



دانشگاه ارومیه

نشریه پژوهش‌های حفاظت آب و خاک

جلد بیست و پنجم، شماره سوم، ۱۳۹۷

<http://jwsc.gau.ac.ir>

DOI: 10.22069/jwsc.2018.14361.2910

کاربرد روش‌های شبکه بیزین و حداقل مربعات ماشین بردار پشتیبان در پیش‌بینی تراز سطح آب دریاچه ارومیه

سجاد کریم‌زادگان^۱، * جواد بهمنش^۲ و حسین رضائی^۲

^۱ دانش‌آموخته کارشناسی ارشد گروه مهندسی آب، دانشگاه ارومیه، دانشیار گروه مهندسی آب، دانشگاه ارومیه

تاریخ دریافت: ۹۶/۹/۱۴؛ تاریخ پذیرش: ۹۷/۲/۲۴

چکیده

سابقه و هدف: دریاچه ارومیه به‌عنوان یک اکوسیستم آبی مهم در شمال‌غرب ایران واقع شده است. در ۱۴ سال اخیر میانگین تراز سطح آب دریاچه ارومیه به ۱۲۷۲/۲ متر تقلیل پیدا کرده و این به این معنی است که اختلاف تراز سطح اکولوژیک دریاچه و تراز سطح کنونی ۲ متر است. خشک شدن دریاچه ارومیه باعث بروز مسائل و بحران‌های جدی برای حوضه، استان‌های مجاور و کشور خواهد شد. در این پژوهش از پارامترهای مؤثر مستقیم و غیرمستقیم در پیش‌بینی تراز سطح دریاچه از جمله تبخیر، رواناب ورودی به دریاچه، بارش، دما، باد، میانگین رطوبت هوا و تراز سطح آب دریاچه در ماه قبل به‌عنوان ورودی‌های مدل استفاده شده است. مقایسه کارایی مدل‌های شبکه بیزین که یک مدل احتمالاتی تحت شرایط عدم قطعیت و الگوریتم ماشینی حداقل مربعات ماشین بردار پشتیبان، هدف اصلی پژوهش حاضر است.

مواد و روش‌ها: در پژوهش حاضر از دو روش حداقل مربعات ماشین بردار پشتیبان و شبکه بیزین در فرایند مدل‌سازی استفاده گردید. در این مطالعه عوامل مؤثر در پیش‌بینی تراز سطح دریاچه در ماه قبل به‌عنوان ورودی و تراز سطح آب دریاچه در ماه کنونی به‌عنوان خروجی مدل‌ها مورد بررسی قرار گرفت. جهت تخمین دما، تبخیر، بارش، باد و میانگین رطوبت هوا بر روی سطح دریاچه از داده‌های پنج ایستگاه سینوپتیک مجاور دریاچه با برآورد ضریب تیسن هر ایستگاه و داده‌های ۱۳ ایستگاه هیدرومتری واقع بر رودخانه‌های منتهی به دریاچه جهت تشکیل پارامترهای ورودی و مدل محاسبه گردید.

یافته‌ها: تحلیل داده‌های ایستگاه‌های هیدرومتری نشان داد که تنها چهار ایستگاه از ۱۳ ایستگاه هیدرومتری از نرمال پیروی می‌کنند. مقایسه و بررسی نتایج دو مدل با بررسی ضرایب R^2 ، RMSE، MBE و ناش ساتکلیف بیانگر برتری عملکرد مدل حداقل مربعات ماشین بردار پشتیبان در مقایسه با شبکه بیزین است. این مقادیر برای مدل برتر به ترتیب برابر با ۰/۹۲/۳، ۰/۰۸۲/۰، ۰/۰۱۲-، متر و ۰/۸۶ به‌دست آمد.

* مسئول مکاتبه: j.behmanesh@urmia.ac.ir

نتیجه‌گیری: در این پژوهش معیارهای ارزیابی نشان داد که مدل حداقل مربعات ماشین بردار پشتیبان نسبت به مدل شبکه بیزین دارای برتری است. نکته دارای اهمیت در مقایسه دو مدل آن است که ماهیت مدل حداقل مربعات ماشین بردار پشتیبان، ماشینی است ولی ماهیت شبکه بیزین یک مدل احتمالاتی تحت شرایط عدم قطعیت است که در آن از توزیع نرمال جهت آموزش متغیرهای شبکه استفاده شده است. بنابراین استفاده از مدل شبکه بیزین نسبت به مدل حداقل مربعات ماشین بردار پشتیبان می‌تواند توصیه گردد.

واژه‌های کلیدی: پیش‌بینی، تراز اکولوژیک، حداقل مربعات ماشین بردار پشتیبان، دریاچه ارومیه، شبکه بیزین

مقدمه

روش ماشین بردار پشتیبان و شبکه‌های بیزین در پیش‌بینی جریان روزانه رودخانه باراندوز چای نشان داد دقت دو مدل در پیش‌بینی به لحاظ آماری بسیار به هم نزدیک است (۲). در مطالعات کمی شبکه تصمیم‌گیری بیزی قادر به تجمیع اهداف و اولویت‌های متفاوت و رویکردهای مختلف مدیریت آب‌های زیرزمینی می‌باشد (۲۳). شبکه بیزین ابزاری قدرتمند برای ایجاد تعادل در تخصیص آب بین دو منبع کشاورزی و محیط زیست در سال‌های ترسالی و خشکسالی تحت سناریوهای مختلف مدیریتی می‌باشد (۲۶ و ۳۱). از شبکه بیزین برای بررسی تغییرات مکانی خشکسالی و توسعه مدل پیش‌بینی با استفاده از شبکه بیزین، بر اساس خشکسالی‌های گذشته برای پیش‌بینی خشکسالی در آینده استفاده شد (۱۷). روش بیزین در کاهش میزان عدم قطعیت در تخمین دبی با دوره‌های بازگشت مختلف مورد استفاده قرار گرفت (۲۷). در طی سال‌های اخیر با توجه به رشد جمعیت حوزه و نیاز به تامین غذا و امنیت غذایی و از طرفی گسترش روزافزون فعالیت‌های صنعتی، کشاورزی و عمرانی و نیاز بیش‌تر به منابع آب سبب برداشت بی‌رویه از منابع آب سطحی و احداث سد بر روی رودخانه‌های حوزه شده است، در حالی که رودخانه‌ها مهم‌ترین عامل تغذیه دریاچه به‌شمار می‌رود. در این پژوهش از تمامی عواملی که به‌صورت مستقیم در پیش‌بینی تراز سطح دریاچه از جمله تبخیر، مجموع

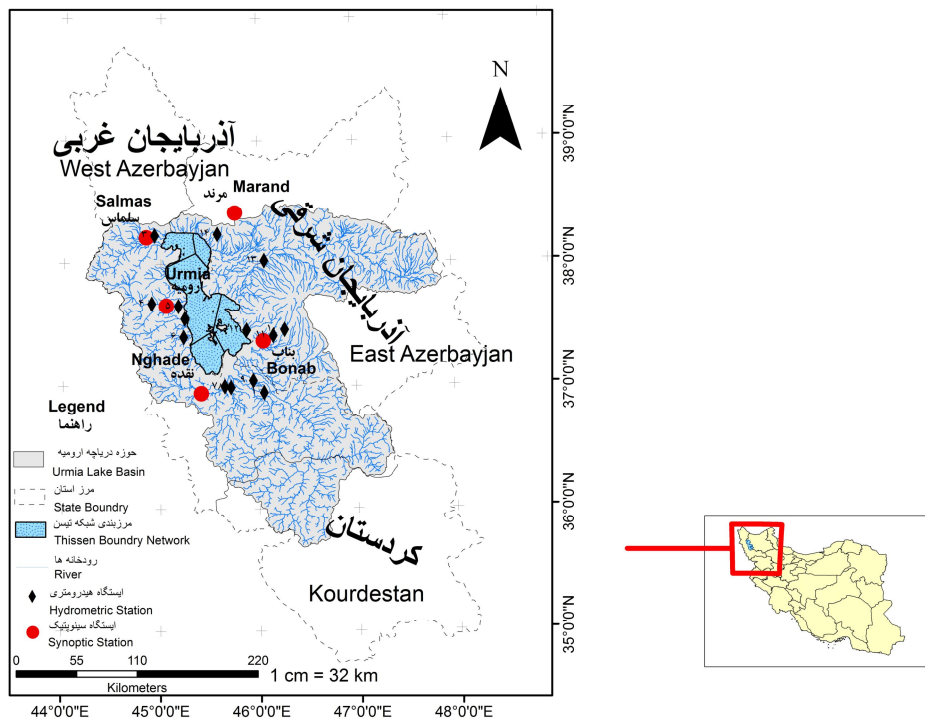
با توجه به افزایش دما، تغییرات زیاد در بارندگی و همچنین وقوع خشکسالی در سطح حوضه، تراز آب دریاچه دچار تغییرات زیادی شده است (۱۹). به‌طورکلی میزان تغییرات تراز سطح هر دریاچه تابع میزان ورودی آب به داخل آن و خروجی از آن است که این امر توسط عوامل طبیعی و یا با دخالت انسان کنترل می‌شود. مقایسه نتایج دو مدل سری زمانی ARMA و مدل تلفیقی AR-ARCH در پیش‌بینی تراز سطح دریاچه ارومیه در مقیاس سالانه نشان داد که مدل تلفیقی از برتری برخوردار است (۱). دو روش برنامه‌ریزی بیان ژن و شبکه‌های بیزین برای تخمین دمای هوا روزانه به‌کار گرفته شد. نتایج نشان داد که هر دو روش دقت خوبی قادر به مدل‌سازی دمای حداقل، متوسط و حداکثر هوا بودند (۲۱). با مطالعه جریان روزانه رودخانه باراندوزچای با استفاده از روش‌های برنامه‌ریزی ژنتیک و شبکه‌های بیزین به این نتایج رسیدند که برای سری روزانه بهترین عملکرد برنامه‌ریزی ژنتیک با سه تاخیر زمانی به‌دست آمد (۳). کاربرد شبکه‌های بیزین در پیش‌بینی تراز سطح آبخوان دشت بیرجند نتایج قابل‌قبولی را نشان داد (۱۳). شبکه‌های عصبی بیزین، ماشین بردار پشتیبان و و برنامه ریزی بیان ژنی در تحلیل بارش- رواناب ماهانه رودخانه کاکارضا واقع در استان لرستان دقت نزدیک بهم داشتند (۱۱). مقایسه عملکرد دو

به درون یک دریاچه مرکزی تخلیه می‌شود و اکوسیستم فعال آن شامل دریاچه و حوضه آبریز آن است (۱۴). تراز کف دریاچه ۱۲۶۸ متر، تراز متوسط سطح آب دریاچه ۱۲۷۵/۵ و دامنه نوسان آن از ۱۲۷۲/۱ تا ۱۲۷۸/۴ متر از سطح دریا واقع شده است. دریاچه ارومیه با حجم حداکثری بین ۲۸ تا ۳۲ میلیارد مترمکعب بزرگ‌ترین دریاچه داخلی ایران و از مهم‌ترین و باارزش‌ترین اکوسیستم‌های آبی ایران به‌شمار می‌رود (۲۴). در این پژوهش از ۱۱ سال داده ماهانه برای آموزش و سه سال داده ماهانه برای صحت‌سنجی هر دو مدل استفاده شد. طول دوره آماری از ابتدای سال آبی ۱۳۸۱-۱۳۸۰ تا انتهای ۱۳۹۴-۱۳۹۳ می‌باشد.

رواناب ورودی به دریاچه، تراز سطح آب دریاچه در ماه قبل مؤثر هستند و عواملی چون بارش، دما، باد، میانگین رطوبت هوا که به‌صورت مستقیم بر تبخیر از دریاچه، مجموع آب رواناب سطحی ورودی به دریاچه تأثیرگذار هستند در قالب دو مدل شبکه بیزین که یک مدل احتمالاتی تحت شرایط عدم قطعیت و الگوریتم ماشینی حداقل مربعات ماشینی بردار پشتیبان ارائه شده است.

مواد و روش‌ها

دریاچه ارومیه در شمال‌غربی ایران واقع در حوزه دریاچه ارومیه بین استان‌های آذربایجان‌غربی و آذربایجان‌شرقی قرار گرفته است (شکل ۱). اکوسیستم دریاچه ارومیه نمونه‌ای شاخص از یک حوضه آبریز بسته است که در آن همه آب‌های سطحی و زیرزمینی



شکل ۱- حوضه آبریز دریاچه ارومیه و ایستگاه‌های سینوپتیک و هیدرومتری مورد مطالعه.

Figure 1. Urmia lake basin and studied hydrometric and synoptic stations.

پلی‌گون‌های تیسن در نرم‌افزار ArcGIS 10.4 استفاده شد. در مدل تیسن هر ایستگاه به ایستگاه مجاور خود وصل شده و با تشکیل مثلث، عمودمنصف هر ضلع رسم و از محل تقاطع پلی‌گون‌های تیسن رسم می‌گردد (۲۹). به کمک مدل تیسن، برای هر ایستگاه با توجه به موقعیت آن، ضریبی از نسبت سطح پوشش ایستگاه به کل سطح دریاچه در نظر گرفته شد. این ضریب وزنی پلی‌گون‌ها باید به گونه‌ای باشند که جمع همه آن‌ها، برابر یک شود. در جدول ۲ و شکل ۱ موقعیت هر ایستگاه آورده شده است.

رودخانه‌ها و ایستگاه‌های منتهی به دریاچه: یکی از مؤثرترین عوامل در تغییرات تراز سطح آب دریاچه ارومیه ورود آب‌های سطحی به این دریاچه است. از مهم‌ترین رودخانه‌های منتهی به دریاچه ارومیه می‌توان به جدول ۱ اشاره کرد. آخرین ایستگاه‌های هیدرومتری موجود واقع بر این رودخانه‌ها نسبت به دریاچه با مشخصات جغرافیایی و محل قرارگیری آن‌ها نسبت به دریاچه در جدول ۱ و شکل ۱ آورده شده است.

ایستگاه‌های سینوپتیک: پس از صحت‌سنجی داده‌ها از نظر همگنی، روند، داده‌های پرت است، از روش وزن‌دهی

جدول ۱- مشخصات ایستگاه‌های هیدرومتری.

Table 1. Hydrometric station characteristics.

| | ۱ | ۲ | ۳ | ۴ | ۵ | ۶ | ۷ | ۸ | ۹ | ۱۰ | ۱۱ | ۱۲ | ۱۳ | ۱۴ |
|---|---------------------------|--------------------|-------------------------|-------------|----------------------|-------------------|------------------------------|---------------------|-----------------------|---------------------|------------------------|-------------------|---------------|---------------|
| نام ایستگاه و مشخصات آن Station name and its characteristics | سیمینه رود- Simineh river | کاشیبان- Kashitban | یاقوزآغاج- YalghozAghaj | تپیک- Tipik | پل اوزبک- Pole Ozbak | بابارود- Babarood | پل بهراملو سنتو- Pole Sentto | گردیغوب- GardYaghob | نظام آباد- Nezam Abad | تازه کند- Taze Kand | خرم‌مازرد- Khorma Zard | شیشوان- Shishivan | آخولا- Akhola | دریان- Daryan |
| طول جغرافیایی Longitude | 46.05 | 45.25 | 44.39 | 44.9 | 45.18 | 45.23 | 46.02 | 45.7 | 45.94 | 45.27 | 46.15 | 45.88 | 46.06 | 45.6 |
| عرض جغرافیایی latitude | 36.95 | 37.55 | 38.23 | 37.67 | 37.65 | 37.4 | 36.52 | 37 | 37.04 | 37.46 | 37.42 | 37.47 | 38.03 | 38.23 |
| سال تاسیس Established | 1343 | 1343 | 1353 | 1329 | 1361 | 1328 | 1336 | 1368 | 1372 | 1352 | 1355 | 1362 | 1362 | 1349 |

شور دریاچه ارومیه به ارتفاع تشت تبخیر آب شیرین (جدول ۳) را برآورد کرد (۱۸).

تخمین مقدار تبخیر از سطح دریاچه: محمودی (۱۳۹۲) در پژوهشی ضریب نسبت ارتفاع تبخیر آب

جدول ۲- مشخصات ایستگاه‌های سینوپتیک مجاور دریاچه ارومیه.

Table 2. Synoptic station characteristics beside Urmia lake.

| نام ایستگاه Station name | مشخصات ایستگاه Characteristic | | | | |
|-----------------------------------|----------------------------------|----------------|---------------|-----------------|-----------------|
| | ارومیه Urmia | مرند Marand | بناب Bonab | نقده Naghade | سلماس Salmas |
| طول جغرافیایی Longitude | 45.05 | 45.76 | 46.05 | 45.41 | 44.84 |
| عرض جغرافیایی Latitude | 37.65 | 38.41 | 37.37 | 36.94 | 38.21 |
| ارتفاع Elevation | 1328 | 1550 | 1281 | 1370 | 1339 |
| سال تاسیس (میلادی) Established | 1951 | 2000 | 1999 | 2001 | 2001 |

جدول ۳- ضریب تبخیر از سطح دریاچه در مقایسه با آب شیرین.

Table 3. Coefficient of evaporation from lake surface comparing to freshwater.

| فروردین Farvardin | اردیبهشت Ordibehesht | خرداد Khordad | تیر Tir | مرداد Mordad | شهریور Shahrivar | مهر Mehr | آبان Aban | آذر Azar | دی Dey | بهمن Bahman | اسفند Esfand | سالانه Annual |
|----------------------|-------------------------|------------------|------------|-----------------|---------------------|-------------|--------------|-------------|-----------|----------------|-----------------|------------------|
| 0.73 | 0.7 | 0.66 | 0.6 | 0.55 | 0.55 | 0.54 | 0.59 | 0.6 | 0.7 | 0.7 | 0.7 | 0.63 |

پراکندگی بیشترین روش‌های آزمون همگنی در پژوهش‌های روز است، استفاده شده است.

معیارهای ارزیابی: در این پژوهش از چهار پارامتر ضریب تبیین^۴ (رابطه ۵)، میانگین مربعات خطا^۵ (رابطه ۶)، میانگین خطای اریب^۶ (رابطه ۷) و ضریب ناش ساتکلیف^۷ (رابطه ۸)، برای بررسی دقت و ارزیابی مدل‌ها استفاده شد.

تحلیل آماری اولیه داده‌ها: سر آغاز هر پژوهش هیدرولوژی صحت‌سنجی داده‌های تاریخی است (۱۰). در این پژوهش از توزیع‌های آماری نرمال و لوگ نپر نرمال سه‌پارامتری^۱ برای توجیه پراکندگی و توزیع داده استفاده شده است. همچنین از آزمون نکوئی برازش کلوموگروف- اسمیرنف^۲ که یک آزمون‌های ناپارامتری نکوئی برازش است برای انتخاب توزیع برتر (۵) و از روش SHNT^۳ (۴) یکی از

4- R²
5- RMSE
6- MBE
7- Nash-Sutcliffe

1- Log-Normal (3P)
2- Kolmogorov-Smirnov (K-S)
3- Standard Normal Homogeneity Test

$$P\langle A|B \rangle = \frac{P\langle B|A \rangle P(A)}{P(B)} \quad (۹)$$

$$R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{\sum (X_i - \hat{X}_i)^2}{\sum (X_i - \bar{X})^2} \quad (۵)$$

در رابطه (۹) $P(B)$ احتمال وقوع پیشامد^۱ B، $P(A)$ احتمال وقوع پیشامد^۲ A یا احتمال پیشین^۳، $P(A|B)$ احتمال پسین^۴ و $P(B|A)$ احتمال B به شرط A یا تابع درستنمایی^۵ تابع می‌باشد (۱۵).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \hat{X}_i)^2} \quad (۶)$$

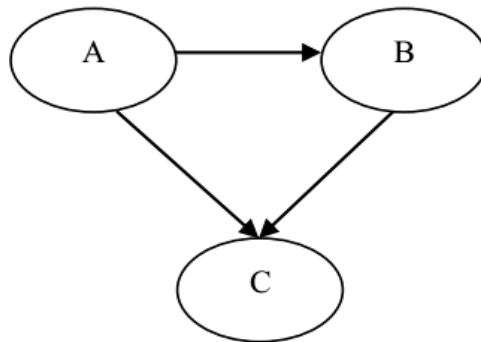
$$MBE = \frac{\sum_{i=1}^n (X_i - \hat{X}_i)}{n} \quad (۷)$$

شبکه بیزین^۶ مدلی است که می‌توان در آن انعکاسی از روابط علی و معلولی اتفاقات پیرامون را نشان داد (۱۷). شبکه بیزین یک گراف احتمالاتی بدون دور^۷ است که شامل مجموعه‌ای از متغیرهاست. در این گراف یک متغیر به‌عنوان والد^۸ و متغیر دیگر که حاصل از متغیر والد است متغیر فرزند نامیده می‌شود (۲۰). در شکل ۲ می‌توان به نمونه‌ای از یک گراف جهت‌دار غیر مدور اشاره کرد:

$$NS = 1 - \frac{\sum_{i=1}^N (X_i - \hat{X}_i)^2}{\sum_{i=1}^N (X_i - \bar{X})^2} \quad (۸)$$

که در آن، X_i مقادیر مشاهداتی، \hat{X}_i مقادیر پیش‌بینی، \bar{X}_i میانگین داده‌های مشاهداتی، n تعداد داده‌ها می‌باشد.

روش بیزین: روش بیزین یک روش احتمالات شرطی می‌باشد که تئوری آن به‌شرح زیر می‌باشد.



شکل ۲- گراف بدون دور شبکه بیزین.

Figure 2. Undirected Bayesian graph.

- 1- Evidence
- 2- Prior Probability
- 3- Posterior Probability
- 4- Likelihood
- 5- Bayesian Network (BN)
- 6- Directed Acyclic Graph (DAG)
- 7- Parent

$$\begin{aligned}
 p(x|D) &= \int p(x|\mu)p(x|D)d\mu \\
 &= \int N(x|\mu, \sigma^2)N(x|\mu_n, \sigma_n^2)d\mu \quad (11) \\
 &= N(x|\mu_n, \sigma_n^2 + \sigma^2)
 \end{aligned}$$

در این مطالعه از نرم افزار Hugin 2016 جهت آموزش شبکه با استفاده از الگوریتم NPC و تخمین پارامترهای توزیع نرمال متغیرها یا گره‌های شبکه مورد استفاده قرار گرفته است.

حداقل مربعات ماشین بردار پشتیبان^۳: این روش ابتدا تحت عنوان ماشین بردار پشتیبان توسط وپنیک بر پایه بر پایه تئوری یادگیری آماری که از حداقل سازی ریسک ساختاری پیروی می‌کند، ارائه شده است (۳۰).

الگوریتم مدل: مجموعه داده‌های $\{X_i, Y_i\}_{i=1}^N$ وجود دارد که $X_i \in R^n$ به عنوان بردار ورودی و $Y_i \in R^n$ به عنوان بردار خروجی معرفی می‌شوند. تابع رگرسیون غیرخطی در مسائل تخمین به صورت رابطه ۱۱ است:

$$y(x) = w^T \varphi(x) + b \quad (12)$$

که در آن، w, b, T به ترتیب مقادیر وزن‌ها، بایاس تابع رگرسیون و نشان‌دهنده ترانهاده هستند. $\varphi(x)$ ترسیم خطی ورودی‌ها در فضای ویژگی با ابعاد بالاست. حال مسأله غیرخطی رگرسیون بر اساس رابطه بهینه‌سازی (۱۲) حل می‌شود:

$$\min_{w, e, b} j(w, e) = \frac{1}{2} w^T w + \frac{\gamma}{2} \sum_{i=1}^N e_i^2 \quad (13)$$

بر اساس محدودیت

$$y(x) = w^T \varphi(x) + b + e_i \quad (14)$$

ایده کلی الگوریتم "چندحالتی غیرقطعی" به دست آوردن ارتباط یا عدم ارتباط بین گره‌های گراف شبکه بیزین با استفاده از آزمون‌های آماری استقلال شرطی برای هر جفت متغیر در شبکه است. از معایب این الگوریتم عدم توانایی آن در تعیین همه جهت‌های کمان‌ها از طریق داده‌های موجود است و بنابراین تعدادی از کمان‌ها به صورت تصادفی جهت‌دهی می‌شوند. می‌توان برای حل این مشکل از الگوریتم NPC^۲ استفاده نمود که در آن، الگوریتم با تعامل با خبره در مورد کمان‌های نامشخص تصمیم می‌گیرد (۶ و ۸). در این پژوهش از الگوریتم NPC در سطح اطمینان ۹۵ درصد برای آموزش شبکه بیزین استفاده شده است.

روش بیزین در تخمین پارامترهای توزیع نرمال: روش تخمین پارامترهای میانگین و واریانس توزیع نرمال از روی داده‌های مشاهداتی و با توزیع پیشین که مقدار میانگین و واریانس آن‌ها به ترتیب یک و صفر لحاظ می‌گردد به صورت رابطه ۱۰ است (۹):

$$\begin{aligned}
 \sigma_1^2 &= \frac{\sigma_0^2 \sigma^2}{\sigma^2 + n\sigma_0^2} = \frac{1}{\sigma^{-2} + n\sigma_0^{-2}} \quad (10) \\
 \mu_1 &= \sigma_1^2 \left(\mu_0 \sigma^{-2} + \sum_{i=1}^n \sigma_0^{-2} x_i \right)
 \end{aligned}$$

که در آن، μ_0 میانگین توزیع پیشین، σ_0^2 واریانس توزیع پیشین، μ میانگین داده‌های مشاهداتی، σ^2 واریانس داده‌های مشاهداتی، μ_1 میانگین توزیع پسین، σ_1^2 واریانس توزیع پسین می‌باشد.

پیش‌بینی شبکه بیزین: معادله پیش‌بینی شبکه بیزین نرمال با استفاده از توزیع پسین (رابطه ۱۱) می‌باشد (۸).

- 1- Path Condition
- 2- Necessary Path Condition (NPC)

نتایج و بحث

در جدول ۴ مقدار ضریب تیسن هر کدام از ایستگاه‌های سینوپتیک برآورد و ارائه گردیده است. در ادامه تحلیل داده‌ای نشان داد که در هر پنج ایستگاه سینوپتیک انتخابی متغیر دما، باد و میانگین رطوبت در طول دوره آماری طبق نتایج آزمون نکویی برازش کلموگروف-اسمیرنوف، از آزمون نرمال پیروی می‌کند ولی دو متغیر اقلیمی بارش و تبخیر از آزمون نرمال پیروی نمی‌کنند. در ادامه نتایج حاصل از تحلیل داده‌های ایستگاه‌های هیدرومتری نشان داد که تنها چهار ایستگاه از ۱۳ ایستگاه هیدرومتری از نرمال پیروی می‌کنند.

نتایج شبکه بیزین: با توجه به انتخاب هفت متغیر دما، تبخیر، بارش، باد، میانگین رطوبت هوا بر روی سطح دریاچه، مجموع تخلیه رواناب آب سطحی به دریاچه و تراز آب دریاچه در ماه قبل به‌عنوان داده‌های ورودی و تراز سطح آب دریاچه در ماه کنونی به شبکه بیزین و آموزش شبکه توسط الگوریتم NPC، در نهایت رابطه علی و معلولی بین متغیرها در گراف بدون دور شکل ۳ حاصل شد. در این گراف EVA (تبخیر)، Temp (دما)، Prec (بارش)، Humidity (میانگین رطوبت هوا)، Wind (میانگین سرعت باد) Discharge (مجموع رواناب سطحی ورودی)، SLT (تراز سطح آب دریاچه) و SL (تراز سطح دریاچه در ماه کنونی) می‌باشد. با توجه به این‌که روش انتخابی جهت آموزش و آزمون نکویی برازش متغیرهای ورودی به شبکه به‌ترتیب آزمون نرمال و آزمون کلموگروف اسمیرنوف می‌باشد. از تبدیل لوگ نپرمال سه پارامتری برای تبدیل داده‌ها به توزیع نرمال استفاده شده است. در جدول ۵ مشخصات جامعه آماری و تبدیل‌های مذکور برای متغیرها آورده شده است. و در جدول ۶ مقادیر تخمینی پارمترهای روش بیزین ارائه گردیده است.

که در آن، γ پارامتر تنظیم بخش خطا است و e میزان خطا را نشان می‌دهد. حل با استفاده از شکل لاگرانژی از تابع هدف اصلی در رابطه ۱۴ آمده است:

$$L(w, b, e, \alpha) = j(w, e) - \sum \alpha_i \{w^T \varphi(x) + b + e_i - y_i\} \quad (15)$$

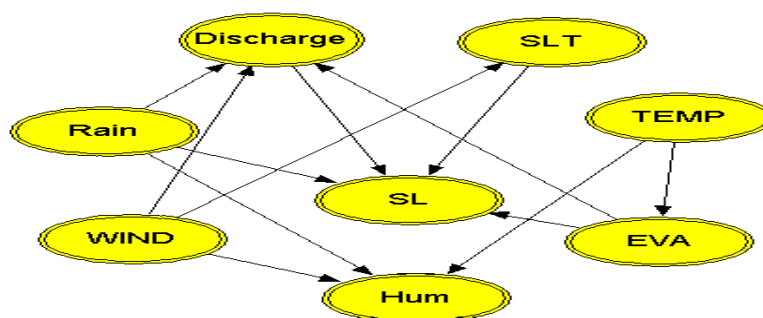
که در آن، α_i ضریب لاگرانژ است. بر اساس شرایط کان-تاگر (KKT)، مدل LS-SVM نتیجه شده در شکل، به‌منظور تخمین تابع به‌صورت رابطه ۱۶ نوشته می‌شود (۲۲):

$$k(x, x_i) y(x) = \sum \alpha_i k(x, x_i) + b \quad (16)$$

تابع کرنل نامیده می‌شود که با پیروی از شرایط Mercer به‌عنوان تابعی با ایجاد ضرب داخلی در فضای ویژگی معرفی می‌شود (۲۸). در این پژوهش با توجه به عملکرد مطلوب آن در مطالعات قبلی چون نوری، همکاران (۲۰۱۴)، لین و همکاران (۲۰۱۱) و رضایی (۲۰۱۳) تابع کرنل پایه شعاعی RBF به همراه γ که در معادله داده شده است به‌عنوان انتخاب تابع کرنل برتر مورد استفاده قرار گرفته است. شایان ذکر است که در معادله γ کنترل‌کننده میزان نوسان تابع گوسی و همچنین کنترل‌کننده نتایج و تعمیم‌دهنده مدل SVM است (۱۶، ۲۵ و ۲۹). برای حل $K(x_i, x_j) = \langle \varphi(x_i), \varphi(x_j) \rangle$ تعیین پارامتر تنظیم‌کننده بخش خطا (γ) و تعیین σ کوچک‌تر از صفر پارامتر مربوط به تابع کرنل پایه شعاعی در رابطه ۱۷ نیاز است:

$$K(x, x_i) = \exp\left(-\frac{|x_i - x_j|^2}{\sigma^2}\right) \quad (17)$$

در این مطالعه از افزونه LS-SVM در نرم‌افزار متلب ۲۰۱۷ جهت انجام این محاسبات استفاده شده است.



شکل ۳- شبکه بیزین.

Figure 3. Bayesian Network.

جدول ۴- ضرایب تیسن ایستگاه سینوپتیک.

Table 4. Thiessen coefficient of synoptic station.

| نام ایستگاه Station | ارومیه Urmia | مرند Marand | بناب Bonab | نقده Naghadeh | سلماس Salmas |
|-----------------------------------|-----------------|----------------|---------------|------------------|-----------------|
| ضریب تیسن Thiessen coefficient | 0.44 | 0.08 | 0.21 | 0.17 | 0.1 |

جدول ۵- مشخصات آماری متغیرهای ورودی به شبکه بیزین.

Table 5. Statistical characteristics of input variables to Bayesian network.

| متغیرهای ورودی Input variable | برازش توزیع Distribution fit | مشخصات آماری متغیر Variable statistical characteristics | | | مقدار آماره آزمون کلموگروف اسمیرنف P-value K-S Test |
|---|---|--|----------------------------------|--------------------------------|---|
| | | میانگین Average | انحراف از معیار St. Diversion | پارمتر گاما (γ) Parameter γ | |
| دما Temperature | نرمال Normal | 12.24 | 9 | - | 0.106**** |
| بارندگی Precipitation | لوگ نرمال سه پارامتری Log- Normal (3P) | 2.68 | 1.21 | -1.123 | 0.33***** |
| تبخیر Evaporation | نرمال Normal | 75.23 | 68.57 | - | 0.0005 |
| باد Wind | نرمال Normal | 2.28 | 0.59 | - | 0.88***** |
| میانگین رطوبت هوا Avg. Humidity | نرمال Normal | 56.87 | 10.1 | - | 0.2***** |
| ورودی آب سطحی Runoff input | لوگ نرمال سه پارامتری Log- Normal (3P) | 3.38 | 1.33 | 0.47 | 0.2**** |
| تراز سطح آب دریاچه Lake water level | نرمال Normal | 1272.7 | 0.97 | - | 0.2** |
| تراز سطح آب دریاچه در ماه آینده Lake water level at next month | نرمال Normal | 1272.7 | 0.97 | - | 0.2** |

***** سطح معنی داری ۰.۰۱، **** سطح معنی داری ۰.۰۵، *** سطح معنی داری ۰.۱، ** سطح معنی داری ۰.۰۲، * سطح معنی داری ۰.۰۵.

***** Significant level 0.01, **** Significant level 0.05, *** Significant level 0.1, ** Significant level 0.02, * Significant level 0.05.

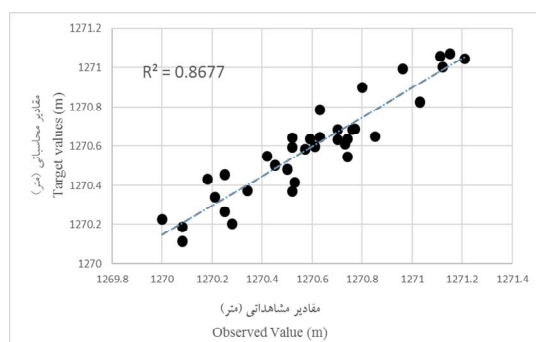
جدول ۶- مشخصات پارامترهای آماری تخمینی روش بیزین.

Table 6. Estimated Statistical parameters characteristics of Bayesian approach.

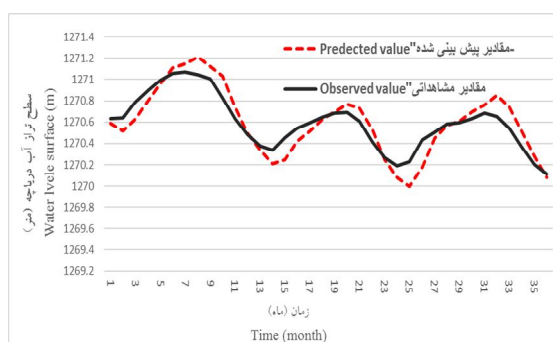
| داده‌های ورودی Input variable | توزیع Distribution | مشخصات آماری متغیر Variable statistical characteristics | |
|---|-----------------------|--|----------------------------------|
| | | میانگین Average | انحراف از معیار St. Diversion |
| دما Temperature | نرمال Normal | 12.24 | 9.03 |
| بارندگی Precipitation | نرمال Normal | 2.69 | 1.22 |
| تبخیر Evaporation | نرمال Normal | 122.68 | 85.12 |
| باد Wind | نرمال Normal | 2.28 | 0.59 |
| میانگین رطوبت هوا Avg. Humidity | نرمال Normal | 3.38 | 1.4 |
| ورودی آب سطحی Runoff input | نرمال Normal | 4.02 | 0.14 |
| تراز سطح آب دریاچه Lake water level | نرمال Normal | 1272.69 | 0.96 |
| تراز سطح آب دریاچه در ماه آینده Lake water level at next month | نرمال Normal | 1272.71 | 0.94 |

به‌دست آمد. در شکل (۴-ب) نمودار تبیین پراکنش نقادیر مشاهداتی اختلاف را با مقادیر مشاهداتی دارد. در گراف شکل ۳ مشاهده می‌شود متغیر رطوبت نسبی به‌عنوان یک گره فرزند نقش دارد با بررسی روابط علی و معلولی گراف می‌توان مصداق و انعکاس این روابط را در محیط پیرامون طبیعی یافت.

پس از آموزش شبکه بیزین طی ۱۱ سال دوره آماری و پیش‌بینی ۳ ساله آن دقت نتایج حاصل با استفاده از ضرایب میانگین مربعات خطا، تبیین، میانگین تغییرات خطا و ضریب ناش- ساتکلیف مورد ارزیابی و بررسی قرار گرفت که برای شبکه بیزین این مقادیر به‌ترتیب برابر با ۰.۱۲٪، متر، ۸۶.۷۷٪، ۰.۰۲٪، متر، ۰.۷۷٪



(ب) (B)



(الف) (A)

شکل ۴- الف): مقادیر محاسباتی و مشاهداتی نسبت به زمان ب): مقادیر پیش‌بینی شده و مشاهداتی شبکه بیزین.

Figure 4. A) Computed and observed values to time B) Predicted and observed values by Bayesian network.

پارامترهای ورودی و مقدار تراز آب سطحی در ماه کنونی به عنوان خروجی مدل در نظر گرفته شد. به دلیل این که تراز سطح در ماه های قبل بیشترین تأثیر را در پیش بینی مدل دارد تاخیرهای متوالی تراز سطح به عنوان ورودی مدل در سناریوهای بعدی در نظر گرفته شد.

نتایج حداقل مربعات ماشین بردار پشتیبان: در این تحقیق سناریوی مختلف (جدول ۷) برای آموزش و پیش بینی توسط مدل مورد ارزیابی قرار گرفت. در اولین سناریو هفت متغیر تراز سطح آب دریاچه، دما، تبخیر، بارش، باد، رطوبت نسبی و مجموع تخلیه رواناب سطحی به دریاچه در ماه قبل به عنوان

جدول ۷- سناریوهای ورودی به مدل حداقل مربعات ماشین بردار پشتیبان.

Table 7. LS-SVM model input scenarios.

| سناریو Scenarios |
|---|
| Prec(t-1), EVA(t-1), Temp(t-1), Discharge(t-1), Humidity(t-1), Wind(t-1), SL(t-1) |
| Prec(t-1), EVA(t-1), Temp(t-1), Discharge(t-1), Humidity(t-1), Wind(t-1), SL(t-1), SL(t-2) |
| Prec(t-1), EVA(t-1), Temp(t-1), Discharge(t-1), Humidity(t-1), Wind(t-1), SL(t-1), SL(t-2), SL(t-3) |
| Prec(t-1), EVA(t-1), Temp(t-1), Discharge(t-1), Humidity(t-1), Wind(t-1), SL(t-2), SL(t-3), SL(t-4) |

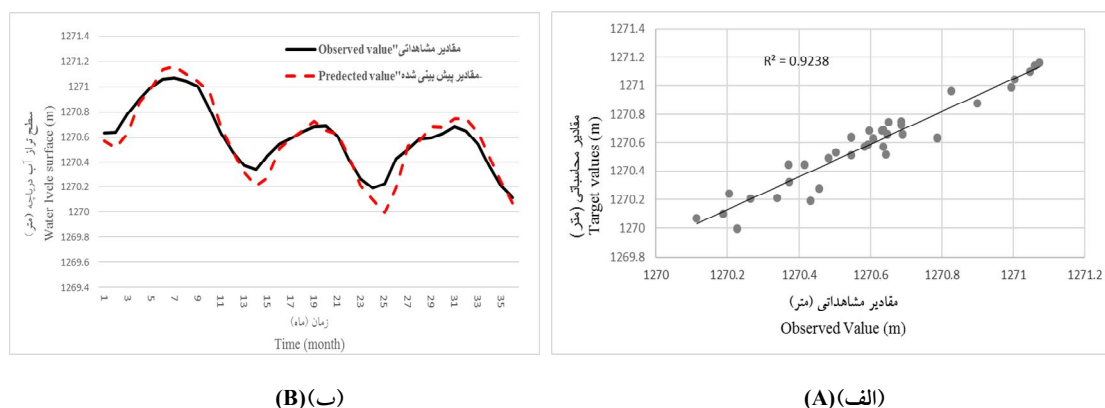
این مدل کمترین میزان را دارد. همان طور که در شکل (۵- الف) ملاحظه می شود مدل در پیش بینی بر خلاف شبکه بیزین تنها در مقادیر حداقل بیشترین اختلاف را با مقادیر مشاهداتی دارد. مقدار گاما برابر 64×10^8 و مقدار σ^2 برابر 14041 برآورد شده است.

نتایج پیش بینی مدل با چهار سناریوی انتخابی در جدول ۸ ارائه شده است. طبق نتایج به دست آمده سناریوهای شماره ۴ به عنوان سناریو برتر انتخاب شده است. در این سناریو مقدار همبستگی نسبت به دو سناریوی دو و سه اختلاف کمتری دارد ولی در میزان میانگین مربعات خطا و میانگین تغییرات خطا

جدول ۸- نتایج تحلیل آماری برای سناریوهای مختلف ورودی مدل.

Table 8. Result of statistical analysis for different scenarios of model input.

| ضریب ناش ساتکلیف Nash | میانگین تغییرات خطا (متر) MBE (m) | میانگین مربعات خطا (متر) RMSE (m) | تنبین R^2 | سناریو Scenarios |
|-----------------------------|---|---|----------------|---|
| 0.76 | -0.057 | 0.152 | 0.808 | Prec(t-1), EVA(t-1), Temp(t-1), Discharge(t-1), Humidity(t-1), Wind(t-1), SL(t-1) |
| 0.84 | -0.029 | 0.099 | 0.932 | Prec(t-1), EVA(t-1), Temp(t-1), Discharge(t-1), Humidity(t-1), Wind(t-1), SL(t-1), SL(t-2) |
| 0.87 | -0.015 | 0.087 | 0.928 | Prec(t-1), EVA(t-1), Temp(t-1), Discharge(t-1), Humidity(t-1), Wind(t-1), SL(t-1), SL(t-2), SL(t-3) |
| 0.86 | -0.012 | 0.082 | 0.923 | Prec(t-1), EVA(t-1), Temp(t-1), Discharge(t-1), Humidity(t-1), Wind(t-1), SL(t-2), SL(t-3), SL(t-4) |



شکل ۵- الف): مقادیر محاسباتی و مشاهداتی نسبت به زمان ب): مقادیر پیش‌بینی شده و مشاهداتی LS-SVM. Figure 5. A) Computed and observed values B) Predicted and observed values by LS-SVM.

نتیجه‌گیری کلی

تاکنون مطالعات زیادی برای پیش‌بینی تراز سطح آب دریاچه ارومیه با استفاده از مدل‌های مختلفی چون سیستم عصبی فازی، بیان ژنتیک، مدل‌های استوکاستیک سری زمانی و دیگر مدل‌ها انجام گرفته است ولی در هیچ‌یک از آن‌ها از مجموعه عوامل مؤثر بر تراز سطح آب دریاچه ارومیه استفاده نشده است. در این پژوهش معیارهای ارزیابی و مقایسه دقت دو مدل در پیش‌بینی با توجه به معیارهای ضرایب میانگین مربعات خطا، تبیین، میانگین تغییرات خطا و ضریب ناش- ساتکلیف بیانگر آن است که مدل حداقل مربعات ماشین بردار پشتیبان نسبت به مدل

شبکه بیزین دارای برتری است. نکته دارای اهمیت در مقایسه دو مدل آن است که ماهیت مدل حداقل مربعات ماشین بردار پشتیبان، ماشینی است ولی ماهیت شبکه بیزین یک مدل احتمالاتی تحت شرایط عدم قطعیت است که در آن از توزیع نرمال جهت آموزش متغیرهای شبکه استفاده شده است چون که ماهیت وقوع رخداد‌های طبیعی تصادفی است، بنابراین استفاده از مدل شبکه بیزین نسبت به مدل حداقل مربعات ماشین بردار پشتیبان از نظر این پژوهش ارجح می‌باشد.

منابع

1. Abbaszadeh, A.A., Khalili, K., and Behmanesh, J. 2015. Application of Combined AR-ARCH model in Forecasting Urmia Lake Water Level. *J. Soil Water Know.* 25: 4/2. 175-186. (In Persian)
2. Ahmadi, F., and Radmanesh, F. 2016. Application of Bayesian Networks and Genetic Programming for Predicting Daily River Flow (Case study: Barandoozchay River). *J. Irrig. Sci. Engin.* 39: 4. 213-223. (In Persian)
3. Ahmadi, F., Radmanesh, F., and Mir Abbasi, N.R. 2016. Comparing the performance of Support Vector Machines and Bayesian networks in predicting daily river flow (Case study: Baranduz Chai River). *J. Water Soil Cons.* 22: 6. 171-186. (In Persian)
4. Alexandersson, H. 1986. A homogeneity test applied to precipitation data. *Inter. J. Climatol.* 6: 6. 661-675.
5. Alizadeh, A. 2010. Principles of Applied Hydrology. Imam Reza University Press, 911p. (In Persian)

6. Anbari, M.J., and Tabesh, M. 2016. Failure Event Probability Calculation in Wastewater Collection Systems Using the Bayesian Network. *J. Water Wastewater*. 3: 48-61. (In Persian)
7. Bishop, C.M. 2006. Pattern recognition and machine learning. Springer.
8. Download.hugin.com. 2017. Introduction to the Hugin Development Environment / Manual. [online] Available at: http://download.hugin.com/webdocs/manuals/Htmlhelp/descr_NPC_algorithm_pane.html [Accessed 31 Aug. 2017].
9. Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., and Rubin, D.B. 2014. Bayesian data analysis (Vol. 2). Boca Raton, FL: CRC press.
10. Ghgajarnia, N., Liyaghat, A., and Danesh Kar Araste, P. 2014. Verifying precipitation data of TAMAB and meteorology institute in Urmia basin. *Water and Soil resources conservation* 4: 1. 91-109. (In Persian)
11. Ghorbani, A.M., and Dehghani, R. 2014. Application of Bayesian Neural Networks, Support Vector Machines and Gene Expression Programming Analysis of Rainfall - Runoff Monthly (Case study: Kakareza River). *J. Irrig. Sci. Engin.* 39: 2. 125-138. (In Persian)
12. Hesar, A.S., Tabatabaee, H., and Jalali, M. 2012. Structure learning of Bayesian networks using heuristic methods. In Proc. of International Conference on Information and Knowledge Management (ICIKM 2012).
13. Kardan Moghadam, H., and Roozbahani, A. 2015. Evaluation of Bayesian networks model in monthly groundwater level prediction (Case study: Birjand aquifer). *J. Water Irrig. Manage.* 5: 2. 139-151. (In Persian)
14. Kelts, K., and Shahrabi, M. 1986. Holocene sedimentology of hypersaline Lake Urmia, northwestern Iran. *Paleogeography, Paleoclimatology, Paleocology*, 54 (1-4): 105-130.
15. Koop, G. 2010. Bayesian econometrics. Chichester [u.a.]: Wiley.
16. Lin, G.F., Chen, G.R., Wu, M.C., and Chou, Y.C. 2009. Effective forecasting of hourly typhoon rainfall using support vector machines. *Water Resources Research*, 45: 8.
17. Madadgar, S., and Moradkhani, H. 2014. Spatio-temporal drought forecasting within Bayesian networks. *J. Hydrol.* 512: 134-146.
18. Mahmoudi, Kh.S., and Rezaie, H. 2013. Investigating and evaluating causes of fluctuations in Lake Urmia water level and providing various solutions. Urmia University, Faculty of Agriculture. 233p. (In Persian)
19. Mahsafar, H., Maknoun, R., and Saghafian, B. 2011. The Impact of Climate Change on Urmia Lake Water Level. *J. Iran Water Resour. Res.* 7: 1. 47-58. (In Persian)
20. Margaritis, D. 2003. Learning Bayesian Network Model Structure. Ph.D. Thesis. School of Computer Science, Carnegie Mellon University.
21. Mehdizadeh, S., Behmanesh, J., and Saadatnejad Gharahassanlou, H. 2017. Evaluation of gene expression programming and Bayesian networks methods in predicting daily air temperature. *J. Agric. Meteorol.* 4: 2. 1-13. (In Persian)
22. Mellit, A., Massi Pavan, A., and Benghanem, M. 2013. Least squares support vector machine for short term prediction of meteorological time series. *Theor. Appl. Climatol.* 111: 297-307.
23. Mohajerani, H., Kholghi, M., Mosaedi, A., Saidodin, A., and Meftah Halaghi, M. 2013. Quantitative Management of Groundwater Using Bayesian Decision Network. *J. Water Soil.* 26: 6. 1522-1534. (In Persian)
24. Moravaj, M., Khalili, K., and Behmanesh, J. 2016. Forecasting Lake Urmia water level using linear time series models. *J. Water Soil Cons.* 22: 5. 287-296. (In Persian)
25. Noori, R., Karbassi, A.R., Moghaddamnia, A., Han, D., Zokaei-Ashtiani, M.H., Farokhnia, A., and Gousheh, M.G. 2011. Assessment of input variables determination on the SVM model performance using PCA, Gamma test and forward selection techniques for monthly stream flow prediction. *J. Hydrol.* 401 (3-4): 177-189.
26. Pang, A.P., and Sun, T. 2014. Bayesian networks for environmental flow decision-making and an application in the Yellow River estuary, China. *Hydrology and Earth System Sciences*, 18 (5): 1641.

27. Parkes, B., and Demeritt, D. 2016. Defining the hundred-year flood: A Bayesian approach for using historic data to reduce uncertainty in flood frequency estimates. *J. Hydrol.* 540, 1189-1208.
28. Seifi, A. 2011. Develop an expert system to predict daily reference evapotranspiration using a backup vector machine (SVM) and compare its results with ANFIS, ANN and empirical methods. Tarbiat Modarres University, Faculty of Agriculture. 153p. (In Persian)
29. Subramanya, K. 2013. *Engineering Hydrology*, 4e. Tata McGraw-Hill Education.
30. Vapnik, V. 1998. *Statistical Learning Theory*. John Wiley&Sons. Inc., New York.
31. Xue, J., Gui, D., Zhao, Y., Lei, J., Zeng, F., Feng, X., and Shareef, M. 2016. A decision-making framework to model environmental flow requirements in oasis areas using Bayesian networks. *J. Hydrol.* 540: 1209-1222.



Application of bayesian network and LS-SVM methods in predicting water surface level of Urmia Lake

S. Karimzadegan¹, *J. Behmanesh² and H. Rezaie²

¹M.Sc. Graduate, Dept. of Water Engineering, University of Urmia, Urmia, Iran,

²Associate Prof., Dept. of Water Engineering, University of Urmia, Urmia, Iran

Received: 12.05.2017; Accepted: 05.14.2018

Abstract

Background and Objectives: Urmia Lake, as an important water ecosystem, is located in the northwest of Iran. Over the past 14 years, the average of Urmia Lake water surface level has decreased to 1272.2 m and it means that the difference between the ecological level of the lake and the present water level is 2-meter. Drying of the Urmia Lake will cause serious problems and crises for the watershed, adjacent provinces and the country. In this research, direct and indirect effective parameters in prediction of the lake level such as evaporation, input runoff to the lake, precipitation, temperature, wind, mean air humidity and water level in the previous month have been employed. The comparison of the efficiency of two models namely the Bayesian network, which is a probabilistic model under conditions of uncertainty and the machine algorithm of Least Square Support Vector Machine (LS-SVM) is the main objective of the present research.

Materials and Methods: In the present research, two methods including least squares support vector machine and Bayesian were employed in the modeling process. In this study, effective factors in predicting Urmia lake water level in the previous month as inputs for the used models and water level in the present month as outputs were investigated.

Results: The analysis of hydrometric stations data showed that the data of only 4 stations among 13 stations are fitted to the normal distribution. Comparison and investigation of the results of two model by examining the coefficients R^2 , RMSE, MBE and Nash-Sutcliff indicated superiority of the least squared vector machine model compared to the Bayesian network. These values for the superior model were 92.3%, 0.082, 0.122 and 0.86, respectively.

Conclusion: In this research, the criteria for evaluating and comparing the accuracy of the two model in prediction based on R^2 , RMSE, MBE and Nash-Sutcliff criteria indicated that the least square vector machine is superior to the Bayesian model. However, the important point in comparing the two model is that the nature of the least squares support vector machine is a machine, however, the nature of the Bayesian network is a probabilistic model under uncertainty conditions in which the normal distribution is used to train network variables because the nature of events is random. Therefore, in this research, the use of Bayesian network model is recommended to the least-squares support vector machine model.

Keywords: Bayesian network, Ecological level, Least Square Support Vector Machine, Prediction, Surface level

* Corresponding Author; Email: j.behmanesh@urmia.ac.ir

Ar