

ارزیابی آماره‌های اندازه‌گیری دقت مدل و استفاده صحیح از آنها (مطالعه موردی: درون‌یابی بارش در استان فارس)

علی اکبر میرشفیعی؛ دانشجو گروه جغرافیای طبیعی، دانشکده جغرافیا، دانشگاه تهران، تهران، ایران.
دکتر سعید بازگیر؛ دانشیار گروه جغرافیای طبیعی، دانشکده جغرافیا، دانشگاه تهران، تهران، ایران.
دکتر قاسم عزیزی؛ استاد گروه جغرافیای طبیعی، دانشکده جغرافیا، دانشگاه تهران، تهران، ایران.

دریافت مقاله: ۱۴۰۲/۰۸/۱۳ پذیرش نهایی: ۱۴۰۳/۰۶/۰۳

چکیده

در بسیاری از تحقیقات علمی، بدون توجه به برخی ملاحظات، از آماره‌های سنجش خطا برای انتخاب مدل یا روش در تحلیل فضایی مخاطرات محیطی استفاده می‌شود. این پژوهش به بررسی دقت روش‌های درون‌یابی بارش در استان فارس پرداخته و هدف آن ارزیابی عملکرد آماره‌های پرکاربرد اندازه‌گیری خطا و ارائه توصیه‌هایی برای استفاده صحیح از آن‌ها بوده است. در این مطالعه، بارش سال ۱۳۹۸ که یک سال پر بارش بود، با استفاده از ۱۶۱ ایستگاه هواشناسی (۲۲ ایستگاه همدیدی و ۱۳۹ ایستگاه باران‌سنجی) و روش‌های وزنی عکس فاصله، کریجینگ، کوکریجینگ و تابع پایه شعاعی درون‌یابی شد. نتایج ارزیابی آماره MBE نشان می‌دهد که به دلیل صفر شدن مجموع مقادیر مثبت و منفی، محققان ممکن است در انتخاب روش درون‌یابی دچار اشتباه شوند. این آماره تنها بیش‌برآوردی و کم‌برآوردی را نشان می‌دهد و نمی‌توان از آن برای ارزیابی دقت و انتخاب روش‌های درون‌یابی استفاده کرد. در مورد ضریب تبیین (R^2)، نتایج نشان داد که به دلیل عدم تطابق دامنه تغییرات این ضریب (صفر تا ۱) با مقادیر خطا (۱۰۰ تا ۴۰۰ میلی‌متر برای درون‌یابی بارش استان فارس)، استفاده از آن برای ارزیابی دقت روش مناسب نیست. همچنین، نتایج NRMSE نشان داد که در نمونه‌های با تعداد کم ($n=3$, $NRMSE=0.35$) در مقایسه با نمونه‌های با تعداد زیاد ($n=20$, $NRMSE=0.097$)، به طور قابل توجهی افزایش می‌یابد و استفاده از این آماره توصیه نمی‌شود. در نتیجه گیری کلی، استفاده از آماره‌های MAE و RMSE برای ارزیابی خطای روش‌های درون‌یابی به دلیل واقعی‌تر نشان دادن مقدار خطا پیشنهاد می‌شود.

واژه‌های کلیدی: بارش، آماره‌های ارزیابی خطا، روش‌های درون‌یابی، استان فارس

مقدمه

تخمین دقیق خصوصیات کمی و کیفی پدیده‌های طبیعی، نیازمند زمان و هزینه‌های زیادی است. در این خصوص، مدل‌های آماری روندی کارآزموده و کارآمد می‌باشد که می‌تواند با تخصیص داده‌های نقطه‌ای به یک صفحه پیوسته، به مدیریت منابع کمک کنند و در کاهش هزینه‌ها و بهره‌وری زمانی مؤثر باشند (زند کریمی، ۱۳۹۸). روش‌های درون‌یابی نقش مهمی در پهنه‌بندی پتانسیل خطر رخداد مخاطرات طبیعی و برنامه‌ریزی برای کاهش خسارت‌های ناشی از این مخاطرات دارند (Kalo et al., 2020). این روش‌ها برای پهنه‌بندی داده‌های نقطه‌ای و تخمین مقادیر در مکان‌های فاقد داده استفاده می‌شوند (Yuval et al., 2017). همچنین روش‌های جدیدی برای درون‌یابی در حوزه‌های مختلف توسعه یافته‌اند، از جمله شبکه‌های عصبی و محاسبات فازی (Alimissis et. al., 2018). خوانندگان محترم می‌توانند برای جزئیات بیشتر در خصوص روش‌های درون‌یابی به مقاله ندیری و همکاران، ۱۳۹۳ مراجعه نمایند.

انتخاب دقیق‌ترین روش درون‌یابی مشروط به داشتن کمترین خطا است (حسینی و همکاران، ۱۴۰۲). با بکارگیری صحیح از روش‌های درون‌یابی جهت پهنه‌بندی مخاطرات محیطی مانند سیلاب، می‌توان به توسعه راه‌حل‌های راهبردی به منظور کاهش خطر سیل کمک کرد (خسروی و همکاران، ۱۴۰۱). پایش خشکسالی در هر منطقه یکی از ارکان اساسی مدیریت منابع آب است. در این راستا، درون‌یابی به عنوان یکی از روش‌های مناسب برای پایش مکانی خشکسالی شناخته می‌شود. برای ارزیابی و مقایسه مقادیر تخمینی با مشاهدات مستقل، می‌توان از آماره‌های اندازه‌گیری خطا به طور گسترده‌تری استفاده کرد (Willmott, ۲۰۰۵). از آماره‌های ارزیابی خطا جهت مشخص کردن اینکه کدام روش بهترین برآورد را ارائه می‌دهد، استفاده می‌شوند (Falivene et.al., ۲۰۱۰). برای بدست آوردن مقدار تخمینی مقادیر در تمام نقاط موجود، از روش اعتبارسنجی متقابل (cross-validation) استفاده می‌شود. در این روش، یک نقطه به صورت موقت حذف می‌شود و سپس توسط دیگر نقاط موجود، مقدار آن نقطه تخمین زده می‌شود. این عمل برای تمام نقاط به ترتیب انجام می‌شود تا تخمین برای تمام نقاط محاسبه شوند. این روش به عنوان یک روش کنارگذاری و تخمین توسط نقاط دیگر، می‌تواند مقادیر تخمینی را برای تمام نقاط ارائه دهد (Davies et.al., ۱۹۸۴).

اگر نقاط به صورت دسته‌ای حذف شوند، این فرآیند اعتبارسنجی^۱ نامیده می‌شود و در صورتی که نقاط به صورت یک به یک حذف شوند، به آن اعتبارسنجی متقابل^۲ گفته می‌شود (Esri, 2020). انتخاب الگوی مناسب درون‌یابی نقش بسزایی در دقت نتایج دارد. در این راستا، در مطالعه‌ای که توسط (Nekoamal and Mirabbasi, ۲۰۱۷) انجام شد، دقت دو روش وزن‌دهی عکس فاصله و کریجینگ معمولی برای درون‌یابی داده‌های سطح آب‌های زیرزمینی دشت سرخون در استان هرمزگان مورد بررسی قرار گرفت. نتایج این تحقیق نشان داد که روش کریجینگ دقت بالاتری نسبت به روش وزن‌دهی عکس فاصله دارد. همچنین، (Garcia-Santos, ۲۰۲۰) در مطالعه‌ای نشان داد که نتایج حاصل از روش‌های درون‌یابی می‌تواند از مکانی به مکان دیگر متفاوت باشد و انتخاب یک روش درون‌یابی به عنوان روش برتر، تنها می‌تواند مختص همان محدوده درون‌یابی باشد. برای ارزیابی آماره‌های سنجش خطا، پژوهش‌های متعددی انجام شده است. به عنوان مثال، در

1-Validation

2 - Cross Validation

مطالعه‌ای توسط ویلیموت (۲۰۰۹) میزان معناداری ضریب همبستگی پیرسون و معیارهای میانگین مجذور مربعات خطا $(RMSE)^3$ ، میانگین خطای مطلق $(MAE)^4$ ، میانگین مربعات خطا $(MSE)^5$ ، میانگین خطای اریب $(MBE)^6$ مورد بررسی قرار گرفته است. این اعتبارسنجی بر روی مقادیر تبخیر و تعرق انجام شده است. نتایج این تحقیق نشان داد که ضریب همبستگی پیرسون در اندازه‌گیری دقت مدل، آماره‌ای گمراه کننده است. همچنین، نتایج تحقیق حاکی از آن است ریشه میانگین مربعات خطا، میانگین خطای مطلق و شاخص توافق (d) برای اعتبارسنجی خطا قابل اعتماد هستند. در مطالعه‌ای ارجحیت استفاده از RMSE نسبت به MAE مورد بررسی قرار گرفت (Chai et al., 2004). ایشان در این مطالعه نشان دادند که بر خلاف آنچه که در تحقیقات قبلی (Willmott and Matsuura, 2005) بیان شده است، نبود علامت قدر مطلق در RMSE از مزایای استفاده از این آماره می‌باشد. دیویس و همکاران^۷ (۱۹۸۴) در پژوهشی به این نتیجه رسیدند که انتخاب RMSE و MBE به طور جداگانه می‌تواند منجر به تصمیم اشتباه در انتخاب بهترین مدل از میان مدل‌های کاندید شود. در این خصوص پیشنهاد شد که آماره (d) باید همراه با آماره‌های RMSE و MBE برای ارزیابی عملکرد یک مدل استفاده شود و این آماره را می‌توان در کنار دو آماره دیگر برای قضاوت بهتر ارزیابی یک مدل بکار برد. در پژوهشی دیگر ماتسورا (۲۰۰۵) نشان داد که RMSE معادل MAE نیست و نمی‌توان به راحتی مقدار MAE را از RMSE استخراج نمود. در این مطالعه نشان داده شد که با وجود یکسان بودن چندین مجموعه RMSE، مقدار MAE از مجموعه‌ای به مجموعه دیگر متفاوت است. در تحقیقی، وضعیت خشکسالی استان گلستان با استفاده از نمایه‌های خشکسالی بررسی شد. برای این منظور، بهترین روش درون‌یابی با معیار کمترین خطا (RMSE) انتخاب گردید (عیوضی و مساعدی، ۱۳۹۰). با توجه به افزایش استفاده کاربران از آماره‌های ارزیابی خطا، این پرسش مطرح است که آگاهی از نحوه عملکرد و بکارگیری صحیح از یک آماره اندازه‌گیری خطا چه کمکی برای پژوهشگر می‌تواند بکند. همچنین، انتخاب مناسب‌ترین آماره‌ها می‌تواند به تصمیم‌گیری صحیح در انتخاب بهترین مدل یا روش منجر شود. بنابراین هدف اصلی این پژوهش ارزیابی آماره‌های اندازه‌گیری خطا در انتخاب روش‌های درون‌یابی برای بارش استان فارس است. از دیگر اهداف این مطالعه، استفاده از نتایج بدست آمده به عنوان الگویی برای انتخاب صحیح مدل‌های پایش مخاطرات محیطی می‌باشد.

داده‌ها و روش کار

قلمرو جغرافیایی مورد مطالعه

استان فارس، در جنوب ایران واقع شده و چهارمین استان بزرگ کشور به شمار می‌رود. آب و هوای این استان به سه اقلیم کوهستانی، معتدل و گرم تقسیم می‌شود. میانگین بارش در این استان ۲۸۸ میلی‌متر است و متوسط دمای آن ۱۶٫۶ درجه سلسیوس می‌باشد. از گرم‌ترین شهرهای استان می‌توان به مهر، لامرد، داراب و قیروکازرون اشاره کرد، در حالی که اقلید و

3-Root Mean Square Error

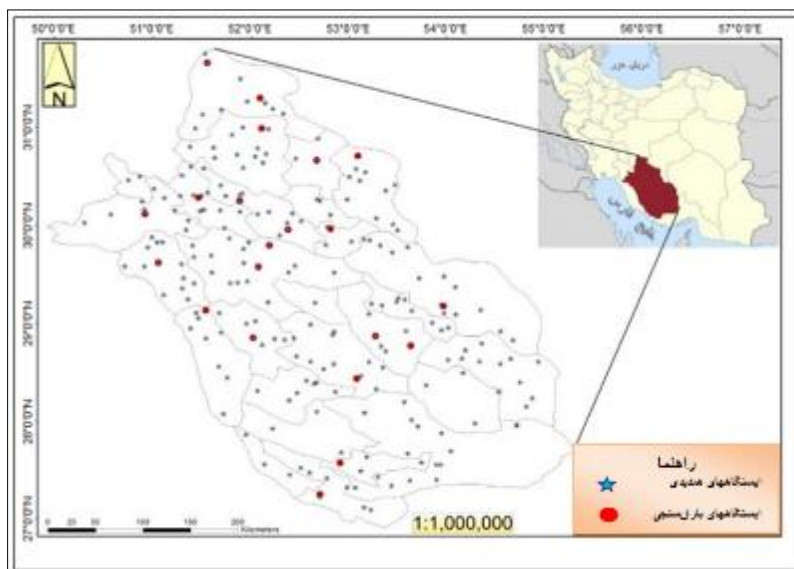
4-Mean Absolute Error

5-Mean Square Error

6-Mean Bias Error

1-Davies et al

سپیدان از سردترین مناطق این استان هستند. همچنین شهرهای مرنجان، ممسنی، فهلیان و بکر از شهرهای پربارش و گردنه کولی‌کش و آباده از کم‌بارش‌ترین شهرها در استان فارس محسوب می‌شوند. در پژوهش‌های مختلف، داده‌های بارش سالیانه از ۱۳۹ ایستگاه باران‌سنجی و ۲۲ ایستگاه همدیدی در سال ۹۸ مورد استفاده قرار گرفته‌اند. انتخاب استان فارس به دلیل توزیع جغرافیایی مناسب و تعداد بالای ایستگاه‌های هواشناسی، همچنین تنوع اقلیمی در این منطقه انجام شده است.



شکل ۱) موقعیت جغرافیایی ایستگاه‌های همدید و باران‌سنجی در استان فارس

برای بهره‌گیری از مدل‌های درون‌یابی در این پژوهش، از نرم‌افزار GIS استفاده شده است. هدف از بکارگیری این مدل‌ها، استخراج مکانی داده‌ها به صورت پهنه‌ای-رستری می‌باشد (شمسی‌پور، ۱۳۹۳). برای انجام صحیح درون‌یابی و استفاده از داده‌های نقطه‌ای و تعمیم آن به سطحی رستری و پیوسته، از سه مرحله ۱- آماده‌سازی داده‌ها شامل شناسایی منابع خطا با استفاده از تجزیه و تحلیل آماری، بررسی توزیع و نرمال‌سازی داده‌ها ۲- اعمال روش‌های درون‌یابی شامل روش‌های جبری و زمین‌آماری ۳- اعتبار سنجی روش‌ها به منظور انتخاب روش مناسب درون‌یابی (Salah, 2009). برای اعمال درون‌یابی، روش‌های مختلفی مانند وزن‌دهی معکوس فاصله^۸ (IDW)، کریجینگ (Kriging)، کوکریجینگ (Co-Kriging)، تابع پایه شعاعی^۹ (RBF) استفاده میشوند. این روش‌ها در تحقیقات پیشین در حوزه‌های متفاوت جغرافیایی به کار رفته‌اند. روش وزن‌دهی معکوس فاصله (IDW)، بر اساس قانون تیلور عمل می‌کند که به عنوان قانون اول جغرافیا نیز شناخته می‌شود. این قانون بیان می‌کند که واحدهای جغرافیایی نزدیک‌تر تأثیر بیشتری بر یکدیگر دارند. این روش بدون توجه به موقعیت و آرایش نقاط، فقط ترکیب خطی از معکوس فاصله‌های موجود را مد نظر قرار می‌دهد (Ding et. al., 2018). در روش تابع پایه شعاعی (RBF) سطحی ایجاد می‌کند که از میان تمامی نمونه‌های اندازه‌گیری شده عبور می‌کند. این روش جزء

8. Inverse Distance Weighting

9. Radial Basis Functions

دسته روش‌های درونیایی دقیق بوده و انواع مختلفی دارد (Esri, 2020). روش تخمین گر عام^{۱۰} (GP)، در مناطق وسیع و جهانی که بررسی‌ها سطحی و گسترده هستند، استفاده می‌شود (Xin, 2009). همچنین روش تخمین گر موضعی^{۱۱} (LP)، فرایندی است که طی آن یک رابطه (غالباً چند جمله‌ای) برای درونیایی تعریف می‌شود. این روش سریع‌تر است اما دقت کمتری دارد و سطحی صاف ایجاد می‌کند. پیشنهاد می‌شود از این روش در مناطق کوچک و با نمونه‌های کم استفاده شود (Fan and Gibels, 1996). یکی از رایج‌ترین روش‌های زمین‌آمار، کریجینگ است (Li et al., 2020). این روش عموماً در تخمین و شباهت توزیع مکانی در متغیرهایی که بین آن‌ها خودهمبستگی مکانی وجود دارد استفاده می‌شود (Belkhir, 2020). از آنجا که روش‌های درونیایی معمولاً با مقادیر متفاوتی از خطا همراه هستند، عدم توجه به میزان خطای مدل‌های درونیایی می‌تواند نتایج را از واقعیت‌ها منحرف کند که این امر قابل قبول نیست (Mitasova and Miats, 1999). بنابراین، برای ارزیابی عملکرد و میزان معناداری آماره‌ها، تعدادی از آنها که شامل میانگین خطای اریب^{۱۲} (MBE)، میانگین مربعات خطا^{۱۳} (MSE)، ریشه میانگین مربعات خطا^{۱۴} (RMSE)، میانگین خطای مطلق^{۱۵} (MAE)، کارایی نش-ساتکلایف^{۱۶} (NSE)، ضریب تبیین^{۱۷} (r^2)، معیار درصد میانگین خطای پیش‌بینی^{۱۸} (MAPE)، ریشه میانگین مربعات خطای نرمال شده^{۱۹} (NRMSE) بررسی و ارزیابی می‌شوند. قابل ذکر است که ضریب تبیین توان دوم ضریب همبستگی پیرسون است (رابطه ۸). روابط هر یک از این آماره‌ها در ادامه آورده شده که در این روابط (e_i) مقدار تخمینی و (o_i) مقدار مشاهده‌ای است.

$$MBE = \frac{\sum_{i=1}^n (e_i - o_i)}{N} \quad \text{رابطه (۱)}$$

$$MSE = \frac{\sum_{i=1}^n (e_i - o_i)^2}{N} \quad \text{رابطه (۲)}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (e_i - o_i)^2}{N}} \quad \text{رابطه (۳)}$$

$$MAE = \frac{\sum_{i=1}^n |e_i - o_i|}{N} \quad \text{رابطه (۴)}$$

10. Global Polynomial

11 Local Polynomial

12-Mean Bias Error

13-Mean Square Error

14-Root Mean Square Error

15-Mean Absolute Error

16-Nash-Sutcliffe efficiency

17-Coefficient of Determination

18-Mean Absolute Percentage Error Loss

19-Normalized Root Mean Square Error

$$MAPE = \frac{\sum_{i=1}^n \frac{(e_i - o_i)}{e_i} \times 100}{N} \quad \text{رابطه (۵)}$$

$$NSE = \frac{\sum_{i=1}^n (e_i - o_i)^2}{\sum_{i=1}^n (e_i - \bar{o})^2} \quad \text{رابطه (۶)}$$

$$NRMSE = \frac{RMSE}{e_i(\max) - e_i(\min)} \quad \text{رابطه (۷)}$$

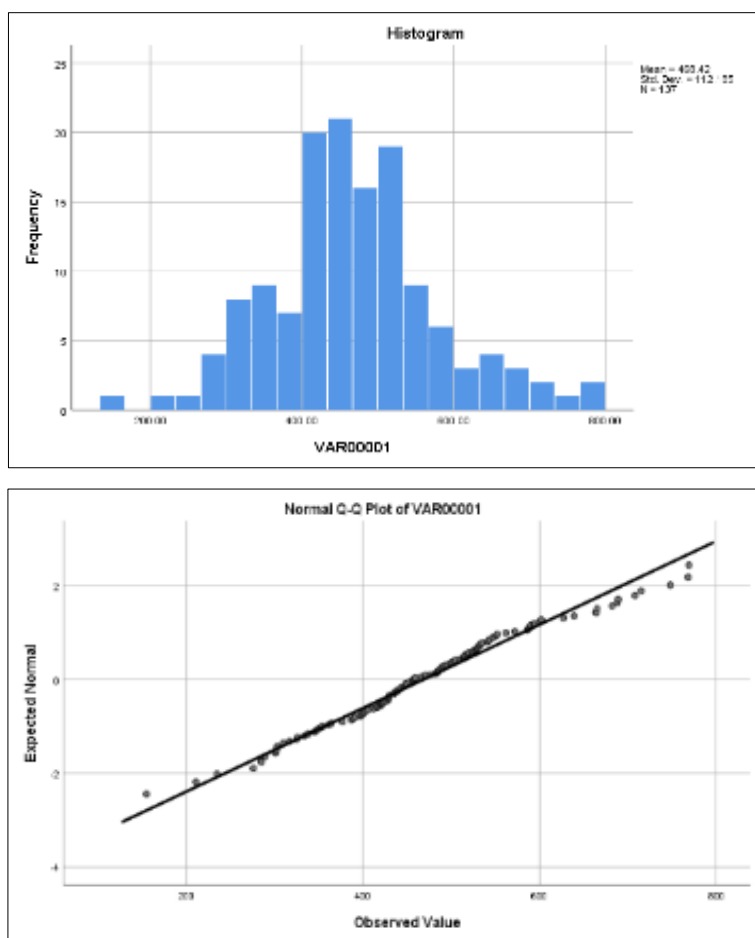
$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad \text{رابطه (۸)}$$

دامنه آماره MBE بین مثبت و منفی بی‌نهایت است که مقادیر مثبت و منفی به ترتیب بیانگر بیش‌برآوردی و کم‌برآوردی هستند. همچنین آماره‌های MSE، RMSE و MAE دارای دامنه‌ای بین صفر تا مثبت بی‌نهایت هستند. هر چه مقادیر به صفر نزدیک‌تر باشند، خطای کمتری در این آماره‌ها وجود دارد. آماره MAPE درصد خطا را نشان می‌دهد و مقادیر نزدیک به صفر دارای خطای کمتری هستند. دامنه آماره NSE بین یک و منهای بی‌نهایت است؛ اگر مقدار آن یک باشد برآورد مدل بالاترین نمره را دارد و بدون خطا است. مقادیر بین صفر و یک بیانگر سطح قابل قبولی از دقت روش و یا مدل مورد نظر هستند و مقادیر کمتر از صفر غیر قابل قبول می‌باشند. در خصوص ضریب تعیین (R^2) دامنه تغییرات آن بین صفر تا مثبت یک است که هر چه مقدار به یک نزدیک‌تر باشد، نشان‌دهنده تخمین دقیق‌تری از روش درونیابی و کمتر بودن خطا است. در نهایت، شاخص توافق (d) دامنه‌ای بین ۰ و ۱ دارد و هر چه مقدار آن به یک نزدیک‌تر باشد، عملکرد مدل بهتر است (Willmott, 1982).

نتایج و بحث

آزمون نرمال بودن داده‌ها

نرمال بودن توزیع داده‌ها با استفاده از آزمون کلموگوروف اسمیرنوف^{۲۰} (K-S) بررسی شد. (حسینی پاک و شرف‌الدین، ۱۳۹۰). نتایج نشان دادند (جدول شماره ۱) که با توجه به مقدار P-value که بزرگتر از ۰/۰۵ است، داده‌ها از توزیع نرمال برخوردار هستند. در شکل (۲) وضعیت نرمال بودن داده‌های بارش به استفاده از هیستوگرام و نمودار چندک چندک نشان داده شده است.



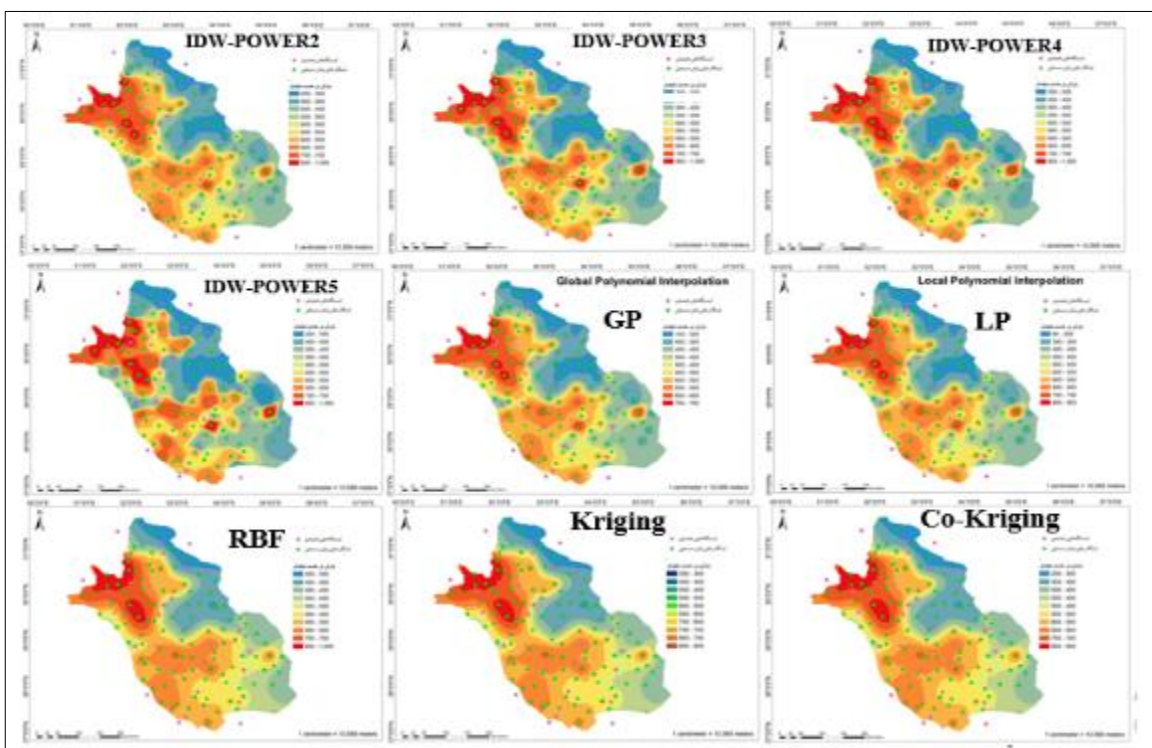
شکل ۲) نمودار بافت نگار (چپ) و نمودار چندک چندک (راست)

جدول ۱) پارامترهای آماری بارش سال ۱۳۹۸ استان فارس به همراه آماره کلموگروف-اسمیرنوف

P-value	کشیدگی	چولگی	انحراف معیار	میانگین	تعداد نمونه	پارامترهای آماری مقادیر
۰/۱۰۷	۰/۷۸۲	۰/۷۴۶	۱۶۱	۴۷۷	۱۳۷	

اعمال روش‌های درون‌یابی

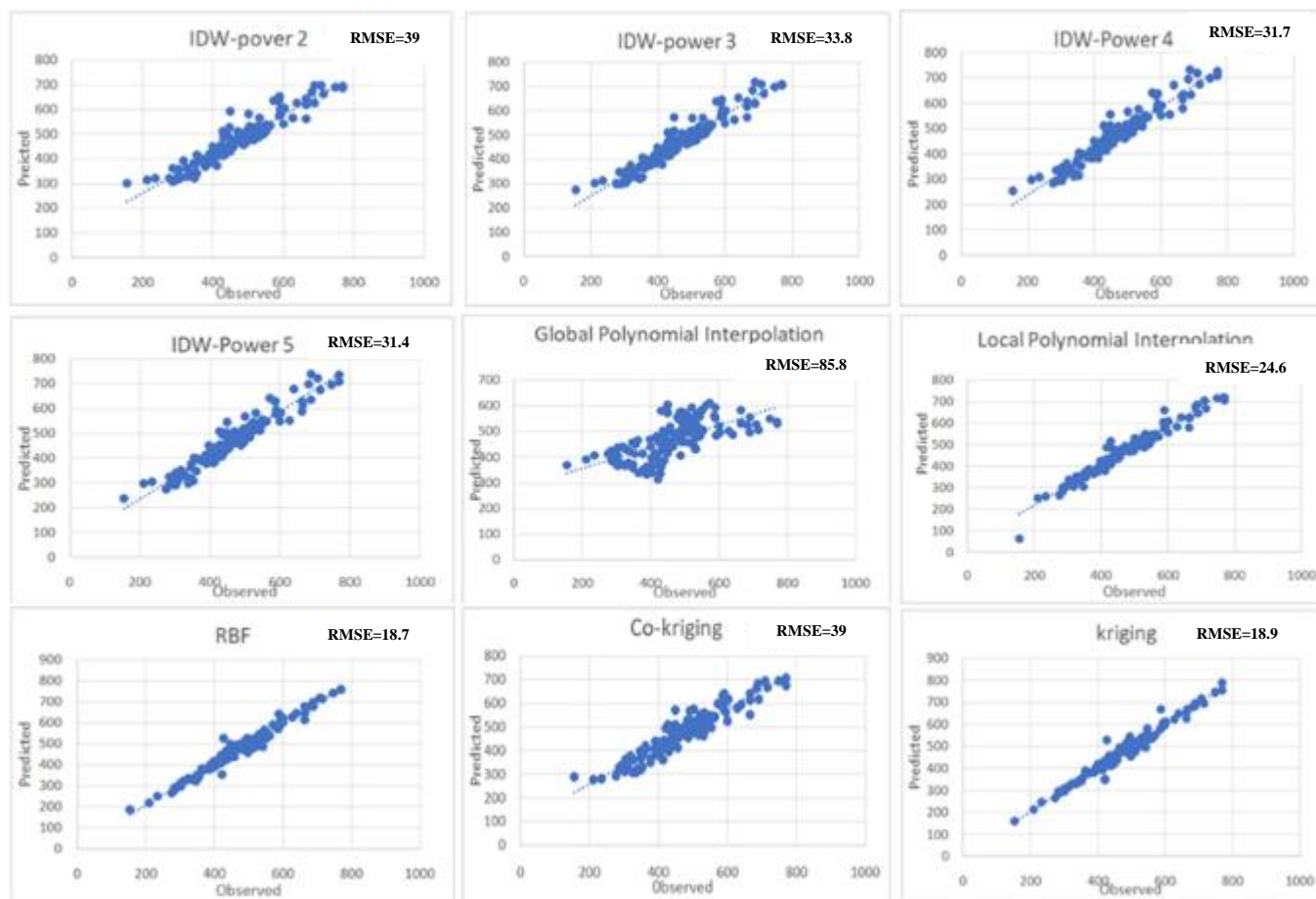
برای اعمال روش‌های درون‌یابی، از روش‌های جبری شامل معکوس وزنی فاصله با توان‌های ۲، ۳، ۴، ۵، توابع پایع شعاعی، چندجمله‌ای‌های محلی و جهانی، و همچنین روش‌های زمین آماری شامل کریجینگ و کوکریجینگ استفاده شده است. در شکل شماره ۳ مقادیر تخمینی و مقادیر مشاهده شده بارش با استفاده از روش اعتبارسنجی متقابل در تمامی مدل‌ها استخراج و مدل‌های درون‌یابی بر روی آنها اعمال گردیده است.



شکل ۳) نقشه پهنه‌بندی بارش سال ۱۳۹۸ استان فارس با استفاده از روش‌های درون‌یابی

ارزیابی دقت روش‌های درون‌یابی

با توجه به نمودارهای شکل شماره (۴) و بررسی اجمالی، می‌توان نتیجه گرفت که روش‌های کریجینگ (میلی‌متر $18/9 = RMSE$)، توابع پایه شعاعی (میلی‌متر $18/7 = RMSE$) و چند جمله‌ای محلی (میلی‌متر $24/6 = RMSE$) مقدار بارش را با دقت بالاتری تخمین زده‌اند. در مقابل، روش‌های چندجمله‌ای جهانی (میلی‌متر $85/8 = RMSE$) و کوکریجینگ (میلی‌متر $39/0 = RMSE$) به ترتیب دقت کمتری داشته‌اند. همچنین، در روش IDW با افزایش توان، دقت درون‌یابی مقادیر بارش افزایش یافته و خطا کاهش می‌یابد (شکل شماره ۴).



شکل (۴). نمودار مقایسه مقادیر مشاهده شده و تخمین‌زده شده بارش سال ۱۳۹۸ به کمک روش‌های مختلف درون‌یابی

سنجش دقت روش‌های درون‌یابی توسط آماره‌های ارزیابی خطا

در مقایسه نتایج بدست آمده از آماره‌های ارزیابی خطا، پس از مشاهده میزان خطا و نمره‌دهی به هر یک از این آماره‌ها، عملکرد و مقایسه‌آماره‌های ارزیابی خطا با یکدیگر مورد بررسی قرار گرفت. نتایج این ارزیابی‌ها در جدول شماره (۲) آورده شده است. در ادامه، ویژگی‌های هر یک از آماره‌ها به صورت جداگانه بررسی شده است.

جدول (۲) مقادیر آماره‌های اندازه‌گیری خطای روش‌های درون‌یابی بارش سال ۱۳۹۸ استان فارس

روش	r^2	d	NSE	SMAPE	MAPE	NRMSE	RMSE	MSE	MBE	MAE
IDW2	۰/۹۴۶	۰/۷۱۷	۰/۸۷۹	۰/۰۶۳	۶/۷	۰/۰۸۴	۳۹	۱۵۲۰	-۶/۴	۲۷
IDW3	۰/۹۵۷	۰/۰۵۲	۰/۹۰۷	۰/۰۵۴	۵/۷۴	۰/۰۴۸	۳۳/۸	۱۱۴۵	-۵/۱	۲۹
IDW4	۰/۹۶۱	۰/۰۴۴	۰/۹۱۸	۰/۰۵۲	۵/۴۸	۰/۰۴۶	۳۱/۷	۱۰۰۸	-۴/۳	۲۳
IDW5	۰/۹۶۷	۰/۴۲۹	۰/۹۲۰	۰/۰۵۳	۵/۵	۰/۰۴۵	۳۱/۴	۹۸۹	-۳/۷	۲۳
RBF	۰/۹۸۶	۰/۱۴۲	۰/۹۷۲	۰/۰۲۸	۲/۸۴	۰/۰۴۲	۱۸/۷	۳۴۸	-۱/۱	۱۲

۶۷	-۰/۱	۷۳۵۷	۵۸/۸	۰/۰۵۸	۱۵/۷	۰/۱۴۸	۰/۴۰۶	۰/۰۴۱۳	۰/۶۳۷	Global
۱۷	۹/۲	۶۰۶	۲۴/۶	۰/۰۳۹	۳/۹۴	۰/۰۴۱	۰/۹۵۱	۰/۰۲۵	۰/۹۷۶	Local
۱۲	۰/۳	۳۵۷	۱۸/۹	۰/۰۴۰	۲/۴۹	۰/۰۲۵	۰/۹۷۱	۰/۱۴۹	۰/۹۸۶	Kriging
۳۰	-۴/۳	۱۵۲۰	۳۹	۰/۰۴۸	۷/۱۲	۰/۰۶۸	۰/۸۷۸	۰/۰۷۰۹	۰/۹۴۲	Co-kriging

آماره MBE

بررسی نتایج حاصل از آماره MBE نشان می‌دهد که به دلیل ماهیت ریاضی این آماره، در مواردی که حاصل جمع مقادیر مثبت (بیش‌برآوردی) و منفی (کم‌برآوردی) صفر می‌شود، محقق ممکن است در انتخاب روش درون‌یابی دچار اشتباه شود و روش با خطای بیشتر را انتخاب کند. به عنوان مثال، در روش کریجینگ، با وجود اینکه مقدار نهایی MBE برابر با ۰/۳ و بسیار نزدیک به صفر است، اختلاف مقادیر واقعی از تخمینی بارش در برخی موارد به ۲۲ و حتی -۳۲ می‌رسد (جدول شماره ۳). بنابراین نمی‌توان از این آماره برای قضاوت در دقت و انتخاب روش‌های درون‌یابی استفاده کرد.

(جدول ۳) مقادیر مشاهده شده و تخمینی بارش (بر حسب میلی‌متر) در روش کریجینگ (MBE=۰/۳)

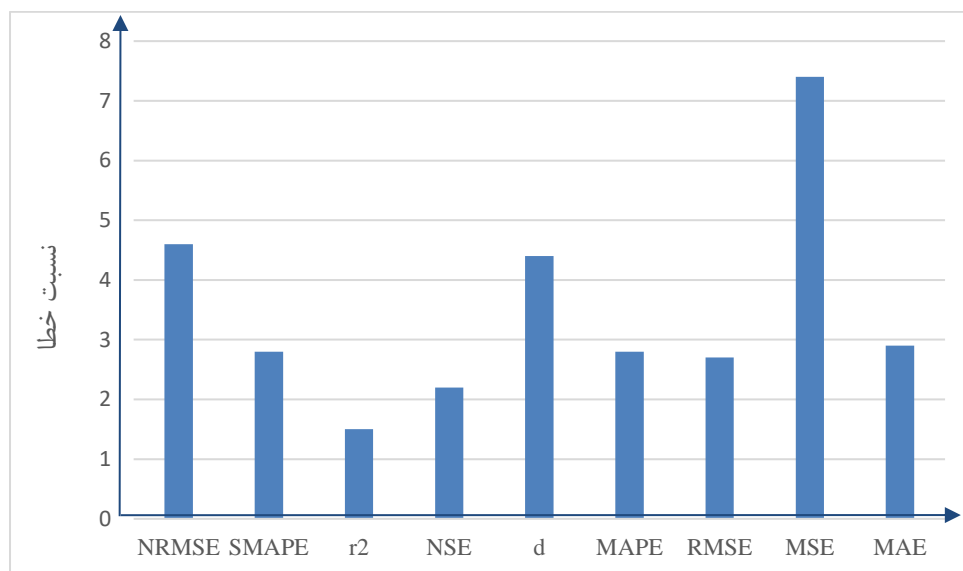
۳۵۱	۳۷۷	۲۸۹	۴۰۱	۵۳۳	۵۵۱	۵۷۲	۷۶۳	۶۶۴	۴۶۶	۴۵۷	۴۸۹	۴۹۳	۳۸۹	مشاهده شده
۳۶۹	۳۸۴	۲۹۰	۳۹۲	۵۲۵	۵۵۶	۵۵۰	۷۵۳	۶۲۸	۴۹۴	۴۵۶	۴۷۱	۵۰۶	۳۸۲	تخمین زده شده

آماره MSE

در بررسی آماره MSE (جدول شماره ۴)، خطای دو روش چندجمله‌ای جهانی و معکوس وزنی فاصله با یکدیگر مقایسه شده است. مشاهده می‌شود که مقدار آماره MSE در روش معکوس وزنی فاصله با توان پنجم، ۹۸۲ می‌باشد و در روش چند جمله‌ای جهانی ۷۳۵۷ است و نسبت خطای این دو روش ۷/۴ می‌باشد (جدول شماره ۴). این در حالی است که نسبت این خطا در دیگر آماره‌ها بسیار پایین‌تر می‌باشد (بین ۱/۵ تا ۴/۶). با توجه به ماهیت رابطه MSE در مواردی که داده‌های پرت وجود دارد، به دلیل مربع شدن خطا، نتیجه نهایی بزرگتر می‌شود (Willmott and Matsuura, 2005). بنابراین در انتخاب روش درون‌یابی به کمک MSE باید دقت بیشتری کرد.

(جدول شماره ۴) مقایسه نسبت خطای روش معکوس وزنی فاصله با توان پنجم و روش چندجمله‌ای جهانی در MSE با دیگر آماره‌ها

روش	NRMSE	SMAPE	r ²	NSE	D	MAPE	RMSE	MSE	MAE
IDW5	۰/۰۶۲	۰/۰۵۳	۰/۹۶۱	۰/۹۲۰	۰/۰۹۲	۵/۵	۳۱/۴	۹۸۲	۲۳
Global	۰/۲۸۷	۰/۱۴۸	۰/۶۳۸	۰/۴۰۶	۱/۱۳	۱۵/۷	۸۵/۸	۷۳۵۳	۶۷
نسبت خطا	۴/۶	۲/۸	۱/۵	۲/۲	۴/۴	۲/۸	۲/۷	۷/۴	۲/۹



شکل ۵) نسبت مقادیر آماره‌های ارزیابی خطا در دو روش معکوس وزنی فاصله و چند جمله‌ای جهانی

آماره‌های (MSE) و (RMSE)

آماره RMSE از جذر MSE به دست می‌آید (رابطه ۳). بنابراین پس از جذر گرفتن از مقادیر MSE عدد کوچکتری به دست می‌آید که همان RMSE است (جدول ۵). تفاوت دیگر این دو آماره در این است که در محاسبه RMSE به دلیل مجذور شدن MSE، نتیجه به مقیاس اصلی متغیر بازگردانده می‌شود و این امر امکان مقایسه منطقی‌تری در تحلیل نتایج فراهم می‌کند.

جدول ۵) مقایسه مقادیر MSE و RMSE در روش‌های مختلف درون‌یابی

MSE	RMSE	روش
۱۵۲۰	۳۹	IDW2
۱۱۴۵	۳۳/۸	IDW3
۱۰۰۸	۳۱/۷	IDW4
۹۸۹	۳۱/۴	IDW5
۳۴۸	۱۸/۷	RBF
۷۳۵۷	۸۵/۸	global
۶۰۶	۲۴/۶	local
۳۵۷	۱۸/۹	kriging
۱۵۲۰	۳۹/۰	Co-kriging

آماره MAE

آماره MAE یکی دیگر از معیارهای پرکاربرد در ارزیابی دقت مدل‌ها است. دامنه مقادیر این آماره، که میانگین مجموع قدر مطلق تفاوت مقادیر مشاهده‌شده و تخمینی است، از صفر تا مثبت بی‌نهایت متغیر است. هر چه مقدار به صفر نزدیک‌تر باشد، نشان‌دهنده خطای کمتر روش درونیابی است. با توجه به جدول شماره (۶) در مقایسه با آماره RMSE، آماره MAE دارای مقادیر خطای کوچک‌تر و حداکثر برابر با آن است. یکی از ویژگی‌های هر دو آماره این است که خطا را با مقیاس متغیر بیان می‌کنند، که این امر از مزایای آن‌ها در تحلیل نتایج است.

جدول (۶) مقایسه مقادیر خطا به کمک آماره‌های MAE و RMSE

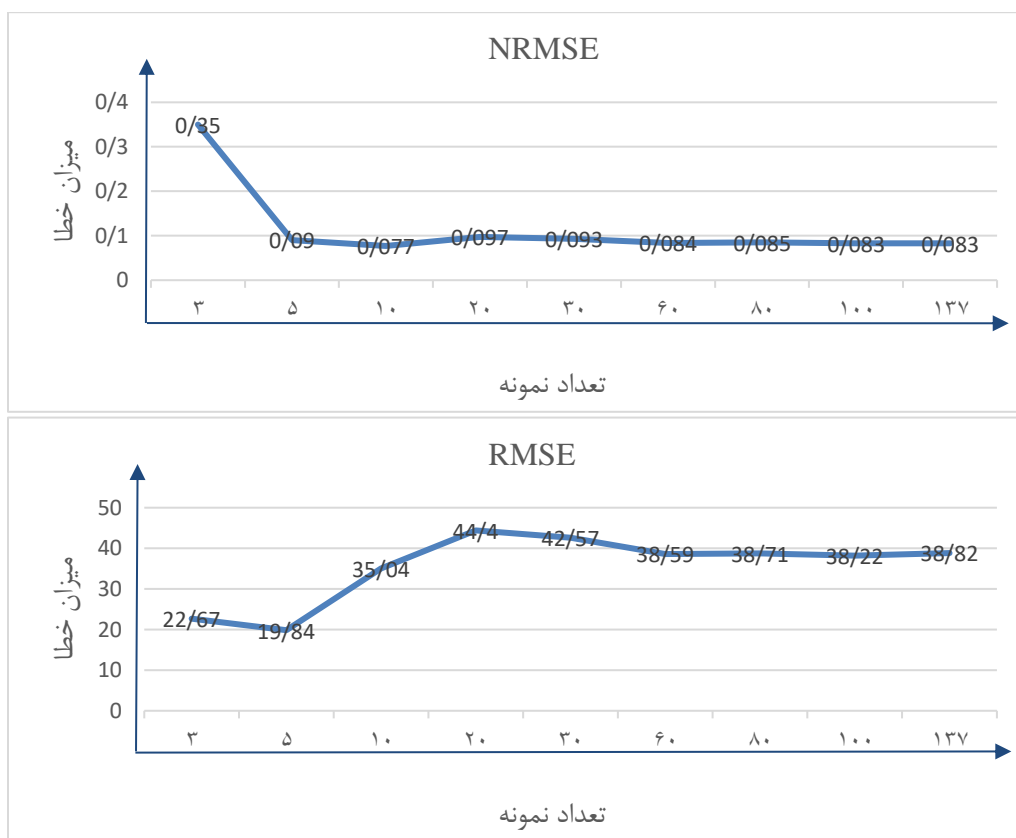
روش	MAE	RMSE
IDW2	۲۷	۳۹
IDW3	۲۹	۳۳/۸
IDW4	۲۳	۳۳/۷
IDW5	۲۳	۳۳/۴
RBF	۱۲	۱۸/۷
Global	۶۷	۸۵/۷
Local	۱۷	۲۴/۶
Kriging	۱۲	۱۸/۹
Co-Kriging	۳۰	۳۹/۰

آماره NRMSE

NRMSE نسخه نرمال شده آماره RMSE است که دامنه تغییرات آن بین صفر و یک قرار دارد. این آماره با هدف سنجش خطای مدل به کار می‌رود و به دلیل بی‌بعد کردن خطا، امکان مقایسه دو متغیر با مقیاس‌های متفاوت را فراهم می‌سازد. در جدول شماره (۷) و شکل شماره (۶) ارزیابی خطا توسط آماره‌های RMSE و NRMSE در تعداد نمونه‌های مختلف محاسبه شده است. همانطور که مشاهده می‌شود (جدول ۷ و شکل ۶)، میزان خطا در روش‌هایی که تعداد نمونه‌های کمتری دارند، به شکل فزاینده‌ای افزایش می‌یابد و استفاده از آماره NRMSE برای نمونه‌های با تعداد کم (کمتر از ۵ عضو) توصیه نمی‌شود. نتایج این مطالعه نشان داد (جدول شماره ۲) که آماره NRMSE، روش IDW با توان دوم را به عنوان روش با دقت پایین‌تر معرفی کرده است، در حالی که سایر آماره‌ها، روش چندجمله‌ای جهانی را به عنوان ضعیف‌ترین روش انتخاب کرده‌اند. بنابراین استفاده از این آماره نتایج قابل اعتمادی به همراه نخواهد داشت.

جدول (۷) ارزیابی خطا به کمک **NRMSE** و **RMSE** در نمونه‌های با تعداد افراد متفاوت

تعداد افراد نمونه	RMSE	NRMSE
۳	۲۲/۶۷	۰/۳۵
۵	۱۹/۸۴	۰/۰۹۰
۱۰	۳۵/۰۴	۰/۰۷۷
۲۰	۴۴/۴۰	۰/۰۹۷
۴۰	۴۲/۵۷	۰/۰۹۳
۶۰	۳۸/۵۹	۰/۰۸۴
۸۰	۳۸/۷۱	۰/۰۸۵
۱۰۰	۳۸/۲۲	۰/۰۸۳
۱۳۷	۳۸/۸۲	۰/۰۸۳




شکل (۶). ارزیابی خطا به کمک **NRMSE** و **RMSE** در نمونه‌های با تعداد افراد متفاوت

آماره MAPE


آماره MAPE که درصد میانگین خطای MAE است، به دلیل بیان نتایج به صورت درصد، می‌تواند خطای دو متغیر با واحدهای مختلف را مقایسه کند. نتایج این مطالعه نشان داد (جدول شماره ۸) که وقتی مقادیر مشاهده شده به صفر نزدیک می‌شوند مقدار خطا افزایش می‌یابد و ممکن است MAPE به طور غیر معمول زیاد شود ($MAPE = 70$) در مقایسه با $0/33$ و $RMSE = 0/18$ و $MAE = 0/18$.

جدول ۸) جدول ارزیابی خطا برای مقادیر مشاهده شده‌ی نزدیک به صفر

مشاهده شده	تخمینی			
0/040	0/0200		MAE	0/1844
0/008	0/0100		MAPE	70/013
0/090	0/0200		RMSE	0/332
0/310	0/0200			
0/120	0/0200			
0/090	0/0010			
0/040	0/0020			
0/060	0/0040			

یکی دیگر نقاط ضعف آماره MAPE این است که اگر مقادیر تخمینی کمتر از مشاهده شده باشد (کم‌برآوردی)، خطای بیشتری را نسبت به زمانی که مقادیر تخمینی بیشتر از مقادیر مشاهده شده باشد (بیش‌برآوردی) نشان می‌دهد. در حالی که آماره MAE در هر دو حالت یک مقدار خطا را نشان می‌دهد (جدول شماره ۹). همچنین، زمانی که مقادیر تخمینی کمتر از مقادیر مشاهده شده است، میزان خطا $6/5$ بوده و در حالت عکس، مقدار خطا کمتر است ($6/0$). بنابراین، در تخمین‌های کمتر، آماره MAPE خطای بیشتری را نشان می‌دهد.

جدول ۹) جدول ارزیابی خطا توسط مقادیر بیش برآوردی و بالعکس

مشاهده شده	تخمینی		مشاهده شده	تخمینی
121	112		112	121
132	139		132	139
124	129		124	129
130	141		130	141
MAPE			MAPE	
6/038			6/458	
MAE			MAE	
8			8	

ضریب تعیین (r^2)

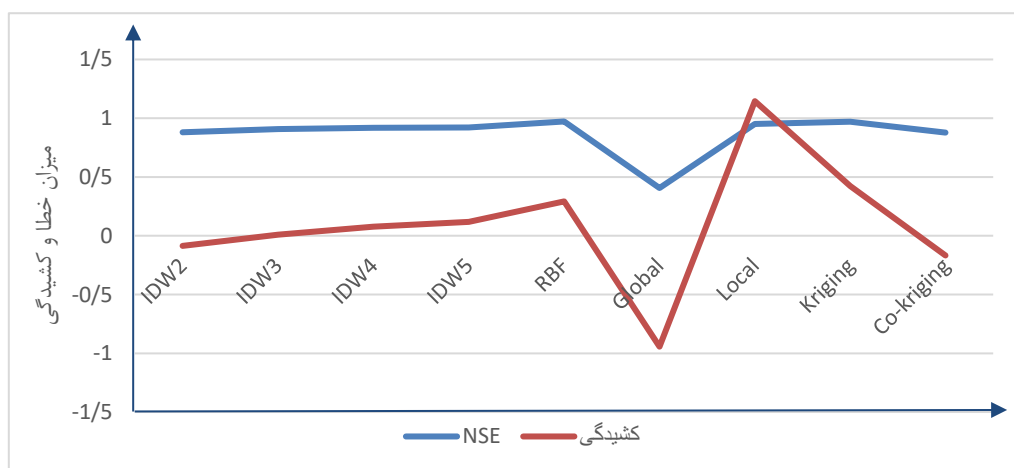
نتایج این مطالعه نشان می‌دهد که مقدار r^2 در دو روش RBF و IDW با توان ۵ مشابه است. با این حال آماره‌های MAE و RMSE در روش IDW با توان ۵ مقدار خطای دو برابری نسبت به روش RBF را نشان می‌دهند (جدول شماره ۱۰). از طرف دیگر بررسی نتایج ارزیابی روش‌های درونیابی نشان داد که استفاده از ضریب تبیین (r^2)، به دلیل عدم تطابق دامنه تغییرات آن (صفر تا ۱) با مقادیر خطای درونیابی بارش استان فارس (۱۰۰ تا ۴۰۰ میلی‌متر)، در ارزیابی دقت روش‌های درونیابی منطقی نیست.

جدول ۱۰ مقایسه نتایج آماره r^2 با دو آماره RMSE و MAE

روش	IDW2	IDW3	IDW4	IDW5	RBF	Global	Local	Kriging	Co-kriging
r^2	۰/۹۴۶	۰/۹۶۱	۰/۹۶۱	۰/۹۶۷	۰/۹۸۶	۰/۶۳۸	۰/۹۷۶	۰/۹۸۶	۰/۹۴۲
MAE	۲۷	۲۹	۲۳	۲۳	۱۲	۶۷	۱۷	۱۲	۳۰
RMSE	۳۹	۳۳/۸	۳۱/۷	۳۱/۴	۱۸/۷	۸۵/۵	۲۴/۶	۱۸/۹	۳۹

شاخص ضریب کارایی نش ساتکلیف (NSE)

با توجه به شکل شماره (۷) و مشاهده نمودار میزان کشیدگی و ارتباط مستقیم آن با خطای NSE می‌توان گفت این آماره نسبت به مقادیر پرت حساسیت بالایی دارد و می‌توان با لگاریتم‌گیری و نرمال کردن داده‌ها این مشکل را رفع کرد. آماره NSE، در واقع نسخه تعدیل یافته و توسعه یافته MSE می‌باشد (زند کریمی و مختاری، ۱۳۹۷). نتایج این تحقیق نشان داد که برخلاف سایر آماره‌ها که برخی روش‌ها را با دقت بالا و برخی را با دقت پایین‌تر ارزیابی کردند، آماره NSE بیشتر روش‌ها را با دقت بالا ارزیابی کرده و همه روش‌های درونیابی در این مطالعه را تایید کرده است ($NSE > 0/90$).



شکل (۷). شاخص ضریب کارایی

فکته وهمکاران^{۲۱} (۲۰۰۴) و ژاکوویدس^{۲۲} (۱۹۹۴) در مطالعه‌ای نشان دادند که آماره MBE اطلاعات مفیدی در خصوص عملکرد یک مدل ارائه می‌دهد. نتایج این مطالعه نشان می‌دهد که مقادیر مثبت و منفی این آماره به ترتیب بیانگر تخمین بیش‌برآوردی و کم‌برآوردی روش درون‌یابی هستند. همچنین مقادیر نزدیک به صفر نشان دهنده عملکرد مناسب روش درون‌یابی است. با مقایسه مطالعه آن‌ها با مطالعه مشخص می‌شود که آماره MBE نمی‌تواند معیار مناسبی برای بررسی دقت یک مدل آماری در تخمین داده‌ها باشد. در تحقیق حاضر نشان داده شد که با توجه به ماهیت رابطه ریاضی این آماره (رابطه ۱) و صفر شدن حاصل جمع مقادیر مثبت و منفی، مقدار MBE ممکن است نزدیک به صفر یا حتی صفر شود، اما این نتیجه (MBE=0) می‌تواند محقق را در انتخاب روش درون‌یابی گمراه کند. بنابراین، این آماره را نمی‌توان به عنوان معیار ارزیابی خطای یک روش درون‌یابی استفاده کرد. در خصوص آماره ریشه میانگین مربعات خطا (RMSE)، این آماره سال‌هاست که به عنوان یکی از معیارهای قابل قبول برای ارزیابی خطای مدل توسط بسیاری از محققان استفاده می‌شود (مک‌کین و همکاران^{۲۳} (۲۰۰۵) و چای و همکاران (۲۰۱۳)). از طرف دیگر نتایج مطالعات ویلموت و ماتسورا (۲۰۰۵) نشان می‌دهد که این آماره نسبت به خطاهای بالا حساسیت بیشتری دارد و خطا را بیشتر از مقدار واقعی آن نشان می‌دهد که هم‌راستا با مطالعه اخیر (بارش استان فارس) می‌باشد.

تحلیل مقایسه‌ای نتایج RMSE و MSE برای درون‌یابی بارش استان فارس نشان داد که به دلیل ماهیت رابطه RMSE و تعدیل خطا، استفاده از این آماره نسبت به MSE ارجحیت دارد. نتایج این مطالعه در رابطه با آماره MAE و مقایسه آن با آماره MBE نشان می‌دهد که مقادیر MAE واقعی‌تر از MBE است. همان‌طور که قبلاً اشاره شد، مقادیر کوچک و حتی نزدیک به صفر میانگین خطای اریب (MBE) می‌تواند محقق را در انتخاب روش مناسب درون‌یابی به خطا و تصمیم‌گیری اشتباه بیندازد. این در حالی است که وجود علامت قدر مطلق در رابطه میانگین خطای مطلق (MAE) این اجازه را نمی‌دهد که حاصل به سمت صفر میل کند و از طرفی مقایسه روش‌های مختلف را منطقی‌تر می‌سازد. با توجه به نتایج تحقیق حاضر، در دو روش با MAE مشابه و نزدیک به یکدیگر، RMSE می‌تواند از روشی به روش دیگر متفاوت باشد. بر اساس نتایج بدست آمده (جدول شماره ۶)، مقادیر MAE در روش‌های درون‌یابی IDW با توان‌های ۲ و ۳ نزدیک به یکدیگر و به ترتیب ۲۷ و ۲۹ میلی‌متر بوده، در حالی که این مقادیر برای RMSE متفاوت و به ترتیب ۳۹ و ۳۳/۸ میلی‌متر به دست آمده است. فکت و همکاران^{۲۴} (۲۰۰۴) نشان دادند که مقادیر RMSE هرگز نمی‌تواند از MAE کوچک‌تر باشد و تفاوت‌های MAE و RMSE به دلیل تأثیر واریانس‌های خطا است. رایپر^{۲۵} (۲۰۰۷) اگرچه سهولت در تفسیر نتایج MAPE را مطرح کرده، اما بیان داشته که این آماره تحت تأثیر مقادیر پرت قرار گرفته و در اغلب موارد خطای محاسبه شده را بیش از واقعیت نشان می‌دهد. نتایج مطالعه حاضر نشان داد که در زمانی که مدل در تخمین خود از کم‌بروردی برخوردار باشد، میزان خطا ۶/۵ بوده و در حالت عکس در بیش‌برآوردی تخمین مدل، مقدار خطا کمتر و برابر با (۶/۰) می‌باشد. بنابراین، در روش‌های

21- Fekete, et al

22-Jacobides

23 -McCain et al

24- Fact

25 -Rayer

درون‌یابی با تخمین‌های کمتر، آماره MAPE خطای بیشتری را نشان می‌دهد (جدول شماره ۹). در خصوص ضریب تبیین (R^2) و کاربرد آن در ارزیابی روش‌های درون‌یابی، نتایج تحقیقات انجام شده توسط (مک لارن لیمیت و همکاران (۱۹۸۰)، ویلموت (۱۹۸۱) و ریوز (۲۰۲۱)) نشان می‌دهد که مقادیر بالای R^2 به دلیل عدم توجه به بزرگی تفاوت مقادیر مشاهده‌شده و تخمینی، نمی‌تواند معیار مناسبی برای ارزیابی این روش‌ها باشد. در مطالعه اخیر، نتایج نشان داد که ارزیابی ۹ روش درون‌یابی با استفاده از آماره R^2 مقادیر بالایی را نشان داده و اغلب روش‌ها را تأیید می‌کند. با این حال، ارزیابی این روش‌ها با استفاده از سایر آماره‌ها تفاوت‌های قابل توجهی را نشان داده و برخی روش‌ها نسبت به دیگران برتری داشته‌اند. بنابراین نمی‌توان به مقادیر بالای R^2 در ارزیابی روش‌های درون‌یابی اعتماد کرد و استفاده از این آماره توصیه نمی‌شود. ویلموت (۲۰۰۵) استفاده از شاخص توافق (d) را که یک معیار توصیفی و نسبی است، پیشنهاد کرده است. بر اساس نتایج تحقیق اخیر (جدول شماره ۲)، تمامی آماره‌های ارزیابی خطا به جز از ضریب توافق، دقت روش درون‌یابی چند جمله‌ای محلی (local) را تأکید کرده‌اند. از سوی دیگر ضریب توافق برای روش IDW با توان ۲ بالاترین نمره را ($d=0/717$) به دست آورده است، در حالی که ارزیابی این روش توسط سایر آماره‌ها از متوسط به پایین بوده است (جدول شماره ۲). بنابراین، بر خلاف نظر ویلموت، این آماره نمی‌تواند برای ارزیابی دقت روش‌های درون‌یابی پیشنهاد شود.

نتایج تحقیق ملسه و ننگیا^{۲۶} (۲۰۰۵) نشان داد که برای مقایسه دو مدل با مقیاس‌های متفاوت، استفاده از جذر میانگین مربع خطای نرمال شده (NRMSE) می‌تواند آماره مناسبی برای انتخاب روش درون‌یابی با دقت بالاتر باشد. با این حال، نتایج تحقیق حاضر نشان می‌دهد که در مقایسه خطای روش‌های درون‌یابی با استفاده از NRMSE باید احتیاط بیشتری کرد. به دلیل وجود دامنه تغییرات خطاها در مخرج کسر NRMSE اگر تعداد افراد نمونه زیاد باشد (جدول ۷)، مقدار خطا کاهش می‌یابد که می‌تواند گمراه‌کننده باشد ($NRMSE=0/085$ برای $n=80$). در مقابل، در نمونه‌های با تعداد کم، مقدار خطا افزایش می‌یابد که در این حالت نیز نتیجه‌گیری نهایی برای انتخاب روش درون‌یابی گمراه‌کننده است ($NRMSE=0/35$ برای $n=3$). در خصوص کارایی نش و ساتکلیف (NSE) نتایج این تحقیق نشان می‌دهد که این آماره به مقادیر پرت حساسیت بالایی دارد و مقادیر خطا را بیش از مقدار واقعی نشان می‌دهد. بنابراین ضمن دقت بیشتر در استفاده از این آماره، باید کارایی آن در مقایسه با دیگر آماره‌ها بررسی شود.

نتیجه‌گیری

ارزیابی دقت، یکی از بخش‌های حیاتی در فرآیند ارزیابی مدل‌های آماری به منظور توصیف عملکرد آنها است. مفهوم اصلی ارزیابی دقت یک مدل، بر پایه آماره‌های خاصی استوار است و میزان خطاها نشان دهنده دقت مدل در پیش‌بینی یا تخمین یک متغیر می‌باشد. روش‌های متعددی برای تعیین میزان خطای مدل‌های آماری پیشنهاد شده‌اند، اما هیچ یک از این روش‌ها بدون محدودیت نیستند و بسیاری از آنها در برخی ویژگی‌ها مبهم می‌باشند. هدف از مطالعه اخیر، دیکته کردن یک آماره به عنوان یک معیار شاخص و برتر برای پژوهش‌های آینده نبوده است. بلکه هدف، ارائه رهنمودها و ارزیابی صحیح از آماره‌های اندازه‌گیری خطا و تفسیر عملکرد آنها بوده تا بتوان گامی مفید در راستای پذیرش یا رد یک آماره بر اساس محاسن و

محدودیت‌های آن برداشت. نتایج این تحقیق نشان می‌دهد که آماره MBE در انتخاب یک روش درون‌یابی کارایی لازم را ندارد. در برخی موارد، جمع مقادیر مثبت و منفی خطا در محاسبات این آماره به گونه‌ای است که نتایج نهایی صفر می‌شود و ما را در انتخاب مدل مناسب گمراه می‌کند. همچنین نتایج آماره در خصوص آماره MSE نشان داد که با افزایش تفاوت مقادیر تخمینی روش درون‌یابی نسبت به مقادیر مشاهده شده، مقادیر این آماره نیز به دلیل ماهیت رابطه و وجود توان دوم در آن، بزرگتر می‌شود و نمی‌تواند به عنوان یک آماره مطمئن مورد استفاده قرار گیرد.

با جذر گرفتن از آماره MSE که حاصل آن RMSE خواهد بود ($RMSE = \sqrt{MSE}$)، مقدار خطا به مقدار واقعی نزدیک‌تر، می‌شود. همچنین هنگام استفاده از آماره NSE که نمونه تعدیل‌یافته و توسعه‌یافته MSE است، باید به نرمال بودن داده‌ها توجه شود تا از میزان حساسیت به مقادیر شدید در این آماره کاسته شود. یکی دیگر از نتایج این مطالعه، حساسیت NRMSE در نمونه‌هایی با تعداد کم است که این موضوع باعث می‌شود مقدار خطا به شدت کمتر از مقدار واقعی نشان داده شود. در خصوص ضریب تبیین (R^2) باید اذعان داشت که به دلیل در نظر نگرفتن مقیاس داده‌ها و دامنه تغییرات آن (از صفر تا ۱)، نمی‌توان درک صحیحی از بزرگی خطا داشت. برای نمونه، نمی‌توان درک صحیح و منطقی در ارزیابی خطای یک روش درون‌یابی برای داده‌های دو رقمی در مناطق کم‌بارش کشور (مانند یزد و بم با حدود ۶۰ میلی‌متر میانگین بارش سالانه) تا چهار رقمی در مناطق پر بارش کشور (مانند انزلی و دیناچال با بارش حدود ۲۰۰۰ میلی‌متر در سال) به کمک ضریب تبیین داشت. نتایج این مطالعه نشان داد که آماره MAE به دلیل استفاده از قدر مطلق و مثبت نمودن تفاوت مقادیر مشاهده شده با برآورد شده، می‌تواند به عنوان معیار مناسبی برای مقایسه و انتخاب مدل دقیق‌تر مورد استفاده قرار گیرد. همچنین، نتایج این تحقیق به عدم توافق شاخص توافق (d) با سایر آماره‌ها اشاره دارد که برای ارزیابی مدل‌های درون‌یابی پیشنهاد نمی‌شود. در نهایت آماره NSE به دلیل حساسیت به مقادیر پرت و نشان دادن مقادیر خطا بیش از مقدار واقعی، پیشنهاد نمی‌شود. در یک جمع‌بندی از نتایج این تحقیق، می‌توان اذعان داشت که استفاده از آماره‌های MAE و RMSE برای ارزیابی خطای روش‌های درون‌یابی به دلیل واقعی‌تر نشان دادن مقدار خطا، پیشنهاد می‌شوند.

منابع

- اصغری مقدم، اصغر؛ وحید نورانی و عطاالله ندیری. ۱۳۸۸. پیش بینی زمانی و مکانی سطح آب‌های زیرزمینی در محدوده متروی شهر تبریز با استفاده از مدل کریجینگ عصبی، تحقیقات منابع آب ایران، ۱۳: ۱۴-۲۴.
- حسنی پاک، علی اصغر؛ محمد شرف الدین. ۱۳۹۰. تحلیل داده‌های اکتشافی، مؤسسه انتشارات دانشگاه تهران.
- حسینی، سیده فاطمه؛ همتی، محمد؛ جعفری، مهتاب؛ استعلاجی، علیرضا. ۱۴۰۲. تحلیل و پهنه‌بندی خطر سیل خیزی و ارتباط آن با پوشش گیاهی در شهرستان قیروکارزین، تحلیل فضایی مخاطرات محیطی، ۱۰(۲): ۷۷-۹۶.
- خسروی، علیرضا؛ اژدری مقدم، مهدی؛ هاشمی فرد، سید آرمان؛ نظری پور، حمید. ۱۴۰۱. مقایسه نتایج تصمیم‌گیری چند معیاره در پهنه‌بندی مناطق مستعد خطر سیلاب با شاخصهای سنجش از دور در حوضه آبریز رودخانه کهیر (بلوچستان جنوبی)، تحلیل فضایی مخاطرات محیطی، ۹(۴): ۲۱-۴۰.
- زندکریمی، آرش؛ داود مختاری. ۱۳۹۷. ارزیابی دقت روش‌های مختلف درون‌یابی در تخمین مقادیر بارش جهت انتخاب بهینه‌ترین الگوریتم در استان کردستان، پژوهش‌های جغرافیایی طبیعی، ۵(۲): ۳۳۸-۳۲۳.
- سازمان هواشناسی کشور. ۱۴۰۰. سالنامه هواشناسی، آمار متغیرهای اقلیمی.
- شمسی‌پور، علی اکبر. ۱۳۹۳. مدل‌سازی آب و هوایی، انتشارات دانشگاه تهران.
- عیوضی، معصومه؛ مشاعدی، ابوالفضل. ۱۳۹۰. پایش و تحلیل مکانی خشکسالی هواشناسی در سطح استان گلستان با استفاده از روش‌های زمین‌آماری، مرتع و آبخیزداری، ۶۴(۱): ۶۵-۷۸.
- مرادی، اسحاق؛ افسانه شهبازی؛ کاظم نصرتی؛ غلامرضا زهتابیان. ۱۳۸۵. ارزیابی روش شبیه‌سازی تصادفی برای تولید داده‌های هواشناسی، پژوهش‌های جغرافیایی، ۶۲: ۹-۱.
- ندیری، عطاالله؛ صدیقه شکور. ۱۳۹۳. ارزیابی انواع روش‌های درون‌یابی، جهت تخمین آلودگی نیترات در منابع آب زیرزمینی، مجله هیدروژئومورفولوژی، ۱: ۹۲-۷۵.

- Alimissis, A.; k. Philippopoulos, C. Gtzanis, and D. Deligiorgi. 2018. Spatial estimation of urban air pollution with the use of artificial neural network models. *Atmospheric Environment*. **191**: 205-213.
- Belkhiri L.; A. Tiri, and L. Mouni. 2020. Spatial distribution of the groundwater quality using kriging and Co-kriging interpolations. *Groundwater for Sustainable Development*. **11**: 100-73.
- Chai, T.; And R. Draxler. 2004. root mean square (RMSE) or mean absolute error (MAE) Arguments against avoiding RMSE in the literature. *Geosci model Dev*. **7**: 1247-1250.
- Davies, j. A.; M. Abdel-Wahab, and D. Makay. 1984. Errors in estimating solar irradiance from a numerical model. *solar Energy*. **32**: 307-309.
- Ding, Q.; Y. Wang, and D. Zhuang. 2018. Comparison of the common spatial interpolation methods used to analyze potentially toxic elements surrounding mining regions. *Journal of Environmental Management*. **212**: 23-31.
- García-Santos, G.; M. Scheiber, and J. Pilz. 2020. Spatial interpolation methods to predict airborne pesticide drift deposits on soils using knapsack sprayers. *Chemosphere*. **258**: 127-231.

- Falivene, O.; R. Cabrera, R. Tolosana-Delgado, and A. Saez. 2010. Interpolation algorithm ranking Using cross-validation and the role of asmoother effect: A coal zone example *Comput. Geosci*, **36**: 512-519.
- Fekete, B.; C. Vörösmarty, J. Roads, and C. Willmott. 2004. Uncertainties in precipitation and their impacts on runoff estimates. *Journal of Clim.* **17**: 294-304
- Fan, J.; and I. Gibels. 1996. Local Polynomial Modelling and Its Applications. *Water Resources Bulletin*. **87**: 998-1004.
- Hyndman, R.; and A. Koehler. 2005. Another look at measures of forecast accuracy. *International Journal of Forecasting*, **22**: 679-688
- Li, J.; H. Wan, and S. Shang. 2020. Comparison of interpolation methods for mapping layered soil particle-size fractions and texture in an arid oasis. *Catena*. **190**: 104-1014.
- Jacovides, C.; G. Papaioannou, and P. Kerkides. 1994. Micro and large-scale parameters evaluation of evaporation from a lake. *Agricultural Water Management*. **13**: 263-27
- Kazemi, S.; and S. Hosseini. 2011. Comparison of spatial interpolation methods for estimating heavy metals in sediments of Caspian Sea. *Expert Systems with Applications*. **38**: 1632-1649.
- Moriasi, D. N.; J. G. Arnold, M. W. Van Liew, R. L. Bingner, R. D. Harmel and T. L. Veith. 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE*, **50**: 885-900.
- Nekoamal, M.; and R. Mirabbasi. 2017. Assessment of interpolation methods in estimation of groundwater level (case study: sarkhon plain). *Journal of hydrogeology*, online publish. **2**: 84-95.
- Nash, J.; and E. Sutcliffe. 1970. River Flow forecasting through conceptual models, part 1-A discussion of principles. *Journal of Hydrology*, **10**: 282-290.
- Salah, H. 2009. Geostatistical analysis of groundwater levels in the south Al Jabal Al Akhdar area using GIS. *GIS Ostrava*, **25**: 1-10.
- Willmott, C. J.; and K. Matsuura. 2005. advantages of the mean Absolute Error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, **30**: 79-82.
- Willmott, C.J. 1982. Some comments on the evaluation of the model performance. *Bulletin of American Meteorological Society*, **63**: 1309-1313.
- Xin, Y.; and G. Xiao 2009. Linear regression analysis: theory and computing .world Scientific Publishin **348**.
- Yuval-Levy, I.; and D. Broday. 2017. Improving modeled air pollution concentration maps by residual interpolation. *Science of The Total Environment*. **598**: 780-788.