پیشگویی فضایی داده‌های سخت و نرم برای تهیه نقشه احتمال یخ زدگی

محسن محمدزاده[i]؛ آزاده کیاپور[ii]

چکیده

از مسایل مهم مطالعات محیطی برآورد مقدار نامعلوم یک پدیده تصادفی در موقعیتهای مشخص بر اساس مشاهدات در منطقه‌ای خاص است. معمولا در آمار فضایی از انواع پیشگوهای کریگیدن برای مشاهدات پیوسته یا گسسته و از روش کریگیدن نشانگر برای تحلیل مشاهدات کیفی استفاده می شود. این مقاله، ضمن معرفی داده‌های سخت و نرم روشی برای تولید داده های نرم فضایی ارائه و برتری کارایی آن نسبت به روش های معمول را نشان می‌دهد. سپس نقشه احتمال یخ زدگی گیاهان با استفاده از دمای هوا به عنوان داده سخت و احتمال پیشین یخ زدگی بعنوان داده نرم به کمک کریگیدن نشانگر تهیه و تاثیر انتخاب مدل برای محاسبه داده‌های نرم و استفاده توام داده های سخت و نرم در افزایش دقت پیشگو مورد بررسی قرار می‌گیرد.

کلمات کلیدی

کریگیدن نشانگر، داده سخت و نرم، رگرسیون لوژستیک فضایی

Spatial Prediction of Hard and Soft Data for Preparing Freezing Probability Map

Mohammadzadeh; M. Kiapour; A.

ABSTRACT

One of the most important factors in environmental studies is the estimation of unknown value of a stochastic phenomenon in certain situations, based on observations in a certain area. In spatial statistics, some types of kriging predictors for continues or discrete observation and the method of Indicator kriging for qualitative observation has been used. This paper is an introduction of hard and soft data, including a method for generating soft spatial data. Its superiority in efficiency relative to common methods has been shown. Afterward, the plan of the probability of plant freezing using indicator kriging is provided via air temperature as hard data, and the prior probability of freezing as soft data. Finally, the effect of model selection on computing the soft data and using both hard and soft data jointly is assessed in the increasing the precision of predictor.

KEYWORDS

Indicator kriging, Hard and soft data, Spatial logistic regression

[i] دانشیار گروه آمار دانشگاه تربیت مدرس، تهران: mohsen_m@modares.ac.ir

[ii] دانشجوی کارشناسی ارشد آمار دانشگاه تربیت مدرس، تهران: azade_kiapoor@yahoo.com

۱- مقدمه

داده‌های فضایی که بر حسب موقعیت قرار گرفتن در فضای مورد مطالعه همبسته هستند، مقادیر یک میدان تصادفی $\{Z(s),\ s \in D\}$ در نظر گرفته می‌شود، که در آن مجموعه اندیس گذار D زیرمجموعه‌ای از فضای اقلیدسی R^d، $(d \geq 1)$ است. در آمار فضایی پیشگویی مقدار نامعلوم میدان تصادفی در موقعیتهای مشخص بر اساس بردار مشاهدات به روش کریگیدن، بعنوان بهترین پیشگوی خطی نااُریب صورت می‌گیرد و انحراف معیار آن در هر موقعیت نیز بعنوان اندازه دقت پیشگو ارائه می‌شود. گاهی در آمار فضایی متغیرهایی وجود دارند که مقادیر صفر و یک را اختیار می‌کنند. در اینگونه موارد، برای پیشگویی فضایی از کریگیدن نشانگر استفاده می‌شود، که اولین بار توسط جورنل [۳] به عنوان یک روش ناپارامتری برای پیشگویی توابع غیرخطی از مشاهدات در آمار فضایی مورد استفاده قرار گرفت. اسمیت [۷] کریگیدن نشانگر را برای تجزیه و تحلیل آماری کیفیت خاک، گوارتز و همکاران [۲] برای برآورد ریسک آلودگی خاک، لین [۴] برای طرح تغییرات و منابع آلودگی فلزات سنگین در زمین‌های کشاورزی و برآورد توزیع احتمال آلودگی فلزات سنگین در زمین و لیون و همکاران [۶] برای تعیین احتمال اشباع و برآورد ریسک رواناب بکار گرفته‌اند.

در مسائل کاربردی مقادیر اندازه گیری شده $Z(s_1),...,Z(s_n)$ در موقعیت های $s_1,...,s_n$ داده‌های سخت نامیده می‌شوند. بعضی اطلاعات محلی یا متغیرهای کمکی نیز گاهی می‌توانند برای تکمیل داده های سخت یا بعنوان نماینده یا جانشین مقدار متغیر مورد نظر بکار گرفته شوند. اینگونه اطلاعات که داده‌های نرم نامیده می‌شوند، بعنوان اطلاعات غیرمستقیم از مقادیر واقعی در تحلیل داده‌های سخت مورد استفاده قرار می‌گیرند. از مسائل مهم در این خصوص نحوه تولید داده‌های نرم است، بگونه‌ای که منجر به نتایج تحلیلی دقیق شوند. ون میرون و گوارتز [۸] با استفاده از تابع توزیع تجمعی نرمال استاندارد اقدام به تولید داده‌های نرم نمودند. لیون و همکاران [۶] رگرسیون لوژستیک را برای تولید داده‌های نرم با استفاده از داده‌های صفر و یک بکار گرفتند. در این مقاله برای تولید داده‌های نرم فضایی رگرسیون لوژستیک همراه با جمله خطایی که از یک ساختار همبستگی فضایی پیروی می‌کند، بعنوان رگرسیون لوژستیک فضایی معرفی می‌گردد و نشان داده خواهد شد این روش در عمل موجب افزایش دقت پیشگوی کریگیدن نشانگر می‌شود. سپس در یک مثال کاربردی تاثیر استفاده توام داده‌های سخت و نرم در بهبود رفتار پیشگو مورد بررسی قرار می‌گیرد. برای این منظور روش پیشگویی کریگیدن نشانگر بطور خلاصه در بخش ۲ ارائه می‌شود. رگرسیون لوژستیک فضایی برای تولید داده‌های نرم در بخش ۳ معرفی و مثال کاربردی در بخش ۴ ارائه می‌گردد.

۲- کریگیدن نشانگر

برای میدان تصادفی $\{Z(s), s \in D\}$ با تابع توزیع $F_{Z(s)}(\cdot)$، در هر موقعیت $s \in D$ متغیر تصادفی نشانگر با آستانه $z \in R$ بصورت

$$I(s,z)=\begin{cases}1 & Z(s)\leq z\\ 0 & Z(s)>z\end{cases} \qquad (۱)$$

با میانگین

$$\mathrm{E}\big(I(s,z)\big)=P\big(Z(s)\leq z\big)=F_{Z(s)}(z) \qquad (۲)$$

و تغییر نگار

$$2\gamma(s_1,s_2,z)=\mathrm{Var}\big(I(s_1,z)-I(s_2,z)\big)$$

تعریف می‌شود، که بیانگر ساختار همبستگی فضایی داده‌های نشانگر است. اگر میانگین میدان تصادفی ثابت و به موقعیت s بستگی نداشته باشد، یعنی $E\big(I(s,z)\big)=\mu$ و تغییرنگار فقط تابعی از $h \in R^d$ به صورت

$$\mathrm{Var}\big(\mathrm{I(s,z)-I(s+h,z)}\big)=2(\mathrm{h})$$

باشد، میدان تصادفی مانای ذاتی نامیده می‌شود و می‌توان آن را به صورت

$$2\hat{\gamma}(h,z)=\frac{1}{N_h}\sum_{i=1}^{N_h}\big(I(s_i,z)-I(s_j,z)\big)^2$$

برآورد نمود، که در آن $N(h)=\{(i,j): s_i-s_j=h\}$ مجموعه تمام زوج موقعیت‌هایی است که در فاصله h از یکدیگر قرار دارند و N_h تعداد اعضای $N(h)$ را نشان می‌دهد. کرسی [۷] نشان داد کریگیدن نشانگر بعنوان بهترین پیشگوی خطی نااریب و براساس داده‌های نشانگر فضایی $I(s_1,z),...,I(s_n,z)$ بصورت

$$\hat{I}(s_0,z)=\sum_{i=1}^{n}\lambda_i(z)I(s_i,z) \qquad (۳)$$

است، بطوریکه بردار ضرایب $\lambda(z)=\big(\lambda_1(z),...,\lambda_n(z)\big)$ با کمینه کردن میانگین توان دوم خطای پیش بینی $\sigma_e^2=E\Big[I(s_0,z)-\hat{I}(s_0,z)\Big]^2$ بصورت

$$\lambda'(z)=\left(\gamma(z)+\underline{1}\frac{\left(1-\underline{1}'\Gamma^{-1}\gamma(z)\right)}{\underline{1}'\Gamma^{-1}\underline{1}}\right)'\Gamma^{-1}$$

حاصل می‌شود، که در آن $\gamma(z)=\left(\gamma(s_0-s_1,z),...,\gamma(s_0-s_n,z)\right)'$، Γ یک ماتریس $n\times n$ با عناصر $\gamma_{ij}(z)=\gamma(s_i-s_j,z)$ و $\underline{1}$ برداری با n عنصر یک است. در اینصورت واریانس کریگیدن نشانگر نیز در موقعیت s_0 بصورت

$$\sigma_k^2(s_0)=\lambda'(z)\gamma(z)-\frac{\left(1-\underline{1}'\Gamma^{-1}\gamma(z)\right)}{\underline{1}'\Gamma^{-1}\underline{1}}$$

است، که براساس آن میزان دقت پیشگویی در موقعیت s_0 قابل محاسبه است.

براساس قضیه پیش بینی (لوئن برگر [۸]) برآورد کمترین توانهای دوم نشانگر $I(s_0,z)$ برابر برآورد امید آن است. بنابراین با توجه به رابطه (۲) یک برآورد برای تابع توزیع تجمعی $F_{Z(s)}(z)$ بصورت $\hat{F}_{Z(s_0)}(z)=\hat{I}(s_0,z)$ می‌باشد.

۳- تولید داده‌های نرم

گاهی در بررسی‌های محیطی، اندازه‌گیری دقیق متغیر مورد نظر (داده سخت)، بطور غیرمستقیم یا به وسیله اطلاعات نرم تکمیل می‌شوند. معمولا چنین داده‌هایی هرچند نادقیق ولی با اندازه گیری‌های دقیق از متغیرهای کمکی دیگر قابل حصول هستند. رابطه (۲) بیانگر آنستکه متوسط مقدار متغیر تصادفی نشانگر فضایی $I(s,z)$ برابر مقدار $P\left(Z(s)\le z\right)$ است، که احتمال پیشین متغیر نشانگر نامیده می‌شود. بنابراین از این مقدار می‌توان بعنوان داده‌های نرم و جایگزین مقادیر متغیر تصادفی نشانگر $I(s,z)$ استفاده نمود. به همین دلیل ون میرونی و کوارتز [۸] تولید داده های نرم را بصورت

$$p_i=\Phi\left(\frac{z-Z(s_i)}{S(s_i)}\right),\quad i=1,...,n \qquad (٤)$$

پیشنهاد نمودند، که در آن Φ تابع توزیع تجمعی نرمال استاندارد و $S(s_i)=CV\cdot Z(s_i)$ می‌باشد، بطوریکه CV ضریب تغییر خطای تحلیلی است. لیون و همکاران [۶] برای تولید داده‌های نرم با استفاده از داده‌های صفر و یک، مدل رگرسیون لوژستیک

$$\pi_i=\beta_0+\sum_{j=1}^{K}\beta_j X_{ij}+\varepsilon_i,\quad i=1,...,n \qquad (٥)$$

را پیشنهاد نمودند، که در آن $X_{i1},...,X_{iK}$ متغیرهای کمکی و ε_iها خطاهای مستقل هستند، بطوریکه ضرایب رگرسیونی $\beta=(\beta_0,...,\beta_K)$ با کمینه کردن میانگین توان دوم خطاها برآورد می‌شوند. سپس احتمال پیشین یا داده‌های نرم را بصورت

$$p_i=\frac{e^{\pi_i}}{1+e^{\pi_i}}\quad i=1,...,n \qquad (٦)$$

معرفی نمودند. اما در عمل موارد زیادی شرط استقلال خطاها در مدل (۵) را برقرار نمی‌سازند، لذا در این مقاله برای لحاظ نمودن همبستگی فضایی داده‌ها در تعیین داده‌های نرم، مدل رگرسیون لوژستیک فضایی بصورت

$$\pi(s_i)=\beta_0+\sum_{j=1}^{K}\beta_j X_j(s_i)+\varepsilon(s_i),\quad i=1,...,n \qquad (٧)$$

پیشنهاد می‌شود، که در آن $X_1(s_i),...,X_K(s_i)$ و $\varepsilon(s_i)$ بترتیب متغیرهای تبیینی و خطا در موقعیت s_i هستند، بطوریکه $\{\varepsilon(s),s\in D\}$ یک میدان تصادفی گاوسی مانای ذاتی و مستقل از X_iها می‌باشند. برای برآورد پارامترهای مدل (۷) ابتدا به روش رگرسیون لوژستیک ضرائب $\beta=(\beta_0,...,\beta_K)$ برآورد می‌شوند. سپس جمله‌های خطا از رابطه

$$\varepsilon(s_i)=\pi(s_i)-b_0+\sum_{j=1}^{K}b_j X_j(s_i),\quad i=1,...,n$$

محاسبه و با استفاده از کریگیدن معمولی (کرسی،[۱])، مقدار خطا در موقعیت s_0 بر اساس $n-1$ خطای دیگر بصورت

$$\hat{\varepsilon}(s_i)=\sum_{j\neq i}\lambda_j(z)\varepsilon(s_j),\quad i=1,...,n \qquad (٨)$$

برآورد می‌شود. اکنون مقادیر $\pi(s_i)$ بصورت

$$\hat{\pi}(s_i)=b_0+\sum_{j=1}^{K}b_j X_j(s_i)+\hat{\varepsilon}(s_i),\quad i=1,...,n \qquad (٩)$$

قابل پیشگویی هستند و احتمال پیشین یا داده‌های نرم به کمک رگرسیون لوژستیک فضایی بصورت

$$\hat{p}(s_i)=\frac{e^{\hat{\pi}(s_i)}}{1+e^{\hat{\pi}(s_i)}}\quad i=1,...,n \qquad (١٠)$$

تولید می‌شوند. در بخش ۴ تاثیر روش پیشنهادی این مقاله برای تولید داده‌های نرم بفرم (۱۰) در تحلیل داده‌های سخت مورد بررسی قرار خواهد گرفت.

٤- مثال کاربردی

یکی از مسائل مهم تحقیقات اقلیمی در کشاورزی، شناخت مکان وقوع سرما و یخبندان در هر ناحیه است که می‌تواند کشاورزان و برنامه‌ریزان بخش کشاورزی را به منظور کاهش خسارت سرما برای گیاهان زراعی راهنمایی و کمک نماید. یکی از عناصر شناخت هوا بعنوان شاخصی از شدت گرما دما می‌باشد. نظر به دریافت نامنظم انرژی خورشید توسط زمین، دمای هوا در سطح زمین دارای تغییرات زیادی است، که به

جدول (۱): مشخصات ایستگاه‌های هواشناسی استان مازندران

نام ایستگاه	مختصات جغرافیایی			دما	انواع داده ها			
	طول	عرض	ارتفاع		نشانگر	نرم۱	نرم۲	نرم۳
رامسر	۴۹۹۹۸۹	۴۳۲۷۶۵۵	-۲۰	۴	۰	۰/۸۸	۰/۵	۰/۵
سیاه بیشه	۶۹۲۵۷۳	۴۳۸۷۳۲۲	۱۸۵۵/۴	-۵/۵	۱	۰/۰۵	۰/۷	۰/۷۲
نوشهر	۵۸۹۸۳۷	۴۰۱۲۰۲۶	-۲۱	۶	۰	۰/۹۶	۰/۵	۰/۵
علوم دریایی نوشهر	۴۹۹۹۷۱	۴۳۸۲۸۶۶	-۲۶	۵/۵	۰	۰/۹	۰/۷۲	۰/۵
چمستان	۷۵۶۱۱۰۱	۴۲۳۱۹۶۷	۷۳	۲/۳	۰	۰/۷	۰/۵۲	۰/۵۳
بلده	۸۵۱۲۶۵	۴۲۱۳۲۶۲	۲۱۲۰	۰	۱	۰/۵	۰/۶	۰/۶
کجور	۷۷۰۴۲۳	۳۹۸۷۹۹۶	۱۵۵۰	۰	۱	۰/۵	۰/۶۲	۰/۶۴
بایعکلا	۷۸۸۵۸۷	۴۲۴۴۱۴۷	۴	۱/۲	۰	۰/۶	۰/۵۱	۰/۵۳
تاله	۶۱۶۷۱۸	۴۰۱۷۸۹۳	۸۴۴	۰	۱	۰/۵	۰/۷۱	۰/۷۲
رینه	۶۸۲۵۷۳	۴۲۸۷۳۲۲	۱۹۵۰	-۶/۵	۱	۰/۰۲	۰/۷۲	۰/۷۲
آمل	۶۲۸۵۳۵	۴۲۲۱۵۱۶	۲۹	۲/۲	۰	۰/۷	۰/۵	۰/۵
بابل کنار	۶۸۲۹۱۹	۴۲۰۷۷۵۴	۹۱	۳/۵	۰	۰/۸	۰/۵۲	۰/۵۵
بابلسر	۵۱۰۲۱۴	۴۲۲۰۵۰۶	-۲۱	۴/۲	۰	۰/۸	۰/۵	۰/۵
قائمشهر	۶۲۷۴۲۰	۴۱۰۳۱۱۶	۱۴/۷	۴/۷	۰	۰/۹	۰/۵۱	۰/۵۲
فراخیل	۷۷۳۶۸۹	۳۸۸۰۷۳۶	۱۴/۷	۲/۴	۰	۰/۷۵	۰/۵	۰/۵
پل سفید	۶۹۸۱۲۷	۳۹۹۳۴۶۷	۶۱۰	-۰/۲۶	۱	۰/۴۶	۰/۵۸	۰/۵۸
ساری	۵۹۸۹۷۱	۴۱۹۵۲۲۷	۲۳	۲/۹	۰	۰/۸	۰/۵۱	۰/۵۱
آلاشت	۶۸۰۲۶۴	۳۹۸۵۶۸۲	۱۹۰۰	۵/۵	۰	۰/۹	۰/۵۱	۰/۵۱
فریم محمودآباد	۶۸۱۸۴۵	۳۹۰۸۰۲۹	۷۲۰	۰	۱	۰/۵	۰/۷۲	۰/۷۲
دشت ناز	۶۸۶۴۷۶	۴۱۸۰۰۷۷	۱۶	۲/۵	۰	۰/۷	۰/۵۱	۰/۵۳
دانشکده کشاورزی	۵۰۱۴۳۲	۴۳۱۶۶۷۱	۱۰۷	۲/۸	۰	۰/۷	۰/۵۲	۰/۷۲
کیاسر	۵۹۳۱۲۷	۳۹۸۴۳۲۶	۱۲۹۴/۳	-۳/۳	۱	۰/۱۶	۰/۶۹	۰/۷
بهشهر	۵۸۸۹۷۱	۴۰۹۵۲۲۷	۱۴	۲/۲	۰	۰/۷	۰/۵	۰/۵
زردگل سرخ آباد	۵۸۹۹۹۰	۳۹۹۷۲۲۶	۱۵۰۰	-۵/۳	۱	۰/۰۵	۰/۷	۰/۷۲
گلوگاه	۴۹۹۹۹۹	۴۴۲۷۶۵۵	-۱۰	۳/۷	۰	۰/۸	۰/۵۱	۰/۵۲

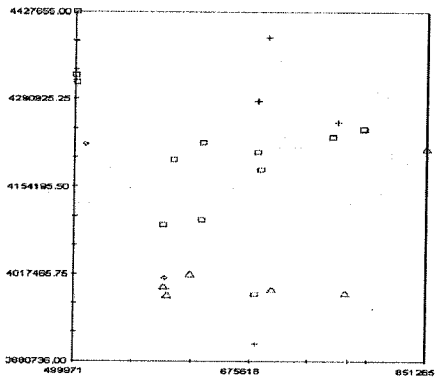
نوبه خود سبب تغییرات گسترده دیگری در سایر موارد هواشناسی می‌شود. البته درجه حرارت تماما تحت تاثیر موقعیت خورشید نبوده بلکه توسط عوامل محلی نظیر ارتفاع، بادهای غالب، ابرها، رطوبت هوا، فاصله از دریا و پوشش طبیعی زمین کنترل می‌شوند، که ارتفاع بعنوان یکی از مهمترین عوامل موثر در دما مطرح می‌باشد.

گیاهان مختلف می‌توانند فقط در محدوده حرارتی معینی رشد نمایند و تحمل پذیری آنها در مقابل دماهای بالا و پایین محدود می‌باشد. کاهش دما از آستانه تحمل گیاهان همه ساله خسارتهای گسترده‌ای به کشاورزان کشور وارد می‌آورد. حداقل دما برای رشد گیاهان بر حسب نوع و مرحله رشد گیاهان متفاوت است. همچنین، احتیاجات دمایی آنها در فصول مختلف نیز متفاوت می‌باشد. ولی در اکثر گیاهان علفی، دمای زیر صفر باعث از بین رفتن گیاه می‌شود. لذا بررسی دما در یک منطقه و احتمال یخبندان در هر نقطه می‌تواند برای پیش‌گیری خسارت کشاورزان موثر واقع شود. از آنجا که احتمال یخزدگی نواحی نزدیک در هر منطقه به یکدیگر مرتبط هستند و میزان این ارتباط در نواحی دورتر از هم کاهش می‌یابد، به گونه‌ای بین آنها همبستگی فضایی برقرار است. بنابراین، لازم است احتمال یخ زدگی در هر موقعیت مشخص بکمک روشهای آمار فضایی پیشگویی شود. جدول (۱) ارتفاع و متوسط دمای روز پانزدهم بهمن ماه ۱۳۸۴ را در ۲۶ ایستگاه

هواشناسی استان مازندران بر حسب سانتیگراد نشان می‌دهد و موقعیت قرار گرفتن ایستگاه‌ها نیز در شکل (۱) رسم شده‌اند. طول و عرض جغرافیایی بر حسب متر در جدول (۱) ارائه گردیده‌اند. در این مثال، داده‌های سخت با تبدیل مقادیر اندازه‌گیری شده دما به نشانگر دودویی، بصورت رابطه (۱) تولید شده‌اند، که در آن $Z(s)$ دما در مکان s و z آستانه دما است. مقادیر نشانگر، کد یک را به ناحیه یخ زده و کد صفر را به ناحیه یخ نزده اختصاص می‌دهند.

همانطور که گفته شد، ارتفاع یکی از عوامل مهم محیطی موثر در دمای هوا در مناطق مختلف می‌باشد. بنابراین ارتفاع هر ایستگاه هواشناسی را بعنوان یک متغیر کمکی در نظر گرفته و بر اساس آن احتمال پیشین یخبندان در موقعیت‌های مختلف بعنوان داده‌های نرم تولید شده‌اند. با توجه به برابری:

$$\hat{F}_{Z(s_0)}(z) = \hat{I}(s_0, z)$$

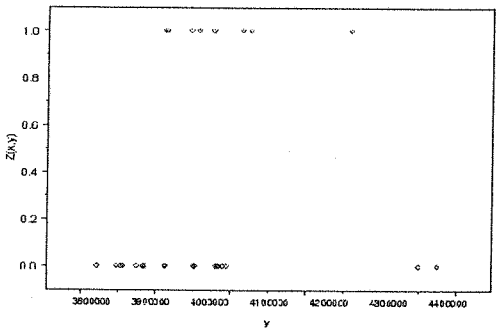


شکل (۱): مختصات جغرافیایی محل قرارگرفتن ایستگاه‌ها

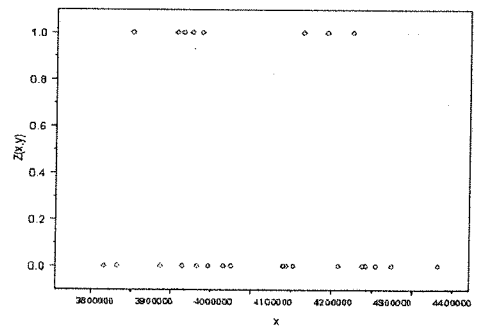
احتمال یخزدگی با آستانه صفر در هر موقعیت معلوم از رابطه (۳) بر اساس داده‌های سخت محاسبه می‌شود. داده های نرم نیز به سه روش معرفی شده در بخش ۲، تولید و احتمال پیشین یخزدگی بر اساس هر سه نوع داده نرم بکمک کریگیدن نشانگر (۳) پیشگویی می‌شود.

داده‌های نرم نوع ۱ یا احتمال پیشین یخزدگی بر اساس تابع توزیع تجمعی نرمال استاندارد با آستانه صفر و ضریب تغییرات ۳/۰۴ بصورت رابطه (۴) تولید شده‌اند. همچنین داده‌های نرم نوع ۲ بر اساس داده‌های نشانگر و ارتفاع، به روش رگرسون لوژستیک به ازای $K = 1$ تولید شده‌اند. برای این منظور ابتدا پارامترهای β_0 و β_1 در مدل (۵) برآورد شده و مقادیر $b_1 = ۰/۰۰۳$ و $b_0 = -۲/۶$ حاصل گردیده‌اند. سپس احتمال یخزدگی در موقعیت نمونه‌گیری s_i، با استفاده از رابطه (۶) تولید شده است. داده‌های نرم نوع ۳ نیز با برآورد پارامترهای مدل (۷) و برآورد جملات خطای (۸) بر اساس رابطه (۹) تولید شده‌اند. انواع داده‌های نرم تولید شده برای تمام ایستگاه‌های هواشناسی در جدول (۱) آورده شده‌اند. سپس اختلاف داده‌های سخت از انواع داده‌های نرم بعنوان داده‌های سخت و نرم تعیین شده‌اند.

برای انجام کریگیدن، ابتدا لازم است با تحلیل اکتشافی داده‌ها وجود داده‌های پرت، مانایی و همسانگردی تغییرنگار انواع داده‌ها مورد بررسی قرار گیرد. شکلهای (۲) و (۳) به ترتیب نمودار داده‌های سخت را در مقابل طول و عرض جغرافیای نشان می‌دهند. همانطور که ملاحظه می‌شود، هیچگونه روندی در جهات مختلف جغرافیایی وجود ندارد. بررسی انجام شده برای انواع داده‌های نرم و همچنین داده‌های سخت و نرم نیز بطور مشابه بیانگر عدم وجود روندی خاص در میانگین داده‌ها است. با رسم تغییرنگار تجربی انواع داده‌ها در چهار جهت ۰، ۴۵، ۹۰ و ۱۳۵ درجه ملاحظه گردید تغییرنگار انواع داده ها تقریبا همسانگرد می‌باشند.



شکل(۲): نمودار $I(x, y)$ در مقابل x

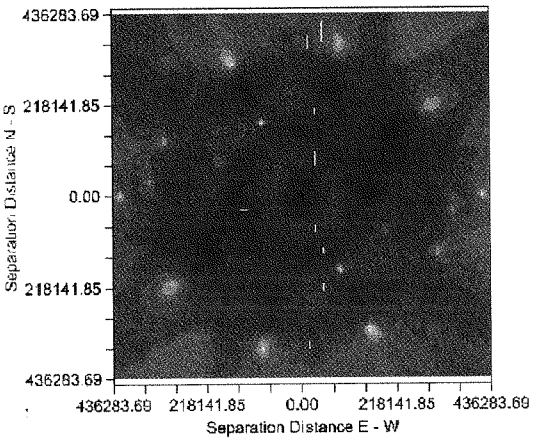


شکل (۳): نمودار $I(x, y)$ در مقابل y

با استفاده از نرم افزار Arc Map 9.1 تغییرنگار نشانگر همسانگرد کروی بصورت

$$\gamma_{sp}(h) = \begin{cases} \theta_0 + \theta_s \left[\frac{3}{2}\frac{\|h\|}{\theta_0} - \frac{1}{2}\left(\frac{\|h\|}{\theta_r}\right) \right] & \theta_s \geq \|h\| \\ \theta_0 + \theta_s & \theta_s \leq \|h\| \end{cases}$$

با پارامترهای اثر قطعه‌ای θ_0، ازاره θ_s و دامنه θ_r بعنوان بهترین تغییرنگار به انواع داده‌های سخت و نرم برازانده شده و مقادیر برآورد پارامترها در جدول (۲) ارائه گردیده‌اند. شکل (۴) رویه تغییرنگار داده‌های سخت را نشان می‌دهد. تقارن این رویه بیانگر مانایی در تغییرنگار است. بطور مشابه رویه تغییرنگار سایر داده‌های نرم و داده‌های سخت و نرم نیز حاکی از مانایی در تغییرنگار آنها می‌باشد.



شکل (۴): رویه تغییرنگار نشانگر داده‌های سخت

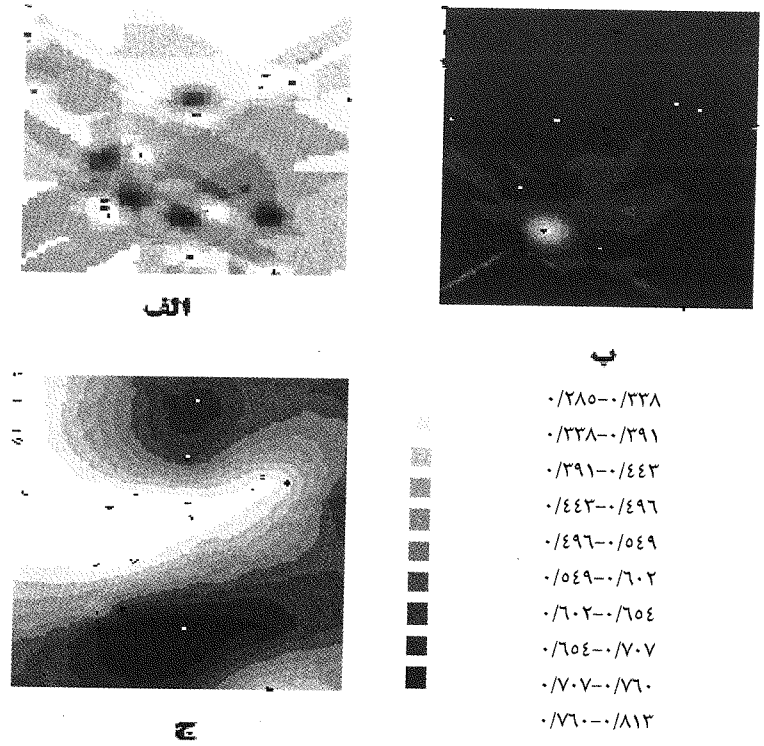
شکل (۵) رویه‌های نقشه احتمال یخ زدگی حاصل از داده‌های سخت، داده های نرم نوع ۳ و ترکیب داده‌های سخت و نرم نوع ۳ را نشان می‌دهد، که در آنها نقاط سیاه مربوط به نواحی یخزده و نقاط سفید مربوط به نواحی یخنزده می‌باشند. در شکل (۵-الف)، ملاحظه می‌شود که بعضی نواحی یخنزده، بعنوان موقعیت یخزده با احتمال بالا پیشگویی شده است. شکل(۵-ب)، احتمال یخزدگی از داده‌های نرم نوع ۳ را نشان می‌دهد. همانطور که ملاحظه می‌شود، بدون درنظر گرفتن داده‌های سخت، احتمال یخزدگی بالایی برای موقعیت‌های یخنزده پیشگویی شده است. شکل (۵-ج)، رویه احتمال یخزدگی حاصل از داده‌های سخت و نرم را نشان می‌دهد، نواحی یخزده با احتمال بالای یخزدگی و نواحی یخنزده با احتمال پایین یخزدگی پیشگویی شده‌اند، که با واقعیت مطابقت بیشتری دارد. در حالیکه بر اساس داده‌های سخت یا داده‌های نرم، بیشتر موقعیت‌های یخزده با احتمال پایین یخزدگی و بیشتر موقعیت‌های یخنزده با احتمال بالایی از یخزدگی پیشگویی شده‌اند.

در بررسی انجام شده نتایجی مشابه برای سایر داده‌های نرم نیز حاصل گردیده است، که در اینجا برای خلاصه گویی فقط به ذکر نتایج پرداخته شده است. بنابراین همواره استفاده توام داده‌های سخت و نرم همراه پیشگویی‌های دقیقتری برای احتمال یخزدگی فراهم می‌سازند.

اکنون دقت پیشگویی‌های فضایی حاصل از داده‌های سخت، داده‌های نرم و داده‌های سخت و نرم بر اساس معیار اعتبارسنجی متقابل مورد مقایسه قرار می‌گیرند. دراین شیوه یک مشاهده را از مجموعه مشاهدات حذف نموده و مقدار آن بر اساس $n-1$ مشاهده باقیمانده پیشگویی می‌شود. سپس از اختلاف مقادیر واقعی و پیشگو، خطای کریگیدن در هر موقعیت محاسبه می‌گردد. این عمل برای تمام n موقعیت تکرار و جذر میانگین توان دوم خطاها به صورت

$$RMSE = \left\{ \frac{1}{n} \sum_{j=1}^{n} \left[Z(s_j) - \hat{Z}_{-j}(s_j) \right]^2 \right\}^{\frac{1}{2}}$$

محاسبه می‌گردد، که در آن $Z(s_j)$ مقدار واقعی مشاهده شده در موقعیت s_j و $\hat{Z}_{-j}(s_j)$ پیشگویی آن براساس بقیه مشاهدات بجز $Z(s_j)$ است. بدیهی است هر اندازه مقدار این معیار به صفر نزدیکتر باشد، پیشگو از دقت بیشتری برخوردار می‌باشد.



شکل (۵): رویه های کریگیدن نشانگر برای (الف) داده‌های سخت، (ب) داده‌های نرم و (ج) داده‌های ترکیبی

مقدار $RMSE$ برای انواع پیشگوها بر اساس داده‌های سخت، داده‌های نرم و داده‌های سخت و نرم محاسبه و در جدول (۲) خلاصه شده‌اند. همانطور که ملاحظه می‌شود وقتی از داده‌های سخت و داده‌های نرم نوع ۳ تولید شده توسط رگرسیون لوژستیک فضایی بصورت توام استفاده شده است، مقدار $RMSE$ کمتر از حالات دیگر می‌باشد. بنابراین روش تولید داده‌های نرم نوع ۳ ارائه شده در این مقاله منجر به پیشگوهای با دقت بیشتر نسبت به سایر روشهای تولید داده‌های نرم می‌گردد.

۵- بحث و نتیجه‌گیری

معمولا برای داده‌های سخت دودویی صفر و یک، داده‌های نرم با استفاده از تابع توزیع تجمعی نرمال استاندارد یا رگرسیون لوژستیک تولید و برای تکمیل و تحلیل داده‌های سخت بکار برده می‌شوند. گاهی در آنالیز رگرسیون، فرض ناهمبسته بودن خطاها در رگرسیون لوژستیک برقرار نیست. بخصوص در تحلیل داده‌های فضایی که بین مشاهدات مجاور همبستگی قوی تری نسبت به مشاهدات دورتر برقرار است، لازم است ساختار همبستگی داده‌ها در تجزیه و تحلیل آنها لحاظ گردد. در این مقاله مدل لوژستیک فضایی برای تولید داده‌های نرم ارائه گردید و در یک مثال کاربردی نحوه کاربست آن نشان داده شد و ملاحظه گردید که با این روش می‌توان داده‌های نرم را برای داده‌های سخت دودویی با دقت بیشتری تولید نمود. سپس به روش کریگیدن نشانگر نقشه احتمال یخ زدگی بر اساس داده‌های سخت، داده‌های نرم و داده‌های سخت و نرم تهیه و میزان دقت پیشگوهای مختلف بر اساس معیار اعتبار سنجی متقابل مورد مقایسه قرار گرفت و نتایج بیانگر آنست که داده‌های سخت و نرم نسبت به داده‌های سخت و داده‌های نرم پیشگوهای دقیق‌تری ارائه می‌دهند.

جدول(۲): برآورد پارامترهای تغییرنگار و جذر میانگین توان دوم خطاها

نوع داده	پارامترهای تغییرنگار			RMSE
	دامنه	ازاره	اثر قطعه‌ای	
سخت	۴۸۱۳۴۰	۰/۱۲	۰/۱۶	۰/۴۷
نرم (۱)	۲۰۱۲۲۰	۰/۰۱۸	۰/۱۴	۰/۹۷
نرم (۲)	۵۴۰۲۴۰	۰/۰۵۸	۰	۰/۴۶
نرم (۳)	۵۱۷۳۸۰	۰	۰/۰۳	۰/۳۹
سخت و نرم (۱)	۴۸۱۳۴۰	۰/۵	۰/۶۳	۰/۴۲
سخت و نرم (۲)	۴۸۱۳۴۰	۰/۱۳	۰/۱۵	۰/۲۲
سخت و نرم (۳)	۲۳۸۶۵۰	۰/۲۲۳۵	۰/۰۵۳	۰/۲۰

۶- تقدیر و تشکر

از هیئت تحریریه محترم مجله به خاطر ارائه نظرات و پیشنهادات ارزنده‌شان در بهبود کیفیت مقاله و حمایت قطب علمی داده‌های ترتیبی و فضایی دانشگاه فردوسی مشهد

قدردانی می‌گردد.

٤- مراجع

[١] Cressie, N.; "Statistics for Spatial Data", Revised Edition, Wiley, New York, 1993.

[٢] Goovarets, P. ; Geostatistics for Natural Resources Evaluation: Oxford University Press, New York, 1997.

[٣] Journel, A. G.; "Nonparametric Estimation of Spatial Distributions", Mathematical Geology, 15, 445-468, 1983

[٤] Lin, Y.; "Factorial and Indicator Kriging Methods Using a Geographic Information System to Delineate Spatial Variation and Pollution Sources of Soil Heavy Metals", Environmental Geology, 42, 900-909, 2002.

[٥] Luenberger, D.; "Optimization by Vector Space Methods". John Wiley Sons, New York, 1969.

[٦] Lyon, S. W., Lembo, A. J., Walter, M. T., and Steenhuis, T. S.; "Defining Probability of Saturation with Indicator Kriging on Hard and Soft Data", Adv. Wat. Resour., 29 2, 181-193, 2006.

[٧] Smith, J.; "Using Multiple Variable Indicator Kriging for Evaluating Soil Quality", Soil. Am. J., 57, 743-749, 1993

[٨] Van Meirvenne, M. and Goovarets, P.; "Evaluating the Probability of Exceeding a Site Specific Soil Cadmium Contamination Threshold", Geoderma, 102, 75 100, 2001