



پارامترهای هواشناسی اثرگذار بر مصرف آب در بخش خانگی شهر قم

نویسندگان :

قاسم امینی، زهره سعیدی

دریافت: ۹۵/۷/۱۴

پذیرش: ۹۵/۱۲/۱۴

ناشر : مهندسين مشاور طرح و تحقيقات آب و فاضلاب

قرارگيری در نوبت چاپ : سال ۹۶

برای ارجاع به این مقاله به صورت زیر عمل شود:

{ اسم نویندگان، (سال)، "عنوان مقاله"، مجله آب و فاضلاب، قرار گرفته در نوبت چاپ سال ۹۶، به نشانی: wwcerd.com }

«حرکت استاده از مطالب این مقاله فقط با ذکر نام مجاز است»

پارامترهای هواشناسی اثرگذار بر مصرف آب در بخش خانگی شهر قم

قاسم امینی^۱، زهره سعیدی^۲

۱- کارشناس ارشد آمار، شرکت آب و فاضلاب شهری استان قم
(نویسنده مسئول) ghasem2094.amini@gmail.com

۲- کارشناس ارشد آمار اقتصادی و اجتماعی، دانشکده علوم ریاضی و کامپیوتر، دانشگاه شهید چمران اهواز

(دریافت ۹۵/۷/۱۴ پذیرش ۹۵/۱۲/۱۴)

چکیده

پیش‌بینی مصرف آب و شناسایی عوامل مؤثر بر آن، از گام‌های مهم در مدیریت بحران آب است. پژوهش‌های انجام شده نشان می‌دهد پارامترهای هواشناسی به عنوان مهم‌ترین گروه در زمینه برآورد کوتاه مدت مصرف آب در نظر گرفته شده‌اند. در این پژوهش با استفاده از روش رگرسیون و ارون قطعه‌ای خوشه‌بندی شده به شناسایی متغیرهای هواشناسی اثرگذار بر مصرف آب خانگی شهر قم پرداخته شد. با استفاده از این روش علاوه بر کاهش بعد، می‌توان مسئله همخطی را نیز رفع نمود. داده‌ها شامل هفت پارامتر هواشناسی و مصرف آب خانگی ماهانه طی سال‌های ۱۳۸۰ تا ۱۳۹۲ بود. تحلیل داده‌ها نشان داد که می‌توان به جای هفت متغیر اولیه تنها از دو مولفه جدید که ترکیبی خطی از متغیرهای مستقل هستند، استفاده نمود. در مؤلفه اول حداکثر سرعت وزش باد و رطوبت نسبی دارای بار منفی (۰/۷۵۷) و (۰/۴) و در مؤلفه دوم میانگین حداقل دما دارای بار منفی (۰/۷۵۳) و متوسط دمای هوا با بار مثبت (۰/۶۳۴) بیشترین تأثیر را بر مؤلفه‌ها داشتند. نتایج رگرسیون روی این متغیرها، معنی‌داری حداقل میانگین دما با ضریب ۰/۱۸ و حداکثر سرعت وزش باد با ضریب ۰/۰۰۴- و ضریب تعیین ۹۲ درصد را نشان داد. مقایسه روش پیشنهادی در این پژوهش با روش معمول آنالیز مؤلفه اصلی در تحلیل داده‌های چند متغیره، نشان از خطای کمتر رگرسیون و ارون قطعه‌ای خوشه‌بندی شده دارد و همچنین با توجه به تأثیر همخطی بر نتایج شبکه عصبی، روش ارائه شده عملکرد بهتری نسبت به روش‌های معمول پیش‌بینی مصرف آب دارد.

واژه‌های کلیدی: کاهش بعد، رگرسیون و ارون قطعه‌ای، خوشه‌بندی، همخطی

۱- مقدمه

تقاضا به تغییرات این عوامل است. در این پژوهش با استفاده از روشی جامع در کاهش بعد رگرسیون به نام رگرسیون و ارون قطعه‌ای بر مبنای خوشه‌بندی به شناسایی پارامترهای جو مؤثر که اثر کوتاه مدت بر مصرف آب برای مشترکان خانگی شهر قم دارند، پرداخته شده است.

در دهه‌های اخیر پژوهش‌های متنوعی در زمینه پیش‌بینی میزان مصرف آب در دوره‌های زمانی کوتاه مدت، میان مدت و بلند مدت بر اساس پارامترهای جمعیتی و جو صورت گرفته است

(Wong, 1972; Young, 1973; Willsie & Pratt 1974).

میانو در سال ۱۹۹۰ در پژوهش خود بیان کرد که بارندگی تأثیری پویا بر مصرف دارد و در زمان بارندگی تقاضای آب کاهش یافته و این تأثیر در طول زمان کم و کمتر خواهد شد (Miaou 1990).

جو و همکاران در سال ۲۰۰۲ و همچنین استارک و همکاران در سال ۲۰۰۰ با استفاده از شبکه عصبی، اثر متغیرهای هواشناسی

استان قم به دلیل موقعیت جغرافیایی خود و واقع شدن در کویر مرکزی ایران با اقلیم خشک و میزان بارندگی کم با مشکل کم آبی مواجه است. از طرفی این استان با توجه به موقعیت خاص استراتژیکی، سیاسی و مذهبی یکی از استان‌هایی است که تمرکز جمعیت در آن به شدت رو به افزایش است، به طوری که پیش‌بینی می‌شود جمعیت آن در افق سال ۱۴۰۰ برابر ۱۵۰۰۰۰ نفر باشد، (PBO, 1998). همین موضوع خود به خود رشد چشمگیر صنعت و در نتیجه تقاضای آب را به دنبال داشته است. به همین دلیل امروزه توجه بیشتر به مسائل اقتصادی آب و همچنین بررسی دقیق‌تر مصرف آب در شهر قم، بسیار بیشتر از گذشته ضروری به نظر می‌رسد. با مدیریت تقاضا از طریق مکانیسم‌های اقتصادی هم چون سیاست‌های قیمتی و غیر قیمتی تا حدودی از زیان‌ها پیشگیری می‌شود. لازمه این امر شناسایی عوامل مؤثر و شدت تأثیرگذاری آنها بر روی مصرف و نیز بررسی چگونگی واکنش

داشته است. در آن پژوهش مشخص شد که با افزایش یک میلی‌متر در متوسط بارندگی سالانه، میزان مصرف سرانه مشتریان در حدود یک درصد کاهش می‌یابد و همچنین افزایش یک درصدی متوسط رطوبت نسبی، افزایش ۰/۷ درصد مصرف را به دنبال خواهد داشت (Amini et al. 2015).

در یک مسئله رگرسیونی با تعداد زیادی متغیر مستقل تحلیل، اغلب با مشکلاتی از قبیل زمان اجرای عملیات و محاسبات، محدودیت اساسی برخی روش‌ها، وجود همخطی به دلیل همبستگی قوی بین متغیرها مواجه خواهد شد. از طرفی به‌منظور نمایش الگوی تصویری از داده‌ها مناسب تر است که بعد مسائل به دو یا سه کاهش داده شود؛ مؤلفه اصلی روش سنتی است که به‌منظور کاهش ابعاد استفاده می‌شود. مؤلفه‌های جدید که ترکیبات خطی از متغیرهای اولیه می‌باشند، متعامد و مستقل از یکدیگر هستند به‌همین دلیل در مسایل رگرسیونی که دچار همخطی چندگانه می‌شوند، استفاده از مؤلفه‌های اصلی به جای متغیرهای اولیه راه حلی مناسب است. ولی در این روش کاهش تعداد متغیرهای مستقل کاملاً از متغیر وابسته y بی‌نیاز است. لذا برای دو y متفاوت، آنالیز مؤلفه‌های اصلی، داده‌ها را به ترکیبات خطی یکسانی از x کاهش خواهد داد. همخطی، روش‌های نوین مانند شبکه عصبی مصنوعی را نیز تحت تأثیر قرار داده و باعث پایین آمدن دقت پیش‌بینی مدل شبکه عصبی شده و برآورد وزن‌ها در لایه‌های مختلف شبکه عصبی در هر بار تکرار الگوریتم آموزش، با تغییرات زیادی مواجه می‌شود که این مسئله ممکن است باعث عدم همگرایی شبکه عصبی مصنوعی شود. این مشکل به ویژه زمانی که تعداد متغیرهای ورودی زیاد باشد، بسیار محتمل است (Menhaj 2005). راه حل دیگر برطرف نمودن همخطی، حذف متغیرهایی است که بیشترین تأثیر را بر همخطی دارند ولی این روش نیز هر چند معمولاً استفاده می‌شود، به دلیل حذف اطلاعات و متغیرها (بعضاً متغیرهای ضروری) مناسب نبوده و باید از رگرسیون مؤلفه‌های اصلی استفاده کرد که از ضعف در نظر نگرفتن مقادیر پاسخ برخوردار است. لذا استفاده از روشی که چنین نقاط ضعفی را پوشش دهد، نتایج دقیق و بهتری را ارائه خواهد نمود.

با توجه به مطالب ذکر شده، پژوهش حاضر به معرفی روشی در زمینه رگرسیون و کاهش بعد به نام رگرسیون وارون قطعه‌ای خوشه‌بندی شده خواهد پرداخت که در آن مولفه‌های جدید که

بر مصرف آب را بررسی کرده‌اند (Stark et al. 2000; Jooet al. 2002).

تابش و دینی در سال ۱۳۸۴ روش منطق فازی و تابش و همکاران در سال ۱۳۸۵ روش نروفازی سوگنو را برای پیش‌بینی کوتاه مدت مصرف آب تهران با استفاده از متغیرهای هواشناسی مورد مطالعه قرار دادند (Tabesh & Dini 2005; Tabesh et al. 2006).

تابش و همکاران در سال ۱۳۸۶ با استفاده از شبکه‌های عصبی مصنوعی و متغیرهایی مانند حداقل و حداکثر دما، بارش روزهای قبل، سرعت باد در فصل بهار، مصرف روز قبل و میزان رطوبت در زمستان به مدل‌سازی مصرف آب تهران پرداخته‌اند (Tabesh et al. 2007).

تابش و همکاران همچنین در سال ۱۳۸۶ با استفاده از روش‌های سنتی و پیشرفته سری زمانی به برآورد مصرف روزانه شهر تهران پرداخته‌اند (Tabesh et al. 2002).

در پژوهش دیگری در سال ۲۰۰۹ با استفاده از رگرسیون به بررسی متغیرهای هواشناسی بر مصرف آب شهر سئول پرداخته شد و مشخص شد متغیرهای بیشینه دما، اثر مثبت و سرعت باد، اثر منفی بر مصرف آب دارند (Praskiewicz & Chang 2009).

تابش و دینی در سال ۱۳۸۹ با استفاده از شبکه عصبی مصنوعی و در نظر گرفتن پارامترهای هواشناسی و مصرف روزهای قبل به برآورد مصرف آب در تهران پرداخته و دریافتند میانگین دمای روزانه و درصد رطوبت نسبی مؤثرترین پارامترهای هواشناسی بر میزان مصرف آب می‌باشند (Tabesh & Dini 2010).

سجادی‌فر و خیابانی در سال ۱۳۹۰ در مدل‌سازی تقاضای آب خانگی اراک اثر منفی بارندگی را بر مصرف آب نتیجه گرفتند (Sajadifar & Khiabani, 2011).

در سال ۲۰۱۲ با استفاده از شبکه عصبی و رگرسیون خطی و غیرخطی به بررسی متغیرهای بیشینه دما و مصرف در سه روز قبل پرداخته شد (Adamowski et al. 2012). در سال ۲۰۱۳ در پژوهشی به تأثیر مثبت سرعت باد بر مصرف آب مشترکان خانگی شهر واشون پرداخته شد (Lott et al. 2013). باکر و همکاران در سال ۲۰۱۴ با استفاده از رگرسیون خطی و متغیرهای متوسط دمای روزانه و بارندگی به مدل‌سازی مصرف روزانه آب پرداخته‌اند (Bakker et al. 2014). امینی در سال ۲۰۱۵ در پژوهش خود به تأثیر منفی بارندگی و اثر مثبت رطوبت نسبی بر مصرف آب اشاره

اغلب کاهش‌های خطی بسنده، توسط ترکیب‌های خطی ساخته شده‌اند یعنی $R(x) = \beta^T x$ است که β یک ماتریس $P \times K$ است؛ که اگر معادله ۱ برقرار باشد، زیر فضای $S(\beta)$ یک زیر فضای کاهش بعد^۲ (DRS) برای رگرسیون y روی x نامیده می‌شود.

۲-۲- رگرسیون وارون^۳

مدل‌های رگرسیون معمولاً به وسیله تجزیه توزیع توأم X و Y به شکل $h(y|x)k(x)$ و مدل‌بندی بر اساس $h(y|x)$ تشکیل می‌شوند. مدل رگرسیون وارون نیز از تجزیه توزیع توأم X و Y به صورت $h(x|y)k(y)$ و مدل‌سازی روی $h(x|y)$ تشکیل می‌شود. وقتی y تغییر می‌کند، $E[X|y]$ منحنی را ترسیم می‌نماید که منحنی رگرسیون وارون در R^P نامیده می‌شود (Li 1991).

$E(x|y) - E(x)$ را منحنی رگرسیون وارون مرکزی می‌گویند. به فضای تولید شده توسط $E(x|y) - E(x)$ ، زیر فضای وارون گفته می‌شود که آن را با $S_{E(x|y)}$ نشان می‌دهند. زیر فضای وارون تحت شرط زیر که معروف به شرط خطی بودن است، در یک زیر فضای K بعدی واقع می‌شود.

به ازای هر b در R^P ، $E(b^T x | \beta_1^T x, \dots, \beta_K^T x)$ در $\beta_1^T x, \dots, \beta_K^T x$ خطی است به عبارت دیگر برای ثابت‌های $c_0, c_1, \dots, c_K$:

$$E(b^T x | \beta_1^T x, \dots, \beta_K^T x) = c_0 + c_1 \beta_1^T x + \dots + c_K \beta_K^T x \quad (2)$$

ایتان در سال ۱۹۸۶ نشان داد این شرط برقرار است اگر و تنها اگر X متغیری تصادفی و بیضوی شکل باشد (Eaton 1986).

لی در سال ۱۹۹۱ نشان داد اگر $z = \Sigma_x^{-\frac{1}{2}}[x - E(x)]$ فرم استاندارد شده x باشد که Σ_x و $E(x)$ به ترتیب ماتریس کواریانس و میانگین x می‌باشند، آنگاه

$$S_{E(z|y)} \subset S\left(\Sigma_x^{-\frac{1}{2}}\beta\right) = S_{y|z} = \Sigma_x^{-\frac{1}{2}}S_{y|x} \quad (3)$$

ترکیب خطی از متغیرهای اولیه می‌باشند، جدا از متغیر پاسخ نبوده و با تصویرسازی بهتر داده‌ها به دلیل کاهش ابعاد بر مشکل همخطی نیز به خوبی غلبه می‌کند و باعث افزایش دقت نتایج خواهد شد. در پژوهشی در سال ۲۰۰۰ با استفاده از شبیه‌سازی نشان داده شد که روش وارون قطعه‌ای نسبت به روش رگرسیون مولفه‌های اصلی کارا تر است و بعد به دست آمده از این روش کمتر از روش رگرسیون مؤلفه‌های اصلی خواهد بود.

۲- مواد و روش‌ها

۲-۱- کاهش بعد بسنده^۱

در نیمه دوم قرن نوزدهم کاهش مشاهدات به عنوان یک هدف اصلی در علم آمار به حساب می‌آمد. فیشر در سال ۱۹۲۲ مفهوم آماره بسنده را به معنای کاهش بدون از دست دادن اطلاعات مطرح کرد (Fisher 1922).

در سال‌های اخیر کاهش بعد بسنده، پایه‌ای برای روش‌های کاهش بعد در رگرسیون به حساب می‌آید. مفهوم کاهش بعد بسنده از اهمیت بسیاری برخوردار است. این مفهوم بیانگر وضعیت مطلوبی است که می‌تواند بعد را کاهش دهد بدون اینکه هرگونه اطلاعات ضروری که در پیش بینی y لازم است را از دست بدهد.

کاهش بعد بدون از دست دادن اطلاعات مفهوم بسندگی را تداعی می‌کند. استفاده از بسندگی در رگرسیون منجر به تعریف کاهش بعد بسنده شده است. استفاده از روش‌های کاهش بعد در رگرسیون، نیازمند دانشی درباره زیر فضای کاهش بعد است که در ادامه به معرفی آن پرداخته شده است.

کووک در سال ۱۹۹۸ نشان داد در فضای رگرسیونی، تابع $R(x) \in \mathbb{R}^K$ ، $(K \leq p)$ ، از پیشگویی P ، بعدی بسنده است اگر توزیع y به شرط x همان توزیع y به شرط $R(x)$ باشد

$$F(y|x) = F(y|R(x)) \quad (1)$$

که در آن $F(\cdot|\cdot)$ تابع توزیع شرطی است (Cook 1998).

این معادله به این معنا است که x و y زمانی که $R(x)$ معلوم باشد از هم مستقل‌اند و با جایگزینی $R(x)$ به جای x اطلاعات رگرسیونی از دست نمی‌روند و Cook آنرا با نماد $Y | \beta^T X \perp X$ نشان داد.

²Dimension Reduction Subspace

³ Inverse regression

¹Sufficient Dimension Reduction

(۵) محاسبه $\hat{\beta}_j = \Sigma_x^{-1} \hat{\eta}_j$, $j = 1, \dots, K$, که $\hat{\eta}_j$ ها بردارهای ویژه متناظر با K مقدار ویژه بزرگ ماتریس $\hat{V}$ می‌باشند.

(۶) محاسبه مؤلفه‌های رگرسیون وارون قطعه‌ای به صورت زیر

$$SIR_j = \hat{\beta}_j^T x, \quad j = 1, \dots, K \quad (۷)$$

۴-۲- تعیین بعد

بعد زیر فضای مرکزی که در اینجا با K نشان داده می‌شود، اغلب نامعلوم است و با آزمون‌های متوالی زیر به ازای مقادیر مختلف d تعیین می‌شود

$$H_0: K = d \quad \text{vs} \quad H_1: K > d \quad (۸)$$

بررسی این آزمون‌ها با $d = 0$ شروع می‌شود و تا زمانی که H_0 رد نشود ادامه خواهد داشت.

آماره‌ی این آزمون‌ها برابر با $\bar{\lambda}_{(P-K)} = n(P-K)\bar{\lambda}_{(P-K)}$ است که در آن $\bar{\lambda}_{(P-K)} = \sum_{i=K+1}^P \frac{\hat{\lambda}_i}{P-K}$ میانگین $P-K$ مقدار ویژه کوچک‌تر ماتریس $\hat{V}$ است. آماره t_K به‌طور مجانبی دارای توزیع کای اسکور^۲ می‌باشد (Becker & Gather 2007)

۵-۲- تعداد قطعه‌ها

لی در سال ۱۹۹۱ بیان کرد که تعداد قطعه‌ها در روش رگرسیون وارون قطعه‌ای اهمیت ندارد اما در مطالعه‌ای شبیه‌سازی شده توسط بکر و گتر نشان داده شد که تعداد قطعه‌ها کاملاً بر برآوردهای رگرسیون وارون قطعه‌ای تأثیر دارد و لذا باید با احتیاط انتخاب شود (Becker & Gather 2007; Li 1991). تعداد قطعه‌ها نباید خیلی بزرگ باشد (Li 1991). طبق پژوهش بکر و گتر ایشان انتخاب $H \approx 0.1n$ را مطلوب می‌دانند و اشاره دارند که هنوز هم انتخاب مناسب تعداد قطعه‌ها نیاز به بررسی بیشتر دارد.

۶-۲- الگوریتم رگرسیون وارون قطعه‌ای بر مبنای خوشه‌بندی

برقراری شرط خطی بودن از پیش نیازهای تعداد زیادی از روش‌های کاهش بعد است. در عمل بررسی این شرط به علت

از معادله ۳ نتیجه می‌شود برای برآورد زیر فضای مرکزی می‌توان از زیر فضای وارون مرکزی استاندارد $S_{E(z|y)}$ استفاده کرد. معادله ۴ یک راه حل برای برآورد $S_{E(z|y)}$ ارائه می‌دهد (Li 1991).

$$S[\text{cov}(E(z|y))] = S_{E(z|y)} \quad (۴)$$

بنابراین می‌توان برای برآورد $S_{E(z|y)}$ از برآورد $\text{cov}(E(z|y))$ استفاده کرد و در نهایت زیر فضای مرکزی را برآورد کرد. بردارهای ویژه متناظر با K مقدار ویژه بزرگ‌تر ماتریس $\text{cov}(E(z|y))$ پایه‌ای برای زیر فضای وارون استاندارد هستند.

۳-۲- الگوریتم رگرسیون وارون قطعه‌ای

لی در سال ۱۹۹۱ با استفاده از قطعه‌بندی متغیر پاسخ، $\text{cov}(E(z|y))$ را برآورد کرد. این روش معروف‌ترین روش کاهش بعد در رگرسیون است که رگرسیون وارون قطعه‌ای نام دارد. اگر $(y_1, x_1), \dots, (y_n, x_n)$ مجموعه داده‌های اولیه با $p+1$ متغیر و n مشاهده باشند، الگوریتم رگرسیون وارون قطعه‌ای شامل مراحل زیر است

(۱) استاندارد کردن مشاهدات به صورت $\hat{z}_i = \Sigma_x^{-1/2}(x_i - \bar{x})$ که $\bar{x}$ و $\hat{z}_i$ به ترتیب میانگین و واریانس نمونه‌ای پیشگوا می‌باشند.

(۲) مرتب کردن داده‌ها بر اساس متغیر y به صورت صعودی و تقسیم داده‌ها، متناظر با تقسیم دامنه y به H قطعه^۱ $I_1, \dots, I_H$.

(۳) محاسبه میانگین $\hat{z}_i$ ها در هر قطعه به صورت

$$\hat{Z}_h = \frac{\sum_{\{i: y_i \in I_h\}} \hat{z}_i}{n_h} \quad h=1, \dots, H \quad (۵)$$

که n_h تعداد مشاهدات در هر قطعه است. به‌طور معمول تعداد مشاهدات در هر قطعه تقریباً برابر گرفته می‌شود یعنی $n_h = \frac{n}{H}$.
(۴) به‌دست آوردن بردارهای ویژه و مقادیر ویژه ماتریس کواریانس وزنی زیر

$$\hat{V} = \frac{1}{n} \sum_{h=1}^H n_h \hat{Z}_h \hat{Z}_h^T \quad (۶)$$

$\hat{V}$ برآوردی از $\text{cov}(E(z|y))$ است.

² Chi- Square

¹ Slice

زیر برای برآورد تعداد بهینه خوشه‌ها استفاده می‌شود (Kuentz & Saracco 2010).

$$\hat{c}^* = \arg \min_{c=1, \dots, C} \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_{i,[c]})^2 \quad (9)$$

که در آن

$$\hat{y}_{i,[c]} = \sum_{j=1}^n y_j \mathcal{K} \left(\frac{x_i^T \hat{b}_{[c]} - x_j^T \hat{b}_{[c]}}{h_c} \right) / \sum_{j=1}^n \mathcal{K} \left(\frac{x_i^T \hat{b}_{[c]} - x_j^T \hat{b}_{[c]}}{h_c} \right) \quad (10)$$

یک برآورد کرنل برای $E(y|x_i^T \hat{b}_{[c]})$ است که h_c پهنای باند پارامتر برای خوشه c است و $\mathcal{K}$ یک تابع کرنل است. برای تعیین بعد در این روش از بعدی که از روش رگرسیون واریون قطعه‌ای به دست می‌آید، استفاده می‌شود. به منظور اخذ نتایج از برنامه‌نویسی در نرم‌افزار آماری R استفاده شد.

۲-۸- داده‌های پژوهش

اطلاعات مورد بررسی در این پژوهش شامل متغیرهای مستقل هواشناسی از قبیل میانگین دمای هوا بر حسب درجه سلسیوس، میانگین حداکثر دما، میانگین حداقل دما، بارندگی بر حسب میلی‌متر، درصد رطوبت نسبی، ساعت آفتابی، حداکثر سرعت وزش باد (متر بر ثانیه) و متغیر پاسخ لگاریتم مصرف مشترکان خانگی شهر قم به صورت ماهانه در سال‌های ۸۰ تا ۹۲ می‌باشند. ماتریس همبستگی متغیرهای مستقل در جدول ۱ نشان داده شده است.

با توجه به جدول ۱ و بالا بودن ضرایب همبستگی خطی بین متغیرها، وجود همخطی قوی بین متغیرها به وضوح مشهود است و در صورت استفاده از روش‌های معمول باید برخی از متغیرها را حذف نمود که منجر به از دست دادن اطلاعات خواهد شد.

۳- نتایج و بحث

۳-۱- مؤلفه‌های رگرسیون واریون قطعه‌ای

با استفاده از نرم‌افزار آماری R و بسته dr که توسط ویزبرگ در سال ۲۰۰۹ ارائه شده است، رگرسیون واریون قطعه‌ای اجرا شد (Weisberg 2009). در جدول ۲ که بر اساس آزمون‌های متوالی است، دو آزمون اول با توجه به مقدار p_value کمتر از ۰/۰۰۱

وابسته بودن به پارامترهای مجهول امکان‌پذیر نیست. از دیدگاه کاربردی و نظری این شرط یک محدودیت به حساب می‌آید. لی و همکاران در سال ۲۰۰۴ با استفاده از مطالعات شبیه‌سازی نتیجه گرفتند که اگر داده‌ها را خوشه‌بندی کنند، شرط خطی بودن در هر خوشه برقرار خواهد بود. آنها روش خوشه‌بندی را برای اولین بار در روش کمترین توان‌های دوم عادی به کار بردند (Li et al 2004). کونتز و ساراکو در سال ۲۰۱۰ از این روش بهره گرفتند و رگرسیون واریون قطعه‌ای بر مبنای خوشه‌بندی را ارائه دادند (Kuentz & Saracco. 2010).

حال چنانچه x به خوشه $x^{(1)}, \dots, x^{(c)}$ افراز شده باشد که c عدد ثابتی است. مطابق با افرازبندی x, y نیز به افرازهای $(x^{(l)}, y^{(l)})$ ، $l = 1, \dots, c$ ، تقسیم می‌شوند. شرط خطی بودن در هر خوشه ثابت است در واقع برای هر خوشه به ازای هر $b \in \mathcal{R}^P$ ، $E(b^T x^{(l)} | \beta_1^T x^{(l)}, \dots, \beta_K^T x^{(l)})$ نسبت به $\beta_1^T x^{(l)}, \dots, \beta_K^T x^{(l)}$ خطی است.

الگوریتم رگرسیون واریون قطعه‌ای بر مبنای خوشه‌بندی شامل دو مرحله اصلی است. ابتدا در هر خوشه یک پایه برای زیر فضای مرکزی توسط رگرسیون واریون قطعه‌ای محاسبه می‌شود و در نهایت با ترکیب آنها پایه‌ای برای زیر فضای مرکزی به دست می‌آید. این دو مرحله به شرح زیر است

در هر خوشه، K بردار ویژه اول ماتریس $(\hat{\Sigma}^{(l)})^{-1} \hat{V}_1^{(l)}$ محاسبه می‌شوند و آن‌ها را با $\hat{b}_1^{(l)}, \dots, \hat{b}_K^{(l)}$ نشان می‌دهند و سپس آن‌ها را در یک ماتریس $B^{(l)} = [b_1^{(l)}, \dots, b_K^{(l)}]$ ذخیره می‌کنند.

$\bar{x}^{(l)} = \frac{1}{n^{(l)}} \sum_{i=1}^{n^{(l)}} x_i^{(l)}$ و $\hat{\Sigma}^{(l)} = \frac{1}{n^{(l)}} \sum_{i=1}^{n^{(l)}} (x_i^{(l)} - \bar{x}^{(l)})(x_i^{(l)} - \bar{x}^{(l)})^T$ می‌باشند. K بردار ویژه اول ماتریس $\hat{B}^{(l)} \hat{B}^{(l)T}$ که با $\hat{b}_1^{(l)}, \dots, \hat{b}_K^{(l)}$ نشان داده می‌شوند، محاسبه می‌شوند. این بردارها در ماتریس $\hat{B}^{(l)} = [\hat{b}_1^{(l)}, \dots, \hat{b}_K^{(l)}]$ ذخیره می‌شوند و دوباره این ماتریس‌ها در یک ماتریس جدید به شکل $\hat{B}^{(c)} = [\hat{B}^{(1)}, \dots, \hat{B}^{(c)}]$ ادغام می‌شوند. K بردار ویژه ماتریس $\hat{B}^{(c)} \hat{B}^{(c)T}$ است. این بردارها که با $\hat{b}_1^{(l)}, \dots, \hat{b}_K^{(l)}$ نشان داده می‌شوند، یک پایه برآورد شده برای زیر فضای مرکزی هستند (Kuentz & Saracco 2010).

۲-۷- تعداد بهینه خوشه

تعداد بهینه خوشه با c^* نشان داده می‌شود. از تعریف معیار تجربی

جدول ۱- ماتریس همبستگی متغیرها

Table 1. Correlation matrix of independent variables

Variable	X	Average temperature	Average maximum temperature	Average minimum temperature	Rainfall	Relative humidity	Sunshine	Maximum wind speed
Average temperature	x ₁	1	-	-	-	-	-	-
Average maximum temperature	x ₂	0.98	1	-	-	-	-	-
Average minimum temperature	x ₃	0.983	0.984	1	-	-	-	-
Rainfall	x ₄	-0.508	-0.546	-0.482	1	-	-	-
Relative humidity	x ₅	-0.832	-0.868	-0.814	0.552	1	-	-
Sunshine	x ₆	0.908	0.915	0.896	-0.632	-0.835	1	-
Maximum wind speed	x ₇	0.075	0.096	0.089	0.076	-0.092	-0.006	1

برای استفاده از رگرسیون وارون قطعه‌ای باید در ابتدا شرط خطی بودن بررسی شود که در عمل غیر ممکن است. اگر شرط خطی بودن برقرار باشد، می‌توان از رگرسیون وارون قطعه‌ای استفاده کرد (Li 1991). ولی اگر یک رابطه غیر خطی آشکار بین متغیرهای مستقل وجود داشته باشد، می‌توان با استفاده از خوشه‌بندی فضای متغیرهای مستقل، به نتایج کاراتری دست یافت (Kuentz & Saracco 2010).

یک راه پیشنهادی برای مشاهده برقراری شهودی شرط خطی بودن، رسم نمودار پراکنش پیشگوهاست. اگر شرط خطی بودن برقرار باشد، این نمودار نباید یک رابطه غیر خطی آشکار را بین پیشگوها نشان دهد. از این نمودار که در شکل ۱ نشان داده شده است، به‌طور شهودی می‌توان نتیجه گرفت که بین برخی متغیرها شرط خطی بودن برقرار نیست؛ بنابراین با خوشه‌بندی فضای پیشگوها می‌توان از برقراری شرط خطی بودن اطمینان داشت و به نتایج مطلوب‌تری دست یافت.

۲-۳- انتخاب تعداد بهینه بعد (مولفه) و خوشه

برای استفاده از روش رگرسیون وارون قطعه‌ای بر مبنای خوشه‌بندی، انتخاب تعداد خوشه بسیار مهم است. برای انتخاب بعد در ابتدا از بعد به‌دست آمده در روش رگرسیون وارون قطعه‌ای استفاده شد و سپس بهینه بودن آن مورد بررسی قرار گرفت. با استفاده از روش رگرسیون وارون قطعه‌ای بعد برابر با دو به دست

معنی‌دار بوده لذا بعد صفر و یک نیست. از طرفی در آزمون سوم مقدار آماره آزمون با ۲۰ درجه آزادی ۲۱/۵۴ و مقدار p-value برابر ۰/۳ است که بیانگر پذیرش فرض صفر است. بنابراین بعد زیرفضای مرکزی (تعداد مؤلفه) برابر با دو خواهد بود. پس می‌توان به جای استفاده از ۷ متغیر مستقل تنها از دو مؤلفه استفاده نمود. این مؤلفه‌ها ترکیب خطی از متغیرهای مستقل و برابر است با

(۱۱)

$$SIR_1 = -0.4 x_1 - 0.19 x_2 - 0.16 x_3 + 0.18 x_4 - 0.41 x_5 - 0.04 x_6 - 0.74 x_7$$

(۱۲)

$$SIR_2 = -0.08 x_1 - 0.63 x_2 + 0.74 x_3 + 0.007 x_4 - 0.121 x_5 + 0.003 x_6 - 0.09 x_7$$

ملاحظه می‌شود که متغیرهای میانگین دمای هوا، رطوبت نسبی و سرعت وزش باد، دارای بیشترین اثر در اولین مؤلفه و متغیرهای میانگین حداکثر دما و میانگین حداقل دما بیشترین اثر را در دومین مؤلفه دارند.

جدول ۲- تعیین بعد زیرفضای مرکزی

Table 2. Determining dimension of central subspace

Hypothesis testing	Statistics	Degrees of freedom	P-value
H ₀ : K = 0 vs H ₁ : K = 1	150.9	42	<.001
H ₀ : K = 1 vs H ₁ : K = 2	52.23	30	<.001
H ₀ : K = 2 vs H ₁ : K = 3	21.54	20	.3

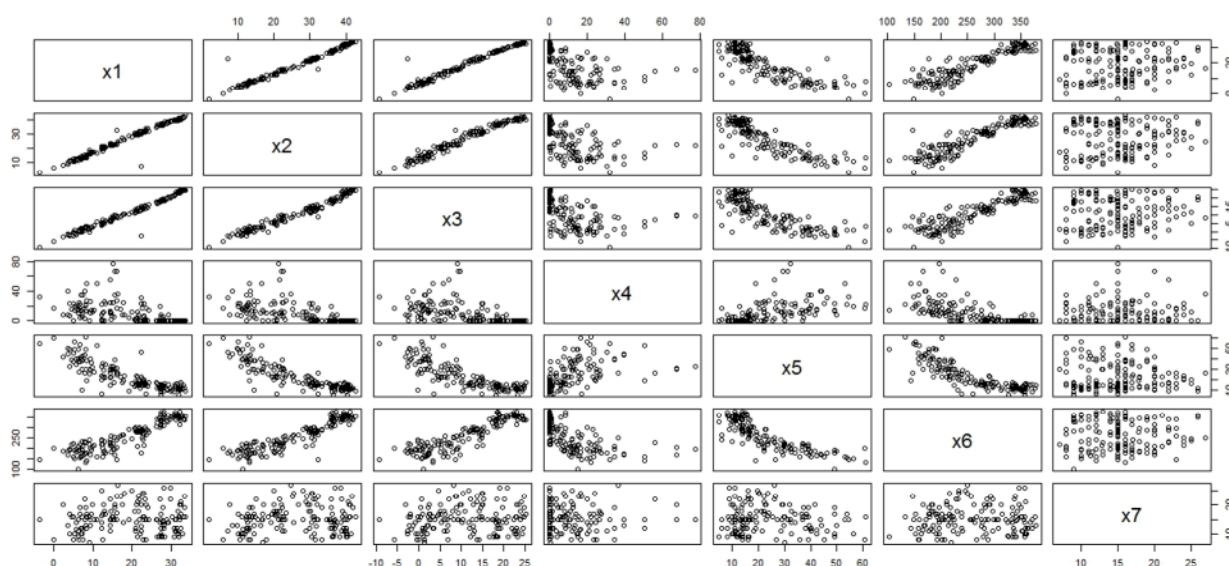


Fig 1. Scattering matrix of independent variables

شکل ۱- ماتریس پراکنش متغیرهای مستقل

جدول ۳- مولفه‌های حاصل از روش CSIR

Table 3. Components of CSIR method

Variable	X	First component	Second component
Average temperature	x ₁	-0.273	0.634
Average maximum temperature	x ₂	-0.326	0.099
Average minimum temperature	x ₃	-0.251	-0.753
Rainfall	x ₄	0.139	-0.013
Relative humidity	x ₅	-0.4	0.119
Sunshine	x ₆	-0.037	-0.003
Maximum wind speed	x ₇	-0.757	-0.088

با توجه جدول ۳ مولفه‌ها برابراند با

$$CSIR_1 = -0.273 x_1 - 0.326 x_2 - 0.251 x_3 + 0.139 x_4 - 0.4 x_5 - 0.037 x_6 - 0.757 x_7 \quad (13)$$

$$CSIR_2 = 0.634 x_1 + 0.099 x_2 - 0.753 x_3 - 0.013 x_4 + 0.119 x_5 - 0.003 x_6 - 0.088 x_7 \quad (14)$$

۳-۳- برازش الگوی رگرسیونی متغیرهای هواشناسی اثرگذار

بر مصرف آب خانگی

با توجه به نتایج به دست آمده ملاحظه شد در اولین مؤلفه حداکثر سرعت وزش باد با ضریب ۷۵۷/۰- و در دومین مؤلفه میانگین

آمد، لذا در ادامه بعد دو در نظر گرفته شد و تعداد بهینه خوشه محاسبه شد. برای تعیین تعداد بهینه خوشه از معادله ۹ استفاده شده است. در بررسی داده‌ها، تعداد قطعه‌ها در هر خوشه برابر با ۷ در نظر گرفته شد. لذا بیشینه تعداد خوشه‌ها برابر با ۵ بود. معیار تجربی برای تعیین تعداد بهینه خوشه محاسبه شد. در شکل ۲ مشاهده می‌شود زمانی که تعداد خوشه‌ها برابر با دو باشد، این معیار کمینه می‌شود، لذا تعداد بهینه خوشه ۲ در نظر گرفته شد.

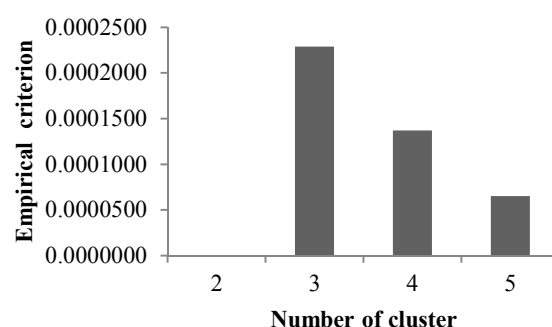


Fig 2. Empirical criterion to determine the optimal number of clusters

شکل ۲- معیار تجربی برای تعیین تعداد بهینه خوشه

مؤلفه محاسبه شده توسط روش رگرسیون وارون قطعه‌ای بر مبنای خوشه‌بندی در حالتی که بعد ۲ و تعداد خوشه نیز ۲ باشد، برآورد شد و نتایج در جدول ۳ نشان داده شد.

میانگین قدر مطلق خطای نسبی (MAPE)، ریشه دوم میانگین مربع خطا (RMSE) و میانگین قدر مطلق خطا (MAD) استفاده شد. نتایج در جدول ۵ نشان داده شده است.

جدول ۵- مقایسه روش رگرسیون وارون قطعه‌ای خوشه بندی شده و

آنالیز مؤلفه اصلی با استفاده از شبکه عصبی

Table 5. Regression analysis for the maximum wind speed and the average minimum temperature variables on consumption logarithm

Measures of accuracy of the fitted model	Method	Test data	Train data
Mean Absolute Percentage Error (MAPE)	CSIR	0.009	0.006
	PCA	0.012	0.007
Root Mean Square Error (RMSE)	CSIR	0.181	0.116
	PCA	0.211	0.128
Mean Absolute Deviation (MAD)	CSIR	0.147	0.091
	PCA	0.19	0.106

با توجه به جدول ۵، در هر سه شاخص خطای اندازه‌گیری شده، مقدار خطای حاصل از روش CSIR کمتر از روش PCA است. بنابراین استفاده از روش رگرسیون وارون قطعه‌ای خوشه بندی شده نتایج دقیق‌تری را نسبت به روش رایج و معمول آنالیز مؤلفه اصلی ارائه می‌دهد. شکل ۳ نمودار مصرف آب خانگی به همراه مقدار برازش یافته توسط روش رگرسیون وارون قطعه‌ای بر مبنای خوشه بندی را نشان می‌دهد.

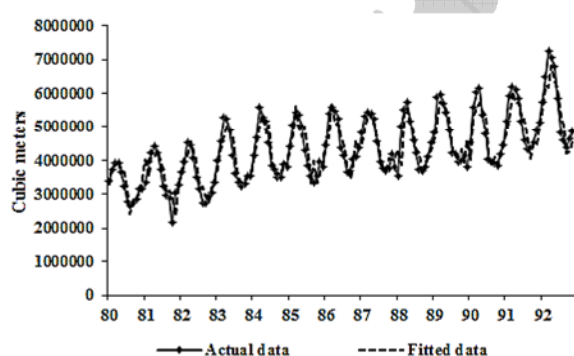


Fig 3. Actual and fitted data for the household water consumption in Qom

شکل ۳- داده‌های واقعی و برازش یافته مصارف خانگی شهر قم

۴- نتیجه‌گیری

به منظور شناسایی متغیرهای جوی اثرگذار بر مصرف آب خانگی در شهر مقدس قم از روش رگرسیون وارون قطعه‌ای بر مبنای

حداقل دما با ضریب 0.753 - از بالاترین ضرایب برخوردار هستند، لذا در ادامه با استفاده از تحلیل رگرسیون به بررسی اثر این دو متغیر بر روی لگاریتم مصرف آب خانگی پرداخته شد. نتایج رگرسیون در جدول ۴ نشان داده شده است. با توجه به جدول ملاحظه می‌شود میانگین حداقل دما با ضریب $0.18/0$ اثر مثبت بر مصرف آب و سرعت وزش باد با ضریب $0.04/0$ - اثر منفی و ناچیزی بر مصرف آب خانگی دارد. مقدار ضریب تعیین مدل برازش یافته 92 درصد است. به عبارتی مدل توانسته 92 درصد از تغییرات لگاریتم مصرف خانگی را به خوبی تبیین نماید. دلیل استفاده از عبارات AR تا مرتبه سه (اتورگرسیو مرتبه سه)، اصلاح خود همبستگی در خطای مدل است که در صورت عدم اصلاح، نتایج فاقد اعتبار است.

جدول ۴- نتایج تحلیل رگرسیون متغیرهای حداکثر سرعت باد و

میانگین حداقل دما بر لگاریتم مصرف آب خانگی

Table 4. Regression analysis for the maximum wind speed and the average minimum temperature variables on consumption logarithm

Model	Coefficient	Standard error	t test	p-value
Intercept	15.23	0.1	142.61	<0.001
average Minimum temperature	0.018	0.009	20.02	<0.001
Maximum wind speed	-0.004	0.001	-3.65	<0.001
AR(1)	0.54	0.12	4.27	<0.001
AR(2)	0.22	0.099	2.28	0.02
AR(3)	0.17	0.08	2.02	0.04

۳-۴- مقایسه روش رگرسیون وارون قطعه‌ای خوشه بندی

شده با روش مؤلفه اصلی

در این بخش به مقایسه روش رگرسیون وارون قطعه‌ای خوشه بندی شده (CSIR) و آنالیز مؤلفه اصلی (PCA) با استفاده از شبکه عصبی پرداخته شد. به این منظور مؤلفه‌های حاصل از دو روش به عنوان ورودی شبکه عصبی در نظر گرفته شدند. سه سال آخر (۹۱ تا ۹۲) به عنوان مجموعه تست و سایر داده‌ها به عنوان مجموعه آموزش لحاظ شده‌اند. در هر دو مقایسه، شبکه عصبی با یک لایه پنهان، دو گره در لایه پنهان، تابع هایپر تانژانت در لایه پنهان و تابع همانی در لایه خروجی با توجه به خطای مجموعه داده تست به عنوان بهترین مدل در نظر گرفته شد. به منظور مقایسه دو روش از سه شاخص

۲ حاصل شد. این دو مؤلفه جدید مستقل از یکدیگر بوده و ترکیب خطی از هفت متغیر مستقل اولیه می‌باشند.

در اینجا علاوه بر رفع مسئله همخطی در متغیرهای اولیه، تعداد متغیرهای جدید نیز به دو کاهش یافته که می‌توانند در تحلیل‌های بعدی استفاده شوند. با توجه به ضرایب بالای میانگین حداقل دما و حداکثر سرعت وزش باد در دو مؤلفه، این دو متغیر اثر قوی بر مؤلفه‌ها دارند لذا به عنوان عوامل اثرگذار بر مصرف می‌توانند در مدل رگرسیونی لحاظ شوند. نتایج تحلیل رگرسیون نشان از اثر مثبت و معنی‌دار میانگین حداقل دما با ضریب $0.1/0$ و اثر منفی و ناچیز و معنی‌دار حداکثر سرعت وزش باد با ضریب $0.04/0$ - بر مصرف آب دارد.

به منظور مقایسه رگرسیون وارون قطعه‌ای خوشه‌بندی شده با روش معمول آنالیز مؤلفه اصلی از روش تلفیقی PCA-ANN استفاده شد. نتایج حاکی از خطای کمتر روش پیشنهادی نسبت به آنالیز مؤلفه اصلی در پیش‌بینی مصرف آب خانگی است. بنابراین در مطالعات با داده چند متغیره که وقوع مسئله همخطی امری اجتناب ناپذیر است، استفاده از روش رگرسیون وارون قطعه‌ای خوشه‌بندی شده نسبت به روش‌های معمول آنالیز مؤلفه اصلی و شبکه عصبی به دلیل عدم تأثیر همخطی بر آن، از عملکرد و کارایی بهتری در حل داده‌های چند متغیره برخوردار است و می‌تواند جایگزین مناسبی برای روش‌های معمول باشد.

خوشه‌بندی استفاده شد. رگرسیون وارون قطعه‌ای یکی از جامع‌ترین روش‌های کاهش بعد در حین کار با داده‌های چند متغیره با ابعاد بالا است. بر اساس مطالعات صورت گرفته کارایی این روش از روش‌هایی مانند رگرسیون مؤلفه اصلی به دلیل استفاده از اطلاعات پاسخ در تشکیل مؤلفه‌ها بسیار بالاتر است.

در این پژوهش متغیرهای جوی میانگین دمای هوا، میانگین حداکثر دما، میانگین حداقل دما، بارندگی، رطوبت نسبی، ساعت آفتابی، سرعت وزش باد، به عنوان متغیرهای توضیحی و مصرف آب خانگی شهر قم به عنوان متغیر پاسخ در نظر گرفته شد. بسیاری از متغیرهای توضیحی از ضریب همبستگی بالایی طبق نتایج پژوهش برخوردار بودند که نشان از مسئله همخطی در داده‌ها دارد. تحلیل چنین داده‌های همخطی بسیار مشکل و در صورت عدم اصلاح نتایج کاملاً مخدوش و نامعتبر به دست می‌آید. استفاده از روش رگرسیون وارون قطعه‌ای بر مبنای خوشه‌بندی در برخورد با چنین داده‌هایی بسیار مناسب است.

در روش رگرسیون وارون قطعه‌ای بر مبنای خوشه‌بندی که از اطلاعات متغیر پاسخ بهره می‌برد، تعداد خوشه نقش مهمی را ایفا می‌کند. با استفاده از معیار تجربی و در نظر گرفتن هفت قطعه در هر خوشه، تعداد بهینه خوشه ۲ و در نهایت با اجرای الگوریتم رگرسیون وارون قطعه‌ای بر مبنای خوشه‌بندی تعداد مؤلفه لازم جهت تحلیل،

References

- Adamowski, J., Fung, C.H., Prasher, S.O., Zielinski, B. & Sliusarieva, A., 2012, "Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal", *Journal of Water Resources Research*, 48, W01528.
- Amini, Gh., 2015, "Estimating household water demand of Qom by artificial neural networks and log linear regression", *1st Water Sciences and Engineering Conference*, Tehran, Iran. (In Persian)
- Bakker, M., Duist, H., Schagen, K., Vreeburg, J. & Rietveld, L., 2014, "Improving the performance of water demand forecasting models by using weather input", *12th International Conference on Computing and Control for the Water Industry*, CCWI2013, Procedia Engineering, 70, 93-102.
- Becker, C. & Gather, U., 2007, "A note on the choice of the number of slices in sliced inverse regression", Working paper, Dept. of Statistic, University of Dortmund, Germany.
- Cook, R.D., 1998, *Regression graphics: Ideas for studying regressions through graphics*. Wiley, New York.
- Eaton, M.L., 1986, "A characterization of spherical distributions", *Journal of Multivariate Analysis*, 20, 272-276.

- Fisher, R.J., 1922, "On the mathematical foundations of theoretical statistics", *Philosophical Transactions of the Royal Statistical Society*, 222, 309-368.
- Joo, C.N., Koo, J.K. & Yu, M.J., 2002, "Application of short-term water demand prediction model to Seoul", *Journal of Water Sciences and Technology*, 46(6-7), 255-261.
- Kuentz, V. & Saracco, J., 2010, "Cluster-based sliced inverse regression", *Journal of Statistical Society*, 39, 251-267.
- Li, K.C., 1991, "Sliced inverse regression for dimension reduction (with discussion)", *Journal of Statistical Association*, 86, 316-327.
- Li, L., Cook, R.D. & Nachtsheim, C.J., 2004, "Cluster-based estimation for sufficient dimension reduction", *Computational Statistics & Data Analysis*, 47, 175-193.
- Lott, C., Tchigriaeva, E. & Rollins, K., 2013, *The effects of climate change on residential municipal water demand in Nevada*, American Geophysical Union.
- Menhaj, M.B., 2005, *Basics of neural network*, Prof Hesabi Inc., Tehran. (In Persian)
- Miaou, S.P., 1990, "A class of time-series urban water demand models nonlinear climatic effects", *Water Resources Research*, 6(2), 169-178.
- Naik, P., Hagerty, M.R. & Tsai, C.L., 2000, "A new dimension reduction approach for data-rich marketing environments: Sliced inverse regression", *Journal of Marketing Research*, 37, 88-101.
- Praskievicz, S. & Chang, H., 2009, "Identifying the relationships between urban water consumption and weather variables in Seoul, Korea", *Journal of Physical Geography*, 30(4), 324-337.
- PBO., 1998, *Qom population projection in the horizon 2021*, Qom, Iran. (In Persian)
- Sajadifar, S.H. & Khiabani, N., 2011, "Modeling of residential water demand using random effect model, case study: Arak city", *Journal of Water and Wastewater*, 22 (3), 59-68. (In Persian)
- Stark, H.L., Stanley, J.S. & Buchanan, I.D., 2000, "The application of artificial neural networks to water demand modeling", *CSCE 28th Annual Conference*, London.
- Tabesh, M. & Dini, M., 2005, "Short-term urban water consumption estimation with fuzzy logic based on temperature and humidity", *5th Iranian Hydraulic Conference, Kerman*. (In Persian)
- Tabesh, M. & Dini, M., 2010, "Forecast daily water demand using artificial neural networks, case study: Tehran city", *Journal of Water and Wastewater*, 21(1), 84-95. (In Persian)
- Tabesh, M., Dini, M., Khoshkholgh, A. J. & Zahraie, B., 2007, "Estimation of tehran daily water demand using time series analysis", *Journal of Iran-Water Resources Research*, 4 (2), 57-65. (In Persian)
- Tabesh, M., Dini, M. & Naseri, M., 2006, "Short-term urban water consumption estimation using neuro-fuzzy sugeno model (two-parameter models)", *7th International Congress of Civil Engineering*, Tehran. (In Persian)
- Tabesh, M., Gooshe, S. & Yazdanpanah, M.J., 2007, "Short-term water demand forecasts in Tehran, using neural networks", *Tehran University, College of Engineering*, 41(1), 11-24. (In Persian)
- Weisberg, S., 2009, "The dr package", <http://www.r-project.org> (May 2016).
- Willsie, R.H. & Pratt, H.L., 1974, "Water use relationships and projection corresponding with regional growth", *Water Resources Bulletin*, 10(2), 360-371.
- Wong, S.T., 1972, "A model on municipal water demand: A case study of northeastern Illinois", *Land Econ*, 48(1), 34-44.
- Young, R.A., 1973 "Price elasticity of demand for municipal water: A case study of Tucson and Arizona", *Journal of Water Resource Research*, 9(4), 1068-1072.

Identification of Meteorological Parameters Affecting Water Consumption in Household Sector of Qom

Gh. Amini¹, Z. Saedi²

1. MSc. of Statistics, Water and Wastewater Co., Qom Province, Qom, Iran
(Corresponding Author) ghasem2094.amini@gmail.com

2. MSc. of Economical & Social Statistis, Department of Math and Computer,
Shahid Chamran University Ahvaz, Iran

(Received Oct. 6, 2016 Accepted March 4, 2017)

Abstract

Prediction of water consumption and its effective factors is an important step in water crisis management. Studies showed that meteorological parameters are considered as the most important factor for short-term prediction of water consumption. In this research, cluster-based sliced inverse regression method was used to identify the meteorological variables affecting the household water consumption in Qom. In addition to dimension reduction, this method can be used to remove collinearity. The data consisted of seven meteorological parameters and monthly household water consumption from 2001 to 2013. Data analysis indicated that instead of seven primary variables, only two new components which are linear combinations of independent variables can be used. The negatively charged maximum wind speed and relative humidity (0.757 and 0.4) of the first component, and the negatively charged average minimum temperature (0.753) and positively charged average air temperature (0.634) of the second component had the greatest impact on the components. The regression analysis indicated that the average minimum temperature coefficient 0.018, the maximum wind speed coefficient -0.004, and determination coefficient 0.92% are significant. Comparing the method proposed in this paper with the usual method of principal component analysis (PCA) for multivariate data analysis indicated that cluster-based sliced inverse regression has fewer errors. Moreover, noticing the impact of collinearity on the outputs of neural networks, the method proposed in this paper had better performance than the usual methods and consequently predicts water consumption.

Keywords: Dimension Reduction, Sliced Inversed Regression, Clustering, Collinearity.