نخستین کنفرانس ملی هیدروانفورماتیک و هوش مصنوعی در مهندسی آب

The First Natinal Conference on Hydroinformatics and Aritificial Inteligence in Water Engineering

خوشه‌بندی و تحلیل عدم‌قطعیت در مصرف آب شهری تحت تأثیر متغیرهای اقلیمی (مطالعه موردی: Sunshine Coast استرالیا)

مرتضی غیب دوست[۱]، زینب دهقانی فیروزآبادی[۲]، علی عباسی[۳*]

۱-دانشجوی کارشناسی ارشد، مهندسی عمران مدیریت منابع آب، دانشگاه فردوسی مشهد، گروه عمران (morteza.gheibdoost@mail.um.ac.ir)

۲-دانشجوی کارشناسی ارشد، مهندسی عمران مدیریت منابع آب، دانشگاه حکیم سبزواری، سبزوار (zynbdf9978@gmail.com)

۳-عضو هیئت علمی، دانشگاه فردوسی مشهد، گروه عمران (aabbasi@um.ac.ir)

چکیده

مدیریت تقاضای آب شهری در شرایط رشد جمعیت و تغییرات اقلیمی مستلزم شناخت دقیق الگوهای مصرف خانوار و عدم قطعیتهای موجود است. در این پژوهش، داده‌های کنتورهای هوشمند آب در منطقه‌ی Sunshine Coast استرالیا طی سه ماه (می–ژوئن–ژوئیه ۲۰۲۲) مورد تحلیل قرار گرفت. پس از پالایش داده‌ها، ۲۳۴ کنتور معتبر انتخاب و الگوهای مصرف روزانه آن‌ها با استفاده از الگوریتم K-means خوشه‌بندی شد. نتایج نشان داد که مصرف‌کنندگان در سه گروه متمایز قابل طبقه‌بندی هستند: خوشه‌ی کم‌مصرف‌ها با الگوی پایدار و محدود، خوشه‌ی پرمصرف‌ها با میانگین بالا و نوسانات شدید، و خوشه‌ی مصرف متوسط با بیشترین ناهمگنی رفتاری. به‌موازات تحلیل مصرف، داده‌های اقلیمی شامل بارش، دما و تابش خورشیدی بررسی و بهترین توزیع‌های احتمالاتی آن‌ها شناسایی شد. توزیع پیرسون تیپ ۳ بهترین برازش برای داده های بارش و تابش، و توزیع ویبول برای دما توزیع مناسب تعیین شدند. سپس با بهره‌گیری از شبیه‌سازی مونت‌کارلو و استفاده از الگوریتم مدل جنگل تصادفی، رابطه‌ی مصرف آب و متغیرهای اقلیمی مدل‌سازی گردید. این رویکرد امکان لحاظ نمودن تاثیر عدم‌قطعیتهای موجود در متغیرهای مستقل در میزان مصرف (تقاضای) روزانه را فراهم می نماید.

واژه‌های کلیدی: عدم قطعیت، مونت کارلو، تقاضای آب روزانه، جنگل تصادفی، آموزش ماشین

۱- مقدمه

مدیریت پایدار منابع آب شهری یکی از چالش‌های اصلی قرن بیست‌ویکم به شمار می‌آید. رشد سریع جمعیت، توسعه شهرنشینی و تغییرات اقلیمی فشار فزاینده‌ای بر سیستم‌های تأمین و توزیع آب وارد کرده است[۱]. در این میان، پایش دقیق و تحلیل الگوهای مصرف آب، ابزار مهمی برای سیاست‌گذاران و مدیران به‌منظور طراحی برنامه‌های مدیریت تقاضا و ارتقای بهره‌وری محسوب می‌شود[۵]. ظهور کنتورهای هوشمند آب در دهه اخیر امکان جمع‌آوری داده‌های با وضوح زمانی بالا را فراهم ساخته و زمینه را برای تحلیل‌های داده‌محور مصرف فراهم کرده است [۲۲]. این فناوری نه‌تنها رفتار مصرف‌کنندگان را در بازه‌های زمانی کوتاه منعکس می‌کند، بلکه می‌تواند ناهمگنی الگوهای مصرف میان خانوارها را آشکار کند. برای مثال، پژوهش‌های انجام‌شده در استرالیا نشان داده‌اند که خوشه‌بندی داده‌های کنتورهای هوشمند قادر است گروه‌های متمایز مصرف‌کنندگان را با ویژگی‌های رفتاری متفاوت شناسایی نماید [۱۵].

علاوه بر عوامل اجتماعی–اقتصادی، متغیرهای اقلیمی مانند دما، بارش و تابش خورشیدی نقش مهمی در تعیین تقاضای آب دارند. مطالعات متعددی رابطه مثبت بین افزایش دما و رشد مصرف آب شهری را گزارش کرده‌اند[۹]. همچنین، بارش باران می‌تواند به‌طور مستقیم تقاضا برای مصارف بیرونی مانند آبیاری فضای سبز را کاهش دهد [۶]. با این حال، ماهیت تصادفی و نوسانی این متغیرها، پیش‌بینی تقاضای آب را با عدم قطعیت قابل توجهی مواجه می‌کند [۳].

در دهه اخیر، استفاده از روش‌های یادگیری ماشین برای پیش‌بینی مصرف آب رشد چشمگیری داشته است. الگوریتم‌هایی مانند Random Forest و Gradient Boosting توانسته‌اند عملکرد بهتری نسبت به مدل‌های کلاسیک رگرسیونی در شناسایی روابط پیچیده میان متغیرهای اقلیمی و الگوهای مصرف نشان دهند [۴]. افزون بر این، استفاده از رویکردهای احتمالاتی و برازش توزیع‌های آماری به داده‌های مصرف، ابزاری کارآمد برای لحاظ عدم قطعیت و ارائه طیف پیش‌بینی‌ها فراهم می‌آورد [۲].

با وجود پیشرفت‌های فوق، همچنان در ادبیات علمی خلأهایی وجود دارد. بسیاری از مطالعات به داده‌های تجمیعی (aggregate) در سطح شهر یا منطقه تکیه کرده‌اند و کمتر به سطح خرد (household-level) و تحلیل ناهمگنی مصرف پرداخته‌اند [۲۲]. همچنین، کمتر پژوهشی همزمان از خوشه‌بندی مصرف‌کنندگان، مدل‌سازی رابطه با متغیرهای اقلیمی و تحلیل عدم قطعیت بهره گرفته است.

در این راستا، پژوهش حاضر با استفاده از داده‌های کنتورهای هوشمند منطقه‌ی Sunshine Coast استرالیا به‌دنبال شناسایی الگوهای مصرف آب در سطح خانوار، بررسی تأثیر متغیرهای اقلیمی و ارائه مدلی احتمالاتی برای پیش‌بینی مصرف روزانه آب است. بدین ترتیب، این مطالعه در پیوند دادن تحلیل داده‌محور مصرف آب با ملاحظات اقلیمی و در نظر گرفتن عدم قطعیت‌ها، گامی در جهت توسعه ابزارهای علمی برای مدیریت تقاضای آب شهری برمی‌دارد.

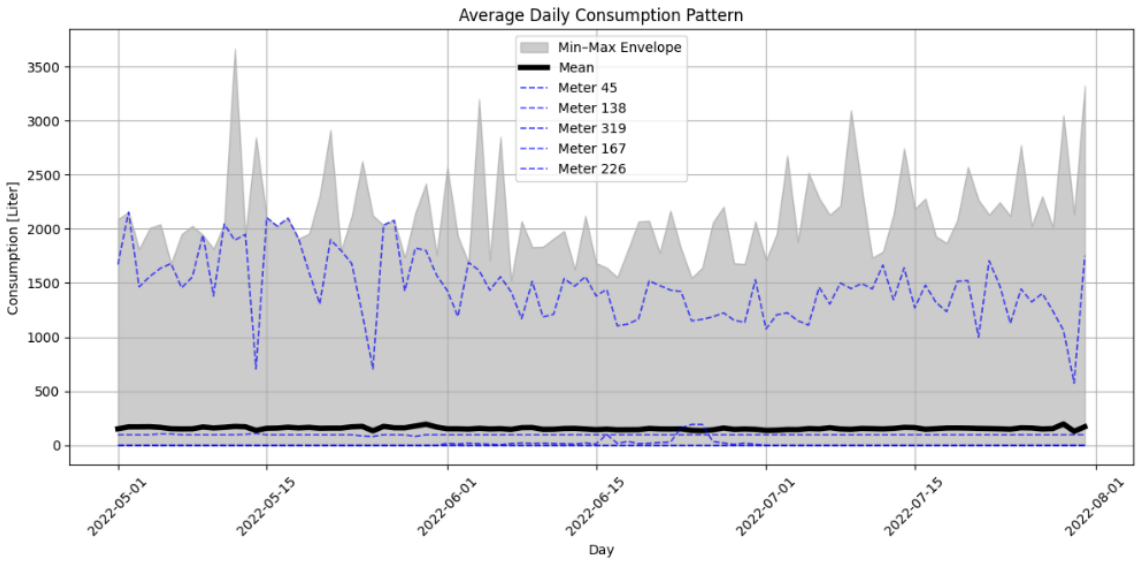
۲- روش تحقیق

در این پژوهش، داده‌های مصرف آب منطقه‌ی Sunshine Coast در ایالت Queensland استرالیا مورد استفاده قرار گرفت. Sunshine Coast منطقه‌ای ساحلی با اقلیم نیمه‌گرمسیری مرطوب است که در فاصله‌ای کمتر از ۱۰۰ کیلومتری از شمال Brisbane واقع شده است. این منطقه به‌طور فزاینده‌ای تحت تأثیر پیامدهای تغییر اقلیم قرار دارد؛ از جمله این پیامدها می‌توان به افزایش خطر فرسایش ساحلی ناشی از بالا آمدن سطح دریا، وقوع طوفان‌ها و بارش‌های شدیدی که به دلیل شرایط توپوگرافی منطقه منجر به سیلاب‌های ناگهانی می‌شوند

و همچنین افزایش دفعات خشکسالی اشاره کرد. علاوه بر این فشارهای اقلیمی، رشد سریع جمعیت و توسعه شهرنشینی، ضرورت انجام تحلیل‌های دقیق و جامع در زمینه‌ی مصرف آب را در منطقه‌ی Sunshine Coast بیش از پیش آشکار می‌سازد[۷].

داده‌های این پژوهش مربوط به سه ماه متوالی می، ژوئن و ژوئیه‌ی ۲۰۲۲ است که به‌صورت رکوردهای ۱۵ دقیقه‌ای از کنتورهای هوشمند آب ثبت شده‌اند. بهره‌گیری از این کنتورها به شرکت‌های آب و پژوهشگران امکان می‌دهد تا رفتار مصرف‌کنندگان را با دقت بالا پایش و تحلیل کنند. وجود چنین داده‌های با وضوح زمانی بالا (High-resolution data) از مهم‌ترین دلایل انتخاب این مجموعه‌داده برای تحلیل بود، چرا که امکان بررسی دقیق تغییرات الگوی مصرف در بازه‌های زمانی کوتاه را فراهم می‌آورد و بستر مناسبی برای انجام تحلیل‌های مبتنی بر داده در مدیریت تقاضای آب ایجاد می‌کند. در ماه می، ژوئن و ژوئیه به‌ترتیب تعداد ۷۸۷، ۷۹۳ و ۷۷۴ کنتور منحصربه‌فرد ثبت شده‌اند که در مجموع، کل تعداد کنتورهای منحصربه‌فرد در این سه ماه برابر با ۷۹۸ واحد بود. با این حال، تمامی این داده‌ها قابلیت استفاده مستقیم در تحلیل را نداشتند و نیازمند مراحل دقیق پیش‌پردازش بودند. در نخستین گام، با انجام اقداماتی شامل حذف داده‌های ناقص، شناسایی و حذف کنتورهایی با مقادیر غیرواقعی یا فاقد داده‌های معتبر، و همچنین همگام‌سازی بازه‌های زمانی، تعداد ۴۱۹ کنتور فاقد صلاحیت کنار گذاشته شد و در نتیجه ۳۷۹ کنتور معتبر باقی ماندند.

از آنجا که داده‌ها در ابتدا با تفکیک زمانی ۱۵ دقیقه‌ای ثبت شده بودند، برای انجام تحلیل‌های بعدی، کلیه رکوردها با روش ریسمپل (Resampling) به بازه‌ی روزانه تبدیل شدند. این کار موجب شد که تغییرات مصرف در سطح روز قابل بررسی باشد و امکان مقایسه الگوهای مصرف میان کنتورها فراهم گردد. با وجود این پالایش اولیه، بررسی‌های بعدی نشان داد که در داده‌های مصرف روزانه همچنان مقادیر پرت (Outliers) وجود دارد که می‌تواند بر نتایج تحلیل اثرگذار باشد. به همین دلیل، مجموعه‌ای از فیلترهای تکمیلی برای شناسایی و حذف این داده‌های پرت اعمال گردید. طی این مرحله، ۱۴۵ کنتور دیگر که دارای الگوهای مصرف غیرعادی بودند حذف شدند. در نهایت، تنها ۲۳۴ کنتور معتبر باقی ماندند که به دلیل کیفیت و ثبات بالای داده‌ها، به‌عنوان مجموعه‌داده نهایی پژوهش انتخاب و مبنای تحلیل‌های بعدی قرار گرفتند. در شکل ۱ سری زمانی مصرف چند کنتور منتخب به همراه مقدار حداقل و حداکثر مصرف ثبت شده در بازه موردنظر نشان داده شده اند.



شکل۱ – مصرف روزانه کلیه کنترها پیش از انجام پیش‌پردازش

به‌منظور شناسایی الگوهای مصرف آب میان خانوارها و کاربران مختلف، داده‌های روزانه استخراج‌شده از کنتورها با استفاده از روش خوشه‌بندی K-means مورد تحلیل قرار گرفتند. فرآیند خوشه‌بندی طی چند مرحله متوالی انجام شد. نخست، داده‌ها بازآرایی شدند به‌گونه‌ای که هر ردیف نمایانگر یک شناسه‌ی کنتور (id meter) و هر ستون نشان‌دهنده‌ی مقدار مصرف در یک بازه‌ی زمانی مشخص باشد. این بازآرایی (Pivoting) امکان مقایسه‌ی الگوهای زمانی مصرف را میان مصرف‌کنندگان مختلف فراهم می‌کرد [۸].

در گام بعدی، داده‌ها به‌وسیله‌ی استانداردسازی Z-score مقیاس‌بندی شدند تا از تسلط یک بازه‌ی زمانی خاص با مقادیر مصرف بالا بر فرآیند خوشه‌بندی جلوگیری شود. این تبدیل بر اساس رابطه‌ی زیر انجام شد:

$$\frac{x-\mu}{\sigma}=z \quad (۱)$$

که در آن x مقدار اولیه، μ میانگین و σ انحراف معیار داده‌ها است[10].

سپس برای تعیین تعداد بهینه خوشه‌ها (K)، از دو شاخص پرکاربرد در ادبیات خوشه‌بندی استفاده گردید: روش Elbow که بر مبنای بررسی تغییرات درون‌خوشه‌ای (Within-Cluster Sum of Squares, WSS) است. در این روش، تابع هدف الگوریتم K-means به‌صورت زیر تعریف می‌شود:

$$WSS(k)=\sum_{i=1}^{k}\sum_{x\in C_i}\| x-\mu_i \|^2 \quad (۲)$$

که در آن Ci خوشه‌ی i ام، μi مرکز خوشه و x داده‌ی عضو خوشه است. است [۱۱]. با تغییر مقدار k، شاخص WSS محاسبه شده و نقطه‌ای که کاهش آن به‌طور محسوسی کند می‌شود، به‌عنوان نقطه‌ی آرنج (Elbow Point) و تعداد خوشه‌ی بهینه در نظر گرفته می‌شود [۱۲].

روش دوم، ضریب سیلوئت (Silhouette Coefficient) است که کیفیت خوشه‌بندی را از منظر جداسازی بین خوشه‌ها و فشردگی درون خوشه‌ها ارزیابی می‌کند [۱۳]. برای هر داده‌ی i، ضریب سیلوئت به‌صورت زیر تعریف می‌شود:

$$s(i)=\frac{b(i)-a(i)}{max\{a(i),b(i)\}} \quad (۳)$$

که در آن $a(i)$ میانگین فاصله‌ی داده i با سایر نقاط خوشه‌ی خودش و $b(i)$ کمترین میانگین فاصله‌ی آن با نقاط سایر خوشه‌ها است. مقدار کلی سیلوئت میانگین این شاخص برای کل داده‌هاست و مقادیر نزدیک به ۱ بیانگر خوشه‌بندی مطلوب است.

به‌موازات تحلیل داده‌های مصرف، داده‌های اقلیمی شامل بارش، تابش خورشیدی و دما نیز مورد بررسی قرار گرفتند. داده‌های اقلیمی ده‌ساله (2010–2020) برای بازه‌ی مشابه زمانی استخراج و به توزیع‌های احتمالی مختلف (Normal, Exponential, Lognormal, Gamma, Weibull, GEV, Pearson III, Kappa-3) برازش داده شدند. آزمون‌های Akaike Information Criterion ، Kolmogorov–Smirnov Test و Bayesian Information Criterion برای انتخاب بهترین توزیع مورد استفاده قرار گرفتند. این ابزارها به‌طور گسترده در پژوهش‌های هیدرولوژی و اقلیم‌شناسی برای انتخاب بهترین توزیع احتمالاتی به کار می‌روند[17,14,16] .

در آزمون Kolmogorov–Smirnov، حداکثر فاصله بین توزیع تجمعی تجربی Fn(x) و توزیع تجمعی نظری F(x) محاسبه می‌شود:

$$D = \sup_x | F_n(x) - F(x) | \quad (۴)$$

که در آن Fn(x) تابع توزیع تجمعی تجربی و F(x) تابع توزیع تجمعی مورد آزمون است. هر چه مقدار D کوچک‌تر باشد، برازش بهتر خواهد بود [18].

برای مقایسه مدل‌ها از معیارهای اطلاعاتی استفاده شد. Akaike Information Criterion به صورت زیر تعریف می‌شود:

$$AIC = 2k - 2\ln(L) \quad (۵)$$

و Bayesian Information Criterion نیز به شکل زیر بیان می‌شود:

$$BIC = k\ln(n) - 2\ln(L) \quad (۶)$$

که در آن k تعداد پارامترهای مدل، n تعداد داده‌ها و L بیشینه‌ی تابع درست‌نمایی (Likelihood) است. توزیعی که مقادیر کمتری از AIC و BIC داشته باشد به عنوان مدل مناسب‌تر انتخاب می‌شود [21]. این رویکرد ترکیبی (استفاده همزمان از آزمون‌های نیکویی برازش و معیارهای اطلاعاتی) موجب افزایش دقت در انتخاب توزیع احتمالاتی و جلوگیری از برازش بیش از حد (Overfitting) می‌گردد [17].

برای بررسی روابط درونی میان متغیرهای اقلیمی، یک ماتریس همبستگی ۳×۳ شامل بارش (Precipitation)، تابش خورشیدی (Solar Radiation)و دما (Temperature) تشکیل شد. هدف از این تحلیل، ارزیابی میزان وابستگی یا استقلال این پارامترها نسبت به یکدیگر بود. بدین منظور، علاوه بر محاسبه ضرایب همبستگی، یک نقشه‌ی حرارتی (Heatmap) نیز ترسیم گردید تا ساختار روابط میان متغیرها به‌صورت بصری نمایان شود. برای سنجش همبستگی، از دو معیار آماری متداول یعنی ضریب همبستگی پیرسون (Pearson correlation) و ضریب همبستگی اسپیرمن (Spearman correlation) استفاده شد. ضریب پیرسون که برای متغیرهای کمی پیوسته تعریف می‌شود، میزان رابطه‌ی خطی میان دو متغیر را نشان می‌دهد [19]. در مقابل، ضریب اسپیرمن که بر پایه‌ی رتبه‌ها محاسبه می‌شود، توانایی ارزیابی روابط غیرخطی یکنواخت را دارد [20].

برای مدل‌سازی رابطه‌ی میان متغیرهای اقلیمی و مصرف روزانه آب، از رگرسیون جنگل تصادفی (Random Forest Regressor) استفاده شد. مقادیر متغیرهای اقلیمی(پارامترهای ورودی مدل یادگیری ماشین و یا پیش بینی کننده ها) با بهره‌گیری از شبیه‌سازی مونت‌کارلو و توزیع‌های احتمالی برازش‌شده تولید شدند تا عدم قطعیت آنها در پارامتر خروجی (خروجی) لحاظ گردد. در این فرآیند، داده‌ها به مجموعه‌های آموزش (۸۰٪) و آزمون (۲۰٪) تقسیم شدند. سپس با اجرای شبیه‌سازی‌های متعدد و استفاده از Kernel Density Estimation (KDE)، توزیع پیش‌بینی مصرف روزانه آب برآورد شد.

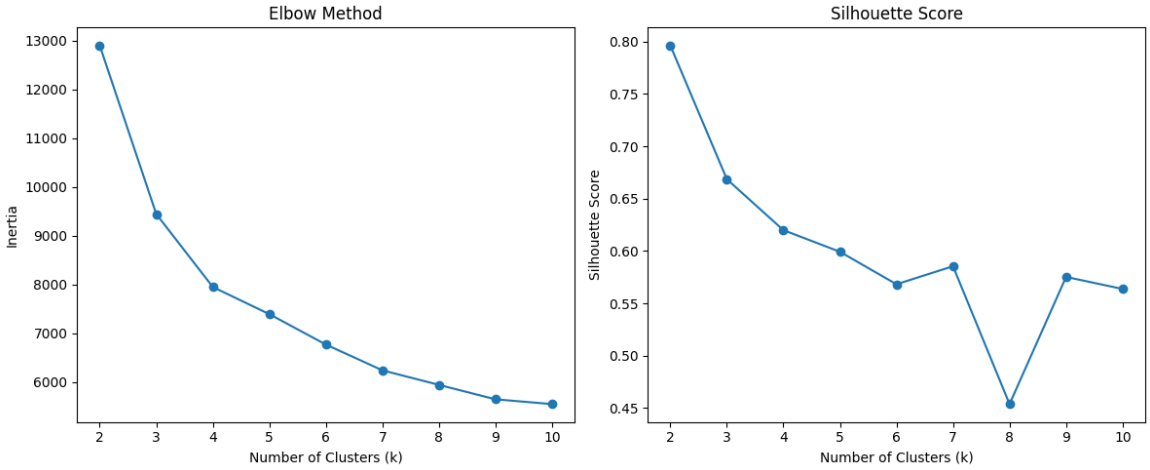
برای توصیف آماری مصرف روزانه آب، مجموعه‌ای از توزیع‌های رایج در هیدرولوژی و مدل‌سازی تقاضا به‌عنوان گزینه‌های کاندید در نظر گرفته شد: لوگ‌نرمال (lognormal)، گاما (gamma)، ویبول (weibull_min)، لوگ‌لاجستیک/فیسک (loglogistic/fisk)، بر۱۲ (burr12)، گاما تعمیم‌یافته (gengamma) و پیرسون نوع سوم (pearson3). این توزیع ها به‌طور گسترده برای کمیت‌های مثبت و

دارای چولگی در منابع آب و اقلیمی کاربرد دارند و در ادبیات، برای دبی/بارش و همچنین اجزای تقاضای آب شهری (حجم و طول کشش‌ها) توصیه شده‌اند.

پس از برازش مجموعه‌ای از توزیع‌های کاندید به داده‌های مصرف روزانه آب در هر خوشه، معیارهای اطلاعاتی (AIC و BIC) به همراه شاخص‌های نیکویی برازش (KS و CvM) محاسبه و بهترین مدل انتخاب شد.

۳- نتایج پژوهش

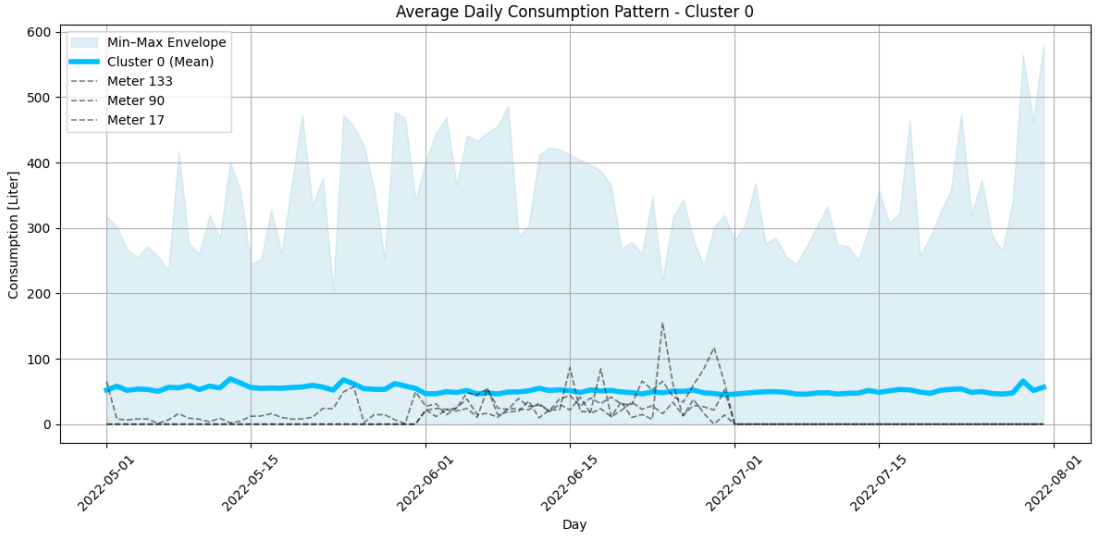
نتایج تحلیل‌ها نشان داد که بر اساس روش Elbow، نقطه‌ی شکست در k=3 ظاهر می‌شود و در شاخص سیلوئت نیز k=2 بهترین گزینه می‌باشد. بنابراین، برای نشان دادن تکثر الگوها، در نهایت k=3 انتخاب شد و داده‌ها در قالب سه خوشه‌ی متمایز دسته‌بندی شدند (شکل ۲).



شکل۲- نتایج آزمون silhouette و Elbow

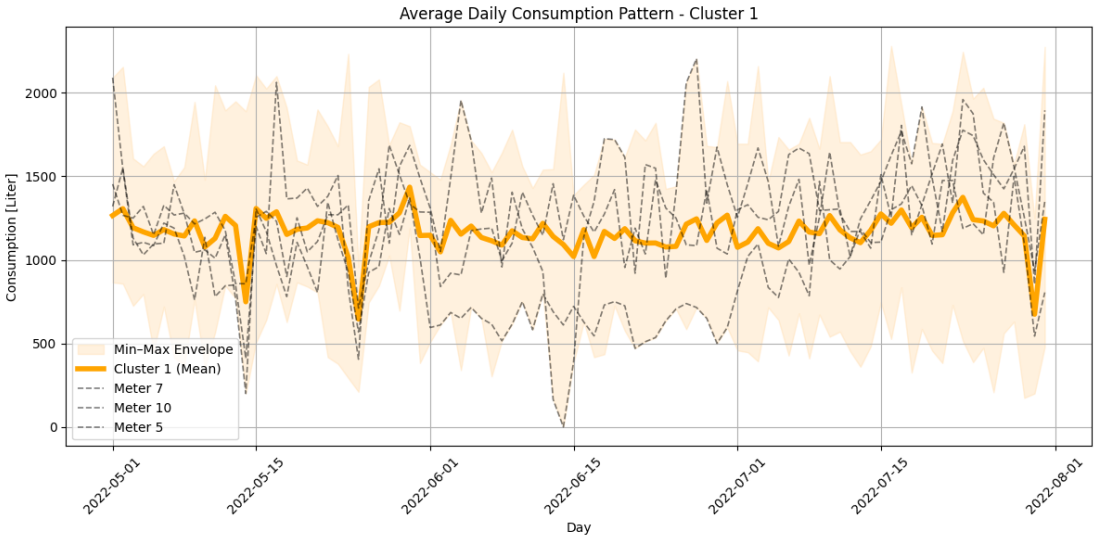
نتایج خوشه‌بندی نشان داد که مصرف‌کنندگان آب در منطقه مورد مطالعه به سه الگوی عمده قابل تفکیک هستند:

الف) خوشه‌ی کم‌مصرف‌ها [0]Cluster که عمدتاً شامل خانوارهای کوچک یا مصرف‌کنندگان با رفتار پایدار و محدود بودند (جمعیت ۱۷۹ کنتور). خوشه‌ی صفر (کم‌مصرف‌ها) نشان‌دهنده‌ی گروهی از مصرف‌کنندگان است که در کل بازه‌ی زمانی مورد مطالعه، الگوی مصرفی نسبتاً پایدار و محدود داشتند. همان‌گونه که در نمودار مشاهده می‌شود، میانگین مصرف روزانه در این خوشه در محدوده‌ی ۵۰ تا ۸۰ لیتر قرار گرفته و نوسانات شدیدی در طول زمان ثبت نشده است. دامنه تغییرات مصرف (Min–Max envelope) نیز محدود بوده و حداکثر مقادیر مصرف به ندرت از ۶۰۰ لیتر در روز فراتر رفته است. این ویژگی‌ها نشان می‌دهد که خوشه‌ی کم‌مصرف‌ها ترکیب یکنواختی از کاربران است که الگوی مصرفی آن‌ها تفاوت چشمگیری با یکدیگر ندارد (شکل ۳).



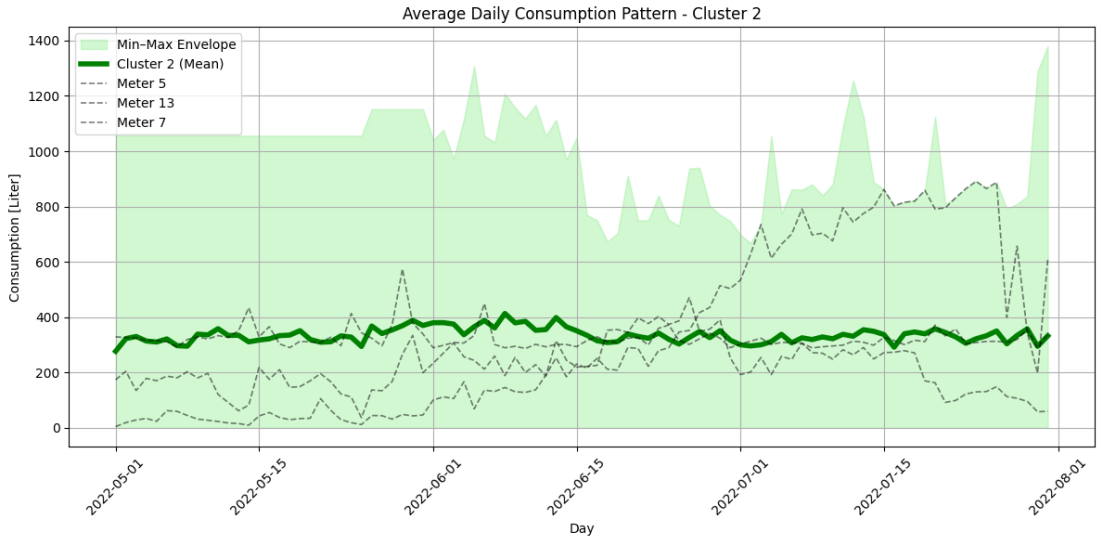
شکل۳- الگوی مصرف آب روزانه برای دسته صفر

ب) خوشه‌ی پرمصرف‌ها [1]Cluster که ویژگی‌هایی همچون نوسانات بالا، پیک‌های مصرف شدید یا خانوارهای بزرگ‌تر را نشان می‌دهد (جمعیت ۱۴ کنتور). خوشه‌ی یک (پرمصرف‌ها) شامل کاربرانی است که میانگین مصرف روزانه‌ی آنان در محدوده‌ی ۱۰۰۰ تا ۱۳۰۰ لیتر قرار دارد و بنابراین به‌عنوان پرمصرف‌ترین گروه در میان خوشه‌ها شناسایی می‌شوند. همان‌گونه که در نمودار شکل ۴ نشان داده شده است، دامنه‌ی تغییرات مصرف در این خوشه بسیار وسیع است؛ به‌طوری که حداقل مصرف روزانه گاهی به کمتر از ۵۰۰ لیتر و حداکثر آن به بیش از ۲۰۰۰ لیتر در روز می‌رسد. این گستره‌ی بالا بیانگر نوسانات شدید در رفتار مصرفی و وجود پیک‌های ناگهانی در طول زمان است (شکل ۴).



شکل۴- الگوی مصرف آب روزانه برای دسته یک

ج) خوشه‌ی مصرف متوسط Cluster[2] که نمایانگر اکثریت مصرف‌کنندگان با الگوی نسبتاً متعادل روزانه است (جمعیت ۴۱ کنتور). خوشه‌ی دوم (مصرف متوسط‌ها) الگوی مصرفی عمده‌ی کاربران را نشان می‌دهد که میانگین مصرف روزانه‌ی آن‌ها در بازه‌ی ۲۵۰ تا ۴۰۰ لیتر قرار دارد. با این حال، این خوشه نسبت به دو خوشه‌ی دیگر دارای بیشترین میزان نوسان است. همان‌گونه که در نمودار مشاهده می‌شود، دامنه‌ی تغییرات مصرف روزانه بسیار گسترده است؛ به‌طوری که حداقل مصرف برخی کاربران به کمتر از ۱۰۰ لیتر و حداکثر آن به بیش از ۱۲۰۰ لیتر در روز می‌رسد(شکل ۵).



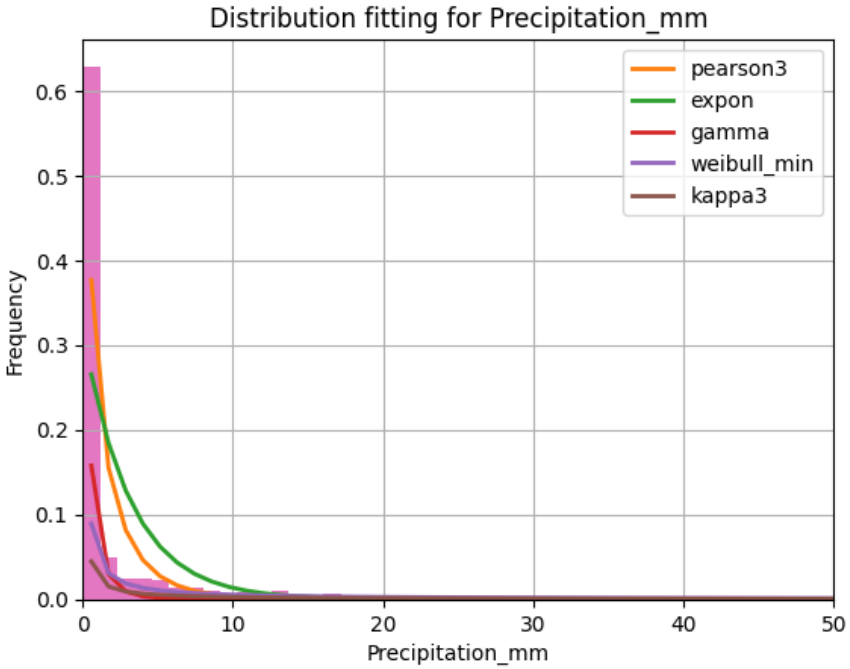
شکل۵- الگوی مصرف آب روزانه برای دسته دو

همچنین بررسی و برازش توزیع‌های احتمالاتی به داده‌های مورداستفاده در این پژوهش در جدول ۲ خلاصه شده است:

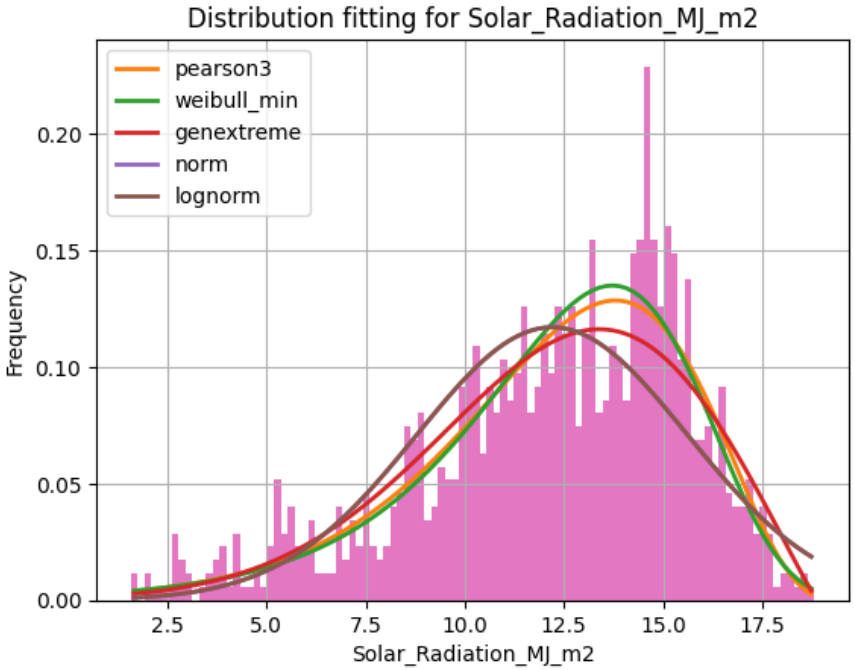
جدول ۱: توزیع احتمالاتی برازش داده شده به پارامترهای اقلیمی

ردیف	پارامتر	توزیع احتمالاتی	پارامترهای توزیع	شکل
۱	بارش (Precipitation_mm)	Pearson Type III	α=2.617, ξ=1.552, β=2.032	شکل ۶
۲	تابش خورشیدی (Solar Radiation, MJ/m²)	Pearson3	α=−0.931, ξ=12.202, β=3.426	شکل ۷
۳	دما(Temperature,°C)	Weibull_min	α=−0.931, ξ=12.202, β=3.426	شکل ۸

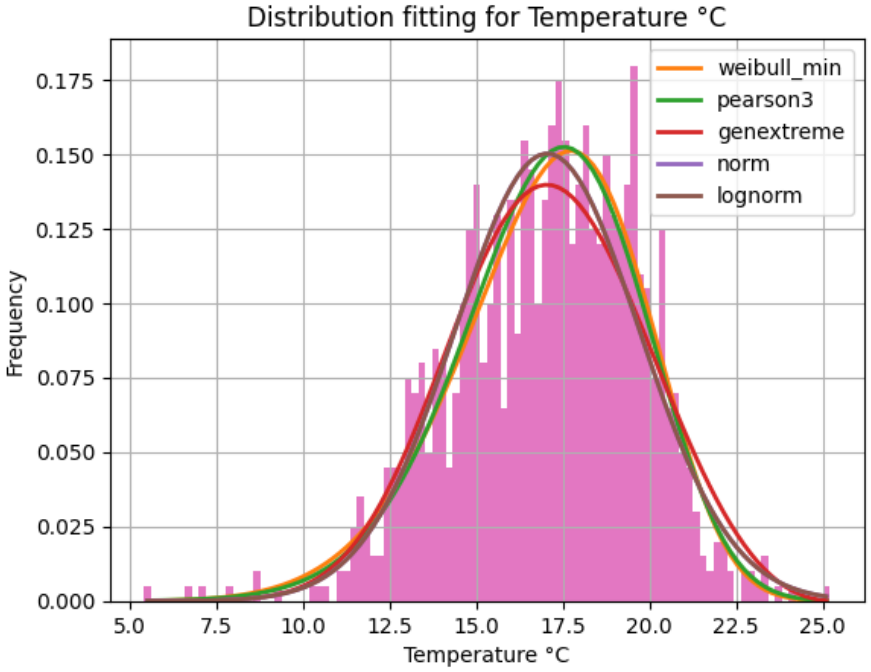
در شکل های ۶ تا ۸ توزیع احتمالاتی پارامترهای موردنظر (مطابق جدول ۱) ارائه شده اند:



شکل۶- برازش توزیع‌های احتمالاتی برای بارش

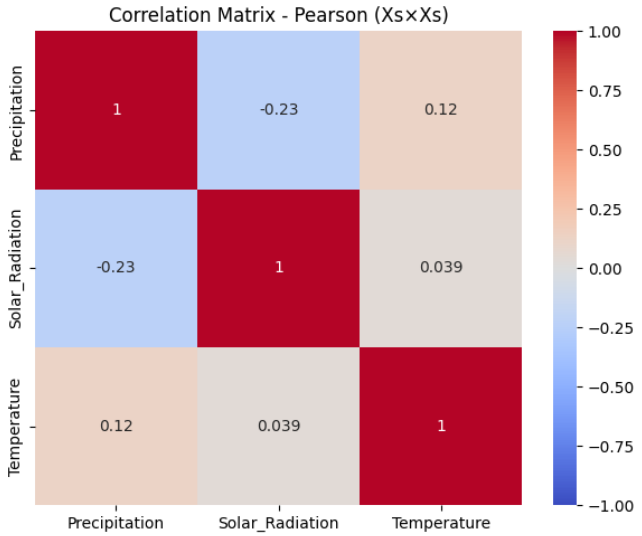


شکل۷- برازش توزیع‌های احتمالاتی برای تابش خورشیدی



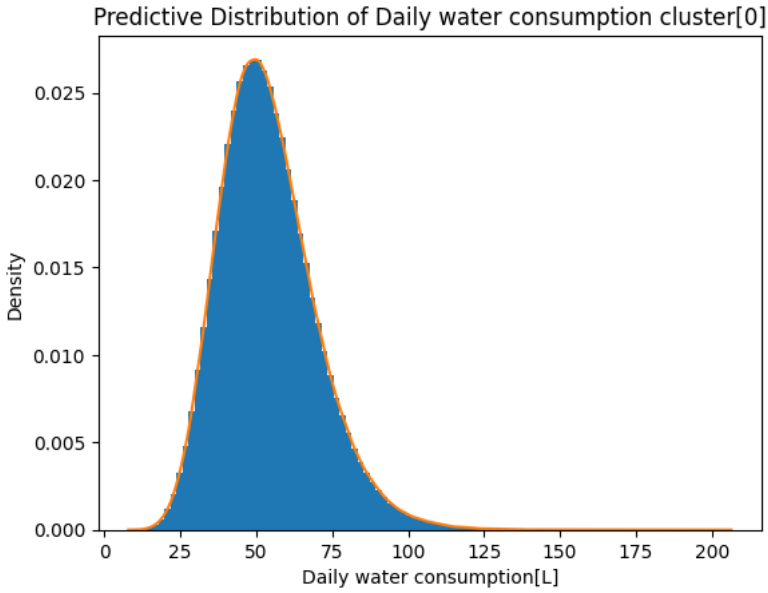
شکل۸- برازش توزیع‌های احتمالاتی بر دما

تحلیل همبستگی (ماتریس همبستگی) میان متغیرهای اقلیمی نشان داد که هیچ رابطه‌ی معناداری، چه خطی و چه غیر خطی، میان بارش، تابش و دما وجود ندارد. . به بیان دیگر، بارش، تابش خورشیدی و دما در این بازه‌ی زمانی به‌صورت آماری مستقل از یکدیگر رفتار کرده‌اند (شکل ۹).

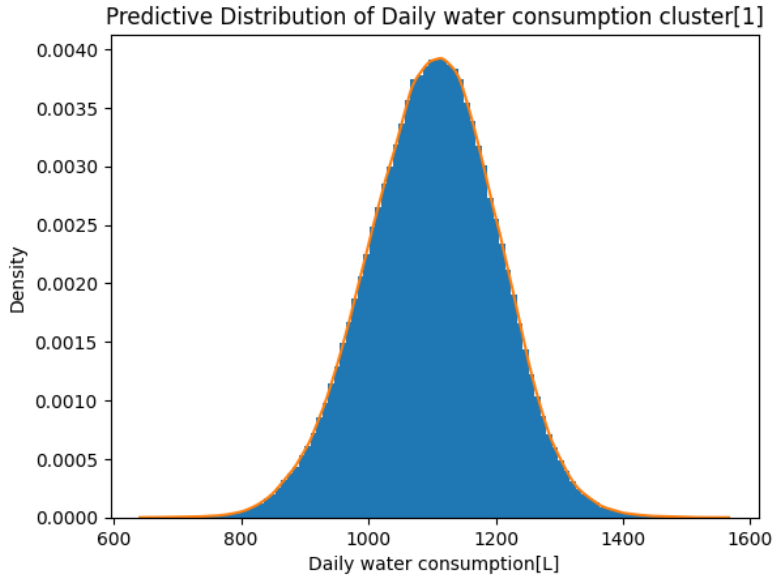


شکل ۹- ماتریس همبستگی بین متغیرهای ورودی مدل آموزش ماشین

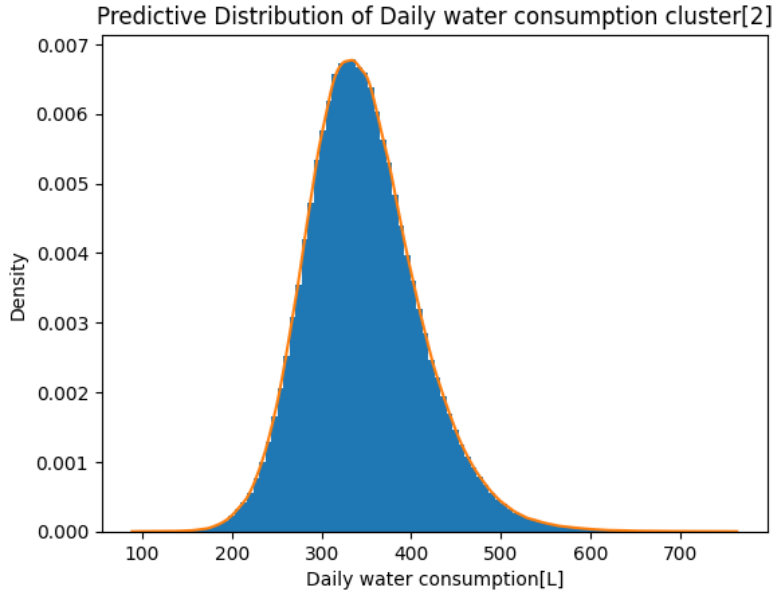
الگوریتم جنگل تصادفی: متغیرهای ورودی مدل (X) شامل سه پارامتر اقلیمی بارش، تابش خورشیدی و دما بودند که پیش‌تر توزیع‌های احتمالی آن‌ها شناسایی و برازش گردیده بود. برای ساخت مجموعه‌ی داده‌ی آموزشی و آزمایشی، مقادیر این متغیرها از توزیع‌های متناظرشان به‌صورت مستقل و بدون در نظر گرفتن همبستگی درونی با استفاده از روش نمونه‌گیری تصادفی مونت‌کارلو تولید شدند. در هر تکرار، نمونه‌هایی به تعداد روزهای دوره‌ی زمانی (۹۲ روز) از توزیع‌های برازش‌شده استخراج گردید و به همراه مقادیر واقعی مصرف روزانه‌ی آب (خروجی مدل) برای آموزش جنگل تصادفی مورد استفاده قرار گرفتند. این فرآیند سبب شد تا اثر عدم قطعیت ذاتی متغیرهای اقلیمی در مدل‌سازی مصرف آب به شکل صریح لحاظ شود. مدل جنگل تصادفی با داده‌های شبیه‌سازی‌شده اقلیمی اجرا شد. خروجی‌ها نشان داد که این مدل قادر است توزیع پیش‌بینی مصرف روزانه آب را با لحاظ عدم قطعیت بازتولید کند. نمودارهای تولیدشده همراه با بازه اطمینان ۹۵ درصد نشان دادند که مدل می‌تواند محدوده‌ی قابل‌قبول برای مقادیر مصرف را تخمین بزند. توزیع احتمالاتی مناسب با میزان مصرف آب (خروجی مدل جنگل تصادفی) برای هر کدام از سه دسته مصرف در شکلهای ۱۰ تا ۱۲ نشان داده شده است:



شکل۱۰- توزیع پیش بینی شده برای مصرف روزانه آب دسته صفر

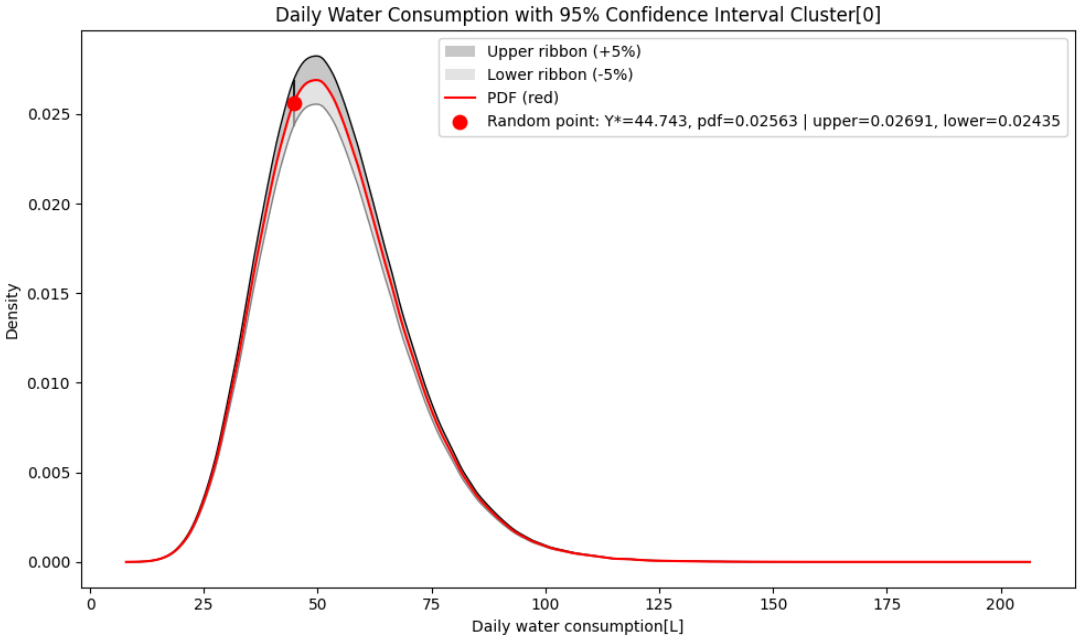


شکل۱۱- توزیع پیش بینی شده برای مصرف روزانه آب دسته ۱

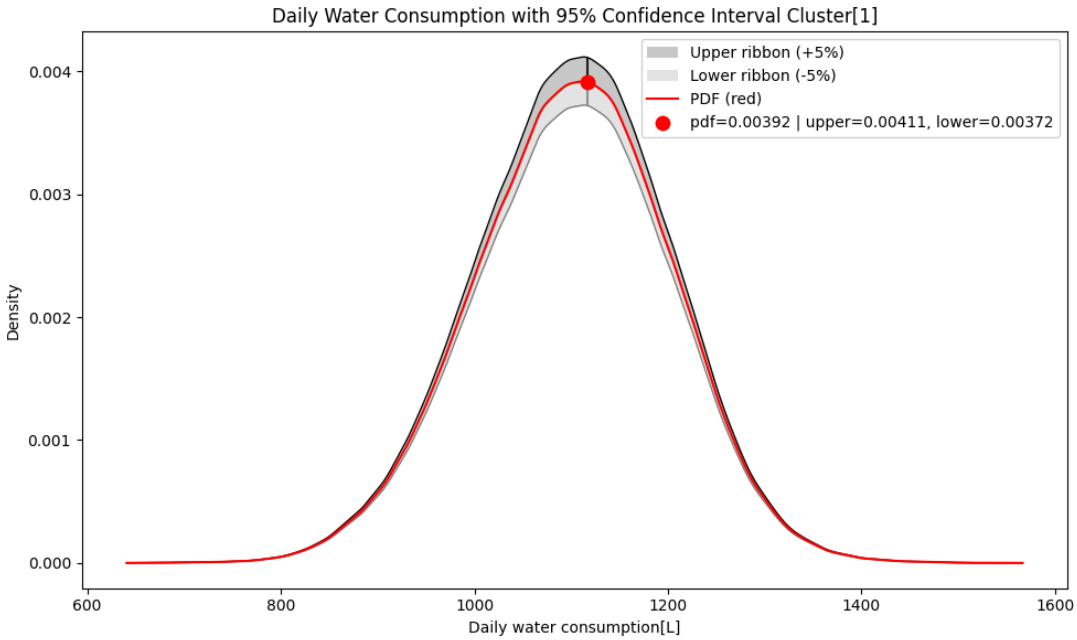


شکل۱۲- توزیع پیش بینی شده برای مصرف روزانه آب دسته ۲

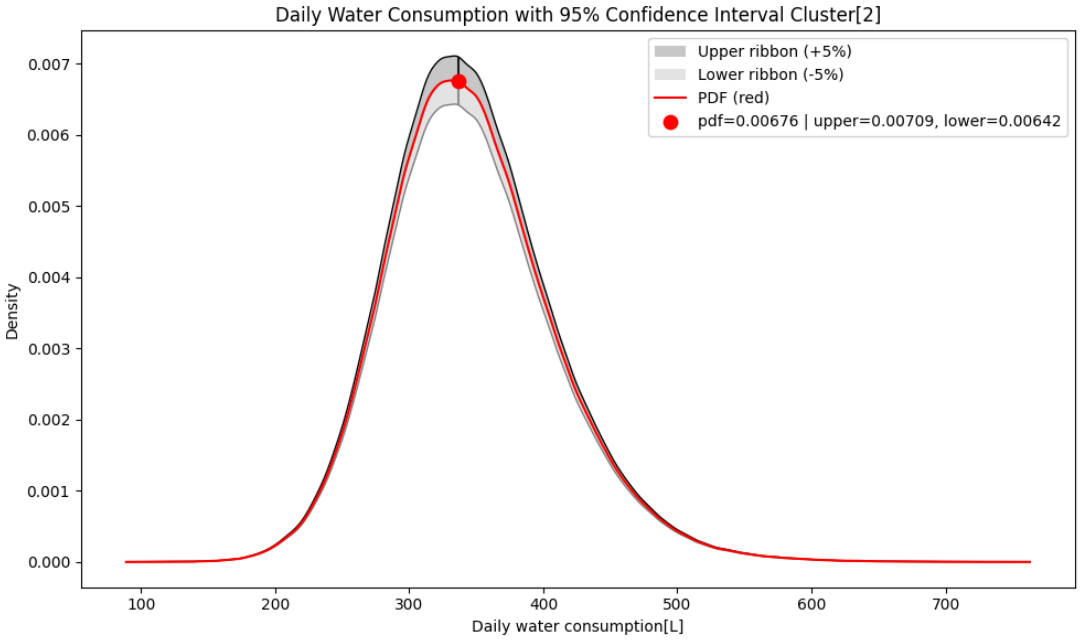
به‌منظور تحلیل عدم‌قطعیت و نمایش بصری نتایج مدل، یک تابع ترسیمی سفارشی توسعه داده شد که توزیع پیش‌بینی‌شده‌ی مصرف روزانه آب را همراه با یک نوار اطمینان به تصویر می‌کشد(شکل ۱۳ تا ۱۵). این تابع ابتدا با استفاده از روش تخمین چگالی کرنل (Kernel Density Estimation, KDE)، تابع چگالی احتمال (PDF) داده‌های پیش‌بینی‌شده را برآورد می‌کند. سپس دو باند بالا و پایین به اندازه‌ی ±۵٪ حول PDF محاسبه شده و به‌صورت یک نوار دو رنگ در اطراف منحنی اصلی ترسیم می‌شوند.



شکل۱۳- تابع چگالی احتمال مصرف روزانه آب با اطمینان ۹۵ درصد برای دسته صفر



شکل۱۴- تابع چگالی احتمال مصرف روزانه آب با اطمینان ۹۵ درصد برای دسته یک



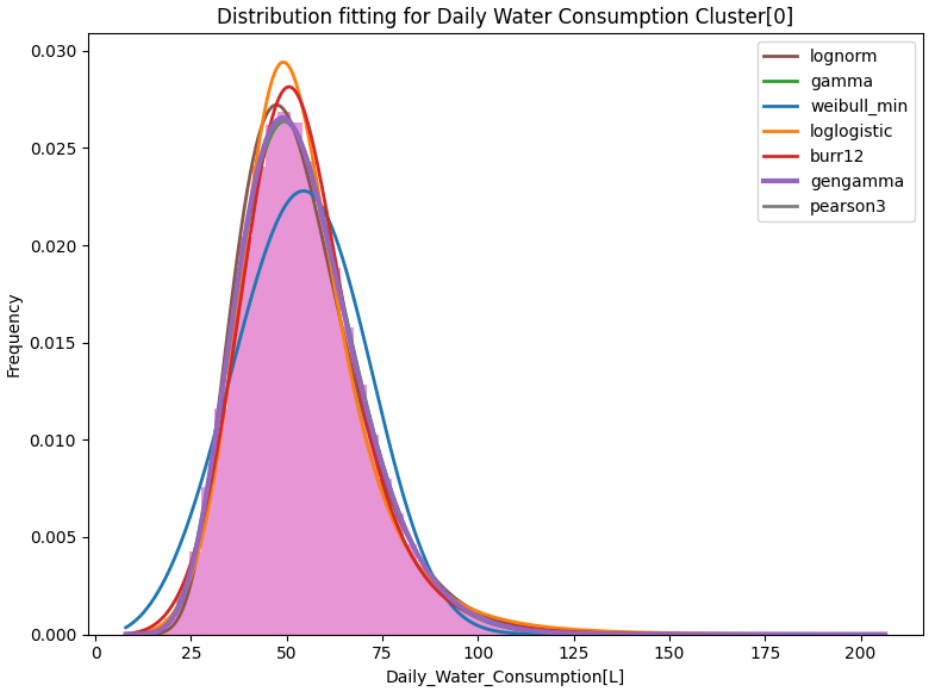
شکل۱۵- تابع چگالی احتمال مصرف روزانه آب با اطمینان ۹۵ درصد برای دسته ۲

برازش توزیع‌های آماری به داده‌های مصرف آب در خوشه‌های مختلف در جدول ۲ ارائه شده است:

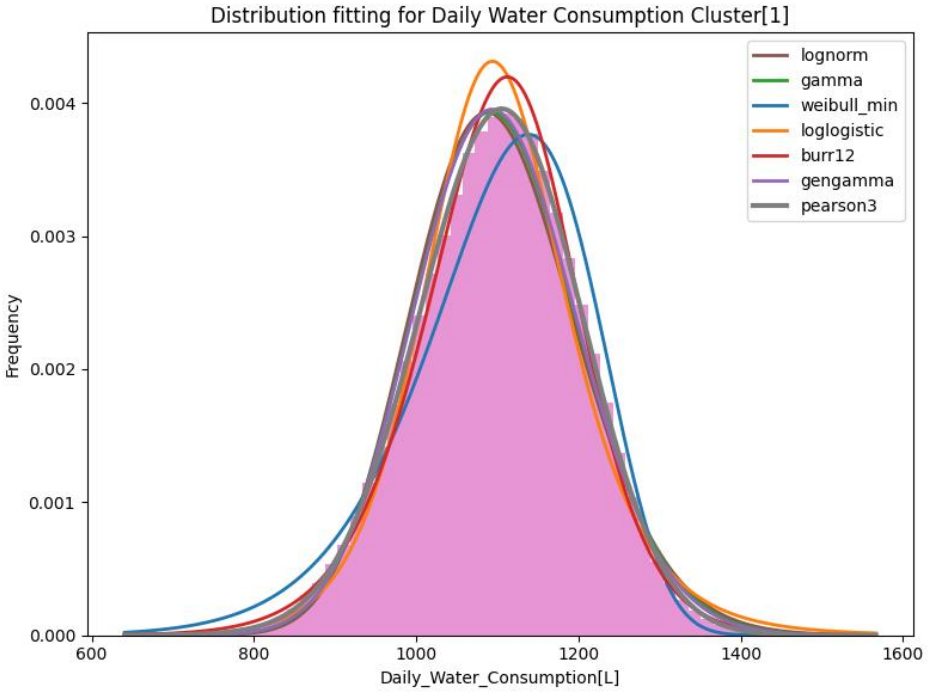
جدول ۲: توزیع احتمالاتی برازش داده شده به میزان مصرف در دسته های مختلف

ردیف	دسته (کلاستر)	توزیع احتمالاتی	پارامترهای توزیع	شکل
۱	دسته صفر	Generalized Gamma (gengamma)	a = 19.4352 , c = 0.7753, loc = 0 , scale = 1.1627	شکل ۱۶
۲	دسته یک	Pearson Type III (pearson3)	skew = 0.0848 , loc = 1101.16 , scale = 100.94	شکل ۱۷
۳	دسته دو	Generalized Gamma (gengamma)	a = 54.9086, c = 0.7490, loc = 0 , scale 1.6406	شکل ۱۸

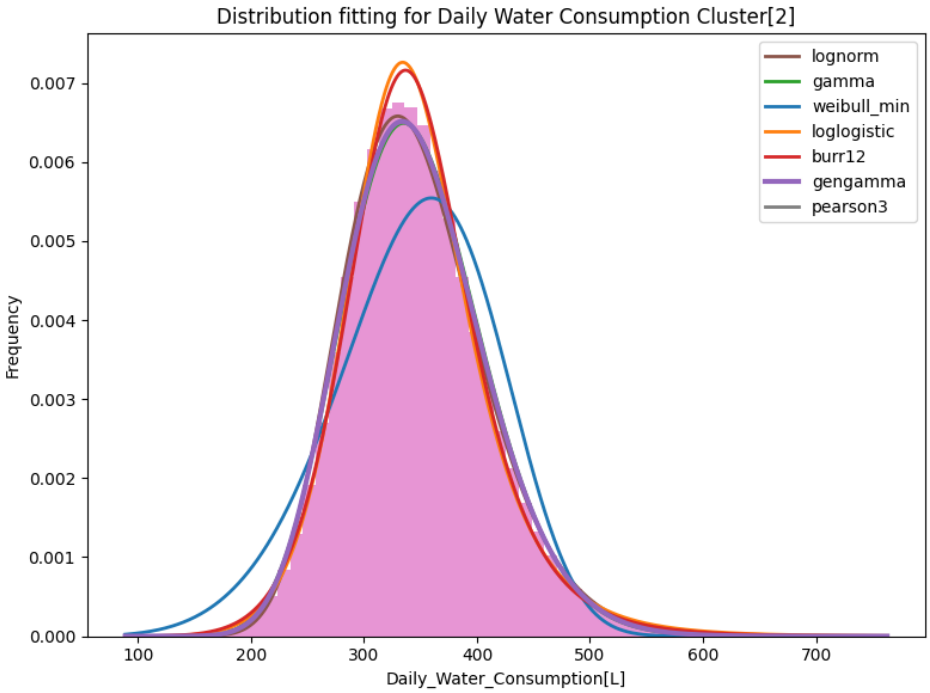
در شکل های ۱۶ تا ۱۸ توزیع احتمالاتی مصرف آب (مطابق جدول ۲) ارائه شده اند:



شکل۱۷- برازش توزیع‌های احتمالاتی بر توزیع پیش بینی شده مصرف روزانه آب برای دسته صفر



شکل۱۷- برازش توزیع‌های احتمالاتی بر توزیع پیش بینی شده مصرف روزانه آب برای دسته ۱



شکل۱۸- برازش توزیع‌های احتمالاتی بر توزیع پیش بینی شده مصرف روزانه آب برای دسته ۲

۴- نتیجه‌گیری و پیشنهادات

یافته‌های این پژوهش نشان داد که بهره‌گیری از داده‌های کنتورهای هوشمند و رویکردهای داده‌محور می‌تواند بینش ارزشمندی در خصوص الگوهای مصرف آب و عدم قطعیتهای موجود در میزان تقاضای آب شرب شهری فراهم سازد. خوشه‌بندی مصرف‌کنندگان به سه گروه اصلی، تصویری روشن از تنوع رفتاری جامعه مصرف‌کننده ارائه کرد. خوشه‌ی کم‌مصرف‌ها (Cluster[0])الگوی پایدار و کم‌ریسکی را نشان داد که احتمالاً به خانوارهای کوچک یا کاربرانی با سبک زندگی کم‌مصرف تعلق دارد. خوشه‌ی پرمصرف‌ها (Cluster[1]) به‌عنوان گروهی با میانگین مصرف بسیار بالا و نوسانات شدید شناسایی شد که بیشترین پتانسیل را برای مداخلات مدیریتی و سیاست‌گذاری دارند. در مقابل، خوشه‌ی مصرف متوسط (Cluster[2]) با ناهمگنی رفتاری قابل توجه، بازتاب‌دهنده‌ی بخش بزرگی از جامعه مصرف‌کننده بود که می‌تواند تمرکز اصلی اقدامات سیاستی در مدیریت تقاضا قرار گیرد.

بررسی نقش متغیرهای اقلیمی نشان داد که دما، تابش خورشیدی و بارش هر یک الگوهای آماری خاص خود را دارند و بهترین برازش برای آن‌ها به‌ترتیب با توزیع‌هایWeibull ، Pearson III وPearson III حاصل شد. این نتایج ضمن تأیید حساسیت مصرف آب به شرایط اقلیمی، نشان می‌دهد که تغییرات کوتاه‌مدت این متغیرها الزاماً همبستگی مستقیم و معنی‌داری با یکدیگر ندارند.

مدل‌سازی رابطه‌ی اقلیم و مصرف با استفاده از الگوریتم جنگل تصادفی، توانست با بهره‌گیری از شبیه‌سازی مونت‌کارلو و برازش توزیع‌های احتمالی، توزیعی از مقادیر پیش‌بینی مصرف را تولید کند. این رویکرد ضمن افزایش دقت پیش‌بینی، امکان لحاظ عدم قطعیت در داده‌های اقلیمی و مصرفی را نیز فراهم آورد. در نتیجه، مدل توسعه‌یافته قادر است طیف احتمالی مصرف روزانه را برآورد کرده و مبنای تصمیم‌گیری‌های مبتنی بر ریسک قرار گیرد. تحلیل توزیع‌های احتمالی مصرف در هر خوشه نیز نشان داد که خانواده‌ی Generalized Gamma انعطاف‌پذیری بالایی در بازنمایی رفتار مصرفی گروه‌های کم‌مصرف و متوسط دارد، در حالی‌که توزیع Pearson Type III بهترین برازش را برای خوشه‌ی پرمصرف‌ها ارائه کرده است.

به‌طور کلی، این پژوهش نشان داد که ترکیب خوشه‌بندی، مدل‌های یادگیری ماشین و تحلیل‌های احتمالاتی می‌تواند ابزار قدرتمندی برای مدیریت تقاضای آب شهری فراهم آورد. نتایج آن می‌تواند برای طراحی سیاست‌های متنوع از جمله تعرفه‌های پلکانی، برنامه‌های صرفه‌جویی هدفمند و مداخلات رفتاری مورد استفاده قرار گیرد. در عین حال، محدودیت‌هایی همچون بازه زمانی کوتاه داده‌های مصرف و عدم دسترسی به داده‌های هم‌زمان اقلیمی، لزوم توسعه پژوهش‌های آتی در بازه‌های زمانی بلندمدت‌تر و با داده‌های جامع‌تر را آشکار می‌سازد.

با توجه به محدودیتهای موجود در این پژوهش، پیشنهادات زیر جهت تکمیل این پژوهش ارائه می شود:

الف) افزایش بازه زمانی داده‌ها: این مطالعه بر مبنای داده‌های سه‌ماهه انجام شد. در صورت دسترسی به داده‌های طولانی‌مدت (مثلاً چند سال متوالی)، امکان بررسی روندهای فصلی، سالانه و میان‌دوره‌ای فراهم می‌شود. چنین داده‌هایی می‌توانند برای تحلیل پایداری خوشه‌ها در بلندمدت و بررسی تأثیر تغییرات اقلیمی یا سیاست‌های مدیریتی ارزشمند باشند.

ب) داده‌های اقلیمی هم‌زمان(Synchronized Climate Data) : در این پژوهش به‌دلیل فقدان داده‌های هم‌زمان، از شبیه‌سازی مبتنی بر توزیع‌های احتمالی استفاده شد. دسترسی به داده‌های واقعی و هم‌زمان اقلیمی (بارش، دما و تابش خورشیدی) در بازه‌ی دقیق ثبت کنتورها، دقت مدل‌سازی را افزایش داده و امکان تحلیل علی–زمانی مستقیم بین اقلیم و مصرف را فراهم می‌آورد.

ج) افزودن متغیرهای اجتماعی–اقتصادی: داده‌هایی همچون اندازه خانوار، نوع مسکن، سطح درآمد یا الگوهای مصرفی روزانه می‌توانند ناهمگنی خوشه‌ها را بهتر توضیح دهند و مدل‌سازی رفتار مصرف‌کنندگان را ارتقا بخشند.

د) بررسی مدل‌های یادگیری عمیق: گرچه جنگل تصادفی نتایج رضایت‌بخشی ارائه کرد، اما مدل‌های یادگیری عمیق مانند LSTM و GRUبا توجه به توانایی در تحلیل وابستگی‌های زمانی می‌توانند برای پیش‌بینی مصرف آب در بازه‌های زمانی طولانی‌تر به‌کار گرفته شوند.

ه) تحلیل سناریوهای مدیریتی: توسعه مدل برای شبیه‌سازی تأثیر سیاست‌هایی مانند تعرفه‌های پلکانی، محدودیت‌های مصرف یا کمپین‌های صرفه‌جویی می‌تواند کاربرد مستقیم در مدیریت شهری داشته باشد.

منابع

1. Gleick, P. H. (2018). The world’s water: The biennial report on freshwater resources. Island Press.
2. Kougias, I., Karampinis, E., & Theodossiou, N. (2018). Probabilistic analysis of water demand in urban areas: A case study. *Urban Water Journal, 15*(3), 181–188.

3. Shabani, A., Klash, A., Elhabishi, R., & Abds, S. (2020). Quantifying uncertainty in urban water consumption modeling under climate variability. *Water Resources Management, 34*(12), 3861–3875.

4. Donkor, E. A., Mazzuchi, T. A., Soyer, R., & Alan Roberson, J. (2014). Urban water demand forecasting: Review of methods and models. *Journal of Water Resources Planning and Management, 140*(2), 146–159.

5. Willis, R. M., Stewart, R. A., Panuwatwanich, K., Williams, P. R., & Hollingsworth, A. L. (2013). Quantifying the influence of environmental and water conservation attitudes on household end use water consumption. *Journal of Environmental Management, 92*(8), 1996–2009.

6. Gato, S., Jayasuriya, N., & Roberts, P. (2007). Forecasting residential water demand: Case study. *Journal of Water Resources Planning and Management, 133*(4), 309–319.

7. Harwood, S., van den Honert, R., & Rissik, D. (2018). Cumulative climate change influences and hazards affecting the Sunshine Coast. Australian Journal of Emergency Management, 33(3), 33–39

8. Jain, A. K. (2010). *Data clustering: 50 years beyond K-means*. Pattern Recognition Letters, 31(8), 651–666

9. Arbués, F., Villanúa, I., & Barberán, R. (2010). Household size and residential water demand: An empirical approach. *Australian Journal of Agricultural and Resource Economics, 54*(1), 61–80.

10. Tan, P. N., Steinbach, M., & Kumar, V. (2019). *Introduction to Data Mining*. Pearson

11. MacQueen, J. (1967). *Some methods for classification and analysis of multivariate observations*. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability

12. Thorndike, R. L. (1953). *Who belongs in the family?* Psychometrika, 18(4), 267–276

13. Kaufman, L., & Rousseeuw, P. J. (2009). Finding Groups in Data: An Introduction to Cluster Analysis. Wiley

14. Coles, S. (2001). An Introduction to Statistical Modeling of Extreme Values. Springer

15. Cominola, A., Giuliani, M., Piga, D., Castelletti, A., & Rizzoli, A. E. (2018). Benefits and challenges of using smart meters for advancing residential water demand modeling and management: A review. *Environmental Modelling & Software, 72*, 198–214.

16. Stedinger, J. R., Vogel, R. M., & Foufoula-Georgiou, E. (1993). Frequency analysis of extreme events. In D. Maidment (Ed.), Handbook of Hydrology (pp. 18.1–18.66). McGraw-Hill

17. Wilks, D. S. (2011). Statistical Methods in the Atmospheric Sciences (Vol. 100). Academic Press

18. Massey, F. J. (1951). The Kolmogorov-Smirnov test for goodness of fit. Journal of the American Statistical Association, 46(253), 68–78

19. Benesty, J., Chen, J., Huang, Y., & Cohen, I. (2009). Pearson correlation coefficient. In Noise reduction in speech processing (pp. 1–4). Springer

20. Hauke, J., & Kossowski, T. (2011). Comparison of values of Pearson's and Spearman's correlation coefficients on the same sets of data. Quaestiones Geographicae, 30(2), 87–93

21. Burnham, K. P., & Anderson, D. R. (2002). Model Selection and Multimodel Inference. Springer

22. Beal, C. D., & Flynn, J. (2015). Toward the digital water age: Progress in smart metering and smart infrastructure. *Journal of Environmental Engineering, 141*(5), 04015003.