مقایسه کارایی RMSE و NRMSE برای محاسبه مقدار مؤثر خطای مدل‌هایی با مجموعه داده‌های آموزش متفاوت

هادی حدادیان سنو [۱]، علیرضا کریمی [۲]، عیسی اسفندیارپور بروجنی [۳]، غلامحسین حق‌نیا [۴]
۱- دانشجوی کارشناسی ارشد گروه علوم خاک دانشگاه فردوسی مشهد، ۲- دانشیار گروه علوم خاک دانشگاه فردوسی مشهد، ۳- دانشیار گروه علوم خاک دانشگاه ولی‌عصر رفسنجان، ۴- استاد گروه علوم خاک دانشگاه فردوسی مشهد

چکیده
هدف از این پژوهش، مقایسه کارایی RMSE و NRMSE برای محاسبه مقدار خطای مؤثر یک مدل زمین‌آماری با دامنه متفاوت مجموعه داده‌های آموزش بود. در این مطالعه، ۱۵۱ نمونه برای آموزش ۱۴۱ مدل با تعداد متفاوت داده‌های آموزش مورد استفاده قرار گرفت. با کاهش تعداد نمونه آموزشی، شاخص RMSE تغییر قابل توجهی نداشت. با توجه به این که با کاهش تعداد نمونه آموزش مدل، دامنه تغییرات مقدار رس موجود در مجموعه داده‌های آموزش کاهش یافت، از شاخص NRMSE به منظور حذف این عامل استفاده شد. شاخص NRMSE نشان داد که با کاهش تعداد نمونه آموزش مدل، پایداری و دقت مدل به صورت لگاریتمی کاهش می‌یابد.

واژه‌های کلیدی: اعتبارسنجی، تغییرات مکانی، مدل‌سازی، خطای مؤثر

مقدمه
در علم آمار، جذر میانگین مربع اعداد، به‌عنوان "مقدار مؤثر"[۱۰۴] شناخته می‌شود. جذر میانگین مربع خطا (RMSE)[۱۰۵]، شاخصی است که مقدار مؤثر خطا را نشان می‌دهد. به عبارت دیگر، این شاخص بیانگر تفاوت میان مقدار پیش‌بینی‌شده توسط مدل و مقدار واقعی یک متغیر است. این تفاوت‌های مجزا را "مانده‌ها"[۱۰۶] می‌نامند و برای جمع‌آوری آن‌ها در قالب یک عدد، کاربرد دارند. از این ابزار برای تعیین مدلی که می‌تواند نزدیک‌ترین برآورد از واقعیت را داشته باشد استفاده می‌شود و هر مدلی که مقدار RMSE کمتری داشته باشد به‌عنوان مدل نزدیک به واقعیت انتخاب می‌شود. با این وجود، RMSE برای مقایسه چند مجموعه داده، مناسب نیست (Hyndman and Koehler, ۲۰۰۶). از معایب دیگر RMSE می‌توان به حساسیت بسیار بالای آن به داده‌های پرت و دخالت داشتن دامنه داده‌ها در آن اشاره کرد. در واقع، مدل‌هایی که برای آموزش آن‌ها از یک مجموعه داده با دامنه کمتر استفاده شود، دامنه مقدار خطای آن‌ها نیز کم خواهد بود و به همین سبب، مقدار RMSE کمتری خواهند داشت (CAO et al., ۲۰۱۲; Melesse et al., ۲۰۰۶; Melesse and Nangia, ۲۰۰۵). در نتیجه، به نظر می‌رسد که برای مقایسه دو مدل با دامنه متفاوت داده‌های آموزش، استفاده از جذر میانگین مربع خطای نرمال‌شده (NRMSE)[۱۰۷]، با توجه به حذف اثر دامنه داده‌ها، تغییرات مقدار مؤثر خطای مدل را بهتر نشان می‌دهد. هدف از پژوهش حاضر، مقایسه نتایج استفاده از شاخص‌های RMSE و NRMSE برای یک مدل زمین‌آماری دارای مجموعه داده‌های آموزشی با تعداد متفاوت نمونه می‌باشد.

مواد و روش‌ها
دشت قره‌میدان واقع در ۷۰ کیلومتری شمال غرب بجنورد برای انجام این پژوهش مد نظر قرار گرفت. وسعت این منطقه بیش از ۳۰۰ هکتار و شیب عمومی آن ۱۵ درصد است. تعداد ۱۵۱ نمونه از عمق صفر تا ۲۵ سانتی‌متری خاک که در قالب یک شبکه منظم ۱۵۰×۱۵۰ متری، از کل مساحت دشت مزبور برداشت شده بود مورد استفاده قرار گرفت. در این پژوهش از داده‌های درصد رس منطقه مورد مطالعه برای دست‌یابی به اهداف استفاده شد.
به‌منظور بررسی تأثیر تعداد نمونه برای آموزش مدل، تمامی عوامل به‌جز این عامل ثابت نگه داشته شد و برای این کار نمونه‌ها طی ۱۵ گام به‌طور مستقل و با اندازه متفاوت برای آموزش یک مدل زمین‌آماری استفاده شدند؛ به‌طوری‌که در گام اول تمامی نمونه‌ها برای آموزش مدل استفاده شد و در گام‌های بعدی، هر بار داده‌ها به دو مجموعه داده آموزش[۱۰۸] و آزمون[۱۰۹] تقسیم می‌شد و در هر گام به‌صورت تصادفی ۱۰ نمونه از نمونه‌های آموزش کم می‌شد و به نمونه‌های آزمون اضافه می‌شد تا زمانی که ۱۰ نمونه برای آموزش و ۱۴۱ نمونه برای آزمون مدل استفاده شد. با توجه به این‌که خارج شدن نمونه‌ها به‌صورت تصادفی از کل نمونه‌ها انجام می‌گرفت و ممکن بود به‌طور تصادفی خطای مدل کم یا زیاد شود، این ۱۴ گام هر کدام با ۱۰ تکرار انجام شد.

---
[۱۰۴] Effective value
[۱۰۵] Root Mean Square Error
[۱۰۶] Residuals
[۱۰۷] Normalized RMSE
[۱۰۸] Training data
[۱۰۹] Testing data

پس از آموزش مدل، مقدار مؤثر خطای مدل با استفاده از شاخص‌های RMSE و NRMSE، یک‌بار بر اساس روش جک‌نایف (کنارگذاشتن یک نمونه از مجموعه داده‌های اولیه و تخمین آن) و بار دیگر با استفاده از روش ارزیابی متقاطع (استفاده از مدل نهایی برای نمونه‌هایی که از ابتدا برای آزمون مدل کنار گذاشته شده بودند)، مورد محاسبه قرار گرفت. با توجه به تکرارهای ۱۰ گانه برای هر گام و برای این‌که بتوان تنها یک عدد را به‌عنوان نماینده خطای محاسبه‌شده برای آن گام معرفی نمود، میانگین خطای هر گام با توجه به معادله $(Max + Min)/2$)، به‌عنوان خطای آن گام انتخاب شد.

NRMSE نیز مانند RMSE به‌منظور کمی‌سازی مقدار مؤثر خطای یک مدل به کار می‌رود، با این تفاوت که این کار را نسبت به دامنه اعداد به‌کار رفته در مدل استانداردسازی می‌کند. در پژوهش حاضر نیز به‌منظور حذف اثر تغییر دامنه داده‌های به کار رفته در آموزش مدل، شاخص NRMSE و RMSE از طریق کاربرد رابطه‌های زیر محاسبه شد (Hyndman and Koehler ۲۰۰۶; Shcherbakov and Brebels ۲۰۱۳):

$$NRMSE = RMSE/R_{Train}$$

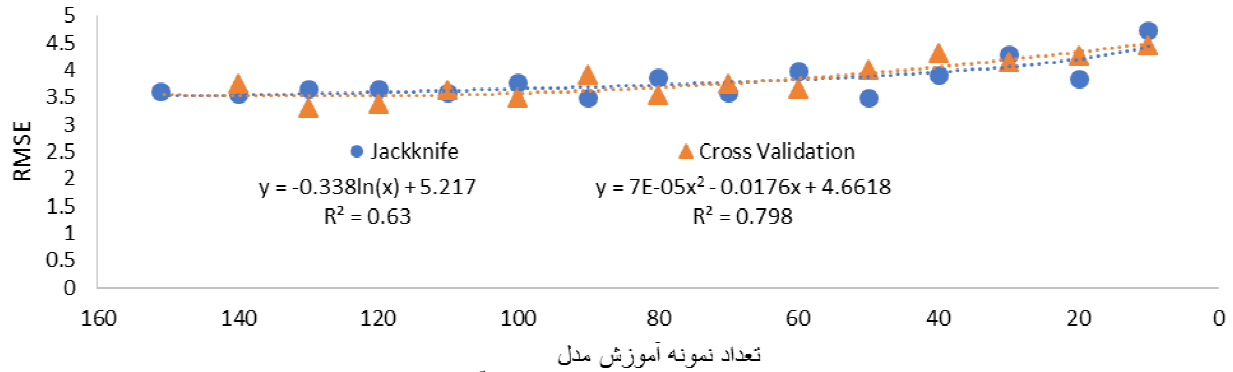
در این معادله، R_{Train}، دامنه اعداد استفاده‌شده در آموزش مدل است و RMSE از رابطه زیر محاسبه می‌گردد:

$$RMSE = \sqrt{Mean((x_{Obs,i} - x_{Per,i})^2)} \quad (۲)$$

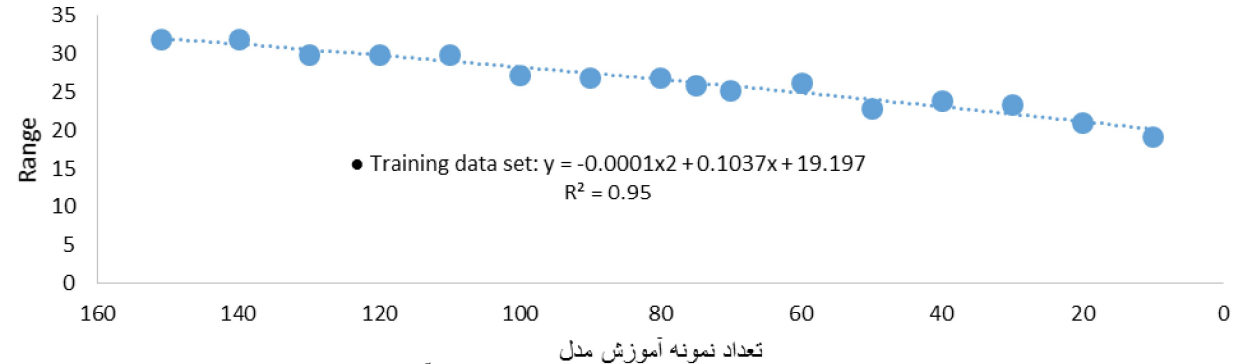
که X_{Obs} مقدار داده واقعی در نقطه i و X_{Per} مقدار داده پیش‌بینی‌شده توسط مدل در نقطه i می‌باشد.

نتایج و بحث

شکل ۱ نشان می‌دهد که با کم شدن تعداد نمونه آموزشی، مقدار RMSE با شیب ملایمی افزایش می‌یابد. با توجه به این‌که با کاهش تعداد نمونه آموزشی، دامنه تغییرات مقدار رس موجود در مجموعه داده‌های آموزش کاهش می‌یابد (شکل ۲)، به‌منظور حذف تأثیر این عامل، از شاخص NRMSE برای اندازه‌گیری مقدار مؤثر خطای مدل استفاده شد.

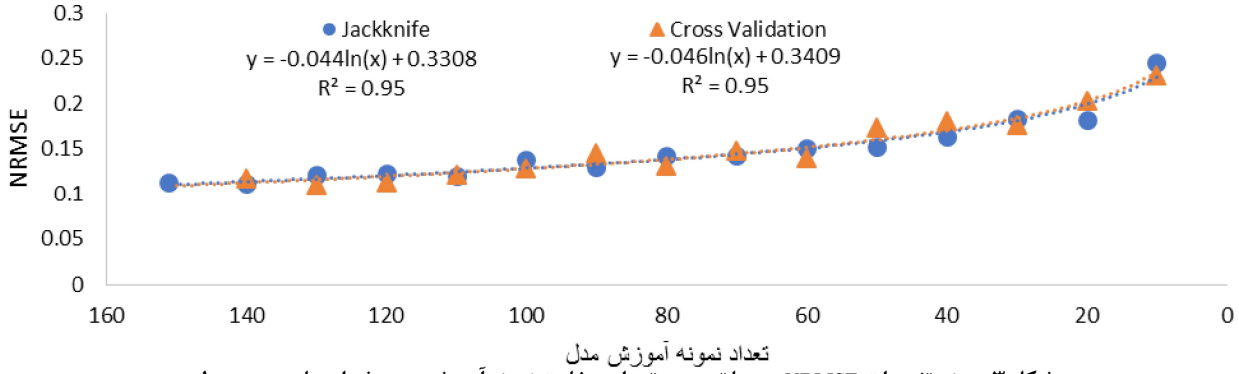


شکل ۱- روند تغییرات RMSE رس با توجه به تعداد متفاوت نمونه آموزشی و روش اعتبارسنجی مدل



شکل ۲- روند تغییرات دامنه رس با توجه به تعداد نمونه آموزش مدل

با توجه به حذف تأثیر دامنه داده‌های به‌کار رفته در آموزش مدل، باکاهش تعداد نمونه‌های آموزش مدل، مقدار مؤثر خطای محاسبه‌شده توسط شاخص NRMSE به صورت لگاریتمی افزایش می‌یابد که نشان از افزایش شدت ناپایداری و عدم اعتبار مدل در تعداد پایین نمونه آموزش مدل است (شکل ۳).



شکل ۳- روند تغییرات NRMSE رس با توجه به تعداد متفاوت نمونه آموزشی و روش اعتبارسنجی مدل

تفاوت نمودارهای RMSE و NRMSE (شکل‌های ۱ و ۳) نشان می‌دهد که استفاده از شاخص NRMSE در کنار شاخص RMSE برای تفسیر روند تغییرات مقدار مؤثر خطای مدل، با تغییر تعداد داده‌های به‌کار رفته در آموزش مدل، می‌تواند کارایی به مراتب بهتری داشته باشد.

همچنین شکل ۳ نشان می‌دهد که مقدار مؤثر خطای محاسبه‌شده توسط شاخص NRMSE در مدل‌هایی با مجموعه داده آموزش متفاوت با دقت بالایی قابل پیش‌بینی است و به دلیل حذف اثر دامنه تغییرات رس در این شاخص، دقت این مدل‌ها با یکدیگر قابل مقایسه می‌شود.

منابع

Shcherbakov M.V., Brebels A., Shcherbakova N.L., Tyukov A.P., Janovsky T.A., Kamaev V.A. ۲۰۱۳. A Survey of Forecast Error Measures. World Applied Sciences Journal, ۲۴: ۱۷۱-۱۷۶.

Hong-xin C., Hanan J.S., Yan L., Yong-xia L., Yan-bin Y., Da-wei Z., Jian-fei L., Jin-ying S., Chun-lin S., Dao-kuo G., Xiu-fang W., An-qing Y., Ping-ping T., Tai-lin B. ۲۰۱۲. Comparison of Crop Model Validation Methods. Journal of Integrative Agriculture, ۱۱: ۱۲۷۴-۱۲۸۵.

Hyndman R.J., and Koehler A.B. ۲۰۰۶. Another Look at Measures of Forecast Accuracy. International Journal of Forecasting ۲۲: ۶۷۹-۶۸۸.

Melesse A.M., Oberg J., Nangia V., Beeri O., Baumgartner D. ۲۰۰۶ Spatiotemporal Dynamics of Evapotranspiration at the Glacial Ridge Prairie Restoration in Northwestern Minnesota. Hydrological Processes ۲۰: ۱۴۵۱-۱۴۶۴.

Melesse A.M., and Nangia V. ۲۰۰۵. Estimation of Spatially Distributed Surface Energy Fluxes Using Remotely-Sensed Data for Agricultural Fields. Hydrological Processes ۱۹: ۲۶۵۳-۲۶۷۰.

Abstract

The objective of this study was to compare efficiency of RMSE and NRMSE for calculating effective error of a geostatistical model with different training data set. In this study, ۱۵۱ soil samples were used to train ۱۴۱ models with different number of training data sets. The results showed there was no significant change in RMSE value by reducing the number of training samples. According to decreasing the range of variation of clay content by reducing the number of training samples, NRMSE index was used to remove the effect of this factor. NRMSE index showed that the stability and accuracy of the model decreases logarithmically by reducing the number of training samples.