

Technical Note

Comparison of Multi Linear Regression, Nonparametric Regression, and Times Series Models for Estimation and Prediction of Evaporation Values

B. Ababaei^{1*}, H. Ramezani Etedali²,
S. Araghinejad³ and A. Liaghat³

Abstract

In order to simulate time series, various methods are presented such as times series models (AR, ARMA and ARMAX), multi-linear regression (MLR), and nonparametric regression (K-NN). In this research, performance of these models for estimation of missing values and prediction of future values of evaporation series (from open water) were assessed. ARMAX model with standardized input time series of T_{min} , T_{max} , T_{av} , Wind, RH, and sunshine hours, outperformed the other models and the K-NN and MLR were in the next ranks, respectively. Also after the principal component analysis, ARMAX model showed noticeable deviation for estimating missing values and MLR and K-NN in calibration and MLR in validation stage performed the best. For short-term predictions, ARMAX model has the best performance, but MLR performed better in long-term predictions, Time series models were not robust for long term predictions.

Keywords: Prediction, Evaporation, Multi-linear regression, Nonparametric regression, Time series models.

Received: February 29, 2012

Accepted: October 09, 2012

یادداشت فنی

مقایسه عملکرد رگرسیون‌های چندخطی، رگرسیون ناپارامتری و مدل‌های سری زمانی در برآورد و پیش‌بینی مقادیر تبخیر

بهنام آبابایی^{۱*}، هادی رمضانی اعتدالی^۲، شهاب عراقی‌نژاد^۳
و عبدالمجید لیاقت^۳

چکیده

برای شبیه‌سازی سری‌های زمانی، روش‌ها مختلفی ارائه شده‌اند که از آن جمله می‌توان مدل‌های سری زمانی AR، ARMA و ARMAX و روش‌های رگرسیون چندخطی (MLR) و رگرسیون ناپارامتری (K-NN) را برشمرد. در این تحقیق، عملکرد این روش‌ها در برآورد داده‌های مفقود و پیش‌بینی مقادیر آبی سری زمانی تبخیر از سطح آزاد آب مورد بررسی قرار گرفت. مدل ARMAX با استفاده از ورودی‌های استاندارد شده دمای کمینه و بیشینه، متوسط دما، سرعت باد، رطوبت‌نسبی و ساعات آفتابی، نسبت به سایر روش‌ها به بهترین نتایج منجر شد و روش‌های K-NN و MLR در رتبه‌های بعدی قرار گرفتند. مشاهده گردید که پس از تحلیل اجزاء اصلی، مدل ARMAX دارای خطای قابل ملاحظه‌ای در برآورد مقادیر مفقود می‌باشد و روش‌های MLR و K-NN در مرحله واسنجی و روش MLR در مرحله اعتبارسنجی به بهترین نتایج منجر می‌شوند. برای پیش‌بینی‌های کوتاه مدت، مدل ARMAX دارای بهترین نتایج بود؛ ولی در پیش‌بینی‌های بلند مدت، روش MLR به بهترین نتایج منجر شد و مدل‌های سری زمانی قادر به پیش‌بینی‌های بلندمدت نبودند.

کلمات کلیدی: پیش‌بینی، تبخیر، رگرسیون چندخطی، رگرسیون ناپارامتری، مدل‌های سری زمانی.

تاریخ دریافت مقاله: ۱۰ اسفند ۱۳۹۰

تاریخ پذیرش مقاله: ۱۸ مهر ۱۳۹۱

1- Young Researchers and Elites Club, Science and Research Branch, Islamic Azad University, Tehran, Iran, Email:Behnam.ab@gmail.com

2- Assistant Professor, Department of Water Engineering, Imam Khomeini International University, Qazvin, Iran.

3- Assistant Professor and Professor (Respectively), Department of Irrigation and Reclamation, Faculty of Agricultural Engineering and Technology, University of Tehran, Karaj, Iran.

*- Corresponding Author

۱- کارشناس ارشد دانشگاه آزاد اسلامی، واحد علوم و تحقیقات، باشگاه پژوهشگران جوان و نخبگان، تهران، ایران.

۲- استادیار گروه مهندسی آب، دانشگاه بین المللی امام خمینی، قزوین، ایران.

۳- به ترتیب، استادیار و استاد گروه آبیاری و آبادانی، دانشکده مهندسی و فناوری کشاورزی، دانشگاه تهران، کرج، ایران.

*- نویسنده مسئول

۱- مقدمه

$$\alpha(Xt, Xi) = \begin{cases} \frac{1}{k} & k \leq K \\ 0 & k > K \end{cases} \quad (۲) \text{ تابع وزن شماره ۱:}$$

$$\alpha(Xt, Xi) = \begin{cases} \frac{1}{d(k)} & k \leq K \\ 0 & k > K \end{cases} \quad (۳) \text{ تابع وزن شماره ۲:}$$

$$\alpha(Xt, Xi) = \begin{cases} \frac{1}{\Delta T(k)} & k \leq K \\ 0 & k > K \end{cases} \quad (۴) \text{ تابع وزن شماره ۳:}$$

$d(k)$ فاصله محاسبه شده نمونه i از نمونه t و $\Delta T(k)$ فاصله زمان وقوع نمونه i از زمان وقوع نمونه t می باشد. در نهایت پس از استانداردسازی وزن ها (مجموع وزن ها برابر با یک)، متغیر وابسته مورد نظر در زمان t از رابطه (۵) محاسبه می شود:

$$Z_t = \sum_{i=1}^K \alpha(X_t, X_i) Z_i \quad (۵)$$

Z_i مقدار متغیر وابسته در همسایه i می باشد. تعداد زیاد K منجر به نرم شدن داده ها شده و کم بودن آن نیز می تواند به جواب های نادرست بیانجامد. برای تعیین مقادیر اولیه K می توان از رابطه \sqrt{N} (Lall and Sharma (1996)) استفاده کرد که N تعداد داده های مشاهده ای است.

۲- مواد و روش ها

۱-۲- رگرسیون ناپارامتری K نزدیکترین همسایه (K-NN)

به منظور تشخیص همسایه ها، لازم است که یک تابع فاصله (یا شباهت) ^۴ مورد استفاده قرار گیرد (Wilson and Martinez (2000)). از جمله توابع فاصله مرسوم، تابع فاصله اقلیدسی می باشد که برای دو نمونه X_t و X_i به صورت زیر بیان می شود:

$$d_{t,i} = \sqrt{\sum_{j=1}^m \{S_j(X_{j,i} - X_{j,t})\}^2} \quad (۱)$$

که $X_{j,i}$ ویژگی یا پیش بینی کننده j از نمونه i و m تعداد پیش بینی کننده ها و d فاصله نمونه i از نمونه t می باشد. S پارامتر بی بعدسازی هریک از پیش بینی کننده ها می باشد که باعث می شود هریک از آنها در انتخاب نزدیکترین همسایه ها، نقش یکسانی داشته باشند (Harrold et al. (2003), Sharma and Lall (1999), Rajagopalan and Lall (1999)). پس از تعیین فاصله تمامی نمونه ها از نمونه مورد نظر، این مقادیر به صورت صعودی مرتب شده و K نمونه اول انتخاب می شوند. برای تعیین وزن هریک از K نمونه، روابطی ارائه شده اند (روابط ۲ تا ۴). رابطه (۲) توسط Lall and Sharma (1996) پیشنهاد شده و رابطه (۴) در این مقاله پیشنهاد گردیده است.

۲-۲- مدل های سری زمانی

مهمترین هدف از تجزیه و تحلیل یک سری زمانی، یافتن مدلی است که بتواند براساس مشاهدات موجود، اندازه کمیت را در چند واحد زمانی آینده پیش بینی کند. از جمله پرکاربردترین مدل ها، مدل های AR، ARMA، و ARMAX می باشند. پارامترهای na ، nb ، nc (مرتبه مدل) و nk (تأخیر زمانی پیش بینی کننده ها نسبت به متغیر وابسته)، پارامترهای ورودی به مدل هستند که لازم است از پیش تعیین شوند.

۲-۳- پیش پردازش اطلاعات

به منظور ارزیابی و مقایسه عملکرد روش های مختلف، از اطلاعات هواشناسی روزانه ایستگاه بندرعباس (حداکثر، حداقل و میانگین دمای روزانه، بارش روزانه، سرعت باد، ساعات آفتابی، درصد رطوبت نسبی هوا) در سال های ۲۰۰۵ و ۲۰۰۶ برای شبیه سازی سری زمانی تبخیر از سطح آزاد آب استفاده شد. پیش پردازش اطلاعات شامل حذف داده های پرت با استفاده از توزیع t و با حد اطمینان ۹۵ درصد، حذف روند سری های زمانی، استانداردسازی و تحلیل اجزاء اصلی (PCA) ^۹ (حذف متغیرهای همبسته و کاهش ابعاد بردار ورودی) می باشد. سری بارش روزانه به دلیل عدم همبستگی با سایر

متغیرهای ورودی، از مجموعه پیش‌بینی‌کننده‌ها حذف گردید.

۳- نتایج و بحث

۳-۱- نتایج مدل سری زمانی

در مرحله واسنجی (برای حفظ اختصار، نتایج مرحله واسنجی ارائه نشده‌اند) در انتخاب بهترین مدل ARMAX، افزایش رتبه مدل تأثیری در بهبود عملکرد آن ندارد. در ارتباط با مدل‌های AR، بهترین مدل، یکی از مدل‌های مرتبه ۱ تا ۳ می‌باشد. جدول ۱ نشان می‌دهد که در روش ۱، مدل ARMAX در تمامی موارد نتایج بهتری نسبت به مدل‌های ARMA یا AR دارد. همچنین هیچیک از مدل‌های سری زمانی، قادر به پیش‌بینی بلندمدت سری زمانی (۳۰ درصد از داده‌ها) نمی‌باشند. پس از فرآیند PCA (روش ۲)، شاخص RMSE در مرحله واسنجی بشدت افزایش یافته و مدل سری زمانی نسبت به مقادیر واقعی، مقادیر بزرگتری برآورد می‌نماید. این مسأله در مرحله اعتبارسنجی دیده نشد. مدل‌های ARMA یا AR، مشابه مدل‌های بدست آمده در روش ۱ خواهند بود. خطای زیاد شبیه‌سازی مدل‌ها در تکرار سوم را می‌توان به تفاوت میانگین و انحراف استاندارد در مراحل واسنجی و اعتبارسنجی که نشان از تغییر روند سری زمانی واقعی دارد مرتبط دانست. همچنین میانبایی داده‌های مفقود در مرحله واسنجی (تکرار ۲)، تأثیر نامطلوبی در نتایج داشته است.

۳-۲- نتایج روش رگرسیون ناپارامتری

در جدول ۲، منظور از اعداد ستون "مدل"، به ترتیب، مقدار K و شماره تابع وزن انتخاب شده می‌باشد. مشاهده می‌شود که در بیشتر موارد، تابع وزن شماره ۳ که در این مقاله پیشنهاد شده است (رابطه ۴) و با حداکثر مقدار K، به کمترین شاخص RMSE منجر شده است. در پیش‌بینی مقادیر آتی، با افزایش تعداد این مقادیر، مقدار کمتر K به بهترین نتیجه منجر شده است (۲۰ در برابر ۳۰). تفاوت انحراف استاندارد روش K-NN با مقادیر واقعی نسبت به نتایج حاصل از سری زمانی مدل ARMAX بیشتر و از نتایج حاصل از سری ARMA یا AR کمتر بوده است. شاخص RMSE در روش K-NN از مدل ARMAX کمی بیشتر، ولی نسبت به مدل‌های ARMA و AR در حد کمتری قرار دارد. در شبیه‌سازی سری‌های زمانی دارای روندهای معنی‌دار، استفاده از روش K-NN با چالش روبروست. در پیش‌بینی‌های کوتاه‌مدت‌تر (۱۰ درصد از کل داده‌ها)، مدل‌های سری زمانی نتایج بسیار بهتری در مقایسه با روش K-NN داشته‌اند. اما این مدل‌ها قادر به پیش‌بینی‌های بلندمدت (۳۰ درصد از داده‌ها) نبوده و روش K-NN نتایج نسبتاً مطلوبی داشته است، هرچند که خطای این روش هم قابل توجه است. فرآیند PCA (روش ۲) باعث افزایش خطای روش K-NN شده است. در روش ۳،

۳-۲- روش ارزیابی مدل سری زمانی

ارزیابی مدل‌های سری زمانی به ۲ روش و در ۳ تکرار انجام شد. در روش اول، مدل سری زمانی بر پیش‌بینی‌کننده‌های استاندارد شده اعمال شد. در روش دوم، از پیش‌بینی‌کننده‌های باقی‌مانده از فرآیند PCA استفاده شد. در هر روش، تکرار اول شامل شبیه‌سازی مقادیر مفقود بدون میانبایی آنها در مرحله واسنجی، تکرار دوم با میانبایی و تکرار سوم مربوط به پیش‌بینی مقادیر آتی می‌باشد. داده‌ها بصورت تصادفی بین دو مجموعه واسنجی^۱ و اعتبارسنجی^{۱۱} (به نسبت‌های ۷۰-۳۰ و ۹۰-۱۰) تقسیم شدند. برای ارزیابی مدل‌های سری زمانی در پیش‌بینی مقادیر آتی، داده‌های اعتبارسنجی از انتهای سری زمانی و پس از داده‌های واسنجی انتخاب شدند.

پس از تعیین بهترین مدل‌ها، کلیه فرآیند محاسبات با استفاده از شاخص‌های ریشه میانگین مربع خطاها (RMSE)، ضریب همبستگی خطی (R)، شیب و عرض از مبدأ خط بین داده‌های واقعی و مقادیر شبیه‌سازی شده (M و B) و میانگین و انحراف استاندارد سری‌های واقعی و شبیه‌سازی شده مورد ارزیابی قرار گرفت.

۳-۵- روش ارزیابی رگرسیون ناپارامتری و رگرسیون چندخطی

به منظور ارزیابی عملکرد روش‌های رگرسیونی، ۳ روش و در ۲ تکرار در نظر گرفته شد. هر سه تابع وزن با محدود کردن مقدار پارامتر K به ۳۰، مورد ارزیابی قرار گرفتند. در روش اول، روش K-NN بر پیش‌بینی‌کننده‌های استاندارد شده اعمال شد. در روش دوم، رگرسیون با استفاده از پیش‌بینی‌کننده‌های باقی‌مانده از فرآیند PCA و بدون اعمال وزن صورت گرفت. در روش سوم (که تنها در مورد روش K-NN مورد استفاده قرار گرفت)، از پیش‌بینی‌کننده‌های باقی‌مانده از فرآیند PCA و با اعمال وزن بر آنها استفاده شد. وزن‌ها، از رابطه (۶) محاسبه شدند:

$$r_i = \frac{\text{Std}_i}{\sum_j \text{Std}_j} \quad (6)$$

Std انحراف استاندارد نمونه هریک از پیش‌بینی‌کننده‌های باقی‌مانده و z تعداد پیش‌بینی‌کننده‌های باقی‌مانده از فرآیند PCA را نشان می‌دهد.

از روش وزن‌دهی به پیش‌بینی‌کننده‌ها استفاده شد (رابطه ۶). رگرسیون حذف شدند. محاسبه وزن پیش‌بینی‌کننده‌ها براساس فرآیند پیش‌بینی‌کننده‌هایی که دارای وزن کمتر از ۲ درصد بودند، از فرآیند PCA، در غالب موارد به افزایش شیب خط برازش داده شده بین

جدول ۱- نتایج مدل‌های سری زمانی پس از استاندارد شدن داده‌ها

روش	تکرار	تعداد	هدف	مدل	میانگین داده‌ها	میانگین مدل	انحراف استاندارد داده‌ها	انحراف استاندارد مدل	RMSE	R	
۱	۱	۱۰٪	مقادیر مفقود	ARMAX	۵/۸۵	۵/۹۵	۲/۹۳	۲/۷۱	۱/۳۱	-۰/۸۹	
				AR		۶/۲۰		۲/۳۱	-۰/۷۹		
	۲	۱۰٪	مقادیر مفقود	ARMAX	۵/۸۵	۵/۹۷	۲/۹۳	۲/۷۰	۱/۲۹	-۰/۹	
				AR		۶/۲۲		۲/۳۱	-۰/۷۹		
	۳	۱۰٪	پیش بینی	ARMAX	۳/۶۱	۳/۶۹	۱/۸۶	۱/۵۰	۱/۳۷	-۰/۶۸	
				AR		۴/۳۹		۱/۳۵	-۰/۵۲		
	۱	۳۰٪	مقادیر مفقود	ARMAX	۶/۱	۵/۶۶	۲/۵۱	۲/۲۴	۱/۳۴	-۰/۸۶	
				AR		۶/۰۰		۲/۱۲	-۰/۷۹		
	۲	۳۰٪	مقادیر مفقود	ARMAX	۶/۱۲	۶/۳۵	۲/۵۷	۲/۳۵	۱/۴۳	-۰/۸۴	
				ARMA		۶/۲۱		۲/۱۹	-۰/۷۷		
	۳	۳۰٪	پیش بینی	ARMAX	۶/۶۴	-----	۲/۷۸	-----	-----	-----	-۰/۱۴
				AR		-----		-----	-----	-۰/۱۵	
۲	۱	۱۰٪	مقادیر مفقود	ARMAX	۵/۸۵	۶/۰۳	۲/۹۴	۲/۶۱	۱/۲۵	-۰/۹۱	
	۲	۱۰٪	مقادیر مفقود	ARMAX	۵/۸۵	۶/۰۰	۲/۹۴	۲/۶۱	۱/۲۵	-۰/۹	
	۳	۱۰٪	پیش بینی	ARMAX	۳/۶۱	۳/۸۷	۱/۸۷	۱/۴۴	۱/۳۴	-۰/۷۱	
	۱	۳۰٪	مقادیر مفقود	ARMAX	۶/۱۰	۶/۰۳	۲/۵۱	۲/۴۴	۱/۲۸	-۰/۸۶	
	۲	۳۰٪	مقادیر مفقود	ARMAX	۶/۱۲	۶/۲۹	۲/۵۷	۲/۲۹	۱/۴	-۰/۸۴	
	۳	۳۰٪	پیش بینی	ARMAX	۶/۶۴	-----	۲/۷۸	-----	-----	-----	-۰/۰۳

جدول ۲- نتایج رگرسیون ناپارمتری پس از استاندارد شدن داده‌ها (بدون وزن‌دهی)

روش	تکرار	تعداد	هدف	مدل	میانگین داده‌ها	میانگین مدل	انحراف استاندارد داده‌ها	انحراف استاندارد مدل	RMSE	R	M	B
۱	۱	۱۰٪	مقادیر مفقود	۲-۲۹	۵/۸۵	۵/۹۵	۲/۹۴	۲/۲۱	۱/۴۹	-۰/۸۷	-۰/۶۵	۲/۱۳
	۲	۱۰٪	پیش بینی	۳-۳۰	۳/۶۱	۵/۷۶	۱/۸۷	-۰/۶۷	۲/۶۹	-۰/۵۱	-۰/۱۹	۵/۰۹
	۱	۳۰٪	مقادیر مفقود	۳-۳۰	۶/۱۰	۵/۸۵	۲/۵۱	۲/۲۸	۱/۴۷	-۰/۸۲	-۰/۷۴	۱/۳۱
	۲	۳۰٪	پیش بینی	۲-۲۰	۶/۶۴	۷/۴۷	۲/۷۸	۱/۹۴	۱/۷۸	-۰/۸۳	-۰/۵۸	۳/۵۹
۲	۱	۱۰٪	مقادیر مفقود	۲-۲۹	۵/۸۵	۵/۹۷	۲/۹۴	۲/۲۱	۵/۳۴	-۰/۸۷	-۰/۶۵	۲/۱۵
	۲	۱۰٪	پیش بینی	۳-۳۰	۳/۶۱	۵/۷۷	۱/۸۷	-۰/۶۹	۶/۸۸	-۰/۵۱	-۰/۱۹	۵/۱۰
	۱	۳۰٪	مقادیر مفقود	۳-۳۰	۶/۱۰	۵/۸۵	۲/۵۱	۲/۲۹	۴/۹۴	-۰/۸۲	-۰/۷۵	۱/۲۷
	۲	۳۰٪	پیش بینی	۲-۲۰	۶/۶۴	۷/۴۷	۲/۷۸	۱/۹۶	۴/۱۴	-۰/۸۳	-۰/۵۹	۳/۵۷
۳	۱	۱۰٪	مقادیر مفقود	۳-۳۰	۵/۸۵	۵/۹۱	۲/۹۴	۲/۳	۵/۳۴	-۰/۸۷	-۰/۶۸	۱/۹۳
	۲	۱۰٪	پیش بینی	۳-۳۰	۳/۶۱	۵/۷۲	۱/۸۷	-۰/۷۰	۶/۸۸	-۰/۶۲	-۰/۲۳	۴/۸۸
	۱	۳۰٪	مقادیر مفقود	۳-۳۰	۶/۱۰	۵/۸	۲/۵۱	۲/۳۴	۴/۹۴	-۰/۸۲	-۰/۷۷	۱/۱۲
	۲	۳۰٪	پیش بینی	۲-۳۰	۶/۶۴	۷/۴۹	۲/۷۸	۱/۹۷	۴/۱۴	-۰/۸۴	-۰/۶	۳/۵۳

مقادیر واقعی و مقادیر شبیه‌سازی شده و کاهش عرض از مبدأ این خط و تا حدودی افزایش مقدار R منجر می‌شود. اما مقدار RMSE را تغییر نمی‌دهد.

۳-۳- نتایج روش رگرسیون چندخطی

جدول ۳، نتایج روش MLR برای تکرارهای مختلف را نشان می‌دهد. پس از استاندارد کردن داده‌ها (روش ۱)، خطای این روش در مرحله واسنجی، تقریباً مشابه با روش K-NN می‌باشد، ولی در تمامی موارد از مدل ARMAX بیشتر است. در پیش‌بینی‌های کوتاه‌مدت، خطای روش MLR از مدل AR و ARMA بیشتر، ولی در پیش‌بینی‌های بلندمدت از آنها کمتر است. تحلیل اجزاء اصلی (روش ۲) تأثیر مثبتی در مقدار پارامترهای R و RMSE روش MLR نداشته و در اکثر موارد باعث افزایش خطا شده است (به ویژه در پیش‌بینی‌های بلندمدت). در مرحله اعتبارسنجی، در بیشتر موارد روش MLR نسبت به روش K-NN به خطای کمتر و مقدار R بیشتری منجر شده است. به ویژه پس از تحلیل اجزاء اصلی. در پیش‌بینی‌های بلندمدت، روش MLR در بین تمامی روش‌ها به بهترین نتیجه منجر شده و حتی در مقایسه با روش K-NN دارای خطای کمتری می‌باشد.

۴- نتیجه‌گیری

نتایج این مطالعه نشان داد که مدل ARMAX عملکرد بهتری نسبت به مدل‌های ARMA و AR دارد و فرآیند PCA در زمان برآورد مقادیر مفقود، باعث افزایش عدم قطعیت و در زمان پیش‌بینی

مقادیر آتی، باعث کاهش عدم قطعیت می‌شود. هیچیک از مدل‌های سری زمانی، برای پیش‌بینی‌های بلندمدت توصیه نمی‌شوند. فرآیند PCA در روش K-NN در اکثر موارد باعث افزایش خطای تمامی مدل‌ها گردید. استفاده از روش K-NN در پیش‌بینی‌های کوتاه‌مدت توصیه نمی‌شود و در این شرایط، مدل‌های سری زمانی به نتایج بهتری منجر می‌شوند. بهتر است در پیش‌بینی‌های کوتاه‌مدت، از مدل‌های سری زمانی و در پیش‌بینی‌های بلندمدت، از روش‌های MLR و K-NN بدون PCA استفاده شود. در پیش‌بینی‌های بلندمدت، روش MLR به بهترین نتایج منجر گردید. در روش K-NN، انتخاب تابع فاصله بیش از تعیین وزن ورودی‌ها اهمیت دارد. تعیین وزن ورودی‌ها پس از فرآیند PCA و براساس میزان انحراف استاندارد آنها، تأثیر چندانی در کاهش مقدار خطا نداشت. استفاده از تابع وزن پیشنهادی در این مقاله، می‌تواند در شبیه‌سازی سری‌های زمانی با استفاده از روش K-NN مورد توجه قرار گیرد.

پی‌نوشت‌ها

- 1- Nearest Neighbour
- 2- Pattern Classification
- 3- Multi-linear Regression
- 4- Distance (or Similarity) Function
- 5- Mahalanobis Distance Function
- 6- Smoothing
- 7- Autoregressive Models
- 8- Autoregressive - Moving Average Models
- 9- Principal Component Analysis (PCA)
- 10- Calibration
- 11- Validation

جدول ۳- نتایج رگرسیون چندخطی پس از استاندارد شدن داده‌ها

روش	تکرار	تعداد اعداد اعتبارسنجی	هدف	میانگین داده‌ها	میانگین مدل	انحراف استاندارد داده‌ها	انحراف استاندارد مدل	RMSE	R	M	B
۱	۱	۱۰٪	مقادیر مفقود	۵/۸۵	۵/۸	۲/۹۴	۲/۳۹	۱/۴۷	۰/۸۷	۰/۷۱	۱/۶۸
	۲	۱۰٪	پیش‌بینی	۳/۶۱	۵/۸۴	۱/۸۷	۱/۱۵	۲/۶۷	۰/۶۱	۰/۳۷	۴/۴۹
	۱	۳۰٪	مقادیر مفقود	۶/۱۰	۵/۹۴	۲/۵۱	۲/۱۷	۱/۵۱	۰/۸۰	۰/۷۰	۱/۷۰
	۲	۳۰٪	پیش‌بینی	۶/۶۴	۷/۳۳	۲/۷۸	۲/۰۴	۱/۶۶	۰/۸۴	۰/۶۲	۳/۲۰
۲	۱	۱۰٪	مقادیر مفقود	۵/۸۵	۵/۷۴	۲/۹۴	۲/۴۲	۱/۴۵	۰/۸۷	۰/۷۲	۱/۵۵
	۲	۱۰٪	پیش‌بینی	۳/۶۱	۶/۱۶	۱/۸۷	۱/۱۴	۲/۹	۰/۶۶	۰/۴۱	۴/۶۹
	۱	۳۰٪	مقادیر مفقود	۶/۱۰	۵/۹۴	۲/۵۱	۲/۱۶	۱/۵۴	۰/۷۹	۰/۶۸	۱/۷۷
	۲	۳۰٪	پیش‌بینی	۶/۶۴	۷/۳۷	۲/۷۸	۱/۹۴	۱/۶۹	۰/۸۵	۰/۵۹	۳/۴۳

- Science Society of America Journal*, 70, pp. 327-336.
- Rajagopalan, B. and Lall, U. (1999). "A k-nearest-neighbor simulator for daily precipitation and other variables", *Water Resources Research*, 35(10), pp. 3089-3101.
- Sharma, A. and Lall, U. (1999). "A nonparametric approach for daily rainfall simulation", *Math Comput Simulat*, 48, pp. 361-371.
- Todini, E. (2000). "Real-time flood forecasting: operational experience and recent advances", In: Marsalek, J., et al. (Eds.), *Flood Issues in Contemporary Water Management*, Kluwer Academic Publisher, The Netherlands.
- Wilson, D. R. and Martinez, T. R. (2000). "Reduction techniques for exemplar-based learning algorithms", *Machine Learning*, 38(3), pp. 257-286.
- Wojcik, R. and Buishand, T. A. (2003). "Simulation of 6-hourly rainfall and temperature by two resampling schemes", *Journal of Hydrology*, 273, pp. 69-80.
- Wu, W., Xing, E.P., Myers, C., Mian, I.S. and Bissell, M.J. (2005). "Evaluation of normalization methods for CDNA microarray data by K-NN classification", *Bioinformatics*, 6, pp. 191-200.
- Young, K. C. (1994). "A multivariate chain model for simulating climatic parameters with daily data", *J. Appl. Meteorol.*, 33(6), pp. 661-671.
- Araghinejad, S., Burn, D. H. and Karamouz, M. (2006). "Long-lead probabilistic forecasting of streamflow using ocean-atmospheric and hydrological predictors", *Water Resour. Res.*, 42, W03431, doi:10.1029/2004WR003853.
- Beersma, J. J. and Buishand, T. A. (2003). "Multi-site simulation of daily precipitation and temperature conditional on the atmospheric circulation", *Clim Res.*, 25, pp. 121-133.
- Chi, M. and Bruzzone, L. (2005). "An ensemble-driven K-NN approach to ill-posed classification problems", *Pattern Recognition Letters*, 27, pp. 301-307.
- Davis, G.A. and Nihan, N.L. (1991). "Nonparametric regression and short-term free way traffic forecasting", *Journal of Transportation Engineering, ASCE*, 117(2), pp. 178-188.
- Harrold, T. I., Sharma, A. and Sheather, S. J. (2003). "A nonparametric model for stochastic generation of daily rainfall occurrence", *Water Resour Res.*, 39, pp. 1-11.
- Lall, U. and Sharma, A. (1996). "A nearest neighbor bootstrap for time series resampling", *Water Resour Res.*, 32, pp. 679-693.
- Nemes, A., Rawls, W.J. and Pachepsky, Y.A. (2006). "Use of the nonparametric nearest neighbor approach to estimate soil hydraulic properties", *Soil*