

استفاده همزمان از آنتروپی و کریجینگ فضایی - زمانی برای تعیین شبکه بهینه پایش کیفی منابع آب زیرزمینی دشت مشهد

مسلم اکبرزاده^{*۱} - بیژن قهرمان^۲

تاریخ دریافت: ۱۳۹۱/۱۰/۲۰

تاریخ پذیرش: ۱۳۹۲/۵/۲۰

چکیده

در زمین آمار، برآورد مقدار نامعلوم میدان تصادفی در موقعیت و زمان مشخص، توسط کریجینگ فضایی-زمانی صورت می پذیرد. در برخی شرایط استفاده از متغیر کمکی دقت برآورد را افزایش می دهد. برای ارزیابی این فرضیه، روش های زمین آمار کریجینگ عام و برونری برای داده های آبخوان شهر مشهد مورد استفاده قرار گرفت. با به کارگیری آنتروپی، شبکه بهینه پایش کیفی منابع آب زیرزمینی ارائه گردید. در این روش ابتدا چاه ها بر اساس معیارهای آنتروپی و اطلاع متقابل، رتبه بندی شد. سپس بر اساس درصدهای اطلاع و با تکیه بر کریجینگ فضایی-زمانی، شبکه بهینه پایش کیفی تعیین گردید. نتایج بدست آمده از مقایسه کریجینگ عام و برونری حاکی از آن است که عامل هدایت الکتریکی در برآورد بهتر نقاط فاقد مشاهده تاثیر بسزایی دارد. در نتیجه کریجینگ برونری که در آن از EC به عنوان متغیر کمکی در برآورد روند استفاده شده است، به عنوان کریجینگ برتر انتخاب شد. همچنین با بهره گیری از کریجینگ برونری و مقایسه آن با کریجینگ عام اثبات گردید استفاده از داده های کیفی شبکه ای با ۱۱۱ حلقه چاه، اطلاعاتی برابر با شبکه پایش اولیه با مجموع ۲۳۷ حلقه چاه، در اختیار قرار می دهد.

واژه های کلیدی: نظریه اطلاع، زمین آمار، واریوگرام، اعتباریابی، نیترات، هیدروژنوشیمی

مقدمه

تحقیقات صنعتی ایران ویرایش ۵ مورخه ۱۳۸۸ حداکثر غلظت مجاز نیترات (بر حسب NO_3) برای تخلیه به آب های سطحی و شرب برابر 50 mg/L می باشد.

در مطالعات محیطی، گاهی با مشاهداتی سروکار داریم که به موقعیت قرارگیری در فضای مورد مطالعه و زمان مشاهده شدن وابسته اند. چنین مشاهداتی، داده های فضایی-زمانی^۳ نامیده می شوند. ساختار همبستگی این داده ها بوسیله کوواریانس^۴ فضایی-زمانی تعیین می شود. این تابع که نقش به سزایی در برآورد موقعیت های فضایی یا زمانی فاقد مشاهده دارد بطور معمول نامعلوم بوده و باید بر اساس مشاهدات برآورد شود. پذیرش شرایط ویژه برای تابع کوواریانس مانند مانایی^۵، همسانگردی^۶، تقارن^۷ و تفکیک پذیری^۸، برازش مدل های معتبر به تابع کوواریانس فضایی-زمانی را تسهیل

آلودگی آب های زیرزمینی یا از طریق منابع نقطه ای مثل چاه های جذبی فاضلاب یا از طریق منابع غیرنقطه ای مانند مصرف کودهای نیتروژنی در مزارع کشاورزی صورت می پذیرد. به علت محلول بودن نیترات این ترکیب به راحتی بوسیله آب از لایه های مختلف خاک عبور کرده و به سفره های آب زیرزمینی راه می یابد (۹ و ۱۶). با توجه به اینکه نیترات بدون رنگ، طعم و بو می باشد، لذا بدون انجام روش های آزمایشگاهی در آب قابل تشخیص نمی باشد. آزمایش منابع آب آشامیدنی از نظر مقدار یون نیترات باید سالی سه بار انجام پذیرد و لازم است تغییرات سالیانه منابع آب شرب مورد پایش کیفی قرار گیرند (۹ و ۱۶). طبق استاندارد شماره ۱۰۵۳ موسسه استاندارد و

- 3- Spatio-Temporal Data
- 4- Covariance
- 5- Stationary
- 6- Isotropy
- 7- Symmetry
- 8- Separability

۲۰۱- به ترتیب دانشجوی دکتری و استاد گروه مهندسی آب، پردیس بین المللی دانشگاه فردوسی مشهد
* - نویسنده مسئول: (Email: civil.environment@yahoo.com)

می کند.

به دلیل وجود دو بعد فضا و زمان، تبیین ریاضی تابع کوواریانس آسان نمی باشد. از این رو، تلاش هایی برای ساده سازی آن صورت پذیرفته است. برای مثال، روحانی و هال (۲۴) نوعی از تابع کوواریانس فضایی-زمانی تفکیک پذیر را پیشنهاد کرده اند که از جمع دو تابع مستقل از کوواریانس فضایی و زمان حاصل می شود. گوتارپ و همکاران (۱۰) نیز کوواریانس های تفکیک پذیری را که حاصل ضرب دو تابع کوواریانس فضایی و زمانی هستند معرفی کردند.

در روش کریجینگ^۱ باید مانایی را مورد بررسی قرار داد. اگر داده ها مانا نبودند، وجود روند در داده ها منجر به اریب^۲ در برآورد هم تغییرنا^۳ می شود. وقتی میانگین میدان تصادفی^۴ تابعی از مختصات فضایی یا زمانی باشد آنگاه داده های موجود مانا نبوده و به عبارت دیگری داده ها دارای روند^۵ هستند. برای تشخیص این روند، داده ها نسبت به مختصات جغرافیایی در زمان های مختلف رسم می شود. اگر داده ها نسبت به مختصات فضایی یا زمانی (یا هر دو) روند افزایش یا کاهش داشتند می توان بر عدم وجود مانایی در داده ها صحنه گذاشت (میانگین میدان تصادفی تابعی از مختصات فضایی یا زمانی است). اگر داده ها مانا نبودند، وجود روند در داده ها منجر به اریب در برآورد هم تغییرنا می شود بنابراین باید با روشی مناسب روند برآورد شود و سپس با کم کردن مقدار برآورد روند در هر موقعیت از مقدار اندازه گیری شده، مانده ها را اندازه گیری کرد و بر اساس مانده ها یا به عبارتی داد های بی روند شده هم تغییرنا را برآورد نمود.

در صورت بروز روند در داده ها از کریجینگ عام^۶ به منظور برآورد استفاده می شود. اما زمانی که داده هایی با همبستگی قابل توجه با داده های اولیه در اختیار باشد می توان روند را جهت بهبود پیشگویی برآورد نمود. براین اساس محمدزاده و شرفی (۲) با استفاده از یک متغیر کمکی^۷ در برآورد روند، از کریجینگ برونی^۸ استفاده کرده اند.

به منظور تحلیل شاخص های کیفی از جمله نیترات در راستای پایش منابع آب زیرزمینی، فراوانی موقعیت نمونه مورد بررسی از اهمیت ویژه ای برخوردار است. از این رو تعیین شبکه پایش براساس موقعیت چاه ها با استفاده از موقعیت های مکانی و زمانی در مطالعات گذشته مورد بررسی قرار گرفته است و مطالعات زیادی به منظور انتخاب موقعیت های با اهمیت بیشتر در موقعیت های فضایی انجام

شده که از آن جمله می توان به پژوهش های موقیر (۱۸) و رید و مینسکر (۲۱) اشاره کرد. دلایلی از جمله هزینه نمونه برداری، منجر به تعیین شبکه بندی جدید با عنوان شبکه بندی بهینه می شود.

به طور کلی فرآیند بهینه سازی یک فرآیند تصمیم گیری است برای داشتن بهترین حالت از حالت های ممکن باشد (۳). شبکه بندی بهینه را شبکه ای ناقص گویند که بهترین عملکرد را در بین شبکه های ناقص داراست. معیارهای ارزیابی برای تعیین شبکه بندی بهینه، مطالعات گذشته را از هم متمایز کرده است. در مطالعات اخیر از روش های آماری برای تعیین و طراحی مجدد شبکه بندی پایش کیفی چاه های آب استفاده شده است. دهار و دتا (۸) از آنالیز آماری برپایه سری زمانی و نظریه تصمیم برای انجام شبکه بندی استفاده نمودند. همچنین رید و مینسکر (۲۱) با روش های آنالیز فضایی همانند کریجینگ، فراوانی انواع شبکه بندی چاه ها را تعیین نمودند.

در پژوهش حاضر، پارامترهای کیفی نیترات و هدایت الکتریکی مربوط به ۳۳۷ حلقه چاه آب شرب شهر مشهد، در ۹ سال آماری مورد بررسی قرار گرفت. نظر به اینکه داده های موجود شامل ۳۶۱ موقعیت فضایی-زمانی فاقد مشاهده است، بوسیله دو نوع کریجینگ عام و کریجینگ با روند برونی، نقاط فاقد مشاهده برآورد گردید. پس از انتخاب کریجینگ برتر داده های مورد مطالعه تصحیح و تطویل شد. از آنجا که نمونه برداری، حمل، تحلیل و سنجش پارامترهای کیفی برای این تعداد حلقه چاه در سطح یک کلان شهر یقیناً هزینه زیادی در بردارد، ارائه روشی علمی که باعث از دست دادن کمترین اطلاعات شود برای کاهش تعداد چاهها ضروری به نظر می رسد.

نظریه اطلاع^۹ با نگرشی آماری به وضعیت چاه های موجود در شبکه پایش و آمار و اطلاعات آنها و همچنین ارتباط ایستگاه های موجود با یکدیگر ضمن بررسی اطلاعات موجود، امکان حذف ایستگاه های مازاد را فراهم می نماید. بنابراین می توان از این نظریه در مدیریت بهتر شبکه مورد نظر و چیدمان مناسب ایستگاه ها را استفاده نمود. بنا بر دانش نویسندگان، تنها روشی که علاوه بر نظریه اطلاع می تواند ضمن بررسی اطلاعات مشترک موجود چاه ها نسبت به حذف برخی از آنها برای داشتن شبکه بهینه کمک کند روش ضریب همبستگی^{۱۰} است که موقیر و سینگ (۱۸) بر برتری نظریه اطلاع صحنه گذاشتند. از این رو نظریه اطلاع به عنوان بهترین روش جهت بهبود شبکه بندی موجود انتخاب شد. نظریه اطلاع، شاخه ای از نظریه آمار و احتمال است که برای هر سیستم آماری قابل کاربرد می باشد. قسمتی از این نظریه مربوط به اندازه عدم حتمیت^{۱۱} می باشد که آن را آنتروپی^{۱۲} می نامند. اندازه های دیگری همانند اطلاع

- 1- Kriging
- 2- Biased
- 3- Covariogram
- 4- Random Field
- 5- Trend
- 6- Universal Kriging
- 7- Covariate
- 8- Kriging with External Trend

- 9- Information Theory
- 10- Correlation Coefficient
- 11- Uncertainty Measure
- 12- Entropy

مواد و روش ها

در این پژوهش، از آنترپوی به عنوان راهکاری نوین برای تعدیل و انتخاب شبکه بهینه پایش کیفی آبخوان دشت مشهد براساس شاخص های نیترا و هدایت الکتریکی استفاده شد. با استفاده همزمان از آنترپوی و کریجینگ روشی برای تعیین شبکه بندی ارائه گردید. این پژوهش در دو مرحله انجام شد که فلوچارت مراحل پژوهش بطور کامل در شکل ۱ نشان داده شده است:

- ۱- نقاط مکانی که در برخی سال ها فاقد مشاهده می باشند بوسیله ۲ نوع کریجینگ عام و برونی برآورد گردید و براساس معیار های ارزیابی، بهترین نوع کریجینگ انتخاب شد.
- ۲- تعیین اولویت بندی چاه ها و معرفی شبکه بهینه که تعدادی از چاه های غیر ضروری در آن حذف شده است.



شکل ۱- فلوچارت مراحل انجام پژوهش

برای شناسایی داده های دور افتاده در مشاهدات از روش هایبینگ (۱۱) در بعد زمان استفاده شد. بر مبنای این روش اگر مشاهده $Z(0, t)$ از یکی از نابرابری های $Z(0, t) > Q_L - 1/5(Q_u - Q_L)$ یا $Z(0, t) > Q_u + 1/5(Q_u - Q_L)$ صدق کند داده دور افتاده محسوب می شود.

متقابل^۱ وجود دارد که به اندازه گیری اطلاعات می پردازند. به طور مثال اطلاع متقابل برای دو ایستگاه اطلاعاتی (مانند دو چاه) اطلاعات مشترک آن دو ایستگاه را اندازه گیری می کند.

بطور کلی این باور وجود دارد که نظریه اطلاع با انتشار مقاله شانون (۲۷) تحت عنوان نظریه ریاضی ارتباطات نظم تازه ای یافت. از دهه ۱۹۸۰ میلادی نظریه اطلاع برای طراحی شبکه جمع آوری داده های هیدرولوژیکی و سنجش عدم حتمیت توسط چپمن (۵)، هارمنکیوگلو و یوجویچ (۱۲)، کاواشی و همکاران (۱۵)، سینگ (۲۸) و یانگ و بورن (۲۹) مورد استفاده قرار گرفته است. همچنین اندازه های نظریه اطلاع برای تعیین شبکه بندی و کیفیت آب زیرزمینی توسط اذکول و همکاران (۲۱) ارائه شد. سرلک و سورمان (۲۶) و کارآموز و همکاران (۱۴) نیز از آنترپوی برای تعیین جهت گسترش آبراهه ها استفاده نمودند. هویی چونگ و همکاران (۱۳) نیز از آنترپوی و کریجینگ برای تعیین شبکه ایستگاه های باران سنجی تایوان استفاده کردند. آواداله (۴) نیز تحقیقات مشابهی به منظور مطالعه و تعیین شبکه ایستگاه های باران سنجی انجام داد. آنها با استفاده از اطلاعات ایستگاه های قدیمی، پس از برآورد چندین موقعیت با استفاده از کریجینگ، به وسیله اندازه آنترپوی، شبکه جدیدی که بیشترین اطلاعات را دارا بود، ارائه نمودند.

روند داده های کیفی چاه های آب شرب در یک پهنه جغرافیایی در زوایای مختلف تغییر می کند. بنابراین وجود روند در چنین داده هایی محتمل است. از این رو مطالعه و استفاده از کریجینگ عام و برونی در این نوع داده ها معقول به نظر می رسد. استفاده از متغیر کمکی باعث ایجاد اطلاعاتی مازاد می شود که داشتن آن برآورد کریجینگ را بهبود می بخشد و با سنجش میزان تاثیر اطلاعات اضافی می توان شبکه بندی بهتر را ارائه نمود.

هدف بنیادی در پژوهش حاضر، تعیین شبکه بهینه پایش کیفی می باشد و از روش های کریجینگ برای برآورد نقاط فاقد مشاهده و اعتبارسنجی شبکه بهینه استفاده می شود. این پژوهش بر خلاف مطالعات گذشته استفاده همزمان از کریجینگ و آنترپوی را برای تعیین شبکه بندی بهینه توصیه می کند. در واقع هدایت الکتریکی یک متغیر کمکی است که مقادیر آن در هر چاه همزمان با نیترا ت اندازه گیری می شود و می توان با این اطلاع اضافی از وجود برخی چاه های دیگر چشم پوشی کرد. در پژوهش حاضر تعداد متغیرهای کمکی و اصلی برابر با هم می باشد. یکی از منافع استفاده از متغیر کمکی، به رغم برابر بودن تعداد دو متغیر، کم کردن واریانس^۲ برآورد می باشد.

- 1- Mutual Information
- 2- Variance

میدان تصادفی در موقعیت فضایی S_0 و لحظه زمانی t_0 بر اساس مشاهدات $Z(s_1, t_1), \dots, Z(s_n, t_n)$ است. میدان تصادفی فضایی-زمانی را می توان به صورت رابطه ۴ تجزیه کرد که در آن $\mu(s, t) = E(Z(s, t))$ تغییرات بزرگ مقیاس یا روند (میانگین) فضایی-زمانی، و $\delta(s, t)$ مانده یا تغییرات کوچک مقیاس میدان تصادفی می باشد.

$$Z(s, t) = \mu(s, t) + \delta(s, t), \quad s \in D, t \in T \quad (4)$$

وقتی میانگین میدان تصادفی تابعی از (s, t) باشد میدان تصادفی دارای روند است و باید به تغییرات آن توجه شود. برای این منظور، معمولاً (اولیاء و همکاران، ۲۰) میانگین فضایی-زمانی، با مدلی خطی (رابطه ۵) تقریب زده می شود که در آن شکل هر یک از توابع $f_j(s, t)$ برای $j = 0, \dots, p$ ، با توجه به ماهیت روند و موقعیت فضایی-زمانی داده ها تعیین می شود و $\beta = (\beta_0, \dots, \beta_p) \in \mathbb{R}^{p+1}$ برداری از پارامترهای نامعلوم است.

$$\mu(s, t) = \sum_{j=1}^{p+1} f_{j-1}(s, t) \beta_{j-1} \quad (5)$$

در این صورت به ازاء n نقطه در مکان و m نقطه در زمان، بردار مشاهدات

$$\mathbf{Z} = ((Z(s_1, t_1), \dots, Z(s_n, t_n)) = (Z(u_1), \dots, Z(u_N))$$

می توان به صورت ماتریسی $\mathbf{Z} = \mathbf{X}\beta + \delta$ نوشت که در آن \mathbf{X} ماتریسی با بعد $N \times (p+1)$ است. پیشگوی کریجینگ عام در هر موقعیت $(s_0, t_0) = u_0$ ، ترکیب خطی ناریب از مشاهدات به صورت $\hat{\mathbf{Z}}(u_0) = \lambda' \mathbf{Z}$ است که بردار ضرایب آن برابر با $\lambda = \{\gamma + \mathbf{X}(\mathbf{X}'\Gamma^{-1}\mathbf{X})^{-1}(x - \mathbf{X}'\Gamma^{-1}\gamma)\}' \Gamma^{-1}$ و $\mathbf{Y} = (\gamma(u_0 - u_1), \dots, \gamma(u_0 - u_N))'$ می باشد و Γ ماتریسی با بعد $N \times N$ با (i, j) امین درایه $\gamma(u_i - u_j)$ و $\mathbf{x} = (f_0(u_0), \dots, f_p(u_0))$ است.

واریانس پیشگوی کریجینگ عام نیز به صورت $\sigma^2(u_0) = 2\lambda' \gamma - \lambda' \Gamma \lambda$ قابل محاسبه می باشد (کرسی، ۶). اگر فرض شود میدان تصادفی به صورت رابطه ۶ تجزیه شود که در آن K و L دو مقدار صحیح است که بر حسب متغیر اصلی و کمکی تعیین می شود و $f_{kl}(u)$ توابعی معلوم است که شکل آنها با توجه به ماهیت

روند داده ها و متغیرهای کمکی تعیین می شود، $\beta = (\beta_0, \dots, \beta_{KL}) \in \mathbb{R}^{(K+1)(L+1)}$ بردار پارامترها، و $\delta(0,0)$ میدان تصادفی مانای ذاتی با میانگین صفر است.

$$Z(u) = \sum_{k=0}^K \sum_{l=0}^L f_{kl}(u) \beta_{kl} + \delta(u), \quad u = (s, t) \in D \times T \quad (6)$$

در این حالت، بهترین پیشگوی خطی ناریب را پیشگوی کریجینگ فضایی-زمانی با روند برونمی می نامند، زیرا متغیرهای کمکی که در برآورد روند به کار گرفته شده اند نیز از بیرون به داده ها تحمیل داده می شوند. براین اساس مدل ۶ را می توان به صورت

که در آن Q_L و Q_U به ترتیب چارک بالا و چارک پایین داده ها در بعد زمان محسوب می شود. براین اساس، داده های دور افتاده با توجه به مقادیر داده های اطراف در موقعیت مکانی همان سال حذف شد تا در گام بعدی با استفاده از کریجینگ برتر سری داده ها تکمیل گردد. همچنین برای بررسی نرمال بودن توزیع داده ها از آزمون کلمگرف اسمیرنوف (K-S) و در صورت نرمال نبودن توزیع داده ها از تبدیل باکس-کاکس^۱ استفاده می شود و در صورت عدم کارایی این روش می توان از روش های دیگری همانند توان دوم یا تبدیلات جانسون را بنا به ماهیت داده ها استفاده نمود.

برآورد نقاط فاقد مشاهده

برای مدل بندی داده های فضایی-زمانی معمولاً از میدان تصادفی $\{Z(s, t); (s, t) \in D \times T\}$ استفاده می شود، که در آن D زیر مجموعه ای از فضای اقلیدسی \mathbb{R}^d با بعد $d \geq 1$ ، $T \subseteq \mathbb{R}$ ، D ، T و R به ترتیب بعد فضایی، زمانی و مجموعه اعداد حقیقی می باشند. تغییرنا و هم تغییرنمای فضایی-زمانی این میدان که بیانگر ساختار همبستگی فضایی-زمانی میدان تصادفی هستند به صورت روابط ۱ و ۲ تعریف می شود:

$$2\gamma(s, s', t, t') = \text{var}(Z(s, t) - Z(s', t')), \quad (s, t), (s', t') \in D \times T \quad (1)$$

$$C(s, s', t, t') = \text{cov}(Z(s, t) - Z(s', t')), \quad (s, t), (s', t') \in D \times T \quad (2)$$

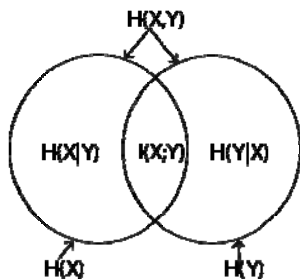
اگر برای $h_t = t - t'$ ، $h_s = s - s'$ که به ترتیب بیانگر تاخیر مکانی و تاخیر زمانی است، بتوان تابع هم تغییرنمای فضایی-زمانی را به صورت $C(h_s, h_t) = C_s(h_s) + C_t(h_t)$ تفکیک کرد که در آن $C_s(h_s)$ ، $C_t(h_t)$ هم تغییرنمای معتبر فضایی و زمانی باشند، این مدل هم تغییرنمای فضایی-زمانی جمعی نامیده می شود. اگر بتوان آن را به صورت $C_{s,t}(h_s, h_t) = C_s(h_s) \cdot C_t(h_t)$ بیان نمود، این مدل ضربی نام دارد. کوواریانس فضایی-زمانی به بعدهای فضایی و زمانی مستقل از هم تفکیک شد (دیاکو و همکاران، ۷) که در آن $\gamma_s(h_s)$ و $\gamma_t(h_t)$ به ترتیب، نیم تغییرنمای معتبر فضایی و زمانی هستند (رابطه ۳).

$$\gamma_{s,t}(h_s, h_t) = k_s \gamma_s(h_s) + k_t \gamma_t(h_t) - k_1 \gamma_s(h_s) \gamma_t(h_t) \quad (3)$$

به منظور برآورد فضایی-زمانی، بنا به شرایط موجود، همانند وجود روند در داده ها، از انواع مختلف کریجینگ می توان استفاده نمود. وجود روند در داده ها برآورد را به سوی استفاده از روش های کریجینگ عام و برونمی سوق می دهند.

یکی از اهداف تحلیل داده های فضایی-زمانی، برآورد مقدار

ماتریس $Z = W\beta + \delta$ نوشت، که در آن W ماتریسی با بعد $(L + 1)(K + 1) \times N$ بوده که درایه های آن به صورت $\omega_{p,q} = f_{kl}(s_i, t_j), p = ij, q = kl, i = 1, \dots, n, j = 1, \dots, m, k = 0, \dots, K, l = 0, \dots, L$ است. بردار β با کمینه کردن میانگین توان دوم خطاهای برآورد به صورت $\hat{\beta} = (W'\Sigma^{-1}W)^{-1}W'\Sigma^{-1}Z$ در آن ماتریس کوواریانس داده های فضایی-زمانی با درایه های $\Sigma_{kk'} = C(u_k - u_{k'})$, $k, k' = 1, \dots, N$ تابع کوواریانس $C(0)$ از آنجا که معمولاً نامعلوم است، باید برآورد شود. از طرفی وجود روند در داده ها منجر به ناریبی برآورد کوواریانس می شود و نمی توان آن را مستقیماً از مشاهدات برآورد کرد. در این حالت با یکی از روش های مدل سازی، مدلی مناسب به روند $\mu(u) = E(Z(u))$ برازش داده می شود. سپس با کم کردن مقدار برآورد روند در هر موقعیت از مقدار اندازه گیری شده، مقدار مانده ها به صورت رابطه γ برآورد می شود که $Z(u)$ مقدار مشاهده شده و $\hat{\mu}(u)$ مقدار تقریبی روند در u می باشد.



شکل ۲- رابطه شماتیک بین آنتروپی توام، شرطی و اطلاع متقابل

بطور کلی اطلاع متقابل کسب شده از مجموعه ای از متغیرهای تصادفی، توسط مجموعه ای دیگر از متغیرهای تصادفی را می توان تعریف کرد. اگر X_1, \dots, X_n و Y_1, \dots, Y_m متغیرهای تصادفی باشند، اطلاع متقابل کسب شده راجع به X_1, \dots, X_n و Y_1, \dots, Y_m به صورت $I(X_1, \dots, X_n; Y_1, \dots, Y_m) = H(X_1, \dots, X_n) - H(X_1, \dots, X_n | Y_1, \dots, Y_m)$ تعریف می شود.

برای تعیین شبکه بهینه پایش لازم است چاه ها را به ترتیب اهمیت دسته بندی نماییم. چاه های منتخب چاه هایی است که نسبت به سایر چاه ها از اهمیت بیشتری برخوردار باشد. در شبکه آبخوان دشت مشهد هر یک از چاه ها در شبکه موجود، مقداری از اطلاع را در اختیار قرار می دهند. سپس براساس چگونگی پراکنش اطلاع، مکان های با اطلاعات بیش تر در اولویت انتخاب چاه های جدید قرار می گیرد. مراحل رتبه بندی چاه ها به شرح زیر می باشد :

- ۱- مقدار آنتروپی هر چاه براساس داده های نیترات ۹ سال آماری (۸۹-۱۳۸۱) محاسبه گردید.
- ۲- چاه ها براساس مقدار آنتروپی به صورت نزولی مرتب شد.
- ۳- چاه یا چاه هایی که دارای بیشترین مقدار آنتروپی هستند رتبه اول را دارا می باشند.
- ۴- آنتروپی متقابل هریک از چاه های دیگر را با چاه (یا چاه های) منتخب محاسبه می شود.
- ۵- اعداد بدست آمده به ترتیب افزایشی مرتب و چاه (یا چاه هایی) با کمترین مقدار به عنوان چاه بعدی انتخاب گردید.
- ۶- مرحله ۴ و ۵ تا انتخاب آخرین چاه ادامه یافت.

اگر $H(X_1, X_2, \dots, X_m)$ آنتروپی توام m چاه باشد چون با افزودن مقادیر هر چاه، اطلاعات از شبکه بندی بیشتر می شود بنابراین $H(X_1, X_2, \dots, X_{237}) \geq H(X_1, X_2, \dots, X_m)$, $(m = 1, 2, \dots, 237)$ به همین منظور معیاری با نام اطلاع نسبی به این صورت تعریف می شود که T_m درصد اطلاع m چاه است اگر

ماتریس $Z = W\beta + \delta$ نوشت، که در آن W ماتریسی با بعد $(L + 1)(K + 1) \times N$ بوده که درایه های آن به صورت $\omega_{p,q} = f_{kl}(s_i, t_j), p = ij, q = kl, i = 1, \dots, n, j = 1, \dots, m, k = 0, \dots, K, l = 0, \dots, L$ است. بردار β با کمینه کردن میانگین توان دوم خطاهای برآورد به صورت $\hat{\beta} = (W'\Sigma^{-1}W)^{-1}W'\Sigma^{-1}Z$ در آن ماتریس کوواریانس داده های فضایی-زمانی با درایه های $\Sigma_{kk'} = C(u_k - u_{k'})$, $k, k' = 1, \dots, N$ تابع کوواریانس $C(0)$ از آنجا که معمولاً نامعلوم است، باید برآورد شود. از طرفی وجود روند در داده ها منجر به ناریبی برآورد کوواریانس می شود و نمی توان آن را مستقیماً از مشاهدات برآورد کرد. در این حالت با یکی از روش های مدل سازی، مدلی مناسب به روند $\mu(u) = E(Z(u))$ برازش داده می شود. سپس با کم کردن مقدار برآورد روند در هر موقعیت از مقدار اندازه گیری شده، مقدار مانده ها به صورت رابطه γ برآورد می شود که $Z(u)$ مقدار مشاهده شده و $\hat{\mu}(u)$ مقدار تقریبی روند در u می باشد.

تابع کوواریانس $C(0)$ از آنجا که معمولاً نامعلوم است، باید برآورد شود. از طرفی وجود روند در داده ها منجر به ناریبی برآورد کوواریانس می شود و نمی توان آن را مستقیماً از مشاهدات برآورد کرد. در این حالت با یکی از روش های مدل سازی، مدلی مناسب به روند $\mu(u) = E(Z(u))$ برازش داده می شود. سپس با کم کردن مقدار برآورد روند در هر موقعیت از مقدار اندازه گیری شده، مقدار مانده ها به صورت رابطه γ برآورد می شود که $Z(u)$ مقدار مشاهده شده و $\hat{\mu}(u)$ مقدار تقریبی روند در u می باشد.

$$\delta(u) = Z(u) - \hat{\mu}(u), \quad u \in D \times T \quad (7)$$

کوواریانس را بر اساس مانده ها، یا به عبارت دیگر داده های بی روند شده، برآورد نموده، با استفاده از این برآورد و کریجینگ معمولی، مقدار $\delta(u_0)$ برآورد می شود. سپس با توجه به رابطه ۶ کریجینگ فضایی-زمانی با روند برونی برای میدان تصادفی در موقعیت u_0 به صورت $\hat{Z}(u_0) = \omega_0' \hat{\beta} + \hat{\delta}(u_0)$ حاصل می شود، که در آن $\omega_0 = [f_0(u_0), \dots, f_{KL}(u_0)]'$ است.

تعیین اولویت چاه ها

اگر X متغیر تصادفی گسسته با احتمال $p_i = p(X=x_i), i=1, \dots, N$ باشد آن گاه آنتروپی X با رابطه ۸ بیان می شود. این رابطه برای n متغیر تصادفی X_1, \dots, X_n آنتروپی توام نام داشته و با رابطه ۹ بیان می شود که $p(x_1, \dots, x_n) = p(X_1=x_1, \dots, X_n=x_n)$ تابع احتمال توأم متغیرهای تصادفی X_1, \dots, X_n است.

$$H(X) = -\sum_{i=1}^N p_i \log p_i \quad (8)$$

$$H(X_1, \dots, X_n) = -\sum_{x_1, \dots, x_n} p(x_1, \dots, x_n) \log p(x_1, \dots, x_n) \quad (9)$$

همچنین اگر X_1, \dots, X_n و Y_1, \dots, Y_m متغیرهای تصادفی باشند، آنتروپی شرطی بیش از یک متغیر تصادفی به صورت رابطه ۱۰ تعریف می شود.

$$H(Y_1, \dots, Y_m | X_1, \dots, X_m) = -\sum_{y_1, \dots, y_m} p(y_1, \dots, y_m | X_1, \dots, X_m) \log p(y_1, \dots, y_m | X_1, \dots, X_m) \quad (10)$$

میزان اطلاع کسب شده راجع به X توسط Y به صورت

رابطه ۱۱ برقرار باشد.

$$T_m = \frac{H(X_1, X_2, \dots, X_m)}{H(X_1, X_2, \dots, X_{237})}, \quad m = 1, 2, \dots, 237 \quad (11)$$

واضح است زمانی $T_m=1$ خواهد بود که $m=237$ باشد. در این حالت بیشترین اطلاع از شبکه بندی موجود است. پس از اجرای مراحل ۱ تا ۶، چاه ها به ترتیب اطلاعاتی که در اختیار قرار می دهند رتبه بندی خواهند شد. بنابراین با وارد نمودن هریک از چاه ها به شبکه بندی بر میزان اطلاع افزوده خواهد شد. در نتیجه T_m نیز نسبت به اطلاعات آن چاه افزایش می یابد.

شبکه بندی جدید بر اساس درصد اطلاع خاص شکل می گیرد. تعیین و ارزیابی شبکه بهینه، با استفاده از داده های سال ۱۳۹۰ انجام شد. این داده ها شامل مقادیر نیترات و هدایت الکتریکی ۱۰۳ حلقه چاه در محدوده مورد مطالعه می باشد. برای سنجش اعتبار شبکه پیشنهادی که با میزان درصد اطلاع خاصی تعیین می شود از معیار های اعتباریابی روابط ۱۲ تا ۱۵ استفاده شد. که در آنها $Z(x_i)$ مقادیر واقعی $Z^*(x_i)$ مقادیر برآورد شده به روش کریجینگ (عام یا برونی) در سال ۱۳۹۰ می باشد.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n \{Z^*(x_i) - Z(x_i)\}^2} \quad (12)$$

$$MBE = \frac{\sum_{i=1}^n \{Z^*(x_i) - Z(x_i)\}}{n} \quad (13)$$

$$MAE = \frac{\sum_{i=1}^n |Z^*(x_i) - Z(x_i)|}{n} \quad (14)$$

$$R^2 = 1 - \frac{ESS}{TSS} \quad (15)$$

MAE و RMSE بیانگر دقت مدل درون یابی بوده و هر چه مقادیر این شاخص ها کمتر باشند و یا به صفر میل کنند، میزان دقت فضایی مدل بالاتر است. MBE نیز اختلاف بین مقدار برآورد شده و مشاهده شده را بیان می کند که هر چه به صفر نزدیکتر باشد برآورد ها از دقت بالاتری برخوردارند. براین اساس کاهش مقادیر RMSE، MBE و MAE دلیلی بر نزدیکی مقادیر برآورد شده و واقعی می باشد و افزایش R^2 نیز بر دقت برآورد انجام شده صحنه می گذارد. با استفاده از معیار های مذکور کریجینگ برتر انتخاب می شود. بنابراین مکان و زمان های فاقد مشاهده توسط کریجینگ برتر برآورد می شود. این بدان معنی است که داده های نیترات ۲۳۷ حلقه چاه در ۹ سال آماری در اختیار است. پس از انتخاب کریجینگ برتر و استفاده مجدد از داده های سال ۱۳۹۰، اعتباریابی به منظور انتخاب شبکه بهینه صورت می گیرد. انتخاب شبکه بهینه با کاهش درصد اطلاع تا رسیدن به خطایی معقول در اعتباریابی انجام شده است.

نتایج و بحث

تحلیل مقدماتی داده ها

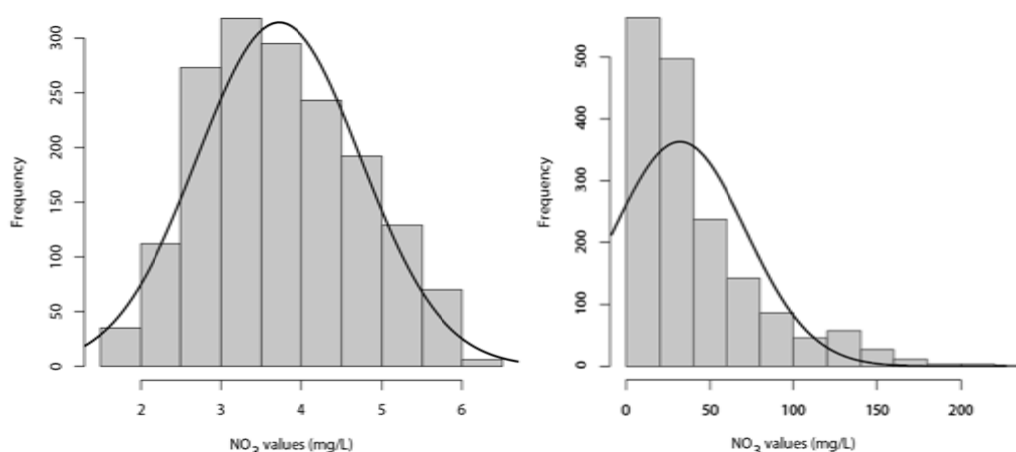
با استفاده از روش هایینگ (۱۱)، ۹۹ داده دور افتاده مشخص و حذف گردید. پرت بودن این داده نسبت به سایر داده های همان گروه (داده های یک چاه در طول دوره آماری) به صورت اتفاقی نیز تایید گردید. همچنین براساس آزمون کلموگورف-اسمیرنوف^۱، نرمال بودن توزیع داده های نیترات موجود بررسی گردید. فرض نرمال بودن داده ها با $p\text{-value}=2.2 \times 10^{-16}$ رد شد و پس از اعمال تبدیل باکس-کاکس با $\lambda = 0.05$ با استفاده از آزمون مذکور نرمال بودن توزیع داده ها با $p\text{-value}=0.1$ و سطح خطای ۰/۰۵ تایید گردید. (شکل ۳) نمودارهای پراکنش داده های نیترات در راستاهای x و y (طول و عرض جغرافیایی) برای ۲ سال نمونه آماری (۱۳۸۸ و ۱۳۸۹) در شکل ۴ نشان داده شده است. این نمودارها بیانگر این است که داده ها در جهت محور x دارای روند افزایشی کمی است و در جهت محور y دارای روند کاهشی است لذا داده های نیترات مانا نیست. برای کاستن از اثر روند بر برآورد تغییرنما و نهایتا کریجینگ، روند مشاهدات تعیین و بر اساس داده های بدون روند نسبت به برآورد تغییرنما اقدام گردید. با توجه به وجود روند در جهت محور x و y و تغییرات آن در سال های مختلف، رگرسیون $\mu(u)$ نسبت به متغیرهای x ، y و t با ضریب تعیین ۰/۷۸۳ به صورت رابطه ۱۶ انجام شد.

$$\mu(u) = 1.96 \times 10^4 - 2.61 \times 10^{-2}(x) - 4.99 \times 10^{-3}(y) - 8.03 \times 10^2(t) + 6.48 \times 10^{-9}(xy) + 1.1 \times 10^{-3}(xt) + 1.99 \times 10^{-4}(yt) - 2.7 \times 10^{-10}(xyt) \quad (16)$$

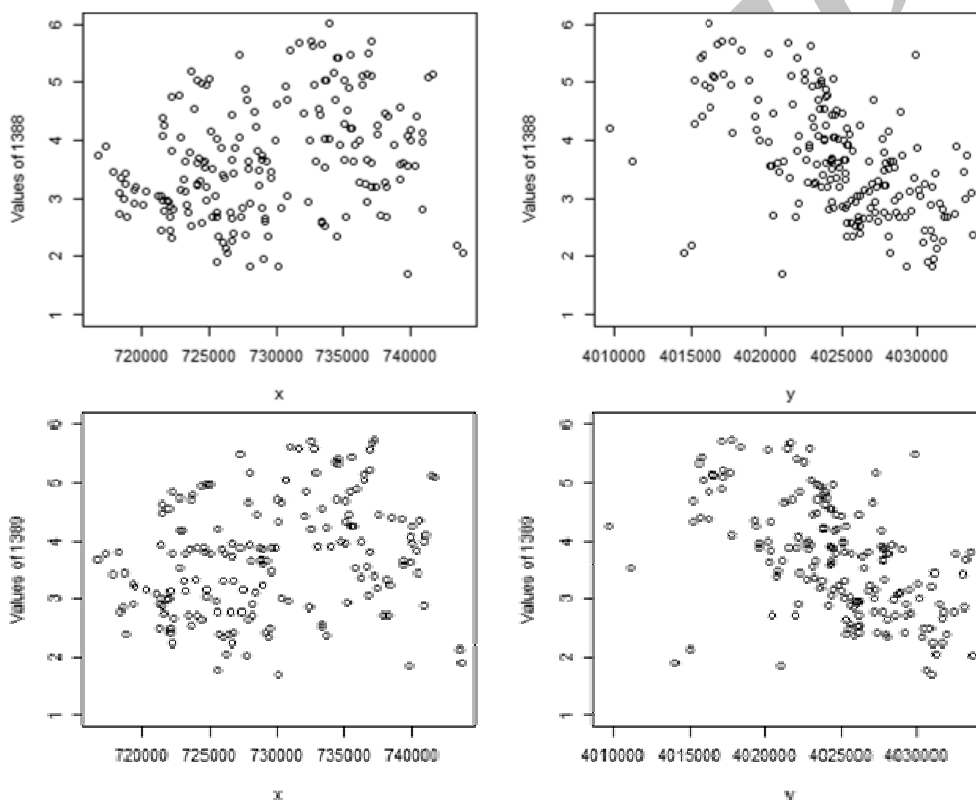
از برآورد پارامترهای رگرسیون مذکور مقدار مانده ها بدست آمد و با استفاده از مانده ها، وجود روند و همسانگردی بررسی شد. مقادیر $\delta(u)$ در برابر راستاهای x و y برای ۲ سال نمونه آماری در نمودارهای شکل ۵ ارائه شده است. همان طور که ملاحظه می گردد، روند موجود در داده ها بطور کامل از بین رفته است. مسلما در سال های مختلف روند شکل یکسانی ندارد.

پس از حذف روند، همسانگردی داده های بدون روند بررسی گردید. در صورتی که تغییرنما در تمام جهات تقریبا بر هم منطبق باشند و تفاوت معنی داری نسبت به هم نداشته باشند، رویه ی تغییرنما متقارن است و همسانگردی نیز برقرار است. بر اساس این تقارن و با توجه به شکل ۶ که رویه ی تغییرنماها مربوط به سال های نمونه آماری مورد مطالعه، از ۱۳۸۶ تا ۱۳۸۹ نشان داده شده است همسانگردی داده های بدون روند مورد تایید است. همچنین براساس نمودار میانگین در برابر واریانس که در شکل ۷ نشان داده شده است، مانایی واریانس نیز برقرار می باشد.

1- Kolmogorov-Smirnov Test



شکل ۳- هیستوگرام داده های نیترات به صورت خام (راست) و تبدیل یافته با توزیع نرمال (چپ)



شکل ۴- نمودارهای پراکنش داده های نیترات در راستاهای x و y برای ۲ سال نمونه آماری (۱۳۸۸ و ۱۳۸۹)

کریجینگ فضایی- زمانی عام

به منظور انجام کریجینگ عام ابتدا مدلی را به داده های تغییرنا برآزش داده و سپس با استفاده از این مدل نقاط فاقد مشاهده برآورد گردید. بدین منظور به هر یک از ابعاد مکان و زمان، مدل های کروی، نمائی، خطی و گوسی برآزش داده شد که در بین مدل ها، مدل نمائی که در بعد مکان و زمان با کمترین RMSE (در

اعتباریابی متقابل)، به ترتیب برابر $2/72$ و $3/67$ بدست آمد، بهترین برآزش برای تغییرنا در هر دو بعد مکان و زمان بود (روابط ۱۷ و ۱۸ و شکل ۸). همچنین با استفاده از آنها مدل مناسب سه بعدی برای کریجینگ بدست آمد. هرچند به نظر می رسد بالا بودن اثر قطعه ای واریوگرام زمانی، استفاده از عامل زمان را تحت تاثیر قرار دهد اما به دلیل اینکه اثر قطعه ای واریوگرام زمانی نسبت به سقف واریوگرام زیاد نمی باشد می توان وجود عامل زمان را توجیه کرد. این بدین

$\gamma(h_s, h_t) = \gamma(h_s, 0) + \gamma(0, h_t) - k\gamma(h_s)\gamma(h_t)$ (۲۰)
 در شکل ۹ (راست) که از داده های واقعی بدست آمده ملاحظه می شود داده های واقعی در موقعیت فضایی-زمانی، شکل نمایی به خود گرفته است بنابراین مدل نمائی به آن برازش گردید (شکل ۹-چپ).

پس از انجام کریجینگ مانده ها و برآورد نقاط فاقد مشاهده، منحنی های هم مقدار نیترات رسم گردید که نتایج برای ۳ سال اخیر یعنی سال های آماری ۱۳۸۷ تا ۱۳۸۹ در شکل ۱۰ نشان داده شده است. همانگونه که از منحنی های هم مقدار نیترات (شکل ۱۰) برمی آید، ۴ جبهه اصلی آلودگی در شمال، غرب، مرکز و جنوب غرب آبخوان دشت مشهد که حضور نیترات بیش از ۵۰ میلی گرم در لیتر است، مشاهده می شود. نتایج نشان می دهد در دوره آماری مورد مطالعه از سال ۱۳۸۱ تا ۱۳۸۹ میزان آلودگی آب های زیرزمینی به نیترات در هر ۴ جبهه اصلی به ویژه جبهه مرکزی و جنوب در حال افزایش می باشد.

مفهوم است که بین داده ها ساختار زمانی مناسبی برقرار بوده و بخش قابل قبولی از واریانس کل به وسیله واریوگرام توجیه می شود.

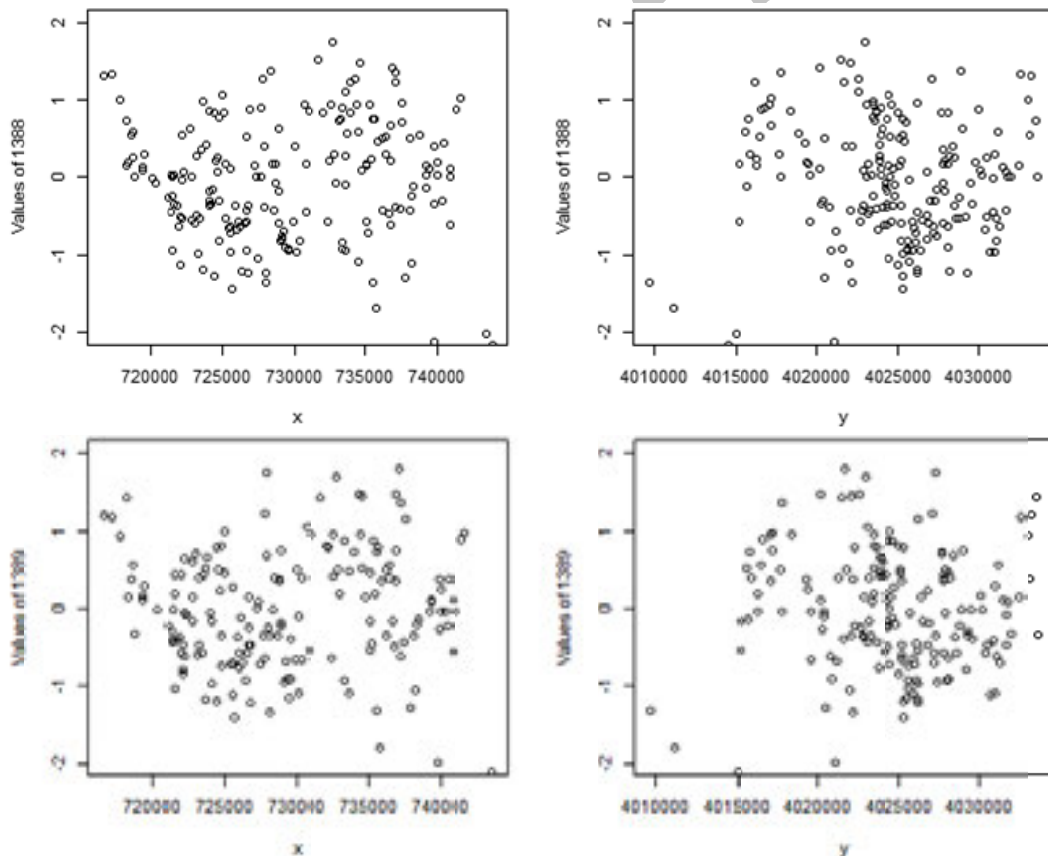
$$\hat{\gamma}(h_s) = 0.0761 + 0.529 \left(1 - \exp\left\{\frac{\|h_s\|}{1235}\right\}\right) \quad (۱۷)$$

$$\hat{\gamma}(h_t) = 0.253 + 0.07(1 - \exp\left\{\frac{\|h_t\|}{750}\right\}) \quad (۱۸)$$

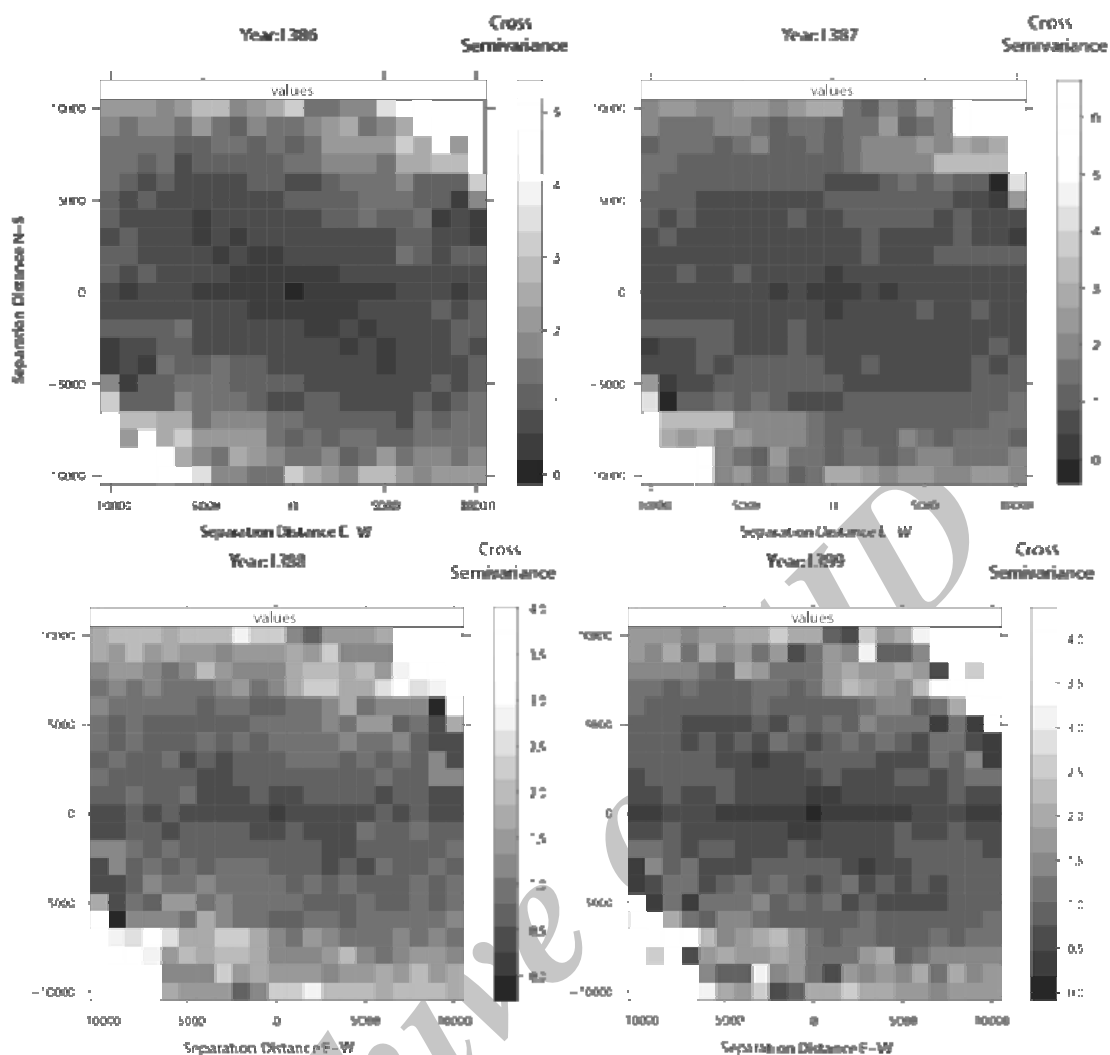
از آنجا که مدل نمایی به هر یک از دو تغییرنمای فضایی و زمانی بطور جداگانه قابل برازش است، از مدل تفکیک پذیر ضربی که توسط دیاکو و همکاران (۷) معرفی شد برای برازش بر تغییرنمای فضایی-زمانی به صورت رابطه ۱۹ استفاده گردید.

$$\gamma(h_s, h_t) = k_s\gamma(h_s) + k_t\gamma(h_t) - k_{st}\gamma(h_s)\gamma(h_t) \quad (۱۹)$$

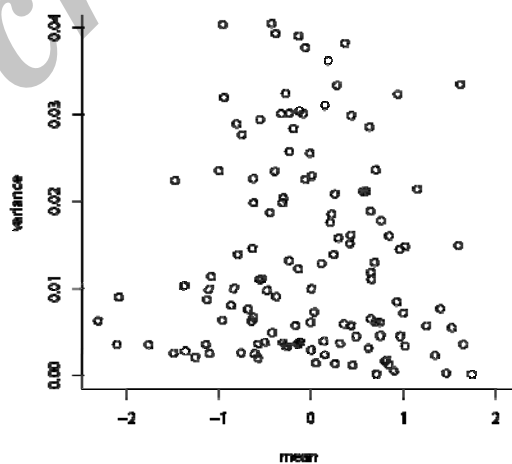
پس از محاسبه و رسم تغییرنمای سه بعدی در نرم افزار R (شکل ۹) تغییرنمای حاشیه ای $\gamma(0, h_t)$ و $\gamma(h_s, 0)$ که مقادیر تغییرنما در یک زمان یا مکان معین می باشند، بدست آمد و سپس با استفاده از معادله ۱۹ مدل ۲۰ که در آن $k=1.881$ محاسبه شد، برازش گردید.



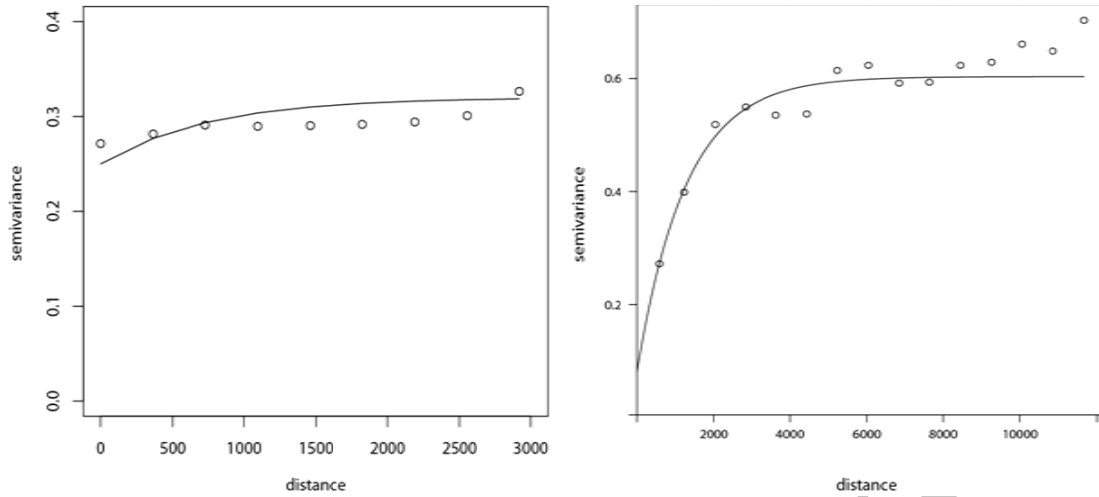
شکل ۵- نمودارهای پراکنش مقادیر $\delta(u)$ در راستاهای x و y برای ۲ سال نمونه آماری (۱۳۸۸ و ۱۳۸۹)



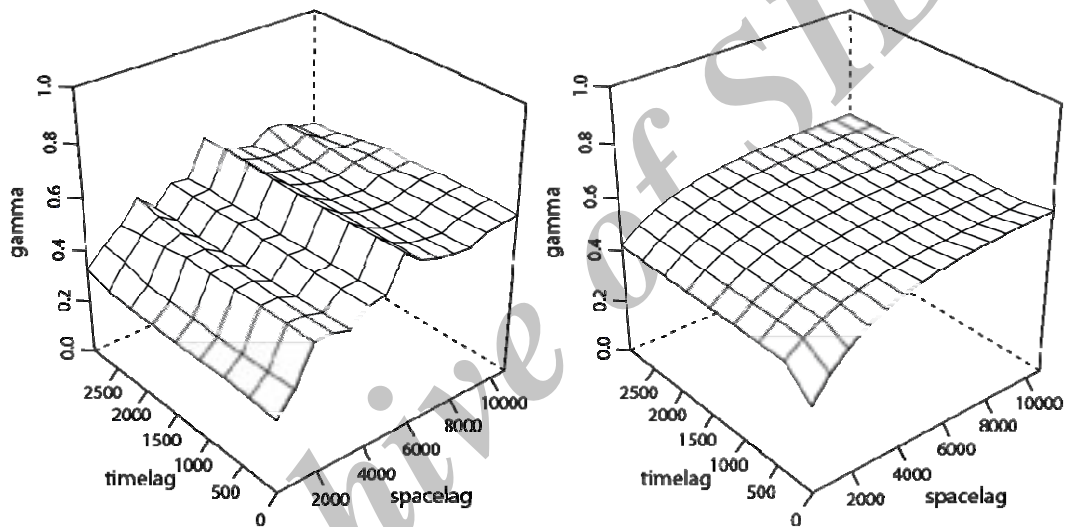
شکل ۶- رویه ی تغییرنماها در سال های نمونه آماری مورد مطالعه (۱۳۸۶ تا ۱۳۸۹)



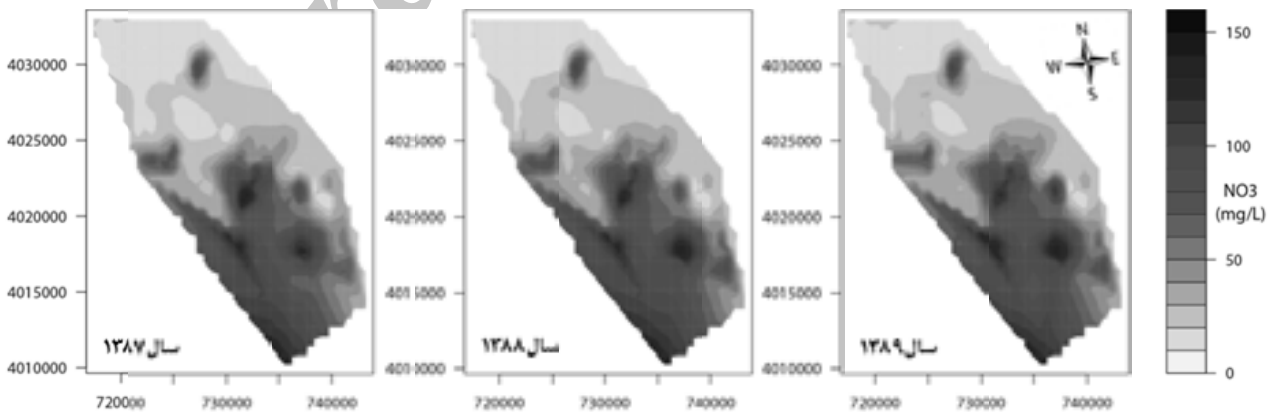
شکل ۷- نمودار پراکنش واریانس در برابر میانگین داده های نیترات



شکل ۸- تغییرنماهای فضایی (راست) و زمانی (چپ) داده های نیترات برای کریجینگ عام



شکل ۹- تغییرنماهای سه بعدی تجربی (راست) و برازش داده شده (چپ) برای کریجینگ عام



شکل ۱۰- منحنی های هم مقدار نیترات حاصل از کریجینگ عام در ۳ سال پایانی طرح (۱۳۸۷ تا ۱۳۸۹)

کریجینگ فضایی- زمانی با روند برون

در پژوهش حاضر، فرض بر این است که استفاده از متغیری که همبستگی زیادی با نیترا دارد می تواند برآورد دقیق تری از مقادیر نیترا را در نقاط فاقد مشاهده فراهم نماید. به همین منظور از پارمتر کیفی هدایت الکتریکی (EC) به عنوان متغیر کمکی استفاده شد. ضریب همبستگی EC و نیترا برابر ۰/۶۵ بدست آمد لذا از پارامتر EC در رابطه رگرسیونی برآورد روند استفاده گردید. براین اساس برآورد روند به صورت رابطه ۲۱ می باشد. همانند بخش کریجینگ عام، برای انتخاب بهترین مدل تغییرنمای فضایی و زمانی مدل های کروی، نمائی، خطی و گوسی برازش داده شد که در بین مدل ها برای هر یک از ابعاد فضا و زمان مدل نمائی با کمترین RMSE (در بعد مکان و زمان به ترتیب برابر ۲/۲۲ و ۲/۵۸) و به صورت معادلات ۲۲ و ۲۳ به مانده ها برازش داده شد. شکل ۱۱ نمودارهای برازش مدل های فضایی و زمانی را نشان می دهد.

$$\mu(u) = 2.1 \times 10^4 - 2.8 \times 10^{-2}(x) - 5.21 \times 10^{-3}(y) - 1.21 \times 10^3(t) + 6.95 \times 10^{-9}(xy) + 1.64 \times 10^{-3}(xt) + 2.99 \times 10^{-4}(yt) - 4.1 \times 10^{-10}(xyt) - 4.1 \times 10^{-10}(Ec) \quad (21)$$

$$\hat{\gamma}(h_s) = 0.05119 + 0.3682 \left(1 - \exp \left\{ -\frac{\|h_s\|}{1404.5} \right\} \right) \quad (22)$$

$$\hat{\gamma}(h_t) = 0.253 + 0.07 \left(1 - \exp \left\{ -\frac{\|h_t\|}{750} \right\} \right) \quad (23)$$

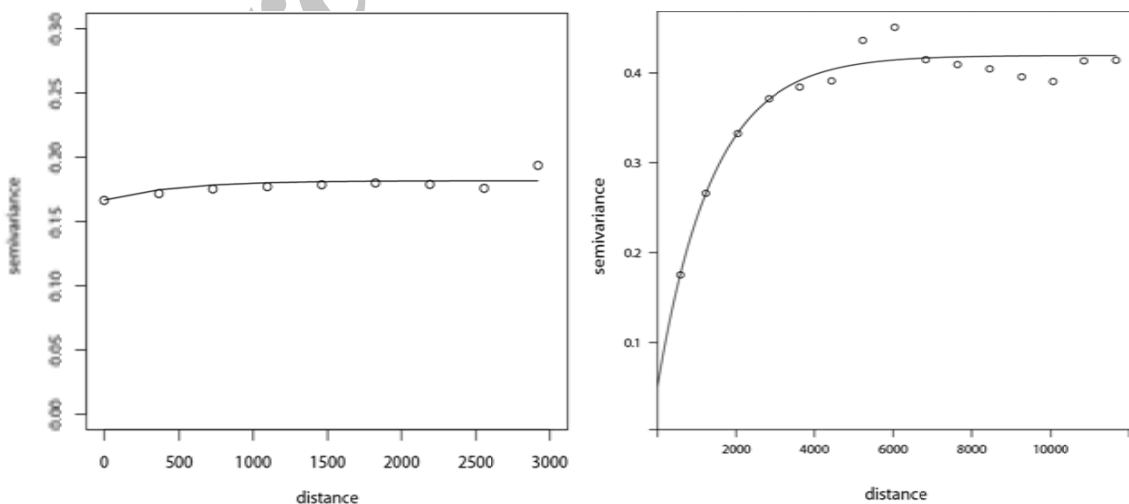
تغییرنمای سه بعدی مانده های حاصل از معادله رگرسیونی در نرم افزار R محاسبه و رسم گردید که در شکل ۱۲ نشان داده شده است. تغییرنماهای حاشیه ای $\gamma(h_s, 0)$ و $\gamma(0, h_t)$ نیز بدست آمد و مدل رابطه ۲۴ که در آن $k=0.881$ محاسبه گردید، بر تغییرنما برازش داده شد. شکل ۱۲ معادله برازش داده شده به تغییرنما را نمایش می دهد.

$$\gamma(h_s, h_t) = \gamma(h_s, 0) + \gamma(0, h_t) - k\gamma(h_s)\gamma(h_t) \quad (24)$$

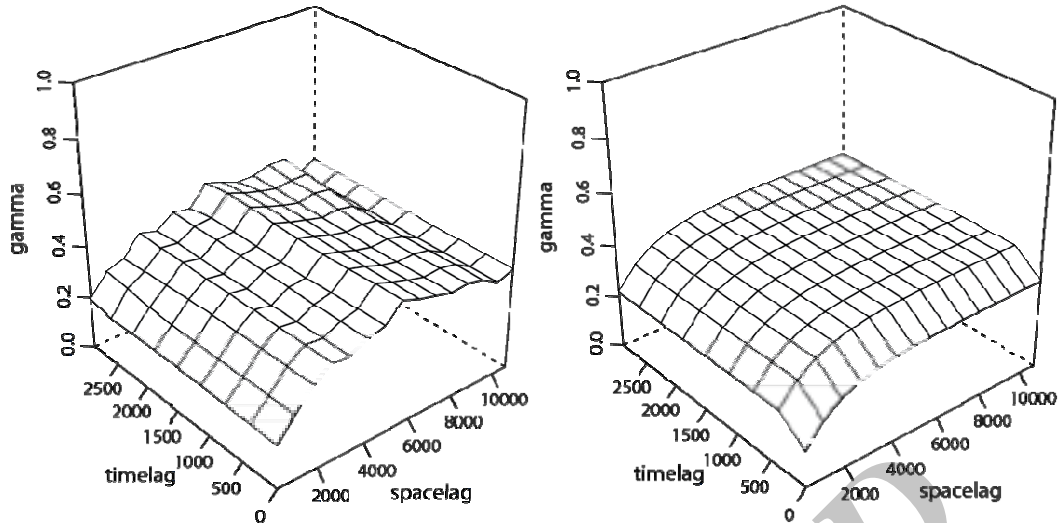
در شکل ۱۲ (راست) که از داده های واقعی بدست آمده ملاحظه می شود داده های واقعی در موقعیت فضایی- زمانی، شکل نمایی به خود گرفته است بنابراین مدل نمائی به آن برازش گردید (شکل ۱۲- چپ).

مقادیر تغییرنما با روند برون در مقایسه با تغییرنمای فضایی- زمانی کاهش مشهودی داشت یعنی با افزودن EC به مدل، تغییرنما در فاصله فضایی- زمانی کاهش یافته و همبستگی فضایی- زمانی را بیش تر نشان داد. پس از انجام کریجینگ مانده ها به روش برون و بدست آوردن برآورد نقاط فاقد مشاهده، نمودار منحنی های هم مقدار نیترا برای ۳ سال ابتدا و انتهای مورد مطالعه یعنی سال های آماری ۱۳۸۱ تا ۱۳۸۳ و ۱۳۸۷ تا ۱۳۸۹ رسم گردید. همانگونه که در منحنی های هم مقدار نیترا ارائه شده در شکل ۱۳ ملاحظه می شود میزان آلودگی آب های زیرزمینی به نیترا در ۴ جبهه اصلی آلودگی در شمال، غرب، مرکز و جنوب آبخوان دشت مشهد در دوره آماری مورد مطالعه از سال ۱۳۸۱ تا ۱۳۸۹ در حال افزایش بوده است.

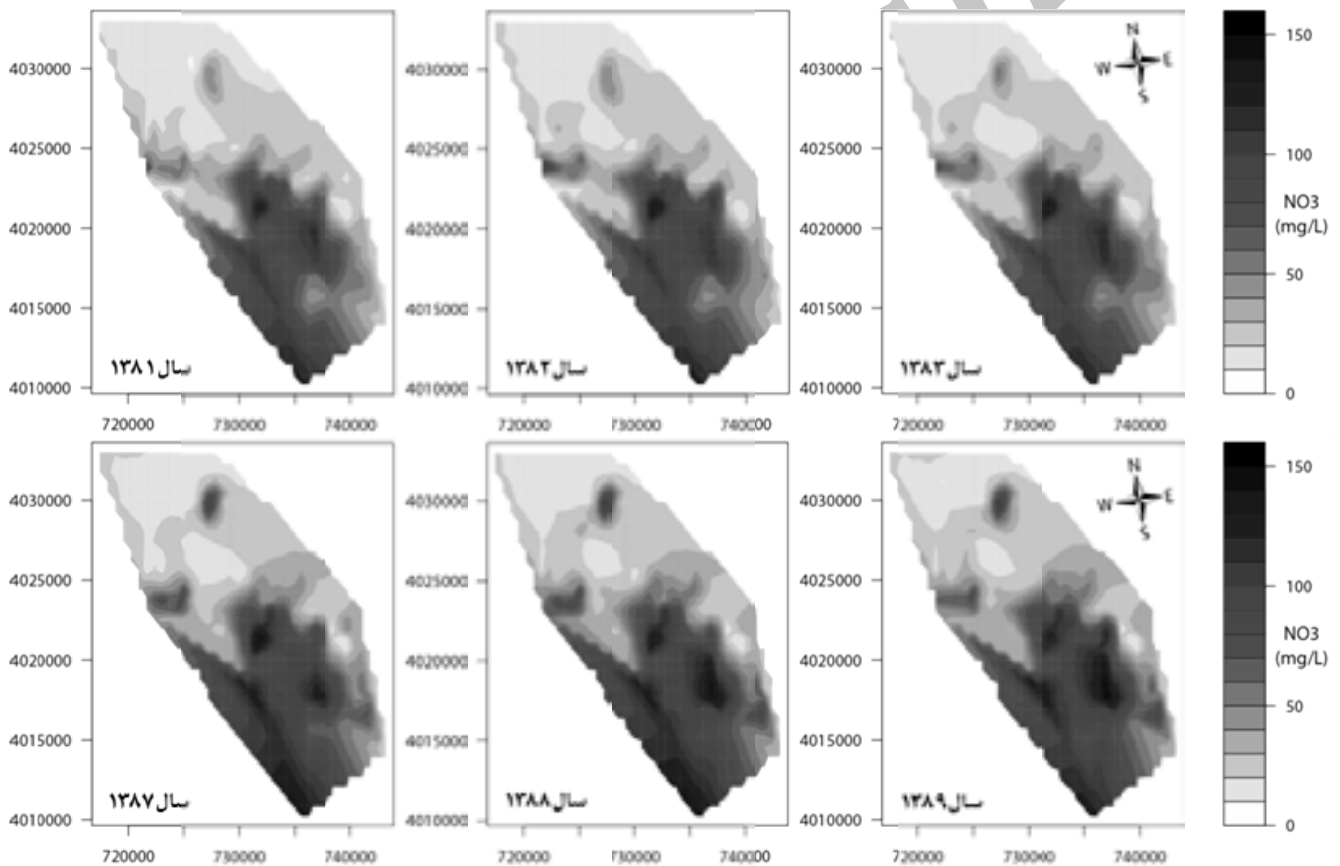
پس از تعیین نقاط فاقد مشاهده با استفاده از کریجینگ عام و برون داده های آماری مورد مطالعه تکمیل شد. انحراف معیار برای کل داده های هر حلقه چاه در دوره آماری مورد مطالعه اعم از داده های اولیه و برآورد شده به روش های کریجینگ، محاسبه گردید. در ادامه منحنی های هم مقدار برای مقادیر انحراف معیار به تفکیک دو روش کریجینگ عام و برون در شکل ۱۴ ارائه شده است. این نمودارها نشان می دهد مقادیر انحراف معیار کریجینگ برون از کریجینگ عام کمتر است.



شکل ۱۱- تغییرنمای فضایی (راست) و زمانی (چپ) مانده ها برای کریجینگ برون



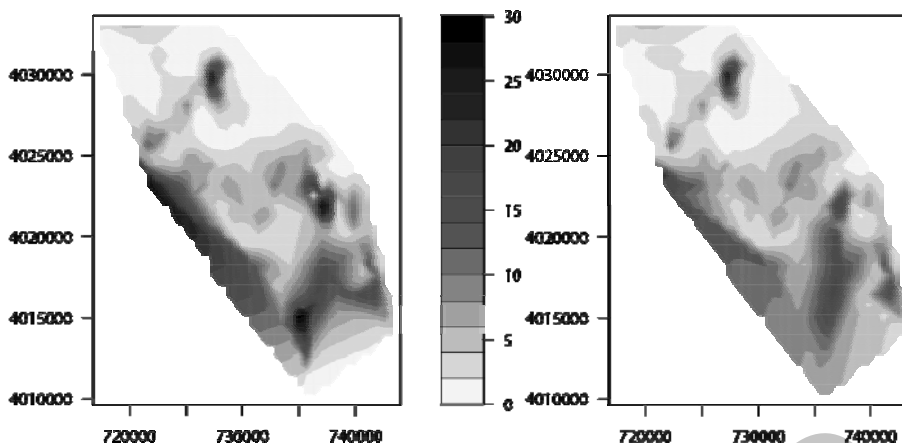
شکل ۱۲- تغییرنماهای سه بعدی تجربی (راست) و برازش داده شده (چپ) برای کریجینگ برون



شکل ۱۳- منحنی های هم مقدار نیترات حاصل از کریجینگ برون برای ۳ سال ابتدا و انتهای طرح

حلقه چاه گزارش شده است. با استفاده از روش های کریجینگ مورد مطالعه برای ۱۰۳ حلقه چاه دارای آمار کیفی در سال ۱۳۹۰، معیارهای اعتباریابی به منظور مقایسه روش های مذکور در جدول ۱ آمده است.

این امر در مطالعه موردی که توسط محمد زاده و شرفی (۲) نیز انجام گردیده، به وضوح دیده می شود. در این پژوهش برای تعیین بهترین روش کریجینگ از داده های موجود در سال مقصد (۱۳۹۰) استفاده شد. این داده ها برای ۱۰۳



شکل ۱۴- مقادیر انحراف معیار براساس کریجینگ عام (راست) و کریجینگ برونی (چپ)

جدول ۱- اعتباریابی روش های کریجینگ عام و برونی

روشن کریجینگ	MAE	R ²	MBE	RMSE
عام	۱/۳۶۳	۰/۷۲۶	-۰/۶۹۴	۲/۱۸۲
برونی	۱/۰۲۹	۰/۸۳۰	-۰/۳۲۸	۱/۷۱۷

را به آن تعمیم دهد، مربوط است. لذا پس از برآورد نقاط فاقد مشاهده توسط کریجینگ فضایی- زمانی برونی برای انتخاب شبکه بهینه ابتدا با استفاده از داده های دوره آماری مورد مطالعه، آنترپوی ۲۳۷ حلقه چاه محاسبه گردید. نظر به اینکه دو حلقه چاه با شماره های ۳ و ۸۴ در داده های مورد مطالعه دارای آنترپوی یکسان و در بین همه چاه ها دارای بیشترین مقادیر آنترپوی هستند به عنوان اولویت اول انتخاب شدند. سپس آنترپوی متقابل دو حلقه چاه مذکور با سایر چاه ها محاسبه گردید. برای اولویت بعدی در شبکه پایش بهینه، چاه شماره ۱۱۶ که دارای کمترین اطلاع متقابل با دو چاه اولیه بوده انتخاب شد. برای انتخاب اولویت بعدی آنترپوی متقابل ۳ حلقه چاه منتخب (داخل شبکه) با سایر چاه ها محاسبه شد و چاه شماره ۲۸ که دارای کمترین اطلاع متقابل است انتخاب گردید. این عملیات تا انتخاب آخرین اولویت ادامه پیدا کرد.

آنترپوی توام شبکه کامل با ۲۳۷ حلقه چاه شامل داده های نیترا ت مربوط به ۹ سال آماری (۸۹-۱۳۸۱) برای هر حلقه چاه، برابر ۷/۴۵ بدست آمد. این بدان معنی است که میزان عدم حتمیتی (اطلاعات) که مقادیر نیترا ت در اختیار قرار می دهد برابر ۷/۴۵ واحد می باشد. براساس درصد اطلاع ۹۵، با انتخاب ۱۶۴ حلقه چاه که در اولویت شبکه بهینه می باشد، آنترپوی شبکه برابر ۷/۱ بدست آمد. به همین ترتیب برای درصدهای اطلاع ۹۰، ۸۹، ۸۸ و ۸۵ مقادیر آنترپوی به ترتیب ۶/۷۲، ۶/۶۲، ۶/۵۹ و ۶/۳۴ و تعداد حلقه های چاه دارای اولویت در هر شبکه به ترتیب ۱۱۱، ۱۰۰، ۹۷ و ۷۷ بدست آمد. به منظور انتخاب شبکه ای بهینه با درصد اطلاع مناسب، داده

با توجه به مستندات ارائه شده در جدول ۱ ملاحظه می شود مقادیر RMSE، MBE و MAE برای کریجینگ برونی نسبت به کریجینگ عام کمتر است و همچنین R² برای کریجینگ برونی مقدار بیشتری است که نشان دهنده برتری کریجینگ برونی نسبت به عام می باشد. به عبارت دیگر عامل کمکی (EC) نقش موثری در کاهش خطای پیش بینی خواهد داشت. محمد زاده و شرفی (۲) نیز در بررسی کریجینگ برونی نسبت به کریجینگ عام، نشان دادند که برای پایش ۳۰ ایستگاه باران سنجی در ۱۲ ماه، میزان واریانس برآورد کاهش چشم گیری دارد. برای کریجینگ داده های نیترا ت دشت مشهد نیز وجود داده های کمکی هدایت الکتریکی علاوه بر کاهش واریانس برآورد، اعتبار بالایی در برآورد آینده خواهد داشت. این بدان معنی است که کریجینگ برونی با دقت بسیار بالاتری می تواند تحولات آینده آلودگی منابع آب زیرزمینی به نیترا ت را پیش بینی نماید.

انتخاب شبکه بهینه براساس آنترپوی

کریجینگ برونی برای برآورد نقاط فاقد مشاهده، به روش اعتباریابی بیرونی به وسیله معیارهای ارزیابی انتخاب شد. این اعتباریابی به قابلیت تعمیم پذیری یافته های پژوهش ارتباط دارد. به این معنی که آیا نتایج آزمایش را می توان به جامعه ای که نمونه از آن انتخاب شده است، تعمیم داد یا نه. به عبارت دیگر، آیا یافته های پژوهش، معرف و بیانگر شرایط و موقعیت های زمانی و مکانی خاص دیگر هستند؟ اعتبار بیرونی با جامعه ای که پژوهشگر قصد دارد نتایج

چاه ها را ارائه می دهد.

نتیجه گیری

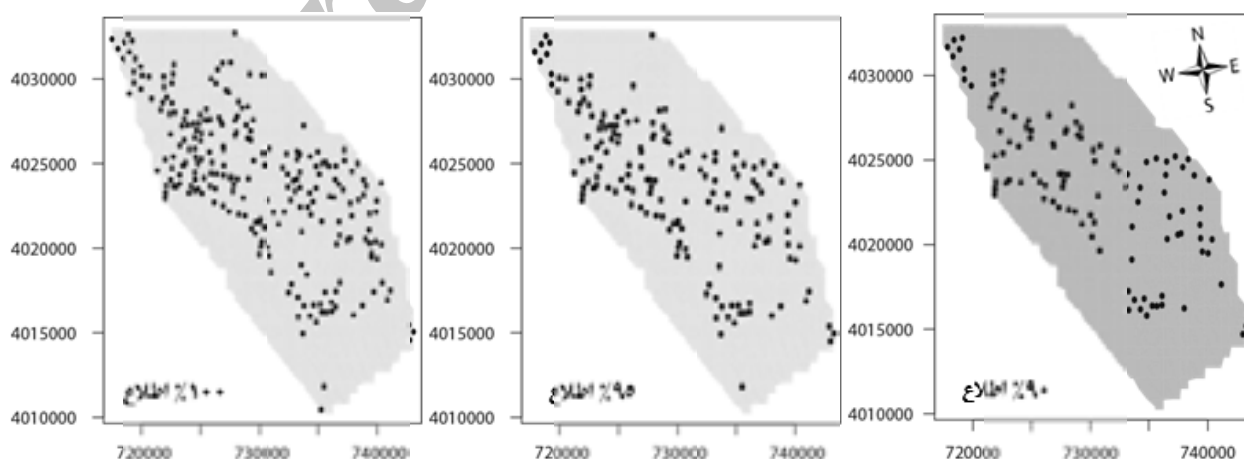
به منظور انتخاب کریجینگ برتر برای برآورد آینده و براساس نتایج بدست آمده، انحراف معیار برای کریجینگ برونمی نسبت به کریجینگ عام کاهش قابل ملاحظه ای داشته است. از طرفی پس از پیش بینی ۱۰۳ حلقه چاه در سال ۱۳۹۰، معیار RMSE برای روش عام و برونمی به ترتیب ۲/۱۸۲ و ۱/۷۱۷ بدست آمد که نشان دهنده برتری کریجینگ برونمی بر عام می باشد. همچنین معیارهای MBE و MAE نیز برای روش عام به ترتیب $-۰/۶۹۴$ و $۱/۳۶۳$ و برای روش برونمی $-۰/۳۲۸$ و $۱/۰۲۹$ بدست آمد. کمتر بودن این دو معیار نیز برتری روش برونمی را نشان می دهد. در انتها با مقایسه معیار R^2 که در روش عام $۰/۷۲۶$ و در روش برونمی $۰/۸۳$ بدست آمد می توان به علت بزرگتر بودن این مقدار، روش برونمی را گزینه برتر دانست. بنابراین روش برونمی هم از لحاظ درون یابی (براساس انحراف معیار) و هم از لحاظ برون یابی (براساس معیارهای اعتبارسنجی) نسبت به روش عام برتری قابل ملاحظه ای دارد.

های ۱۰۳ حلقه چاه مورد استفاده در اعتباریابی روش های کریجینگ، مجددا در اعتباریابی شبکه های پیشنهادی نیز بکار گرفته شد. براین اساس پس از پیش بینی مقادیر نیترا ت برای ۱۰۳ حلقه چاه مربوط به سال مقصد (۱۳۹۰) بوسیله هریک از شبکه های بهینه پیشنهاد شده بر مبنای درصد های اطلاع، معیارهای اعتباریابی محاسبه گردید که نتایج محاسبات در جدول ۲ ارائه شده است. همچنین توزیع چاه ها در شبکه پایش اصلی و شبکه بندی های پیشنهادی بر اساس درصد های اطلاع ۱۰۰، ۹۵ و ۹۰ ارائه شده در جدول ۲ در شکل ۱۵ مشاهده می شود.

با مقایسه جداول ۱ و ۲ ملاحظه می شود که تا ۹۰ درصد اطلاع (۱۱۱ حلقه چاه) معیارهای اعتباریابی شبکه نسبت به معیارهای متناظر آن در کریجینگ عام که به وسیله شبکه کامل بدست آمده، مقادیر بهتری اختیار نموده است. با کاهش درصد اطلاع از ۹۰ به ۸۹ درصد ملاحظه می شود که معیارهای اعتباریابی کریجینگ عام شبکه کامل نسبت به کریجینگ برونمی با شبکه نوع ۳ بهتر می باشد. اولیاء و همکاران (۲۰) نیز برای تعیین شبکه بندی از آنتروپی استفاده نمود ولی با این تفاوت که ابتدا با استفاده از خوشه بندی فازی پهنه مورد مطالعه تقسیم بندی شد. سپس با برازش منحنی T-D نشان دادند که برای داده های آبخوان تهران-کرج، آنتروپی حاشیه ای بهترین شبکه

جدول ۲- معیارهای اعتباریابی انواع شبکه بندی براساس درصد اطلاع با کریجینگ برونمی

نوع شبکه	درصد اطلاع	تعداد چاه	آنتروپی	RMSE	MBE	R^2	MAE
شبکه کامل	۱۰۰	۲۳۷	۷/۴۵	۱/۷۱۷	-۰/۳۲۸	۰/۸۳۰	۱/۰۲۹
شبکه نوع ۱	۹۵	۱۶۴	۷/۱۰	۲/۰۴۹	-۰/۵۱۰	۰/۷۵۸	۱/۲۳۶
شبکه نوع ۲	۹۰	۱۱۱	۶/۷۲	۲/۱۷۷	-۰/۵۹۶	۰/۷۲۷	۱/۳۱۷
شبکه نوع ۳	۸۹	۱۰۰	۶/۶۲	۲/۲۱۴	-۰/۶۳۹	۰/۷۱۸	۱/۳۵۱
شبکه نوع ۴	۸۸	۹۷	۶/۵۹	۲/۲۳۰	-۰/۶۶۵	۰/۷۱۳	۱/۳۷۰
شبکه نوع ۵	۸۵	۷۷	۶/۳۴	۲/۲۶۵	-۰/۸۰۸	۰/۷۰۴	۱/۴۲۷



شکل ۱۵- پراکنش چاه ها در شبکه های پایش پیشنهادی بر اساس درصد های اطلاع

نمونه برداری، انتقال مواد و تجهیزات بین محل های نمونه برداری و آزمایشگاه مرکزی و آنالیز آزمایشگاهی پارامترهای کیفی، کاهش می یابد، در نتیجه شبکه نوع ۲ با ۱۱۱ حلقه چاه که هزینه های به مراتب کمتری برای پایش کیفی منابع آب زیرزمینی تحمل می کند، می تواند مدلی مناسب برای شبکه بندی و پایش کیفی آبخوان دشت مشهد باشد.

در این پژوهش، راهکاری جهت استفاده از آنتروپی برای تعیین شبکه بهینه پایش کیفی منابع آب زیرزمینی ارائه گردید. همچنین اثبات گردید که با استفاده همزمان از کریجینگ فضایی- زمانی و آنتروپی می توان شبکه بهینه پایش کیفی منابع آب زیرزمینی را تعیین نمود. در واقع اعتبارسنجی کریجینگ و درصد اطلاع به طور همزمان در تعیین شبکه بندی موثر می باشد که این موضوع در مطالعات مشابه منابع آب زیرزمینی دیده نشده است. اگرچه این روش به طور خاص برای داده های نیترات آبخوان دشت مشهد به کار رفته است، برای شرایط مشابه (وجود روند) نیز می توان به کار برد. به عبارت دیگر این راهکار را می توان برای مطالعات دیگر به کار برد ولی باید شرایط داده های موجود را داشته باشد به عنوان مثال اگر داده ها دارای روند نباشد نمی توان از کریجینگ برونوی و عام استفاده نمود و سپس با استفاده از درصد اطلاع تا رسیدن به شبکه بهینه پیش رفت.

لذا با انتخاب روش کریجینگ برونوی، نسبت به پایش کیفی شبکه مبادرت گردید. هرچند روش دقیقی برای ارجحیت کریجینگ عام بر برونوی به اثبات نرسیده ولی در مطالعه های موردی گذشته، محمدزاده و شرفی (۲) نیز بر ارجحیت کریجینگ برونوی بر عام صحه گذاشته اند.

براساس منحنی های هم مقدار نیترات بدست آمده از کریجینگ برونوی، میزان آلودگی آب های زیرزمینی به نیترات در ۴ جبهه اصلی آلودگی در شمال، غرب، مرکز و جنوب آبخوان دشت مشهد در دوره آماری مورد مطالعه از سال ۱۳۸۱ تا ۱۳۸۹ در حال افزایش می باشد. آب انتقالی از سد دوستی به مناطقی از شهر مشهد که میزان آلودگی آب های زیرزمینی به نیترات در حال افزایش است اختصاص یابد و این چاه ها از مدار بهره برداری خارج شوند تا با رقیق شدن آبخوان از شدت آلودگی در این مناطق به تدریج کاسته شود.

نتایج نشان داد مقادیر RMSE، MBE و MAE (مربوط به کریجینگ برونوی) با کاهش درصد اطلاع، افزایش یافته و مقدار R^2 نیز کاهش پیدا کرده است. این امر طبیعی به نظر می رسد زیرا با کاهش درصد اطلاع تعداد چاه های موجود کاهش یافته و خطای برآورد نیز به همین سبب افزایش می یابد. اما مطابق داده های جدول ۲ این مقادیر در شبکه بندی نوع ۱ و ۲ هنوز هم از مقادیر متناظر در برآورد با شبکه ی کامل کریجینگ عام کمتر می باشد. لذا با کاهش قابل توجه تعداد چاه های درمدار بهره برداری، هزینه های مربوط به

منابع

- ۱- ریواز ف، محمدزاده م. و خالدی م. ۱۳۸۵. برآورد فضایی- زمانی بیزی برای میدانهای تصادفی تفکیک پذیر. مجموعه مقالات هشتمین کنفرانس آمار ایران، دانشگاه شیراز، ۴۲-۵۱.
- ۲- محمدزاده م، و شرفی م. ۱۳۸۶. کریجیدن فضایی- زمانی با روند برونوی. مجله پژوهش های آماری، جلد ۷، شماره ۳، ۱۵-۲۸.
- ۳- وصالی ز، زارع م، واعظی ع، حیدری ح. ر. ۱۳۹۱. معرفی نرم افزار SORAM جهت طراحی بهینه شبکه پایش آب زیرزمینی: مطالعه موردی آبخوان جایگاه پالایشگاه شیراز. شانزدهمین همایش انجمن زمین شناسی ایران.
- 4- Awadallah A.G. 2012. Selecting Optimum Locations of Rainfall Stations Using Kriging and Entropy. International Journal of Civil & Environmental Engineering, IJCEE-IJENS, Vol:12, No:01,126301-7575
- 5- Chapman T.G. 1986. Entropy as a measure of hydrologic data uncertainty and model performance. J.Hydrol. 85:111-126.
- 6- Cressie N. 1993. Statistics For Spatial Data. Biometrical Journal, John Willey, New York, 35(2):192.
- 7- De Iaco S., Myers D.E., and Posa D. 2001. Space-time analysis using a general product-sum model. Statist.Probab.Lett. 52:21-28.
- 8- Dhar A., and Datta B. 2007. Multiobjective design of dynamic monitoring for detection of groundwater pollution. Journal of Water Resources Management, 133(4), 329-338.
- 9- Eckenfelder W.W. 1999. Industrial Water Pollution Control., 3rd ed., McGraw-Hill, Inc., New York.
- 10- Guttrop P., Sampson P.D. and Newman K. 1992. Non-parametric estimation of spatial covariance with application to monitoring network evaluation. Statistics in the Environmental and Earth Sciences, Edward Arnold, London.
- 11- Haining R. 1991. Spatial data analysis in the social and environmental sciences. Cambridge University Press.
- 12- Harmancioglu N., and Yevjevich V. 1987. Transfer of hydrologic information among river points. J.Hydrol. 91:103-118.

- 13-Hui-Chung Y., Yen-Chang C., Chiang W., and Ru-Huei C. 2011. Entropy and kriging approach to rainfall network design. *Paddy and Water Environment* September 2011, Vol:9, No:3, 343-355.
- 14-Karamouz M., Khajezadeh Nokhandan A., Kerachian R., and Maksimovic C. 2008. Design of on-line river water quality monitoring systems using the entropy theory: A case study. *Environmental Monitoring and Assessment*, 155(1-4), 63-81.
- 15-Kawachi T., Maruyama T., and Singh V.P. 2001. Rainfall entropy for delineation of water resources zones in Japan. *J.Hydrol.* 246:36-44.
- 16-Metcalf & Eddy, Inc. 2003. *Wastewater Engineering: Treatment and Reuse*. 4th ed., McGraw-Hill, Inc., New York.
- 17-Mogheir Y. 2003. Assessment and redesign of groundwater quality monitoring networks using the entropy theory. Gaza Strip case study. Ph.D. thesis, Coimbra University, Portugal.
- 18-Mogheir Y., and Singh V.P. 2003. Specification of information needs for groundwater management planning in developing country. *Groundwater Hydrology*. Balema Publisher, Tokyo. 2:3-20
- 19-Mogheir Y., Lima J.L.M.P., and Singh V.P. 2003. Spatial structure assessment of groundwater quality variables based on the entropy theory. *Hydrology and Earth System Sciences*, 7(5), 707-721.
- 20-Owlia R.R., Abrishamchi A., and Tajrishy M. 2011. Spatial-temporal assessment and redesign of groundwater quality monitoring network: a case study. *Environ Monit Assess*, 172:263-273.
- 21-Ozkul S., Harmancioglu N.B., and Singh V.P. 2000. Entropy-based assessment of water quality monitoring networks. *Journal of Hydrologic Engineering*, ASCE, 5(1), 90-100.
- 22-Reed P.M., and Minsker B.S. 2004. Striking the balance: Long term monitoring design for conflicting objectives. *Journal of Water Resources Planning and Management*, 130(2), 140-149.
- 23-Rivaz F., Mohammadzadeh M., and Khaledi M.J. 2009. Bayesian prediction for spatio-temporal Gaussian random fields. *Proceedings of the 7th Seminar on Probability and Stochastic Processes*, Isfahan, Iran.
- 24-Rivaz F., Mohammadzadeh M., and Khaledi M.J. 2011. Spatio-temporal modeling and prediction of CO concentrations in Tehran city. *Journal of Applied Statistics*, Published Online DOI, 10.1080/02664763.2010.545108.
- 25-Rouhani S. and Hall T.J. 1989. Space-time kriging of groundwater data. *Armstrong, M. Geostatistics*, Kluwer Academic Publishers, Dordrecht, 2:639-651.
- 26-Sarlak N., and Sorman A.U. 2006. Evaluation and selection of stream flow network stations using entropy methods. *Turkish Journal of Engineering and Environmental Sciences*, 30, 91-100.
- 27-Shanon C.E. 1948. A mathematical theory of communication. *Bell.Syst.Tech.J.* 27:623-656.
- 28-Singh V.P. 1997. The use of entropy in hydrology and water resources. *Hydrol.Process.* 11(6):587-626.
- 29-Yang Y., and Burn D.H. 1994. An entropy approach to data collection network design. *J.Hydrol.* 157:307-324.



A Combined Strategy of Entropy and Spatio-Temporal Kriging in Determining Optimal Network for Groundwater Quality Monitoring of Mashhad Basin

M. Akbarzadeh^{1*} - B. Ghahraman²

Received:09-01-2013

Accepted:11-08-2013

Abstract

In geo-statistics, prediction of an unknown value of random field has been performed in specified time and position, using spatio-temporal Kriging. In some circumstances, a suitable covariate increase the estimation prediction. Geo-statistical methods of Universal Kriging (UK) and Kriging with External Trend (KwET) were applied to Mashhad plain water quality data sets. The optimal network to monitor groundwater quality was presented, using Entropy. All wells ranked based on the criterion of Entropy and mutual information. Then, the optimal network was determined based on the percentages of acquired information and relying on the spatio-temporal Kriging. Based on UK and KwET, electrical conductivity (EC) was the best covariate. KwET with EC as a covariate was the superior Kriging method. A network covering 111 wells showed to be as informative as the existing monitoring network with a total of 237 wells.

Keywords: Information Theory, Geostatistics, Variogram, Validation, Nitrate, Hydrogeochemistry

Archive of SID

1,2- PhD Student and Professor, Department of Water Engineering, International Branch, Ferdowsi University of Mashhad, Iran

(*- Corresponding Author Email: civil.environment@yahoo.com)