

## پیش‌بینی مقادیر دبی روزانه رودخانه تالار با استفاده از مدل یادگیری کندرو

زهرا عبداللهی<sup>1</sup> - عطااله کاویان<sup>2\*</sup> - کاکا شاهی<sup>3</sup> - ندا عبداللهی<sup>4</sup> - محمد جعفری<sup>5</sup>

تاریخ دریافت: 1394/06/16

تاریخ پذیرش: 1395/01/23

### چکیده

توسعه تکنیک‌های مختلف جهت پیش‌بینی کوتاه‌مدت و بلندمدت دبی ساعتی، روزانه، ماهانه و سالانه جریان به‌منظور مدیریت منابع آب و پروژه‌های عمرانی سابقه طولانی دارد. در دهه‌های اخیر، تکنیک‌های یادگیری متعددی به‌طور گسترده جهت پیش‌بینی دبی و سایر متغیرهای هیدرولوژیک مورد توجه قرار گرفته‌اند. پژوهش حاضر با هدف پیش‌بینی مقادیر دبی روزانه با استفاده از بهترین‌ترین تعداد داده‌های ورودی در رودخانه منتهی به رود تالار واقع در حوزه آبخیز کسلیان انجام گرفت. بدین منظور از سه مدل ثابت (conll.c)، خطی (linll.c) و درجه دو (quall.c) الگوریتم یادگیری محلی کندرو که از کارایی قابل توجهی در شبیه‌سازی پارامترهای متغیر با داده‌های ورودی کم برخوردار هستند و هم‌چنین با استفاده از سری داده‌های ورودی 6، 8، 10، 15، 20 روز قبل، یک و دو ماه قبل، یک، دو و سه فصل قبل و در نهایت یک و دو سال قبل استفاده گردید. نتایج به‌دست آمده از چندین هزار مدل تعلیمی تهیه شده نشان داد که مدل ثابت با داده‌های دبی روزانه‌ی 60 روز گذشته با حداقل خطای 0/001 به مقدار واقعی نزدیک‌تر بوده است. نتایج به‌دست آمده از مقادیر محاسبه شده RMSE و MAE حاکی از آن است که الگوریتم یادگیری محلی کندرو علی‌رغم محدودیت‌های موجود از جمله حساسیت بسیار بالا به همسایگی، از کارایی بالایی در شبیه‌سازی سری‌های زمانی با خطای نسبتاً پایین (RMSE کمتر از 0/06) برخوردار می‌باشد.

**واژه‌های کلیدی:** حوزه آبخیز کسلیان، دبی روزانه، مدل یادگیری محلی کندرو، نزدیک‌ترین همسایگی

### مقدمه

فرآیندهای طبیعی، محققان را ملزم به استفاده از روش‌های متعددی جهت توصیف متغیرهای هیدرولوژیک و پیش‌بینی گام‌های زمانی آن نموده است. توسعه تکنیک‌های مختلف جهت پیش‌بینی کوتاه‌مدت و بلندمدت دبی سالانه، ماهانه، روزانه و ساعتی جریان با افق زمانی کم‌تر از هفت روز تا چندین ماه سابقه طولانی دارد (6 و 14). طبق نتایج گزارش شده، پیش‌بینی‌های کوتاه‌مدت در مقابل پیش‌بینی‌های بلندمدت قابلیت شبیه‌سازی بهتر با خطای تخمین کم‌تری داشته و از پیچیدگی کم‌تری در مدل‌سازی برخوردارند (12). در دهه‌های اخیر، تکنیک‌های یادگیری متعددی از قبیل شبکه عصبی مصنوعی (7، 13، 17 و 22)، منطق فازی، الگوریتم ژنتیک، آنالیز موجک (21)، بردار پشتیبان و یادگیری مبتنی بر الگوریتم کرنل (1) به‌طور گسترده در مدل‌سازی‌ها و پیش‌بینی متغیرهای هیدرولوژیک مورد توجه قرار گرفته‌اند. روش‌های یادگیری به دو بخش اصلی کلی<sup>4</sup> و محلی<sup>5</sup> قابل تقسیم‌اند که در یادگیری‌های کلی پس از بررسی کل داده‌های آموزشی، یک رابطه از بین آن داده‌ها استخراج و سپس بر اساس رابطه به‌دست آمده برای ورودی‌های جدید پاسخ مناسب پیش‌بینی

دبی جریان رودخانه به عنوان یکی از مؤلفه‌های مهم در چرخه هیدرولوژی که در ارتباط مستقیم با محیط زیست قرار دارد، نقش اساسی در فرآیندهای فیزیکی، اکولوژیکی، اجتماعی و اقتصادی ایفا می‌نماید (1). لذا، پیش‌بینی و تخمین دقیق و قابل اعتماد دبی جریان رودخانه‌ها جهت برنامه‌ریزی و مدیریت آب‌های سطحی، طراحی سازه‌های هیدرولیکی، پیش‌بینی احتمال وقوع سیلاب و مطالعات اکولوژیکی در مقیاس گسترده مکانی و زمانی از دیرباز مورد توجه مهندسان و محققان زیادی قرار گرفته است (16). اهمیت پیش‌بینی جریان رودخانه در مدیریت منابع آب حوزه‌هایی که حجم بیش‌تر بارش در فصول خاصی اتفاق افتاده و باقی‌ماه‌های سال بدون بارندگی می‌باشد چشم‌گیرتر است. از طرفی، سرشت متغیر و پیچیده‌ی

1، 2 و 3- به ترتیب دانشجوی دکتری مهندسی آبخیزداری، دانشیار و استادیار گروه مهندسی آبخیزداری، دانشگاه کشاورزی و منابع طبیعی ساری  
(\*) - نویسنده مسئول: Email: a.kavian@sanru.ac.ir

4- مربی گروه مهندسی نرم‌افزار، دانشگاه صائب

5- مدرس گروه مهندسی نرم‌افزار، دانشگاه آزاد اسلامی واحد زنجان

DOI: 10.22067/jsw.v30i6.47522

4- Global learning  
5- Local learning

در حالی است که تاکنون در مطالعات مربوط به علوم طبیعی به‌طور گسترده مورد توجه محققان قرار نگرفته و تنها سوابق تحقیقی اندکی در کاربرد این روش در علوم محیطی موجود است. کرانی (2005) در میلان، کارایی دو روش شبکه عصبی FFNN و PNN را در تخمین مقدار دو آلاینده هوا، اوزن و  $PM_{10}^2$ ، با مدل محلی L.L مقایسه نمود. او در نتایج خود بیان داشت که مدل L.L در تخمین مقادیر آلاینده‌های مورد نظر از عملکرد بهتری برخوردار بوده است (8).

والز و همکاران (2004)، از استراتژی یادگیری کندرو جهت انتخاب مناسب‌ترین الگوهای تعلیمی به‌منظور پیش‌بینی نوسانات سطح خلیج و نیز با روش شبکه عصبی RBNN، استفاده نمودند. نتایج ارائه شده حاکی از آن بود که از بین مجموعه 3000 داده در دسترس مربوط به طغیان دریا، شبیه‌سازی با استفاده از الگوی تعلیمی با مجموعه شش داده گذشته متوالی بهترین نتیجه را داشته است (20).

واضح است که پارامترهای متعددی از قبیل بارش، تبخیر و تعرق، دما، سطح آب زیرزمینی، محتوی رطوبتی اولیه خاک و غیره در میزان دبی رواناب روز آینده دخیل بوده و ثبت حجم زیاد داده‌های ورودی و همچنین تهیه مدل‌های پیچیده‌ی هیدرولوژیک غیرممکن نیست، اما به لحاظ اقتصادی مدلی مطلوب است که قادر به شبیه‌سازی تغییرات جریان با حداقل داده‌های ثبت شده و در دسترس کارشناسان، مدیران و تصمیم‌گیران حوزه آبخیز باشد. کیشی در سال 2004 با استفاده از دو مدل یادگیری کلی FNN و AR و تنها با کمک داده‌های دبی دو سال گذشته اقدام به پیش‌بینی دبی روزانه در سه رودخانه بلک‌واتر و گیلا در آمریکا و فیلیوس در ترکیه نمود (13). پولیدو-کالوو و پرتلا نیز (2007) جهت پیش‌بینی دبی یک روز آینده در رودخانه توآ در شمال کشور پرتغال با استفاده از مدل CNN، تنها از داده‌های دبی گذشته استفاده نمودند (16). همچنین هاگوات و میتی (2012) با استفاده از داده‌های روزانه ایستگاه هیدرومتری رودخانه نارمادا در هندوستان به کمک مدل دو LS-SVR و ANN دبی پنج روز آینده رودخانه مذکور را پیش‌بینی نمودند (1).

بنابراین با توجه به سوابق تحقیقی مطالعه شده، پژوهش حاضر با هدف بررسی شبیه‌سازی دبی جریان یک روز آینده در رودخانه منتهی به رود تالار واقع در حوزه آبخیز کسلیان با حداقل داده‌های ممکن انجام گرفت به‌گونه‌ای که مناسب‌ترین سری داده‌های ورودی جهت تعلیم بهترین مدل تعیین گردد. بدین‌منظور از الگوریتم یادگیری محلی L.L که از کارایی قابل توجهی در شبیه‌سازی پارامترهای متغیر با داده‌های ورودی کم برخوردار است (9) استفاده گردید. همچنین دبی‌های ثبت شده گذشته به‌عنوان داده ورودی در مطالعه حاضر در نظر گرفته شد.

می‌شود (9). در چنین روش‌های یادگیری پس از استخراج رابطه، دیگر به داده‌های آموزشی نیازی نبوده و تنها رابطه استخراج شده برای پاسخ‌گویی نگهداری می‌شود. ولی در روش‌های یادگیری محلی، پیش‌گویی به صورت محلی و بر اساس بخش کوچکی از داده‌های آموزشی انجام می‌شود در این روش‌های یادگیری هیچ رابطه خاصی از بین داده‌ها استخراج نمی‌گردد و برای پیش‌گویی همواره به داده‌های آموزشی نیازمند است و این داده‌ها به صورت مداوم و همیشگی نگهداری می‌شوند. در یادگیری‌های محلی اگر داده‌های ورودی از مجموعه داده‌های آموزشی در فاصله دورتری قرار داشته باشند ممکن است پاسخ به‌دست آمده با مقدار واقعی آن بسیار متفاوت باشد ولی با توجه به ماهیت متغیر و نوسانی داده‌های رواناب امکان استخراج یک رابطه دقیق امکان‌پذیر نخواهد بود. ولی شاید بتوان بر اساس روش‌های یادگیری محلی از روی موقعیت‌های قبلی متغیر و براساس ویژگی خاص این شبکه‌ها موقعیت بعدی متغیر را پیش‌گویی کرد.

روش‌های یادگیری سراسری مبتنی بر شبکه‌های عصبی، در سال‌های اخیر به‌طور چشم‌گیری در مطالعات و پیش‌بینی مؤلفه‌های هیدرولوژیک مورد استفاده قرار گرفته‌اند (14 و 19). ولی معمولاً شبکه‌های عصبی بسیار به داده‌های آموزشی وابسته است. لذا مقادیر پیش‌بینی شده برای داده‌هایی که مشابه و یا نزدیک به داده‌های آموزشی می‌باشند خطای کم‌تری نشان می‌دهد ولی برای داده‌هایی که از داده‌های آموزشی فاصله بیشتری دارند ممکن است نتایج خوبی به‌دست نیاید. بنابراین، با توجه به ماهیت متغیر و نوسانی مقادیر رواناب که از الگوی خاصی تبعیت نمی‌کنند پیش‌بینی دقیق مقادیر روزانه رواناب با استفاده از روابط و الگوریتم‌های یادگیری سراسری ممکن است همواره با عدم قطعیت‌هایی همراه باشد. لذا در این مطالعه تلاش گردید یک الگوریتم محلی جهت پیش‌بینی مقادیر روزانه رواناب ارائه گردد.

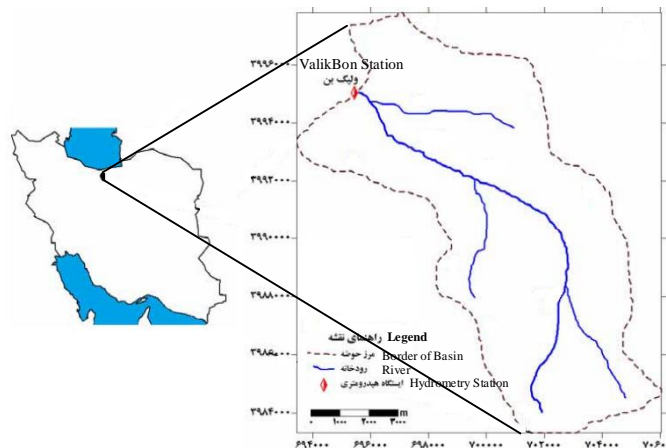
الگوریتم یادگیری کندرو<sup>1</sup> (L.L)، یک رویکرد مدل‌سازی محلی است و ویژگی اصلی آن استفاده از نزدیک‌ترین همسایگی‌های ممکن به نقطه مورد درخواست از بین داده‌های تعلیمی جهت ایجاد یک مدل محلی بوده و به محض دریافت یک درخواست جدید مدل محلی جدیدی بر اساس زیر مجموعه‌ی دیگری از نزدیک‌ترین همسایگی‌ها به نقطه مورد درخواست جدید، ایجاد می‌شود. تحقیقات نشان می‌دهند که این رویکرد، عملکرد بهتری نسبت به سایر روش‌ها از قبیل روابط رگرسیونی و شبکه‌های عصبی در بسیاری از مطالعات مربوط به پیش‌بینی سری‌های زمانی داشته (4) و علاوه بر آن رتبه دوم (از بین 17 رویکرد مدل‌سازی و پیش‌بینی کننده) در رقابت بین‌المللی در تخمین سری‌های زمانی را به خود اختصاص داده است (3 و 5). این

## مواد و روش‌ها

## منطقه مورد مطالعه

عرصه انجام تحقیق، حوزه آبخیز کسلیان می‌باشد که با مساحتی بالغ بر 66/75 کیلومتر مربع در موقعیت جغرافیایی بین  $35^{\circ}58'30''$  تا  $36^{\circ}7'15''$  عرض شمالی و  $53^{\circ}08'44''$  تا  $53^{\circ}15'42''$  طول شرقی واقع شده است. شکل 1 موقعیت جغرافیایی حوزه را نشان می‌دهد. آب رودخانه اصلی حوزه از جنوب به سمت شمال غربی حوزه جریان دارد و در محلی به نام ولیک بن از حوزه

خارج و به رودخانه تالار می‌ریزد. ایستگاه هیدرومتری حوزه در ارتفاع 1200 متر از سطح دریا و در نزدیکی روستای ولیک بن قرار دارد و مجهز به پارشال فلوم بتنی، اشل ولیمنوگراف است که می‌تواند دبی تا حدود 20 متر مکعب بر ثانیه را اندازه‌گیری نماید (شکل 1). در پژوهش حاضر، با توجه به کیفیت داده‌های موجود، از داده‌های رواناب روزانه ده ساله مربوط به سال‌های 1381 تا 1391 استفاده گردید. بر اساس آمار موجود، متوسط آبدهی سالانه رودخانه برابر با 0/441 متر مکعب بر ثانیه محاسبه گردید.



شکل 1- موقعیت جغرافیایی حوزه آبخیز کسلیان و ایستگاه ولیک بن  
Figure 1- Geographic location of Kasilian Basin and Valicbon station

لازم به ذکر است که در تعریف سری‌های زمانی فوق، آن مجموعه‌هایی جهت ارائه نتایج انتخاب گردیدند که نتایج حاصل از شبیه‌سازی دبی آن‌ها با نتایج حاصل از شبیه‌سازی دبی مجموعه‌های دیگر اختلاف معنی‌داری نداشته‌اند.

جهت بررسی بهترین مدل تعلیمی، از جذر میانگین مربعات خطا و درصد میانگین مطلق خطا استفاده گردیده و نتایج به‌طور خلاصه در جدولی گردآوری گردید. لازم به ذکر است که در این مطالعه دبی متوسط روزانه 29 شهریور 1391 به عنوان نقطه مورد درخواست در نظر گرفته شد.

## الگوریتم یادگیری کندرو

یادگیری کندرو یک تکنیک مبتنی بر حافظه می‌باشد که به محض دریافت یک درخواست بر اساس فاصله نقطه درخواست با همسایه‌های اطراف آن نسبت به پیش‌گویی مقدار مناسب اقدام می‌کند. در این روش یادگیری، انتخاب تعداد همسایه‌هایی که در محاسبات دخیل هستند به صورت بهینه انجام می‌گیرد. لازم به ذکر است که پس از انتخاب اولیه از همسایه‌های نقطه مورد درخواست و

## روش تحقیق

در تحقیق حاضر مجموعه 720 داده تعلیمی مربوط به دبی روزانه 2 ساله (سال 1390 و 1391) در ایستگاه ولیک بن حوزه کسلیان به‌عنوان داده‌های در دسترس انتخاب شدند. جهت پیش‌بینی مقادیر روزانه دبی در این پژوهش از الگوریتم یادگیری محلی کندرو استفاده گردید. بدین‌منظور کدهای پیاده‌سازی هر یک از ابزارهای یادگیری کندرو به زبان C نوشته شده و توابع در نرم افزار متلب فراخوانی گردید (<http://iridia.ulb.ac.be/~lazy>).

برای دستیابی به بهترین مدل محلی، حالت‌های مختلف و ممکن از الگوهای مختلف تعلیمی، تعریف و وارد مدل گردید. سپس تمام الگوهای تعلیمی با سه مدل ثابت conLL.c، خطی linLL.c و درجه دو quall.c مورد بررسی قرار گرفتند و در تست هر الگو دامنه همسایگی‌های مورد محاسبه در برنامه تغییر داده شده تا در نهایت بهترین الگوی محاسبه همسایگی مشخص گردید. سپس جهت یافتن بهترین سری داده‌های ورودی، مدل‌های مختلف به‌ترتیب با داده‌های ورودی 6، 8، 10، 15، 20 روز قبل، یک و دو ماه قبل، یک، دو و سه فصل قبل و در نهایت یک و دو سال قبل مورد بررسی قرار گرفت.

درخواست به‌صورت افزایش تعداد آن‌ها انجام می‌دهد و در هر مرحله یک مقدار برای نقطه مورد درخواست محاسبه می‌کند. تابع چند جمله‌ای درجه یک linLL.c و دو QUAL.L.C برای هر درخواست محاسبات خود را با استفاده از همسایگان اطراف نقطه درخواست به صورت افزایش تعداد آن‌ها به‌ترتیب در تابع درجه یک و تابع درجه دو انجام می‌دهد و در هر مرحله یک مقدار برای نقطه مورد درخواست محاسبه می‌کند. فرمت کلی فراخوانی این توابع به صورت زیر است:

>>[h,s,k]= conll/linll/quaLL(X,Y,Q,id-par);  
 که در آن X[n,m] ماتریس داده‌های آموزشی ورودی، Y[n,1] بردار خروجی برای داده‌های آموزشی، Q[q,m] ماتریس نقاط مورد درخواست و [id\_par,2,1] بردار حداقل و حداکثر همسایگی برای نقاط مورد درخواست می‌باشد. هم‌چنین، h[q,1] نمایان‌گر مقدار پیشگویی شده برای نقطه مورد درخواست q، S[q,1] مقدار خطا برای پیشگویی انجام شده و [k[q,1] تعداد همسایگی استفاده شده است. توابع و روابط ریاضی مدل‌های محلی ثابت، خطی و درجه دو استفاده شده در این پژوهش در منابع به تفصیل معرفی شده‌اند (Birattari و Bontempi, 1999).

### انتخاب مدل بهینه

سپس جهت انتخاب بهترین مدل تخمین‌گر از بین سه مدل استفاده شده در این مقاله و هم‌چنین انتخاب بهترین بازه‌ی زمانی جهت تخمین از شاخص خطای ریشه میانگین مجذور خطا (RMSE) و میانگین قدر مطلق خطا (MAE) مطابق رابطه 2 و 3 استفاده گردید.

$$RMSE = \sqrt{\frac{\sum_{t=1}^m (y_t - \hat{y}_t)^2}{m}} \quad (2)$$

$$MAE = \frac{\sum_{t=1}^m |y_t - \hat{y}_t|}{m} \quad (3)$$

به‌طوری که در آن  $y_t$  مقادیر واقعی و  $\hat{y}_t$  مقادیر پیش‌بینی و m طول دوره پیش‌بینی است که از T + 1 تا T + m می‌باشد. در نهایت پس از تولید کلیه خروجی‌ها، نتایج به‌دست آمده و مقادیر خطای محاسبه شده به‌طور خلاصه در جدولی ارائه گردید.

### نتایج و بحث

این تحقیق با هدف بررسی شبیه‌سازی دبی جریان یک روز آینده با کمک مجموعه داده‌های دبی روزهای گذشته در رودخانه منتهی به رود تالار واقع در حوزه آبخیز کسلیان با استفاده از مدل یادگیری محلی L.L. انجام شد. شکل 2 تغییرات روزانه‌ی دبی را طی دو سال گذشته نسبت به روز مورد نظر (29 شهریور 1391) در ایستگاه ولیک‌بن حوزه آبخیز کسلیان نشان می‌دهد. از آن‌جایی که مهم‌ترین مرحله شبیه‌سازی با استفاده از مدل مذکور، تعیین ورودی‌های مناسب

انجام عمل پیش‌گویی، در هر مرحله به تعداد همسایه‌های انتخاب شده یک همسایه دیگر اضافه شده و مجدداً محاسبات جهت پیش‌گویی از سر آغاز می‌شود. عمل انتخاب همسایه جدید به همان ترتیب آنقدر تکرار می‌شود تا به یک نقطه بهینه دست یابد و در نهایت با استفاده از دو الگوریتم ثابت و خطی یا درجه دو پیش‌گویی نهایی انجام می‌گیرد (9). توابع و روابط ریاضی مربوط به مدل‌های مذکور استفاده شده در این پژوهش در منابع به تفصیل معرفی شده‌اند (2). نکته قابل تعمق در روش‌های یادگیری محلی و الگوریتم مذکور این است که به داده‌های همسایگی بسیار حساس بوده و فقط با داده‌های ورودی نزدیک پیش‌بینی خوبی ارائه می‌کند. لازم به ذکر است که در مرحله کدنویسی، به‌منظور پیش‌گیری از تکرار ضرب تعداد زیادی از ماتریس‌ها در هر مرحله، از روش بازگشتی استفاده گردیده و توابع مربوط در برنامه اعمال گردیدند تا از محاسبات مرحله قبل برای محاسبه مرحله بعد استفاده گردد. بنابراین، برای هر مقدار درخواستی مورد نظر مجموعه‌ای از مدل‌های محلی با درجه تابع و تعداد همسایگی‌های مختلف تعریف می‌شود.

برای هر یک از بازه‌های تعریف شده در تحقیق، چندین مدل محلی با هر سه تابع مذکور بر اساس حداقل و حداکثر همسایه‌های ممکن ایجاد گردید. به این صورت که در هر مرتبه از مدل‌سازی مجموعه مشخصی از همسایگی‌ها به دو زیر مجموعه تعلیمی و اعتبارسنجی تقسیم و پارامترها بر اساس مجموعه تعلیمی تخمین زده شد. سپس مدل‌های کاندید شده با استفاده از مقدار خطای تخمین، بر اساس مجموعه‌ی اعتبارسنجی با استفاده از روش آماری PRESS (15)، طبق رابطه زیر بر اساس ماتریس ورودی ( $X_j$ ) و خروجی ( $y_i$ ) و پارامتر تخمین‌زده شده ( $\beta$ ) مورد ارزیابی قرار گرفتند.

$$e_i^{cv} = y_i - x_i^T \beta_i \quad (1)$$

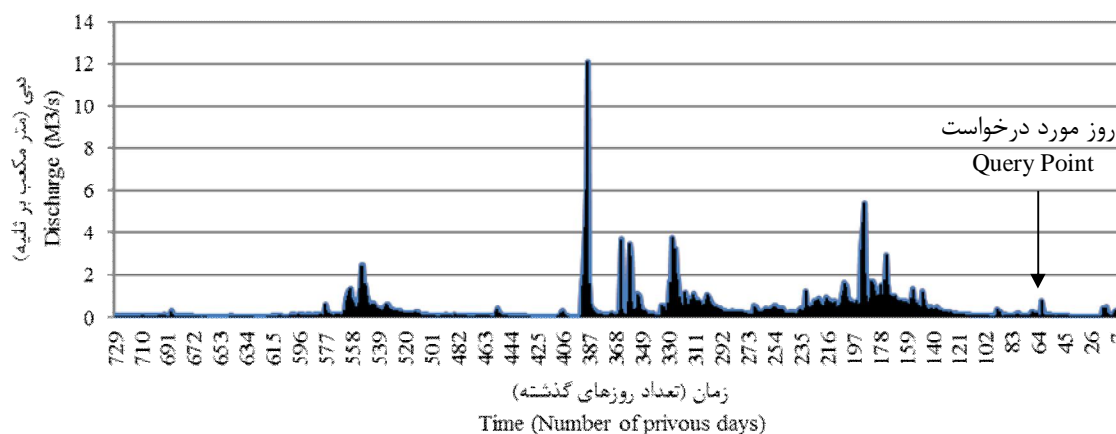
در این روش مقدار خطا با استفاده از رابطه بازگشتی از روی مقادیر قبلی به صورت بازگشتی محاسبه می‌گردد (9). از این‌رو وقتی در هر مرحله یک نقطه جدید به مجموعه‌ی نقاط آموزشی قبلی اضافه می‌شود میزان خطا مجدداً محاسبه گردیده تا در صورت افزایش خطا نقطه اضافه شده به‌عنوان همسایگی در نظر گرفته نشود و عملیات محاسبه مقدار ( $y_i$ ) در همان مرحله قبلی متوقف گردد. در نهایت بر اساس متوسط کمترین خطای نزدیک‌ترین همسایه، مدل محلی مناسب انتخاب شد.

### مدل محلی ثابت، خطی و درجه دو

یکی از مدل‌های کارآمد الگوریتم L.L.، مدل محلی ثابت (conLL.c) می‌باشد. تابع چند جمله‌ای درجه صفر conLL.c برای درخواست محاسبات خود را با استفاده از همسایگان اطراف نقطه

شده و نتایج حاصل از آنالیز آماری هر مجموعه به صورت جداگانه در جدول 1 خلاصه گردیده است.

برای مدل بود لذا، سری‌های مختلف زمانی از داده‌های روزانه گذشته تعریف و مورد بررسی قرار گرفت. سری‌های مختلف زمانی تعریف



شکل 2- تغییرات روزانه دبی را در ایستگاه ولیک‌بن حوزه آبخیز کسلیان طی دو سال 1391 و 1390  
Figure 2- Daily discharge variation in Valicbon Station of Kasilian Basin during 2011 and 2012

جدول 1- پارامترهای آماری داده‌های دبی روزانه ایستگاه ولیک‌بن حوزه آبخیز کسلیان طی دو سال 1391 و 1390

Table 1- Statistical parameter of daily discharge in Valicbon station of Kasilian Basin during 2011 and 2012

پارامترهای آماری Statistical parameter	روز Day						فصل Season					سال Year	
	6	8	10	15	20	30	45	60	1	2	3	1	2
حداقل Minimum	0.12 1	0.121	0.121	0.121	0.054	0.054	0.054	0.054	0.054	0.054	0.054	0.054	0.033
حداکثر Maximum	0.32	0.34	0.34	0.34	0.49	0.49	0.49	0.49	0.798	2.97	5.43	5.43	12.1
میانگین Mean	0.22	0.25	0.26	0.23	0.24	0.18	0.15	0.14	0.16	0.37	0.48	0.54	0.41
میانه Median	0.21	0.30	0.30	0.23	0.23	0.15	0.07	0.21	0.21	0.18	0.30	0.34	0.17
چولگی Skewness	- 0.05	-0.77	-0.85	0.07	0.52	0.89	1.54	1.81	2.81	2.43	4.28	3.64	7.82
کشیدگی Kurtosis	- 1.91	-1.11	10.44	-1.16	-0.51	-0.19	1.53	3.01	11.30	8.03	27.27	18.35	96.85
ضریب تغییرات Variationcoefficient	0.41	0.37	0.34	0.35	0.54	0.74	0.81	0.72	0.72	1.16	1.22	1.18	1.82

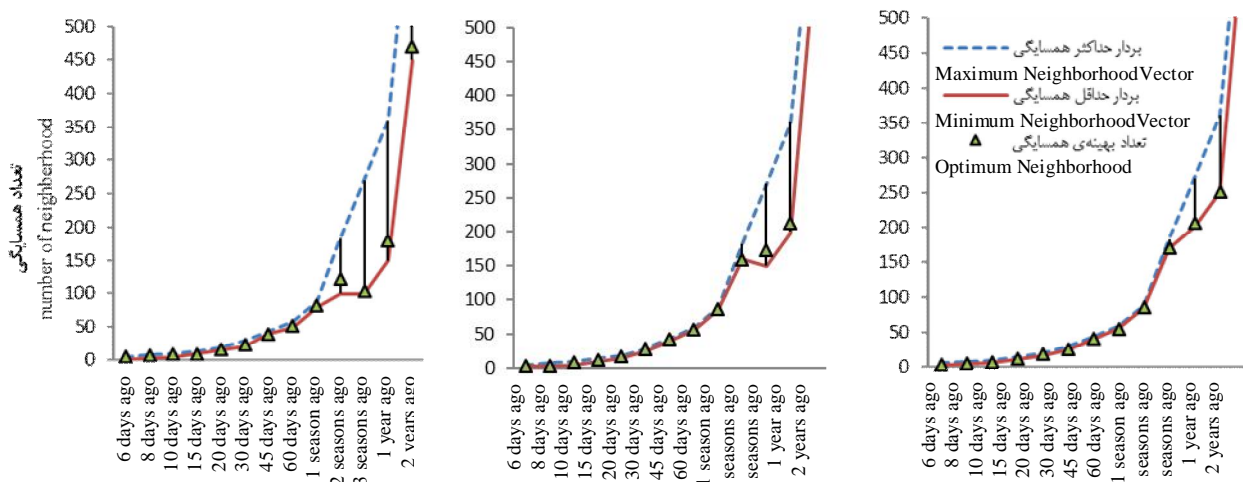
همان‌طور که اشاره شد، الگوریتم L.L. برای هر نقطه مورد درخواست مجموعه‌ای از مدل‌های محلی تعریف می‌کند که شامل پارامترها و تعداد همسایگی‌های مختلف می‌باشند. با توجه به جداول 2 تا 4 مشاهده می‌شود که برای هر بازه زمانی تعریف شده در هر سه مدل مذکور، دامنه‌ای از حداقل و حداکثر همسایگی‌های ممکن تعیین گردید. با تغییر دامنه‌ی مدل در روش یادگیری، سایز مدل و به تبع آن تعداد دفعات تعلیم برای داده‌های ورودی داده شده تغییر نمود.

همان‌طور که پیش‌تر به تفصیل توضیح داده شد، این روش بسیار به همسایگی‌ها و داده‌های ورودی حساس بوده و لذا در هر مدل‌سازی تعیین دامنه‌ی ورودی‌های مدل بسیار مهم می‌باشد. در این مطالعه، به منظور دست‌یابی به بهترین تعداد داده‌های ورودی جهت شبیه‌سازی دبی روزانه، هر یک از سری‌های زمانی تعریف شده به عنوان ماتریس ورودی وارد تابع گردید. نتایج حاصل از شبیه‌سازی دبی روز مورد نظر در این پژوهش با حالت‌های مختلف ماتریس ورودی، با استفاده از سه تابع linLL.c، conLL.c و quall.c از الگوریتم L.L. در جدول 2 تا 4 گردآوری شد.

جدول 2- شبیه‌سازی دبی روزانه مدل conLL.c با استفاده از مجموعه داده‌های ورودی مختلف در ایستگاه ولیک‌بن (90-91)

Table 2- Simulating of daily discharge using conLL.c model by input data set in Valicbon station (2011- 2012)

داده‌های ورودی Input data	تعداد داده‌های ورودی Number of input data	بردار حداکثر همسایگی Max Neighborhood Vector	بردار حداقل همسایگی MinNeighborhood Vector	تعداد بهینه‌ی همسایگی Optimum Neighborhood	مقدار تخمین Estimated value ( $M^3S^{-1}$ )	RMSE	MAE (%)
6 روز قبل 6 days ago	5	5	4	5	0.26	0.139	1.93
8 روز قبل 8 days ago	7	7	4	7	0.27	0.149	2.22
10 روز قبل 10 days ago	9	9	5	9	0.26	0.139	1.93
15 روز قبل 15 days ago	14	14	9	9	0.21	0.089	0.792
20 روز قبل 20 days ago	19	19	15	16	0.21	0.089	0.792
30 روز قبل 30 days ago	29	29	20	23	0.14	0.017	0.028
45 روز قبل 45 days ago	44	44	40	40	0.13	0.005	0.002
60 روز قبل 60 days ago	59	59	50	52	0.12	0.001	0.0001
یک فصل قبل 1 season ago	89	89	80	82	0.14	0.020	0.040
دو فصل قبل 2 seasons ago	183	183	100	122	0.15	0.029	0.084
سه فصل قبل 3 seasons ago	271	271	100	104	0.17	0.053	0.280
یک سال قبل 1 year ago	361	361	150	180	0.19	0.069	0.474
دو سال قبل 2 years ago	725	725	450	471	0.14	0.020	0.040



شکل 3- دامنه‌ی مدل‌سازی و نزدیک‌ترین همسایه‌ها به ترتیب در توابع conLL.c، LinLL.c و QuaLL.c (از چپ به راست)

Figure 3- Range of modeling and nearest neighborhood in conLL.c, LinLL.c and QuaLL.c functions respectively (left to right)

جدول 3- شبیه‌سازی دبی روزانه مدل یادگیری linLL.c با استفاده از مجموعه داده‌های ورودی مختلف در ایستگاه ولیک‌بن (90-91)

Table 3- Simulating of daily discharge using LinLL.c model by input data set in Valicbon station (2011- 2012)

داده‌های ورودی Input data	تعداد داده‌های ورودی Number of input data	بردار حداکثر همسایگی Max Neighborhood Vector	بردار حداقل همسایگی MinNeighborhood Vector	تعداد بهینه‌ی همسایگی Optimum Neighborhood	مقدار تخمین Estimated value (M <sup>3</sup> S <sup>-1</sup> )	RMSE	MAE (%)
6 روز قبل 6 days ago	5	5	4	4	0.363	0.242	5/856
8 روز قبل 8 days ago	7	7	4	4	0.298	0.177	3.133
10 روز قبل 10 days ago	9	9	5	9	0.234	0.113	1.276
15 روز قبل 15 days ago	14	14	10	12	0.201	0.080	0.640
20 روز قبل 20 days ago	19	19	15	18	0.211	0.090	0.810
30 روز قبل 30 days ago	29	29	25	28	0.190	0.069	0.476
45 روز قبل 45 days ago	44	44	40	43	0.182	0.061	0.372
60 روز قبل 60 days ago	59	59	55	57	0.178	0.057	0.324
یک فصل قبل 1 season ago	89	89	85	87	0.181	0.060	0.340
دو فصل قبل 2 seasons ago	183	183	160	160	0.184	0.063	0.397
سه فصل قبل 3 seasons ago	271	271	150	173	0.187	0.066	0.435
یک سال قبل 1 year ago	361	361	200	212	0.191	0.070	0.490
دو سال قبل 2 years ago	725	725	600	604	0.192	0.071	0.504

اظهارات بنتمپی (1999) کل مجموعه‌ی داده‌ها در این یادگیری محلی همواره در حافظه نگهداری شده و بر خلاف سایر الگوریتم‌های یادگیری بعد از تعلیم کنار گذاشته نمی‌شوند (4). لذا، فرآیند یادگیری محلی شامل محاسبات ریاضی پیچیده و قابل توجهی است، ولی علی‌رغم این محاسبات زیاد و پیچیده، سرعت مدل‌سازی به‌طور قابل توجهی بالا بوده به‌گونه‌ای که مجموعه‌ی شبیه‌سازی‌های نزدیک به 700 داده در عرض چند ثانیه انجام می‌گیرند. کرانی (2004) در مطالعات خود به‌منظور شبیه‌سازی آلاینده‌های هوا با استفاده از حدود 1000 داده، ادعای فوق را تأیید می‌نماید. شکل 4 نمودار تغییرات نزدیک‌ترین همسایگی را در سه مدل conLL.c، LinLL.c و quaLL.c در بازه‌های زمانی مختلف نشان می‌دهد. با توجه به شکل مشاهده می‌شود که مدل conLL.c با استفاده از تعداد کمتری از نزدیک‌ترین همسایگی‌های موجود موفق به شبیه‌سازی مقدار خواسته شده گردیده است.

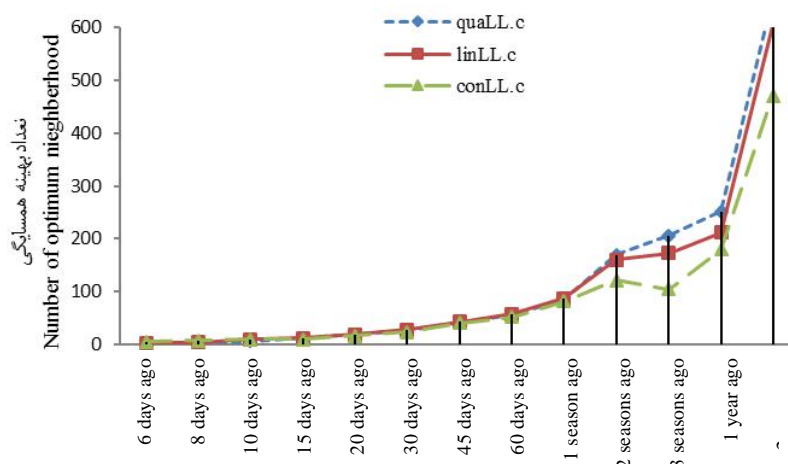
بدین معنی که هر تابع مدل‌سازی در یک دامنه‌ی مشخص، مجموعه‌ای از بهترین مدل‌ها را با درجه توابع و همسایگی‌های مختلف ساخته و همسایگی‌ها بر اساس مقدار فاصله خود نسبت به نقطه مورد درخواست رتبه‌بندی گردیدند. کیشی در سال 2004 ارتباط دامنه‌ی مدل‌های یادگیری را با تعداد دفعات تعلیم تأیید نموده است (13).

شکل 3 بهترین دامنه‌ی مدل‌سازی را برای هر یک از توابع به‌صورت جداگانه نشان می‌دهد. سپس، با استفاده از فرآیند اعتبارسنجی بر اساس متوسط کم‌ترین خطای تعریف شده در دامنه‌ی انتخاب‌شده (شکل 3) نزدیک‌ترین همسایگی‌های ممکن (KNN) به‌عنوان تعداد بهینه‌ی همسایگی مشخص شده و بر اساس آن بهترین مدل برای پیش‌بینی درخواست مورد نظر ساخته شد. همان‌طور که از شکل 3 بر می‌آید، هنگامی که بازه زمانی برای نقطه مورد درخواست تغییر کرده فاصله همسایگی‌ها مورد ارزیابی مجدد قرار گرفته‌اند. بنابراین طبق

جدول 4- شبیه‌سازی دبی روزانه مدل یادگیری quall.c با کمک مجموعه داده‌های ورودی مختلف در ایستگاه ولیک‌بن (91-90)

Table 4- Simulating of daily discharge using quall.c model by input data set in Valicbon station (2011– 2012)

داده‌های ورودی Input data	تعداد داده‌های ورودی Number of input data	بردار حداکثر همسایگی Max Neighborhood Vector	بردار حداقل همسایگی MinNeighbor hood Vector	تعداد بهینه‌ی همسایگی Optimum Neighborhood	مقدار تخمین Estimated value ( $M^3S^{-1}$ )	RMS E	MAE (%)
6 روز قبل 6 days ago	5	5	4	4	0.362	0.241	5.808
8 روز قبل 8 days ago	7	7	4	4	0.275	0.154	2.372
10 روز قبل 10 days ago	9	9	5	6	0.297	0.174	3.028
15 روز قبل 15 days ago	14	14	10	12	0.203	0.082	0.672
20 روز قبل 20 days ago	19	19	15	18	0.201	0.080	0.64
30 روز قبل 30 days ago	29	29	25	26	0.234	0.113	1.277
45 روز قبل 45 days ago	44	44	40	41	0.211	0.090	0.81
60 روز قبل 60 days ago	59	59	55	55	0.187	0.066	0.434
یک فصل قبل 1 season ago	89	89	85	85	0.192	0.071	0.504
دو فصل قبل 2 seasons ago	183	183	170	170	0.180	0.059	0.348
سه فصل قبل 3 seasons ago	271	271	200	206	0.182	0.064	0.410
یک سال قبل 1 year ago	361	361	250	252	0.188	0.067	0.449
دو سال قبل 2 years ago	725	725	650	657	0.197	0.076	0.578



شکل 4- نمودار تغییرات نزدیک‌ترین همسایگی در سه مدل conLL.c، LinLL.c و quaLL.c در بازه‌های زمانی مختلف  
Figure 4- Variation of nearest neighborhood in conLL.c, LinLL.c and QuaLL.c functions in different periods

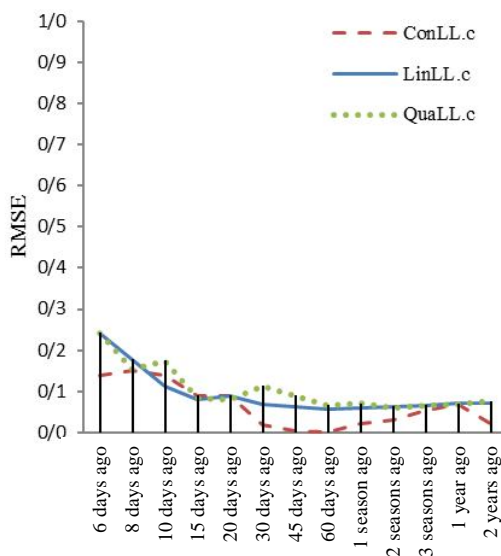


افزایش قابل توجهی یافته است. از طرفی نتایج شبیه‌سازی‌ها حاکی از آن است که با افزایش بازه‌ی زمانی از 6 روز به 60 روز، مقدار خطای شبیه‌سازی در هر سه مدل روند کاهشی داشته است. هاگواتو میتی (2012) در مطالعات خود با مدل‌های یادگیری سراسری به نتایج مشابهی دست یافته‌اند. بنابراین می‌توان نتیجه گرفت که هر چه میزان تغییرات در بازه‌ی زمانی افزایش یابد، الگوریتم یادگیری الگوی تغییرات را به‌منظور ساختن یک مدل کلی بهتر دریافت کرده و در تعقیب نقطه مورد درخواست بهتر عمل می‌کند. لذا با توجه به یافته‌های موجود و نتایج ارائه شده توسط سایر محققین (1 و 11) پیش‌بینی می‌شود که جهت پیش‌بینی دبی روزانه در دوره کم‌آبی که نوسان‌ها و تغییرات جریان در رودخانه به نسبت دوره‌ی پرآبی کمتر است، بازه‌ی زمانی بزرگ‌تری جهت انتشار مناسب موقعیت نقاط شبکه و به تبع آن درک بهتر تغییرات احتیاج است.

این مطالعه با هدف بررسی و مقایسه‌ی کارایی سه مدل از الگوریتم یادگیری کندرو در بازه‌های مختلف زمانی انجام گرفت. نتایج حاصل از این پژوهش نشان داد که مدل conLL.c پیش‌بینی دقیق‌تری را نسبت به دو مدل ارائه شده دیگر در دوره آماری کم‌آبی و نسبتاً پایدار ارائه می‌دهد (شکل 5). هم‌چنین با دقت در نتایج گزارش شده در جدول 2 مشاهده می‌شود که با افزایش بازه‌های زمانی و به بیان دیگر با افزایش تعداد همسایگی‌ها مقادیر تخمین به مقدار واقعی نزدیک‌تر شده به‌گونه‌ای که در بازه‌ی زمانی 60 روز به بهترین تخمین رسیده است. اما بعد از آن با افزایش داده‌های ورودی بیشتر، دقت تخمین کمتر می‌شود. به‌طوری که مقدار خطا از 0/001 در بازه‌ی زمانی دو ماه قبل به 0/020، 0/029 و 0/053 به ترتیب در بازه‌های زمانی یک، دو و سه فصل گذشته رسیده است. بنابراین نتایج به‌دست آمده حاکی از این واقعیت‌اند که لزوماً داده‌های ورودی بیشتر، نتایج بهتری را در بر نداشته و حتی ممکن است بر درک درست تغییرات توسط مدل اثر منفی داشته باشند. به‌طور کلی از نتایج به‌دست آمده استنباط می‌شود تعداد داده‌های ورودی در مدل‌های هوشمند بر دقت و کارایی مدل اثر می‌گذارند. در این پژوهش با افزایش تعداد داده‌های دبی روزانه به‌عنوان داده‌های ورودی مدل، دقت تخمین افزایش یافت و این افزایش دقت تا دو ماه روند صعودی داشته و سپس به آهستگی کاهش یافت. دلیل این امر را می‌توان چنین توجیه کرد که دبی روزانه مورد نظر که در انتهای ماه شهریور فصل تابستان در نظر گرفته شده بود تا حد زیادی تحت تأثیر دبی رودخانه در همان فصل قرار داشته است. به‌طوری که دبی‌های روزانه ماه‌های شهریور و مرداد به‌عنوان داده‌های ورودی پیوستگی خوبی با نقطه مورد تقاضا داشتند.

با مقایسه مقادیر پیش‌گویی شده برای دبی روزانه با استفاده از سه روش مورد بررسی در این پژوهش و مقادیر خطای محاسبه شده در جداول 2 تا 4، مشاهده گردید که نتایج به‌دست آمده از مدل conLL.c با داده‌های دبی روزانه‌ی 60 روز گذشته با حداقل خطای 0/001 به مقدار واقعی نزدیک‌تر بوده، هم‌چنین نتایج نشان داد که مدل linLL.c نیز با حداقل خطای 0/057 در بازه‌ی زمانی 60 روز بهترین تخمین را دارا بود. در حالی که مدل محلی درجه دو quaLL.c با حداقل خطای 0/059 بهترین مقدار را با استفاده از داده‌های دبی روزانه‌ی دو فصل گذشته تخمین زد. نتایج ارائه شده در جدول 4 نشان می‌دهد که زمانی که تعداد داده‌های آموزشی افزایش می‌یابد مقادیر پیش‌گویی شده با کمک مدل quaLL.c بهبود می‌یابد. چراکه این مدل برای به‌دست آوردن نتایج بهتر به نمونه‌های بیش‌تر و به تبع آن به محاسبات بیش‌تری نیاز دارد. هم‌چنین با دقت در شکل 5 مشاهده می‌شود که جذر میانگین مربعات خطا در روش یادگیری محلی linLL.c نسبت به دو روش دیگر در بازه‌های زمانی مختلف نوسانات کم‌تری دارد که نشان‌دهنده‌ی حساسیت کم این مدل به همسایگی‌ها می‌باشد. بنابراین بروز خطاهای جزئی در روند شبیه‌سازی با این مدل تأثیر زیادی در نتایج پیش‌گویی شده نداشته و لذا کمتر دچار مشکل تجمیع خطا می‌گردد. از طرفی بر خلاف مدل linLL.c، جذر میانگین مربعات خطا در مدل quaLL.c دارای بیش‌ترین نوسان در بازه‌های مختلف زمانی بوده و بنابراین حساسیت بالای این مدل را نسبت به همسایگی‌ها بیان می‌کند. به این مفهوم که با اندک خطایی که در پیش‌گویی‌های اولیه انجام می‌پذیرد، تخمین‌های بعدی سریعاً دچار خطا شده و کم‌کم از مقادیر واقعی دور می‌شوند. جعفری و خانتیموری (2011) در یافته‌های خود ادعای فوق را تأیید می‌کنند (10). هم‌چنین مشاهده می‌شود که در بازه‌های اولیه زمانی مقدار خطای تخمین توسط دو مدل linLL.c و quaLL.c بسیار بیش‌تر از مقدار خطای تخمین توسط مدل conLL.c می‌باشد که گویای این حقیقت است که دو مدل مذکور به تعداد همسایگی نیز حساس بوده و کفایت داده‌ها در دقت تخمین توسط این دو مدل نقش اساسی ایفا می‌نماید. شکل 5 تغییرات جذر میانگین مربعات خطا را در سه مدل conLL.c، LinLL.c و quaLL.c در بازه‌های مختلف زمانی نشان می‌دهد. به‌طور کلی با توجه به نتایج ارائه شده، هر چند دقت روش یادگیری L.L برای مدل‌های مختلف متفاوت است اما میزان متوسط خطا در این روش یادگیری پایین است که با نتایج ارائه شده توسط جعفری و خانتیموری (2011) مطابقت دارد.

از طرف دیگر، با دقت در نتایج ارائه شده در جدول 1 مشاهده می‌شود که با افزایش افق زمانی موردنظر، دامنه‌ی تغییرات دبی



شکل 5- نمودار تغییرات جذر میانگین مربعات خطا در سه مدل conLL.c، LinLL.c و QuaLL.c در بازه‌های مختلف زمانی  
Figure 5- Variation of root-mean square error in conLL.c, LinLL.c and QuaLL.c functions in different periods

انتخاب گردید. بدین‌صورت که حداقل داده‌های ورودی مورد نیاز در مدل یادگیری کندرو 6 داده می‌باشد لذا از تعداد شش داده ورودی (دبی روزانه شش روز قبل) آغاز و سپس در هر بار اجرای مدل یک داده به داده‌های ورودی در مرحله قبل اضافه گردید. این روند تا 2 سال ادامه پیدا کرد. به دلیل وسعت نتایج و تعدد مدل‌سازی، در این مقاله تنها به ارائه سری‌های زمانی که نتایج حاصل از مدل‌سازی آن‌ها با حالت قبل اختلاف معنی‌داری داشتند بسنده گردید. بنابراین تنها حالت‌های 6، 8، 10، 15، 20 روز قبل، یک و دو ماه قبل، یک، دو و سه فصل قبل و در نهایت یک و دو سال جهت مقایسه و ارائه نتایج در مقاله حاضر انتخاب شدند.

نتایج به‌دست آمده از چندین هزار مدل تعلیمی تهیه شده نشان داد که مدل conLL.c با داده‌های دبی روزانه‌ی 60 روز گذشته با حداقل خطای 0/001 به مقدار واقعی نزدیک‌تر بوده است. با توجه به این‌که نقطه مورد درخواست در این پژوهش، پیش‌بینی دبی متوسط روزانه 29 شهریور 1391 بوده است، بنابراین لازم به ذکر است که نتایج به‌دست آمده و تعداد داده‌های بهینه مورد نیاز جهت شبیه‌سازی برای یک دوره آماری کم‌آبی و نسبتاً پایدار پیش‌بینی شده است. نتایج به‌دست آمده حاکی از آن است که الگوریتم یادگیری محلی L.L علی‌رغم محدودیت‌های موجود از کارایی بالایی در شبیه‌سازی سری‌های زمانی با خطای نسبتاً پایین برخوردار بوده و قادر به مدل‌سازی با استفاده از حداقل 5 داده آموزشی می‌باشد. اگرچه پر واضح است که با افزایش تعداد داده‌های آموزشی مقدار پیش‌گویی شده بهبود یافته و به مقدار واقعی نزدیک‌تر می‌شود ولی به دلیل

در حالی که با وارد کردن داده‌های دبی روزانه مربوط به تیرماه کم پیوستگی بین داده‌ها و بنابراین دقت مدل کاهش یافت. دلیل این امر را می‌توان به تحت تاثیر قرار داشتن شرایط آب و هوایی تیرماه با فصل قبل نسبت داد. لذا ورود داده‌های با پیوستگی پایین با نقطه مورد نظر بی‌ربطی را در مدل افزایش داده است. به بیان دیگر در این پژوهش داده‌های دبی روزانه‌ی دو ماه گذشته در تخمین مناسب دبی روز آینده در یک دوره کم‌آبی با استفاده از مدل یادگیری محلی conLL.c بیش‌ترین تأثیرگذاری را داشته‌اند و ورود داده‌های بیش‌تر، بی‌ربطی را در مدل افزایش داده و منجر به تجمیع خطا و فاصله گرفتن از مقدار واقعی گردد.

## نتیجه‌گیری

از آنجایی‌که دسترسی به داده در مدل‌سازی به عنوان یکی از مشکلات پیش رو برای پژوهشگران به حساب می‌آید، لذا این مطالعه با هدف تخمین دبی روزانه رودخانه‌ی تالار، بر روی تعداد داده‌های مورد نیاز در درک درست تغییرات دبی جریان متمرکز گردید. بدین‌منظور، از مجموعه‌ی دبی‌های روزانه سال 1390 و 1391 ایستگاه ولیک‌بن حوضه کسلیان استفاده گردید و الگوهای تعلیمی مختلف با سری داده‌های ورودی 6، 8، 10، 15، 20 روز قبل، یک و دو ماه قبل، یک، دو و سه فصل قبل و در نهایت یک و دو سال قبل با سه مدل conLL، linLL و quall مورد بررسی قرار گرفتند. دامنه سری‌های زمانی فوق با توجه به کارایی مدل و مقایسه دقت تخمین

می‌توان در کلیه پروتکل‌های مسیریابی که مبتنی بر موقعیت نود می‌باشند مورد استفاده قرار داد. هر چند که کاربرد این روش در شبکه‌های مختلف تأثیرات مختلفی داشته که می‌بایست از جنبه‌های متفاوت مورد بررسی قرار گیرد. همچنین اجرای این مدل در دوره‌های پربابی و مقایسه نتایج آن با نتایج ارائه شده در این پژوهش و استفاده از مدل معرفی شده در شبیه‌سازی دبی‌های اوج پیشنهاد می‌گردد. علاوه بر آن می‌توان در شبکه‌های مختلف نیز کاهش یا افزایش بهره‌وری این روش یادگیری را مورد بررسی و تحلیل قرار داده و مناسب‌ترین محیط‌های عملیاتی را که الگوریتم L.L بهترین کارایی را در آن محیط‌ها داراست معرفی نمود. در بررسی‌های دقیق‌تر تعیین یک حالت بهینه از نظر تعداد و تراکم نقاط همسایگی، به‌گونه‌ای که با در نظر گرفتن انرژی مصرفی، حجم اطلاعات و پیچیدگی محاسباتی مقرون به‌صرفه‌ترین حالت ممکن مطلوب می‌باشد.

### تشکر و قدردانی

از اداره آب منطقه‌ای استان مازندران که در جمع‌آوری داده‌های مورد نیاز در این پژوهش ما را یاری نمودند صمیمانه سپاسگزاریم. همچنین این دستاورد را مدیون پروژه بزرگ IRIDIA هستیم که با اختیار قرار دادن یافته‌های نرم‌افزاری ارزشمند خود دامنه‌ی علم را گسترش می‌دهند.

حساس بودن این روش به همسایگی‌های ورودی به مدل، همواره این موضوع صادق نبوده و در مواردی با افزایش تعداد داده‌های آموزشی که در محاسبه پاسخ دخیل می‌باشند، نتیجه به‌دست آمده از مقدار واقعی فاصله می‌گیرد. از این رو در صورتی که نقطه‌ی مورد درخواست به‌خوبی توسط داده‌های آموزشی پوشش داده شود، مقدار پیش‌گویی به مقدار واقعی بسیار نزدیک می‌شود. از طرفی این روش برای پیش‌گویی مقدار خواسته شده، نیاز به مجموعه داده‌های تعلیمی زیادی ندارد. همچنین، از دیگر نقاط قوت این روش یادگیری، می‌توان به دقت بالا و قابلیت پردازش موازی آن اشاره نمود. به این مفهوم که محاسبه هر نقطه مستقل از نقاط بعدی بوده و همه محاسبات به صورت هم‌زمان و جدا از هم انجام می‌گردند. در کاربرد این الگوریتم یادگیری، یک نکته حائز اهمیت این است که حساسیت مدل بیش از حد افزایش داده نشود چرا که همواره مقادیر پیش‌گویی شده انطباق کاملاً دقیقی با مقدار واقعی نداشته و لذا افزایش بیش از حد حساسیت مدل ممکن است باعث افزایش خطا گردد و مقادیر پیش‌گویی شده در گام‌های بعد، از مقدار واقعی به‌طور قابل توجهی فاصله گیرد. از این رو، در صورت پذیرش مقدار خطای قابل قبول در مدل LinLL.c که نسبت به دو مدل دیگر از حساسیت کم‌تری نسبت به همسایگی برخوردار است می‌توان آن را به‌عنوان یک مدل موفق در تعقیب نقطه مورد درخواست دانست.

با توجه به یافته‌های حاضر در این پژوهش این روش یادگیری را

### منابع

- 1- Bhagwat P., and Maity R. 2012. Multistep-ahead river flow prediction using LS-SVR at daily scale. *Journal of Water Resource and Protection*, 4:528-539.
- 2- Birattari M., and Bontempi G. 1999. Lazy learning VS. Speedy Gonzales: A fast algorithm for recursive identification and recursive validation of local constant models. Technical reports: Iridia. Universite' Libre de Bruxelles. TR/IRIDIA/99-6.
- 3- Birattari M., Bontempi G., and Bersini H. 1998. Lazy learning for iterated time-series prediction. In: J.A.K. Suykens, J. Vandewalle (ed.). *Proceedings of the International Workshop on Advanced Black-Box. Techniques for Nonlinear Modeling*.
- 4- Bontempi G. 1999. Local learning techniques for modeling, prediction and control. Ph.D. thesis. Universite' Libre de Bruxelles, Belgium.
- 5- Bontempi G., Birattari M., and Bersini H. 1999. Lazy learning for local modeling and control design. *International Journal of Control*, 72(7/8): 643-658.
- 6- Cannas B., Fanni A., See L., and Sias G. 2006. Data preprocessing for river flow forecasting using neural networks: Wavelet transforms and data partitioning. *Physics and Chemistry of the Earth*, 31(18):1164-1171.
- 7- Chang F-J., and Chen Y-C. 2001. A Counterpropagation fuzzy-neural network modeling approach to real time streamflow prediction. *Journal of Hydrology*, 245:153-164.
- 8- Corani G. 2005. Air quality prediction in Milan: feed-forward neural networks, pruned neural networks and lazy learning. *Ecological Modelling*, 185:513-529.
- 9- Jafari M. 2011. Predicting the position of special-purpose nodes in mobile networks using Lazy Learning Model. MSc Thesis. Islamic Azad University. 107 Pages (In Persian with English abstract).
- 10- Jafari M., and Khanteymooori A.R. 2011. Lazy Learning in optimization problems using PRESS statistics. *Proceedings of the the application of ai in industry. Bardsir*.
- 11- Jafari M., Abdollahi N., and Mohammadi H. 2011. Predicating the location of nodes in AD Hoc network by lazy learning method. *Innovative Computing Technology, Communications in Computer and Information Science*,

- 241:336-345.
- 12- Karamuz M., and Araghinejad Sh. 2005. Advanced hydrology. AmirKabir University. Iran.
  - 13- Kişi O. 2004. River flow modeling using artificial neural networks. ASCE Journal of Hydrologic Engineering, 9(1):60-63.
  - 14- Kişi O. 2008. Stream flow forecasting using neuro-wavelet technique. Journal of Hydrological Process, 22:4142-4152.
  - 15- Myer R.H. 1990. Classical and modern regression with applications. Boston. M.A: PWS-KENT.
  - 16- Pulido-Calvo I., and Portela. M.M. 2007. Application of neural approaches to one-step daily flow forecasting in Portuguese water sheds, Journal of Hydrology, 332: 1 – 15.
  - 17- Sivakumar B., Jayawardena A.W., and Fernando T.M.K.G. 2002. River flow forecasting: use of phase space reconstruction and artificial neural networks approaches. Journal of Hydrology, 265:225-245.
  - 18- Suri A. 2014. Advanced econometrics. Farhangshenasi. Iran.
  - 19- Turan M.E., and Yurdusev M.A. 2009. River flow estimation from upstream flow records by artificial intelligence methods. Journal of Hydrology, 369:71–77.
  - 20- Valls J.M., Galván I.M., and Isasi P. 2004. Lazy learning in radial basis neural networks: a way of achieving more accurate models. Neural Processing Letters, 20(2):105-124.
  - 21- Wang W., and Ding J. 2003. Wavelet network model and its application to the prediction of hydrology. Nature and Science, 1:67-71.
  - 22- Zealand C.M., Burn D.H., and Simonovic S.P. 1999. Short-term streamflow forecasting using artificial neural networks. Journal of Hydrology, 214:32–48.



## Daily Discharge Estimation in Talar River Using Lazy Learning Model

Z. Abdollahi<sup>1</sup> - A. Kavian<sup>2\*</sup> - K. Shahedi<sup>3</sup> - N. Abdollahi<sup>4</sup> - M. Jafari<sup>5</sup>

Received: 07-09-2015

Accepted: 11-04-2016

**Introduction:** River discharge as one of the most important hydrology factors has a vital role in physical, ecological, social and economic processes. So, accurate and reliable prediction and estimation of river discharge have been widely considered by many researchers in different fields such as surface water management, design of hydraulic structures, flood control and ecological studies in spatial and temporal scale. Therefore, in last decades different techniques for short-term and long-term estimation of hourly, daily, monthly and annual discharge have been developed for many years. However, short-term estimation models are less sophisticated and more accurate. Various global and local algorithms have been widely used to estimate hydrologic variables. The current study effort to use Lazy Learning approach to evaluate the adequacy of input data in order to follow the variation of discharge and also simulate next-day discharge in Talar River in Kasilian Basin where is located in north of Iran with an area of 66.75 km<sup>2</sup>. Lazy learning is a local linear modelling approach in which generalization beyond the training data is delayed until a query is made to the system, as opposed to in eager learning, where the system tries to generalize the training data before receiving queries.

**Materials and Methods:** The current study was conducted in Kasilian Basin, where is located in north of Iran with an area of 66.75 km<sup>2</sup>. The main river of this basin joins to Talar River near Valicbon village and then exit from the watershed. Hydrometric station located near Valicbon village is equipped with Parshall flume and Limnograph which can record river discharge of about 20 cubic meters per second. In this study, daily data of discharge recorded in Valicbon station related to 2002 to 2012 was used to estimate the discharge of 19 September 2012. The mean annual discharge of considered river was also calculated by using available data about 0.441 cubic meters per second. To estimate the discharge of considered day, three methods of constant, linear and quadratic functions controllers based on the local linearization provided by the lazy learning algorithm were considered. Lazy learning is a memory-based linear technique for local modeling approach which is reported as a high-efficient algorithm for simulating variables with low input data. The series of input data was categorized into previous 6, 8, 10, 15 and 20 days, 1 and 2 months, 1, 2 and 3 seasons and also 1 and 2 years to evaluate which series is appropriately enough to predict next-day discharge in the river. Then, mean absolute error and root-mean square error were calculated for all series and models in order to find the best estimator model and the most appropriate series of input data.

**Results:** Results showed that constant and linear model had the minimum root-mean square error of 0.001 and 0.057 respectively with previous 60 days' data series. While the quadratic model had its best estimation with previous 2 season data series with the minimum root-mean square error of 0.059. The result indicated that the more input data increase, the best quadratic model estimate until 60 days. But after 60 days, estimation error gradually increased. Consequently, not more data but adequate are needed for accurate estimation. Also, RMSE in linear model had less fluctuation and therefore less sensitivity compared with other models. And quadratic model had less fluctuation and sensitivity to neighborhoods. Also, according to results, the more variation in each period increase, the better estimation is accrued by lazy learning algorithm. Hence, it was expected that next-day discharge prediction in low-water period needs longer data series than high-water period.

**Conclusion:** Regarding to thousands of prepared training models, constant model with previous 60 days' data and minimum error of 0.0001 was selected as the most accurate estimate for next-day river discharge in Talar River. Results showed that despite of some limitation and demerits, the local Lazy Learning algorithm has significant efficiency in time series simulating. Although the accuracy of simulation increase with more input data, but this algorithm can run by at least 5 training data. However we find lazy learning to be the best performing approach on average goodness indicators (such as mean absolute error and Root-mean square error).

1, 2 and 3- PhD Candidate, Associate Professor and Assistant Professor of Watershed Science and Engineering, College of Agricultural Sciences and Natural Resources, Sari University, Respectively

(\* - Corresponding Author Email: a.kavian@sanru.ac.ir)

4- Lecturer of Software Engineering, Saeb University, Zanjan

5- Lecturer of Software Engineering, Islamic Azad University, Zanjan

On the other hand, the lazy learning predictor can be quickly developed and easily kept up-to-date by adding new data to its database. Also, it does not face with overfitting problems which are common in global modeling approaches. According to some noteworthy features of lazy learning noticed in this regards, this approach will have good performance for time-series studies.

**Keywords:** Daily discharge, Kasilian Basin, Local lazy learning model, Nearest Neighborhood