش متغیر محیطی می‌باشد.

توابع مورد استفاده در پهنه بندی به دو صورت ریاضی و آمار است. در حالت آماری دموگرافی از داده‌ها و در حالت ریاضی بیان عددی و محاسباتی از داده‌ها را نشان می‌دهد. برای تهیه نقشه پهنه بندی از طریق درون یابی (Interpolation) که طبق تعریف پیش بینی و محاسبه مقادیر نقاط و سطوح نامعلوم با استفاده از نقاط معلوم می‌باشد که این پیش بینی بر اساس خود همبستگی مکانی تعیین می‌شود.

روشهای موجود برای درون یابی به دو صورت قطعی (Detereministic) و غیر قطعی (Geostatistic) است. در روش قطعی از توابع ریاضی بر حسب فاصله دوری یا نزدیکی نقاط نمونه نسبت به همسایه ها(IDW) و یا درجه صافی و همواری (Spline) مورد استفاده قرار می‌گیرد. در روش غیر قطعی علاوه بر استفاده از توابع ریاضی از آمار برای پیش بینی نقاط نامعلوم اسفاده می‌شود. در این روش میزان سطح اطمینان و درستی پیش بینی نقاط محاسبه می‌شود. تعاریفی از این روش آمده که به عنوان مثال به مطالعه پدیده‌ای که در فضا و زمان دارای تغییر می‌باشد(Deutsch,۲۰۰۲) و زمین آمار توصیف پیوستگی مکانی پدیده طبیعی و مطابق با تکنیک‌های آمار کلاسیک است (Isaaks and et al، ۱۹۸۹). نکته حائز اهمیت این است که برای حالات مختلف از نظر توزیع مکانی هیچ گاه از یک روش به عنوان مدل بهینه پیشنهاد نمی‌گردد (رنگزن و دیگران، ۱۳۸۴)، چون از یک مدل خاص در شرایط مختلف دقتهای متفاوتی بدست می‌آید. دقت و صحت درون یابی را پراکندگی نقاط و توزیع مکانی آنها، تراکم و چگالی نقاط و توانایی روش واسطه یابی مشخص می‌کند.

روش IDW (Inverse Distance Weight)

کلیه روشهای درون یابی بر مبنای این فرضیه توسعه یافته‌اند که نقاط نزدیک‌تر به یکدیگر نسبت به نقاط دورتر همبستگی و تشابه بیشتری دارند. در روش IDW فرض اساسی بر این است که میزان همبستگی و تشابه بین همسایه‌ها با فاصله بین آنها متناسب است، که می‌توان آن را به صورت تابعی با معکوس از فاصله هر نقطه از نقاط همسایه تعریف کرد. لازم به یادآوری است که تعریف شعاع همسایگی و توان مربوط به تابع عکس فاصله از مسایل مهم در این روش محسوب می‌شود. این روش در حالتی که نقاط نمونه به‌اندازه کافی با پراکنش مناسب در سطوح مقیاس محلی باشد، مورد استفاده قرار می‌گیرد. یکی از مزایای این روش در نشان دادن خطوط نا پیوسته Barriers مانند شکستگی ها، خطواره ها، گسل ها، Levees و رودخانه‌ها که ایجاد شکستگی و ناپیوستگی در سطح می‌کنند، مناسب می‌باشد.

$$Z_o = \frac{\sum_{i=1}^{N} z_i d_i^{-n}}{\sum_{i=1}^{N} d_i^{-n}}$$

Z_o= مقدار تخمین متغیر z در نقطه i

Z_i= مقدار نمونه در نقطه i

d_i= فاصله نقطه نمونه تا نقطه مورد تخمین

n= ضریبی که وزن را براساس فاصله تعیین می‌کند

Geostatistics[۲]

روشهای زمین آمار بر اساس مدلهای آماری که شامل خود همبستگی (ارتباط آماری نقاط اندازه گیری شده) است، محاسبه می‌شود. برای انجام روشهای زمین آمار، داده‌های مورد استفاده باید دو ویژگی داشته باشند، یکی توزیع نرمال فضایی داده‌ها و دیگری میانگین و واریانس در مکان تغییر معنی داری نداشته باشند(Stationary). این دو ویژگی در هیستوگرام توزیع نرمال و واریانس داده‌ها قابل مشاهده است. بیان ریاضی روش زمین آمار در رابطه زیر آمده است:

$$Z(u) = \mu(u) + \varepsilon(u)$$

$Z(u)$ = مقدار متغیر ناحیه‌ای

$\mu(u)$ = روند قطعی و مولفه ساختاری $Z(u)$

$\varepsilon(u)$ = خطای نمونه نسبت به $\mu(u)$ یا مولفه تصادفی

$\varepsilon(u)$ = ساختار مولفه تصادفی اصطلاحا فضایی می‌نامند که تابعی از فاصله و مستقل از مختصات است بطوریکه ارزش نمونه‌ها نسبت بهم نزدیک باشد و دور شدن فاصله نقاط از هم، تفاوت ارزش آنها افزایش یابد نشان دهنده پیوستگی یک متغیر ناحیه‌ای است.

محاسبه فرمول $\mu(u)$ بر حسب ویژگی‌های محیطی به صورت یک تابع چند جمله‌ای زیر می‌باشد.

$$\mu(u) = ß_0 + ß_1x + ß_2y + ß_3x^2 + ß_4y^2 + ß_5xy$$

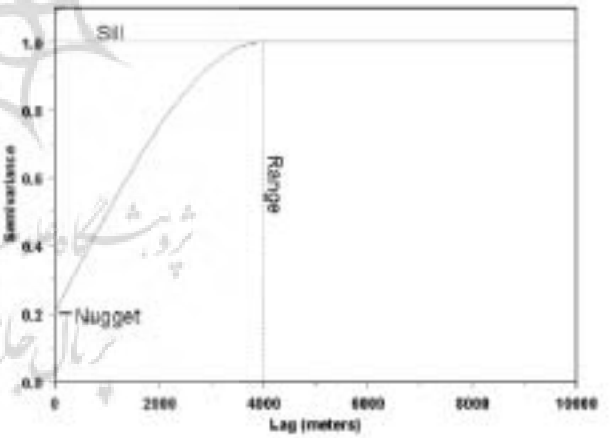
برای تخمین زمین آماری مقادیر نامعلوم در مکان مورد نظر دو مرحله باید انجام گردد. ابتداء شناخت و مدلسازی ساختار فضایی نمونه‌ها که در این مرحله حالت همگنی یا نا همگنی نمونه‌ها در نموداری بنام واریوگرام مورد بررسی قرار می‌گیرد. در مرحله دوم تخمین مقادیر به روش Kriging است.

در مرحله اول بررسی ساختار مکانی نمونه‌ها در منحنی‌های واریوگرام می‌باشد. مشخصات این منحنی در محور y مقدار واریانس مکانی نمونه‌ها نسبت بهم و در محور x مقدار فاصله بر حسب تعداد گام (Lag) که در فرمول به h نشان داده شده است(نمودار شماره۱).

ابتدا با افزایش hمقدار واریوگرام زیاد ولی در یک نقطه‌ای به حد اشباع می‌رسد که به آن Sill که برابر $Z(x+h)$ است.

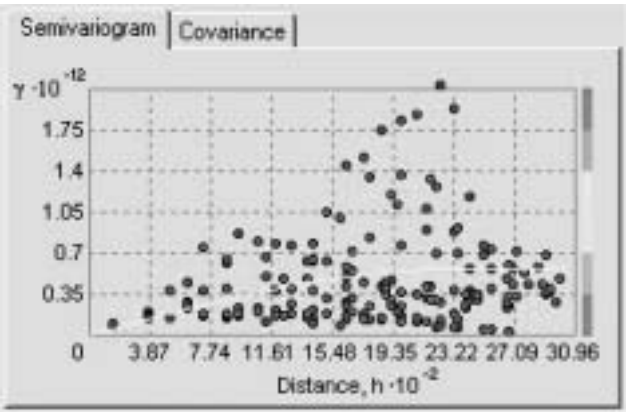
nugget بنام واریانس تصادفی و بدون ساختار که تابع جهت و فاصله نیست. فرمول محاسبه واریوگرام در زیر آمده است.

$$\lambda^{*}(h) = \frac{1}{2(N-h)}\sum_{i=1}^{N-h}\left[Z(x+h) - Z(x)\right]^2$$



نمودار شماره ۱: مشخصات منحنی واریوگرام

مدلهای تئوری واریوگرام بر دو نوع است، یکی واریو گرامهایی که در آنها با افزایش h مقدار $\lambda(h)$ افزایش می‌یابد و تمایلی به نزدیک شدن به یک سقف ثابتی را ندارد. دوم واریوگرامهایی که ابتدا با افزایش hمقدار $\lambda(h)$ افزایش می‌یابد و سپس به حد ثابتی (که همان سقف است) می‌رسد. در مدلهای سقف دار، سقف برابر واریانس مقدار متغیر در جامعه نمونه‌های برداشت شده است.



نمودار شماره ۲: منحنی واریوگرام نمونه‌ها در این تحقیق

مرحله دوم تخمین مقادیر نا معلوم با استفاده درون یابی که در این مرحله از روش Kriging استفاده شده است. روشهای مختلفی برای تخمین وجود دارد یکی به روش آمار کلاسیک و دوم روش زمین آمار که بر اساس ساختار فضایی موجود در محیط و بر منطق میانگین متحرک وزن دار استوار است. یکی از ویژگیهای این روش آن است که به ازای هر تخمین خطای مرتبط با آن را می‌توان محاسبه کرد. برای اجرای این روش از تخمین اول متغیر z توزیع نرمال داشته باشد در غیر اینصورت باید از Kriging غیر خطی یا داده‌ها را به نرمال تبدیل که در آن مرحله از لگاریتم برای تبدیل استفاده می‌کنیم.

فرمول محاسبه درون یابی به روش Kriging

$$Z_v = \sum_{i=1}^{n} \lambda_i Z_{vi}$$

Z_{vi} مقدار تخمین و λ_i وزن کمیت وابسته به نمونه و Z_v ارزش اولیه نمونه i می‌باشد

برآورد مقدار از معادله زیر بدست می‌آید:

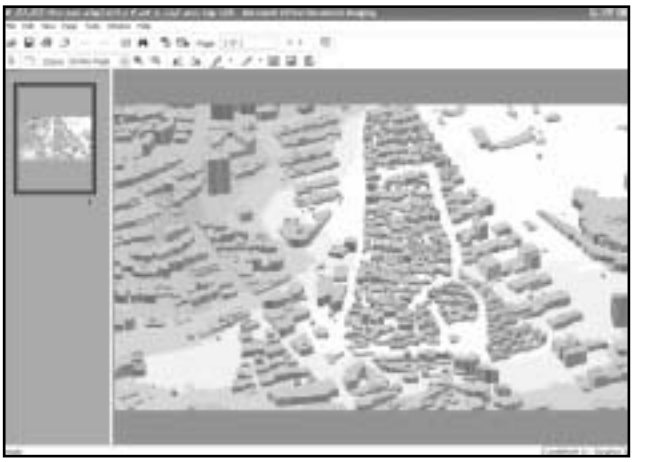
$$\begin{bmatrix} c_{11} & c_{12} & . & . & c_{1n} \\ . & . & . & & . \\ . & . & . & & . \\ . & . & . & & . \\ c_{n1} & . & . & . & c_{nn} \end{bmatrix} \begin{bmatrix} \lambda_1 \\ . \\ . \\ . \\ \lambda_n \end{bmatrix} = \begin{bmatrix} c_{o1} \\ . \\ . \\ . \\ c_{on} \end{bmatrix}$$

c_{ij} میانگین کوواریوگرام برای حالتی که بردار h در دو نمونه i و j تغییر کند.

c_{oi} میانگین کوواریوگرام برای حالتی که بردار h در نمونه i و نمونه پیش بینی o قرار گیرد.

λ_i ضرایب Kriging

از ویژگیهای مهم این روش اینست که واریانس تخمین را در تمام سطح برای هر نمونه محاسبه می‌کند و این به ما کمک می‌کند که



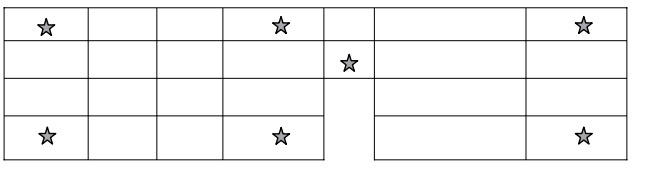
نقشه پهنه بندی قیمت به همراه ساختمانهای جدید محله کاشانک

نمونه‌ها که واریانس تخمین در آنها بالا باشد نشان دهنده نقاط پرت در مکان مورد نظر می‌باشد. دیگر ویژگی آن نرم شدن تغییرات (کاهش نوسانات) یعنی کاهش تغییرات بین مقدار پیش بینی با واقعی می‌شود
(حسنی پاک، ۱۳۷۷).

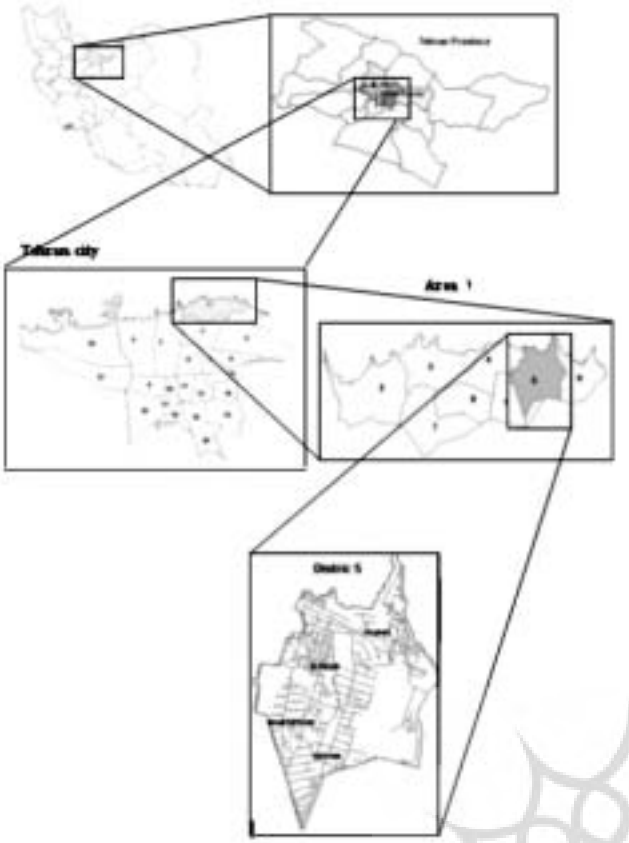
روش تحقیق

مکان مورد مطالعه ناحیه ۵ منطقه یک شهرداری تهران واقع در شمال شرق تهران شامل محله‌های اقدسیه، کاشانک، دارآباد، صاحبقرانیه و آجودانیه می‌باشد(نقشه شماره۱).

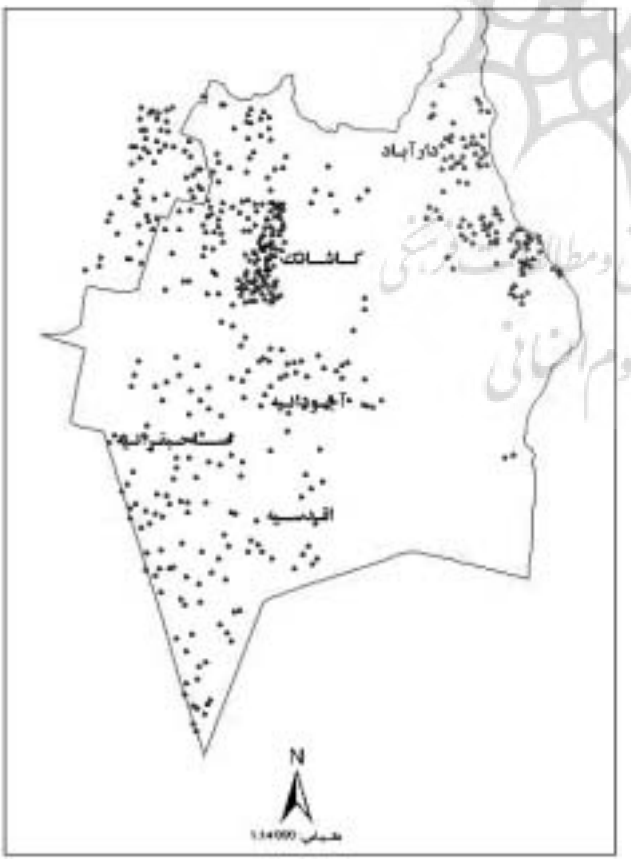
برای تهیه نقشه پهنه بندی قیمت از ۴۸۰ نمونه برداشت زمینی قیمت عرصه انجام گردید که این نمونه‌ها در برخی نقاط بعلت نظامی‌بودن از پراکنش خوبی برخوردار نیستند (نقشه شماره۲). از تعداد ۳۷۶۳ پارسل در ناحیه ۵ منطقه ۴۸۰ نمونه یعنی حدود ۱۲/۷۵٪ از کل پارسلهای ناحیه انتخاب شدند. ملاک انتخاب نمونه براساس شکل بلوک و تعداد پارسلها در بلوک، تعداد تقاطع‌های اطراف بلوک و بن بست‌ها می‌باشد. به طوری که پارسلهای واقع در سر تقاطع، انتهای بن بست و وسط بلوک به عنوان پارسلهای نمونه بر گزیده شدند (شکل شماره۱). ارزش قیمت عرصه در این ناحیه بین ۶۰۰٫۰۰۰ تومان تا ۳٫۰۰۰٫۰۰۰ تومان و میانگین ارزش قیمت نمونه‌ها ۱۳۳۶۵۱۴ تومان می‌باشد.



شکل شماره ۱: نحوه انتخاب پارسلهای نمونه در بلوک



نقشه شماره ۱: موقعیت ناحیه ۵ در منطقه ۱



نقشه شماره ۲: پراکنش نقاط نمونه در ناحیه ۵

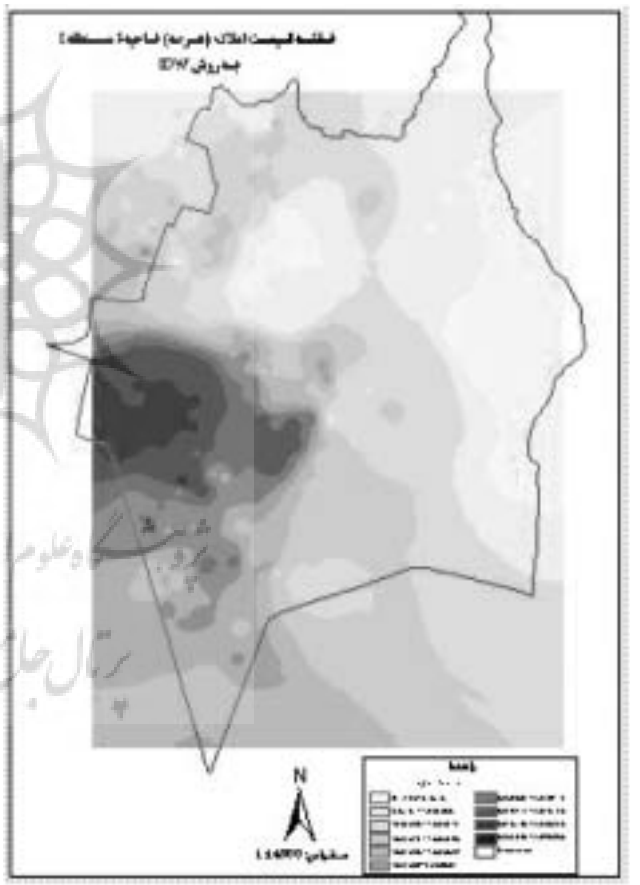
برای انجام روشهای درون یابی و تهیه نقشه پهنه بندی از نرم افزار ArcGIS v.۹ برنامه Spatial Analyst و Analyst Geostatistical استفاده شد. نقاط نمونه پس از برداشت زمینی بر روی نقشه پارسلهای ناحیه ۵ جانمایی گردیدند.

برای محاسبه درون یابی روش IDW، ابتداء مقدار ضریب n را که کمترین خطا RMS را برای درون یابی داشته باشد، تعیین شد.

n	۱	۲	۳	۴	۵	۶
RMS	۱۷۸۵۰۰	۱۶۷۳۰۰	۱۶۴۵۰۰	۱۶۵۸۰۰	۱۶۸۲۰۰	۱۷۰۶۰۰

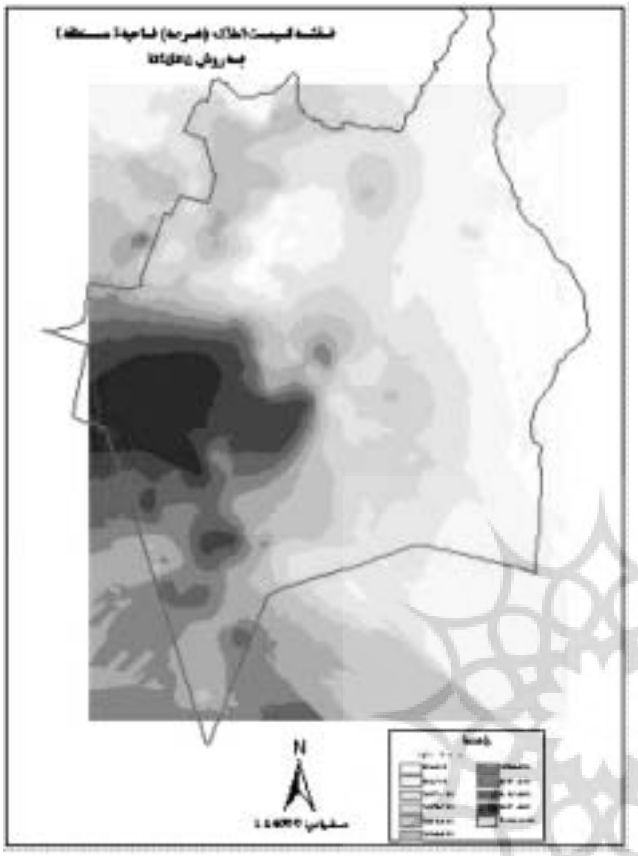
جدول شماره۱: محاسبه ضریب بهینه در روش IDW

با توجه به میزان خطای محاسبه شده RMS، ضریب nبرابر ۳ مناسب برای درون یابی به روش IDWدر نظر گرفته شد.



نقشه شماره ۳: نقشه پهنه بندی قیمت به روش IDW

برای تهیه نقشه پهنه بندی به روش Kriging ابتدا نمونه‌ها از نظرتوزیع آماری مورد بررسی قرار گرفتند، که دارای توزیع آماری غیرنرمال بودند و برای رفع این مشکل از نرمال سازی به روش لگاریتمی استفاده شد. در روش Kriging، مدل واریوگرامی Sphericalکه نسبت به دیگر مدلها برای تخمینگرهای خطی و با نمونه‌های غیر نرمال و بخاطر دارا بودن کمترین خطای تخمین(K.Krivoruchk and et al,۲۰۰۴)، انتخاب شد.



نقشه شماره ۴: نقشه پهنه بندی قیمت به روش Kriging

برای تعیین قیمت عرصه برای هر پارسل ابتدا نقشه پهنه بندی از ساختار رستری به وکتوری تبدیل و سپس از هم پوشانی نقشه‌های پارسل ناحیه بروی نقشه وکتوری پهنه بندی، نقشه قیمت عرصه هر پارسل استخراج شد. نقشه تهیه شده حاکی از بالا بودن قیمت عرصه در محله صاحبقرانیه به مبلغ حدود ۳ میلیون تومان و کمترین قیمت هم به محله‌های دارآباد و کاشانک به مبلغ ۷۰۰ تا ۹۰۰ هزار تومان بوده است(نقشه شماره-۵).

$$p^n = a + bo_i$$

۵- شاخص مقبولیت (d)

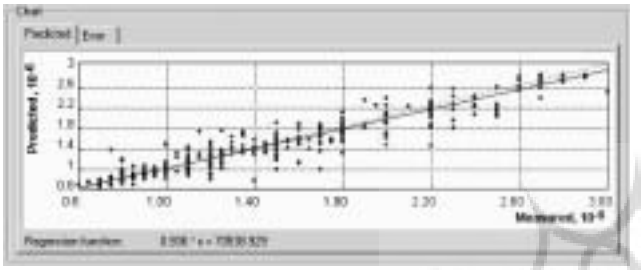
$$d = 1 - \frac{N \times RMSE^2}{PE} \quad ، PE = \sum_{i=1}^{N} \left(\left| p_i - o^- \right| + \left| o_i - o^- \right| \right)^2$$

o_i = نمونه برداشت شده

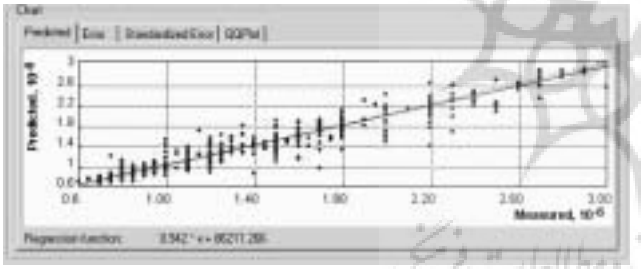
p_i = مقدار تخمین نمونه ها

N = تعداد کل نمونه ها

$a، b$ = رگرسیون خطی بین o و p



نمودار شماره ۴: رگرسیون خطی بین مقادیر مشاهده شده و پیش بینی در مدل IDW



نمودار شماره ۵: رگرسیون خطی بین مقادیر مشاهده شده و پیش بینی در مدل KRIGING

مدل	a	b	R
IDW	۰/۹۳۶	۷۰۵۳۸/۹	۰/۹۵۸
KRIGING	۰/۹۶۰	۸۶۲۱۱/۲	۰/۹۴۲

جدول شماره ۲: ضرا یب رگرسیون و همبستگی بین مقادیر مشاهده شده و پیش بینی در دو مدل

مدل	MAE	RMSE	$RMSE_S$	RMSEu	d
IDW	۱۱۰۸۲۹	۱۱۰۸۳۱	۳۰۲۰۹	۱۰۶۱۲۳	۰/۹۹۰۴۵
KRIGING	۱۱۰۶۴۵	۱۱۰۶۴۷	۳۰۴۶۵	۱۰۶۴۶۱	۰/۹۹۰۵

جدول شماره ۳: برآورد میزان خطا دو روش درون یابی

نقشه شماره ۵: نقشه قیمت عرصه پارسلها در ناحیه ۵

ارزیابی و نتیجه گیری

برای ارزیابی صحت و بررسی میزان خطای روش‌های درون یابی ذکر شده از روش Willmott(۱۹۸۴) که از پنج شاخص استفاده می‌شود.

۱- Mean Absolute Error(MAE)

$$MAE = \frac{\sum_{i=1}^{n} |p_i - o_i|}{N}$$

۲- Root Mean Sqaure(RMS)

$$RMSE = \frac{\sum_{i=1}^{n} \left[(p_i - o_i)^2 \right]^{0.5}}{N}$$

۳- Systematic Root Mean Sqaure(RMSES)

$$RMSE_S = \frac{\sum_{i=1}^{n} \left[(p^n - o_i)^2 \right]^{0.5}}{N}$$

۴- Un Systematic Root Mean Sqaure(RMSEu)

$$RMSE_u = \frac{\sum_{i=1}^{n} \left[(p - p^n)^2 \right]^{0.5}}{N}$$

پرتال جامع علوم انسانی

نتایج بدست آمده به شرح زیر است:

۱-برای انجام درون یابی به روش IDW، محاسبه ضریب بهینه nضرورت دارد که به ازای مقادیر مختلف nمقدار RMSE متفاوت خواهد بود. در این تحقیق ضریب بهینه n که دارای کمترین خطای RMSE بوده برابر ۳ برآورد شده است.

۲-طبق جدول شماره ۲ مقدار Rهمبستگی بین مقادیر مشاهده شده و پیش بینی در دو روش IDW و KRIGING نشان می‌دهد که روش KRIGING نسبت به IDW اندکی بالاتر می‌باشد.

۳-مطابق جدول شماره ۳ ارزیابی خطای محاسبه شده به پنج روش فوق نشان می‌دهد که در دو روش بررسی MAE، RMSE و شاخص مقبولیت d، روش KRIGING دقت بالاتری نسبت به روش IDWداشته است.

۴-با نرمال سازی و حذف نمونه‌های پرت با استفاده از رابطه

$$(Q3\text{-}Q1) - \mu \leq \chi_i \leq (Q3\text{-}Q1) + \mu$$

که Q۱چارک اول، Q۳چارک سوم و میانگین، حدود ۶۸ نمونه پرت از مجموع ۴۸۰ نمونه بدست آمد و ارزش قیمت عرصه در این ناحیه بین ۶۰۰٫۰۰۰ تومان تا ۲٫۱۸۶٫۵۱۴ تومان و میانگین ارزش قیمت نمونه‌ها ۱٫۱۴۶٫۳۵۹ تومان شده است. بررسی خطای دو روش MAE وRMSE نمونه‌های پرت در برابر مقادیر واقعی آنها در جدول شماره ۴ آمده است. با ارزیابی نتیجه جدول شماره ۴ می‌توان به دقت بالاتر روش درون یابی KRIGING پی برد.

با کاهش تعداد و توزیع نرمال نمونه ها، شعاع تاثیر نیز کاهش و این باعث افزایش دقت روش درون یابی می‌شود.

مدل	MAE	RMSE
IDW	۱۶۴۱۱۳	۱۶۴۱۱۵
KRIGING	۱۵۱۷۰۶/۶	۱۵۱۷۰۸/۶

جدول شماره ۴: برآورد میزان خطا دو روش درون یابی با نمونه‌های پرت

در نهایت اینکه یک روش بهینه برای درون یابی پیشنهاد نمی‌شود، بلکه بستگی به شرایط مکانی و نحوه نمونه برداری و توزیع مکانی نمونه‌ها و تراکم آنها دارد. پیشنهاد می‌شود که دقت مدلهای درون یابی با انجام ارزیابی خطا مورد بررسی قرار گیرد و مدلی که کمترین خطا را دارد انتخاب شود.■

پی نوشت:

۱- روش وزن دهی بر مبنای معکوس فاصله

۲-روش زمین آمار

منابع:

۱- حسنی پاک، علی اصغر. زمین آمار. ۱۳۷۷. انتشارات دانشگاه تهران.

۲- رنگزن، کاظم و دیگران. ارزیابی دقت مدلهای IDW و Kriging جهت درون یابی داده‌های سطح آب زیرزمینی دشت میان آب شوشتر ۱۳۸۴. دانشگاه شهید چمران اهواز

3-Anderson,Sharolyn(2004).An Evaluation of Spatial Interpolation Methods on Air Temperature in Pheonix Az.

4-C.V Deutsch,2002,Geostatistical Reservior Modeling,Oxford University Press,376 pages.

5- E.H.Isaaks and R.M.Srivastava,1989,An Introduction to Applied Geostatistics,Oxford University Press,483 pages.

6- K.Krivoruchko and C.A.Gotway 2004, Public Health GIS News and Information.Environmental Systems Research Institue,380 New York Street,Redlands,CA 92373-8100,USA.