



Determining the intensity of rainfall using the analysis of sound frequencies the impact of raindrops resulting from

Habib Palizvan Zand^{1✉} , | Abbas Ahmadi²  | Ali Kadkhodaie³ 

1. Corresponding Author, Department of Soil Science, Faculty of Agriculture, University of Tabriz, Tabriz, Iran. E-mail:

habib.palizvan@gmail.com

2. Department of Soil Science, Faculty of Agriculture, University of Tabriz, Tabriz, Iran. E-mail: a_ahmadi@tabrizu.ac.ir

3. Department of Earth Sciences, Faculty of Natural Sciences, University of Tabriz, Tabriz, Iran. E-mail: kadkhodaie_ali@tabrizu.ac.ir

Article Info

Article type: Research Article

Article history:

Received: Sep. 20, 2022

Revised: March. 21, 2023

Accepted: Apr. 15, 2023

Published online: Apr. 21, 2023

Keywords:

Audio Signals,
Clustering,
Rain Simulator,
Size Frequency.

ABSTRACT

Knowing the intensity and duration of rainfall can be useful in many environmental analyses, including the estimation of rain erosivity and soil erosion. There are various devices to record the intensity and duration of rainfall, but purchasing and maintaining them are costly and often requires an operator to take care of them. The present research deals with the feasibility of using the analysis of sound signals caused by the collision of droplets with surfaces and objects in nature to determine the intensity and duration of rainfall. For this purpose, in the laboratory of the Department of Soil Science, Faculty of Agriculture, University of Tabriz, in 2022, rain simulators were designed to produce rains of different intensities, then, the sound signals caused by the impact of raindrops with the metal tray that was placed under the rain were recorded and transferred to the computer for processing. Then, the frequency size of audio files was extracted in MATLAB software. The results showed that with the increase in rainfall intensity, the audio amplitude and frequency size of the audio signals increased. Then, the frequency measurements were automatically placed in two clusters in SPSS software using the two-stage clustering method. Then the mean and standard deviation of each cluster were calculated and according to the correlation of each with each other and with the intensity of rainfall, and in order to avoid the multi-collinearity phenomenon, only the average of the second cluster was used as the input of gene expression programming and linear regression models. In order to test the accuracy and correctness of the results obtained from the models, the coefficient of determination (R^2), root mean square error (RMSE), geometric mean of error ratio (GMER), geometric standard deviation of error ratio (GSDER) statistics were used. The values of R^2 , RMSE (mm/h), GMER (mm/h) and GSDER (mm/h) for the gene expression programming model in the training series data were 0.97, 1.85, 1.11 and 1.09 respectively and for the validation series data were 0.96, 2.05, 1.14 and 1.12 respectively. While the values of the above criteria in the regression model were 0.94, 2.74, 1.25 and 1.34 respectively for the training series data and 0.92, 2.91, 1.28 and 1.37 respectively for the validation series data. The results of the above statistics indicate that the gene expression programming model is relatively more accurate than the regression and overestimation model, and the estimated data of the regression model is relatively more spread than the gene expression programming model.

Cite this article: Palizvan Zand, H., Ahmadi, A., & Kadkhodaie, A. (2023). Determining the intensity of rainfall using the analysis of sound frequencies resulting from the impact of raindrops. *Iranian Journal of Soil and Water Research*, 54 (2), 319-335. <https://doi.org/10.22059/ijswr.2023.348946.669367>

© The Author(s).

Publisher: University of Tehran Press.

DOI: <https://doi.org/10.22059/ijswr.2023.348946.669367>



تعیین شدت بارندگی با استفاده از تجزیه و تحلیل فرکانس‌های صوتی حاصل از صدای برخورد قطرات باران

حبيب پاليزوان زند^۱ | عباس احمدی^۲ | علی کدخدائی^۳

۱. نویسنده مسئول، گروه علوم خاک، دانشکده کشاورزی، دانشگاه تبریز، تبریز، ایران. رایانامه: habib.palizvan@gmail.com

۲. گروه علوم خاک، دانشکده کشاورزی، دانشگاه تبریز، تبریز، ایران. رایانامه: a_ahmadi@tabrizu.ac.ir

۳. گروه علوم زمین، دانشکده علوم طبیعی، دانشگاه تبریز، تبریز، ایران. رایانامه: kadkhodaie.ali@tabrizu.ac.ir

اطلاعات مقاله	چکیده
نوع مقاله: مقاله پژوهشی	اطلاع از شدت و مدت بارندگی می‌تواند در بسیاری از تحلیل‌های زیست محیطی از جمله برآورد فرسایش خاک و فرسایش خاک مفید باشد. دستگاه‌های مختلفی برای ثبت شدت و مدت بارندگی وجود دارد. اما خرید و نگهداری آن‌ها هزینه‌بر بوده و اغلب نیاز به متصدی جهت مراقبت از آن‌ها دارد. تحقیق حاضر به امکان‌سنجی استفاده از تحلیل سیگنال‌های صوتی ناشی از برخورد قطرات با سطوح و اجسام موجود در طبیعت برای تعیین ثبت شدت و مدت بارندگی پرداخته است. برای این منظور در آزمایشگاه گروه علوم خاک دانشکده کشاورزی دانشگاه تبریز در سال ۱۴۰۰، باران‌سازهایی طراحی شد که باران‌هایی با شدت‌های متفاوت ایجاد نماید. سپس سیگنال‌های صوتی ناشی از برخورد قطرات باران با سینی فلزی که در زیر باران قرار داده شد، توسط رکودر ضبط و جهت پردازش به رایانه انتقال داده شد. سپس در نرم افزار MATLAB، اندازه فرکانسی فایل‌های صوتی استخراج گردید. نتایج نشان داد که با افزایش شدت بارندگی، دامنه صوتی و اندازه فرکانسی سیگنال‌های صوتی افزایش یافت. سپس اندازه‌های فرکانسی در نرم‌افزار SPSS به روش خوشه‌بندی دو مرحله‌ای به‌طور خودکار در دو خوشه قرار گرفته شد. سپس میانگین و انحراف معیار هر خوشه محاسبه شده و با توجه به همبستگی هر کدام با یکدیگر و با شدت بارندگی، و جهت جلوگیری از پدیده چند هم‌خطی شدن تنها از میانگین خوشه دوم به عنوان ورودی مدل‌های برنامه‌ریزی بیان ژن و رگرسیون خطی استفاده شد. جهت آزمون دقت و صحت نتایج حاصل از مدل‌ها، از آماره‌های ضریب تبیین (R^2)، ریشه میانگین مربعات خطا (RMSE)، میانگین هندسی نسبت خطا (GMER) و انحراف استاندارد هندسی نسبت خطا (GSDER) استفاده شد. مقادیر R^2 ، RMSE (mm/h)، GMER (mm/h) و GSDER (mm/h) برای مدل برنامه‌ریزی بیان ژن در داده‌های سری آموزش به ترتیب برابر ۰/۹۷، ۱/۸۵، ۱/۱۱ و ۱/۰۹ و برای داده‌های سری اعتبارسنجی به ترتیب برابر ۰/۹۶، ۲/۰۵، ۱/۱۴ و ۱/۱۲ بدست آمد. در حالی که مقادیر معیارهای فوق در مدل رگرسیونی، برای داده‌های سری آموزش به ترتیب برابر ۰/۹۴، ۲/۷۴، ۱/۲۵ و ۱/۳۴ و برای داده‌های سری اعتبارسنجی به ترتیب برابر ۰/۹۲، ۲/۹۱ و ۱/۲۸ و ۱/۳۷ بدست آمد. نتایج آماره‌های فوق حاکی از دقت و صحت نسبتاً بیشتر مدل برنامه‌ریزی بیان ژن نسبت به مدل رگرسیونی و بیش‌برآوردی و پخشیدگی نسبتاً زیادتر داده‌های تخمینی مدل رگرسیونی نسبت به مدل برنامه‌ریزی بیان ژن می‌باشد.
تاریخ دریافت: ۱۴۰۱/۶/۲۹	
تاریخ بازنگری: ۱۴۰۲/۱/۱	
تاریخ پذیرش: ۱۴۰۲/۱/۲۶	
تاریخ انتشار: ۱۴۰۲/۲/۱	
واژه‌های کلیدی: اندازه فرکانسی، باران‌ساز، خوشه‌بندی، سیگنال‌های صوتی.	

استاد: پالیزوان زند؛ حبيب، احمدی؛ عباس، کدخدائی؛ علی، (۱۴۰۲). تعیین شدت بارندگی با استفاده از تجزیه و تحلیل فرکانس‌های صوتی حاصل از صدای برخورد قطرات

باران. مجله تحقیقات آب و خاک ایران، ۵۴ (۲)، ۳۱۹-۳۳۵. <https://doi.org/10.22059/ijswr.2023.348946.669367>



© نویسنندگان.

ناشر: مؤسسه انتشارات دانشگاه تهران.

DOI: <https://doi.org/10.22059/ijswr.2023.348946.669367>

مقدمه

بارندگی یکی از پارامترهای مهم هواشناسی و هیدرولوژیکی است (Hu *et al.*, 2019). همچنین باران یکی از پارامترهای مهم چرخه آب به شمار می‌رود که با ایجاد خشکسالی و سیل، خسارات زیادی را بر جای می‌گذارد (Dang, 2020). داده‌های دقیق بارندگی در پیش‌بینی‌های بلند مدت و کوتاه مدت هواشناسی از اهمیت بالایی برخوردار است چون که با اقدامات پیش‌گیرانه کاهش خطرات طبیعی را امکان‌پذیر نموده است (Melese *et al.*, 2018). بارش، پدیده‌ای غیرخطی و متغیر با زمان و مکان به شمار می‌رود (Dang, 2020). بنابراین دسترسی به داده‌های قابل اعتماد بارندگی با وضوح زمانی- مکانی بالا به منظور پایش دقیق شدت بارندگی جهت پیش‌بینی سیلاب و کاهش بلایا حائز اهمیت فراوان است (Cristiano *et al.*, 2019).

روش‌های سنتی که برای اندازه‌گیری بارش مورد استفاده قرار می‌گیرند، معمولاً مبتنی بر ایستگاه‌های اقلیم‌شناسی، سینوپتیک زمینی و باران‌سنج‌ها بوده و دارای مشکلاتی از قبیل هزینه‌بر بودن، کمبود ایستگاه و عدم امکان استقرار دستگاه‌های ثبت‌کننده در مناطق سخت‌گذر می‌باشند (Beritelli *et al.*, 2018). برای بهبود دقت اندازه‌گیری، بایستی باران‌سنج‌ها از تعداد کافی برخوردار بوده و پراکنش مناسبی داشته باشند، اما این امر مستلزم هزینه زیادی است.

در صورتی که تعدادی باران‌سنج ثابت در منطقه وجود داشته باشد، می‌توان با روش‌های متداول میان‌یابی، اطلاعات مورد نیاز را برای سایر نقاط فاقد باران‌سنج ثابت استخراج کرد. ولی عموماً این مقادیر توأم با خطا و عدم قطعیت خواهد بود (Beritelli *et al.*, 2018). برای کاهش چنین عدم قطعیت‌هایی، محققین معادلات تجربی شدت-مدت- فراوانی (IDF) را پیشنهاد کرده‌اند (Jaleel & Farawn, 2013; Rasel & Islam, 2015; Al-Amri & Subyani, 2017; Kyaw *et al.*, 2022).

در اغلب موارد آمار و اطلاعات ناحیه مورد مطالعه در اختیار محقق نبوده و یا در صورت وجود از طول دوره آماری کافی برخوردار نمی‌باشد و این امر باعث ایجاد خطا و ناطمینانی در روابط تجربی خواهد شد. همچنین با توجه به تغییرات مکانی- زمانی بارندگی‌ها و فقدان و تراکم محدود ایستگاه‌های زمینی، استفاده از رادارهای آب و هواشناسی و ماهواره‌ها راه‌حل خوبی برای مواجهه با این مشکل است. رادار، یکی از ابزارهای سنجش از دور در هواشناسی برای شناسایی و پایش پدیده‌های جوی است. این ابزار با ارسال موج الکترومغناطیسی با توان مشخص به سوی جو، میزان توان برگشتی در اثر برخورد به یکی از پدیده‌های جوی (مانند باران) را ارزیابی نموده و با بررسی اختلاف میان توان موج ارسالی و برگشتی، به سنجش شرایط جوی و پدیده‌های وضع هوا از جمله بارش می‌پردازد. میزان بارش اندازه‌گیری شده توسط رادار با میزان بارش دریافت شده در سطح زمین تفاوت دارد. این اختلاف علل متعددی دارد که برخی از آن‌ها به ماهیت رادار و برخی دیگر به شرایط آب و هوایی هر منطقه و ویژگی‌های سطح زمین مربوط می‌شود (Uijlenhoet, 2001).

با توجه به پیشرفت‌های زیادی که درباره افزایش دقت رادار و گسترش شبکه ایستگاه‌های رادار در نقاط مختلف دنیا و افزایش علم هواشناسی و محققان انجام گرفته است، باز هم مقادیر مجموع بارش برآورد شده رادار با مقادیر اندازه‌گیری شده توسط ایستگاه‌های زمینی ناسازگاری دارد. بایستی توجه داشت که باران‌سنج‌ها در یک نقطه ثابت قرار داشته و داده‌های همان مکان را برداشت می‌کند و جانمایی نادرست آن موجب اشتباه می‌شود. همچنین نابرابری مقیاس اندازه‌گیری دلیل دیگری بر ناسازگاری است؛ زیرا ابعاد دهانه باران‌سنج ۲۰۰ تا ۳۰۰ سانتی‌متر مربع است در حالی که یک پیکسل رادار دارای ابعاد ۱×۱ یا ۲×۲ کیلومتر است (Prateek, 2017).

ماهواره‌ها، یکی دیگر از ابزارهای سنجش از دور در هواشناسی برای شناسایی و پایش پدیده‌های جوی هستند. در چند دهه گذشته، با در هم آمیختن دانش و فن‌آوری سنجش از دور با پدیده‌های هواشناسی، مدل‌های پیشرفته‌ای برای برآورد ویژگی‌های بارش ارائه شده‌اند که دارای توانایی بالایی در آشکارسازی زمانی و مکانی هستند. از جمله این ماهواره‌ها می‌توان به داده‌های ماهواره TRMM^۱ اشاره کرد. این ماهواره در ۲۷ نوامبر ۱۹۹۷ با همکاری سازمان فضایی ایالات متحده آمریکا و ژاپن به فضا پرتاب شد و مأموریت آن اندازه‌گیری باران‌های مناطق حاره و نیمه‌حاره از طریق سنجده‌های ماکروویو، مادون قرمز و نور مرئی است. بارندگی مناطق حاره بیش از ۷۵ درصد از بارندگی‌های کل جهان را شامل می‌شود (Liang *et al.*, 2019). از دیگر داده‌های ماهواره‌ای معروف می‌توان به داده‌های ماهواره CMORPH^۲ (Joyce *et al.*, 2004) و داده‌های ماهواره PERSIANN^۳ (Hong *et al.*, 2004) اشاره کرد.

Tramblay *et al.* (2016) داده‌های بارش ۵ ماهواره RFE 2.0, PERSIANN-CDR, CMORPH1.0 version 0.x و

1 Tropical Rainfall Measuring Mission

2 Climate Prediction Center Morphing Algorithm

3 Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks

TRMM-3B42 v6 & v7 را برای مدل‌سازی هیدرولوژی در مراکش بررسی کردند. داده‌های بارش این ماهواره‌ها با ۱۰ ایستگاه باران‌سنجی و ۴ روش درون‌یابی مختلف (inverse distance, nearest neighbor, ordinary kriging and residual kriging with) مقایسه شدند. پژوهش آن‌ها در مقیاس روزانه و ماهانه به بررسی داده‌های بارش در حوضه آبریزی با مساحت ۱۷۸۵ کیلومتر مربع در مراکش انجام شده است. نتایج حاصل از پژوهش آن‌ها نشان داد که به جزء روش nearest neighbor، سایر روش‌های درون‌یابی بررسی شده، با داده‌های مشاهده‌ای دارای اختلاف کمی بوده و داده TRMM-3B42 v7 بهترین عملکرد را نسبت به دیگر داده ماهواره‌ها دارد. علیرغم عملکرد ضعیف مقیاس روزانه، نتایج داده‌های ماهانه بارش خوب ارزیابی شد و مناسب برای انجام مدل‌سازی می‌باشد. با وجود این که ماهواره‌ها یکی از گزینه‌های مناسب برای مشاهده بارندگی در مقیاس‌های وسیع هستند که اندازه‌گیری‌ها را با قدرت تفکیک زمانی و مکانی کم انجام می‌دهند ولی استفاده از اندازه‌گیری‌های انجام شده توسط ماهواره‌ها برای مساحت‌های کوچک (کمتر از ۱۰۰۰ کیلومتر مربع) سخت است. علاوه بر این دقت بارندگی تخمین زده شده از این روش با کاهش مقیاس زمانی (مانند تبدیل ماهانه به روزانه) کاهش می‌یابد (Kizza et al., 2012).

Beritelli و همکاران (2018) برای تخمین شدت بارندگی از روشی براساس شدت امواج سیگنال رسیده از ترمینال‌های موبایل (LTE/4G) استفاده کردند. آن‌ها از روش کلاس‌بندی جدید شبکه عصبی احتمالی^۱ برای دستیابی به یک کلاس‌بندی دقیق بین چهار شدت بارندگی (بدون بارندگی، بارندگی ضعیف، بارندگی متوسط و بارندگی شدید) و بر اساس سه سطح سیگنال دریافتی (RSL)، ویژگی‌های محلی 4G/LTE شامل: RSL لحظه‌ای، RSL میانگین و واریانس RSL ارائه کردند. روش کلاس‌بندی آن‌ها برای تخمین شدت بارندگی عملکرد خوبی از خود نشان داد. آنان میزان دقت کلاس‌بندی را ۹۶/۷ درصد به دست آوردند. بطور کلی کاربرد داده‌های صوتی در بسیاری از مباحث مرتبط با منابع آب و خاک هنوز به طور جدی مورد استفاده قرار نگرفته است، به ویژه در ایران تحقیقات امواج صوتی در علوم منابع طبیعی و محیط زیست، تحقیقاتی جدید به شمار می‌روند. بنابراین لازم است تحقیقات بیشتر و متنوع‌تری در ارتباط با بکارگیری این روش در شاخه‌های مختلف مدیریت جامع منابع آب و خاک صورت پذیرد تا در نهایت بتوان با صرف وقت و هزینه کمتر راه‌حل‌های دقیق‌تر و صحیح‌تری را در موضوعات مرتبط با آن به دست آورد و دقت پیش‌بینی‌ها و مدل‌سازی‌ها را افزایش داد. در این پژوهش روش جدید و نوین برای تخمین شدت بارندگی بر اساس جمع‌آوری اطلاعات داده‌های صوتی و تجزیه و تحلیل فرکانس‌های صوتی ارائه شده است.

مواد و روش‌ها

برای ایجاد شدت‌های مختلف بارندگی، ابتدا ته ظروفی پلاستیکی با استفاده از سوزن‌هایی با ضخامت متفاوت سوراخ گردید و سپس نخ‌هایی با قطر مختلف از آن عبور داده شد (شکل ۱).



شکل ۱. ابزار مورد نیاز جهت تولید بارش‌هایی با شدت‌های مختلف

سوراخ‌های ایجاد شده بوسیله سوزن‌ها، و نخ‌های عبور کرده از آن‌ها که دارای قطرهای مختلف بود، اندازه قطرات مختلف را بوجود آورد. حجم قطرات سقوط کرده به مدت ۱ دقیقه در ظرفی استوانه‌ای شکل با قطر ۲۰ سانتی‌متر جمع‌آوری شد. سپس حجم آب جمع‌آوری

1 probabilistic neural network

2 received signal level

شده در استوانه مدرج اندازه گیری گردید. سپس شدت ریزش قطرات با استفاده از رابطه (۱) محاسبه گردید:

$$I = \frac{V}{A \cdot t}$$

طبق رابطه (۱) I ، شدت ریزش قطرات بر حسب سانتی متر بر دقیقه؛ V ، حجم آب جمع آوری شده بر حسب سانتی متر مکعب؛ A ، سطح مقطع ظرف جمع آوری بر حسب سانتی متر مربع و t ، مدت زمان جمع آوری قطرات بر حسب دقیقه می باشد.

جمع آوری داده های صوتی

داده ها در سال ۱۴۰۰ در آزمایشگاه گروه علوم خاک دانشکده کشاورزی دانشگاه تبریز جمع آوری شد. در این راستا برای جمع آوری داده های صوتی ناشی از بارندگی، یک سینی فلزی برای تشدید صدای ناشی از برخورد قطرات در روی زمین قرار داده شد. سیگنال های صوتی ایجاد شده در ۴۰ شدت مختلف بارش به مدت ۱ دقیقه در ۳ تکرار توسط رکوردر REMAX مدل RPI با فرمت wav ضبط گردید.

پردازش سیگنال های صوتی

تمامی پردازش ها در نرم افزارهای شبیه ساز به صورت گسسته در زمان انجام می گیرد؛ زیرا سیگنال ها در حافظه کامپیوتر بصورت گسسته در زمان نشان داده می شوند. بنابراین نمی توان یک سیگنال پیوسته در زمان واقعی را در کامپیوتر تولید کنیم و تنها کاری که باید انجام داد این است که از طریق اتخاذ نرخ نمونه برداری، یک سیگنال شبه پیوسته ایجاد کنیم. بر اساس قضیه نایکوئیست-شانون، یک نرخ نمونه برداری کمینه وجود دارد که در آن می توان یک سیگنال پیوسته را به نمونه های آن تبدیل و بدون اعوجاج آن ها را بازیابی کرد. طبق این قضیه بایستی رابطه زیر در نمونه برداری برقرار باشد (Oppenheim et al., 2006):

$$f_s \geq 2f_m \quad \text{رابطه (۲)}$$

طبق رابطه (۲) فرکانس نمونه برداری (f_s) بایستی بزرگتر و مساوی با دو برابر بزرگترین مولفه فرکانسی سیگنال اصلی (f_m) باشد. بر این اساس نرخ نمونه برداری سیگنال ورودی ۴۴۱۰۰ هر تتر انتخاب شد. سپس برای تحلیل خصوصیات فرکانسی، با تبدیل فوریه، سیگنال های صوتی از حوزه زمانی به حوزه فرکانسی انتقال داده شدند و جهت نرم شدن خصوصیات فرکانسی از فیلتر میانگین متحرک استفاده گردید. سپس از تبدیل فوریه قدرمطلق گرفته شده و اندازه فرکانس ها بدست آمد و از آن برای مدل سازی استفاده گردید.

رابطه (۳) تبدیل فوریه سیگنال زمان گسسته و رابطه (۴) تبدیل معکوس فوریه سیگنال زمان گسسته را نشان می دهد (Oppenheim et al., 2006):

$$X(j\Omega) = \sum_{n=-\infty}^{\infty} X[n]e^{-jn\Omega} \quad \text{رابطه (۳)}$$

$$X[n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(j\Omega)e^{jn\Omega} d\Omega \quad \text{رابطه (۴)}$$

در روابط ۳ و ۴ به ترتیب $X[n]$ ، $X(j\Omega)$ و Ω نشان دهنده سیگنال گسسته در زمان، حوزه فرکانسی سیگنال گسسته در زمان و فرکانس زاویه ای است.

از الگوریتم تبدیل فوریه سریع (FFT) جهت انتقال سیگنال های صوتی از حوزه زمانی به حوزه فرکانسی استفاده گردید چون الگوریتم تبدیل فوریه سریع (FFT) حجم محاسبات را به شدت کاهش داده و در نتیجه سرعت پردازش را افزایش می دهد (Weeks, 2010).

مدل سازی

خوشه بندی:

تحلیل خوشه ای، شامل مجموعه ای از الگوریتم ها است که جهت گروه بندی موضوعات یا اشیاء مشابه در طبقه بندی های مشابه استفاده می شود (Neal & Wurst, 2001). تحلیل خوشه ای می تواند جهت کشف ساختارهای میان داده ها به کار گرفته شود. به عبارت دیگر، این روش به سادگی ساختارها را در داده ها کشف می کند (Chiu et al., 2001).

پس از آن که اندازه فرکانس سیگنال های صوتی ۴۰ شدت بارندگی به دست آمد، با بهره گیری از نرم افزار SPSS از تکنیک خوشه بندی دو مرحله ای جهت تحلیل و گروه بندی این داده ها که دارای مجموعه داده های بزرگ می باشد، استفاده شد. یکی از مزیت های عمده و مهم تحلیل خوشه ای دو مرحله ای توان کار با داده های بزرگ است زیرا این داده ها نیاز به صرف فرآیند زمان بیشتری دارند و این روش نسبت به سایر روش های خوشه بندی زمان کمتری را صرف می نماید (Schioapu, 2010).

در این روش در گام اول که به مرحله پیش کلاستر^۱ معروف است، رکوردها به چندین خرده کلاستر تبدیل می‌شود. روش پیش کلاستر با ایجاد یک ساختار داده به نام درخت CF^2 یا (ویژگی‌های خوشه)، که شامل مراکز خوشه می‌باشد اجرا می‌شود (رابطه ۵)

$$CF_j = \{N_j, S_{Aj}, S_{Aj}^2, N_{Bj}\} \quad \text{(رابطه ۵)}$$

در رابطه (۵) N_j تعداد رکورد داده‌های خوشه j ، S_{Aj} مجموع شاخص‌های پیوسته از رکوردهای داده N_j ، و S_{Aj}^2 مجموع مربعات شاخص‌های پیوسته از رکوردهای خوشه j یا (N_j) ، و همچنین

$$N_{Bj} = (N_{Bj1}, N_{Bj2}, \dots, N_{BjK_B}) \quad \text{(رابطه ۶)}$$

برابر با:

$$N_{BjK_B} = \sum_{K=1}^{K_B} (L_K - 1) \quad \text{(رابطه ۷)}$$

رابطه ۷ مجموع ابعاد بردار دسته‌بندی‌ها برای متغیر طبقه‌ای را نشان می‌دهد و L_K در آن تعداد دسته‌بندی‌ها برای متغیر طبقه K است که از طریق تعداد رکورد داده‌ها در خوشه j که K امین شاخص طبقه‌ای گرفته شده از دسته L است، به دست می‌آید. هنگامی که دو کلاستر با هم ترکیب می‌شوند بدین معنی است که دو مجموعه متناظر از داده‌ها، نقاط جمع شده از یک واحد هستند (رابطه ۸).

$$CF_{\langle i,j \rangle} = \{N_j + N_i, S_{Aj} + S_{Ai}, S_{Aj}^2 + S_{Ai}^2, N_{Bj} + N_{Bi}\} \quad \text{(رابطه ۸)}$$

درخت CF شامل سطوح گره‌هایی است که هر گره یک عدد از ورودی‌ها را در بر دارد. هر برگ نیز به عنوان ورودی یک زیر سیستم نهایی است. الگوریتم برای هر رکورد، از ریشه گره شروع و نزدیک‌ترین گره فرزند به صورت نزولی و بازگشتی در طول درخت CF پیدا می‌شود. هنگام رسیدن به یک برگ گره، الگوریتم نزدیک‌ترین برگ ورودی در گره برگ را پیدا می‌کند. اگر رکورد داخل فاصله آستانه نزدیک‌ترین برگ ورودی است، پس آن رکورد به برگ ورودی اضافه شده و درخت CF به روز می‌شود. در غیر این صورت، آن برای گره برگ ارزش‌های جدید را ایجاد می‌کند. سپس در گام دوم این خرده کلاسترها به تعداد کلاستر مطلوب طبقه‌بندی می‌شوند. یکی از مزیت‌های این روش این است که اگر تعداد کلاسترهای مطلوب برای محقق مشخص نباشد خود روش به صورت خودکار تعداد مناسب این کلاسترها را پیدا می‌کند برای مناسب بودن تعداد خوشه‌ها نیز از دو شاخص BIC و AIC استفاده می‌شود (Şchiopu, 2010).

در این روش فاصله بین خوشه‌ها از فاصله لگاریتم احتمالی استفاده می‌شود و فاصله بین دو خوشه به میزان کاهش لگاریتم طبیعی تابع احتمالی بستگی دارد تا بتواند در یک خوشه گروه‌بندی شوند. فاصله بین دو خوشه i و j به صورت زیر تعریف می‌شود (Şchiopu, 2010):

$$d(i, j) = \xi_i + \xi_j - \xi_{\langle i,j \rangle} \quad \text{(رابطه ۹)}$$

$$\xi_S = \left[\sum_{K=1}^{K_A} \frac{1}{2} \log(\sigma_K^2 + \sigma_{SK}^2) + \sum_{K=1}^{K_B} E_{SK} \right] \quad \text{(رابطه ۱۰)}$$

$$E_{SK} = - \sum_{L=1}^{L_K} \frac{N_{SKL}}{N_S} \log \frac{N_{SKL}}{N_S} \quad \text{(رابطه ۱۱)}$$

در روابط بالا:

$d(i,j)$: فاصله بین دو خوشه i و j

$\xi_{\langle i,j \rangle}$: نشان دهنده خوشه تشکیل شده با ترکیب خوشه‌های i و j

K_A : تعداد کل متغیرهای پیوسته

K_B : تعداد کل متغیرهای طبقه‌ای

L_K : تعداد دسته‌بندی‌ها برای متغیر طبقه K ام

N_S : تعداد کل رکوردهای داده در خوشه S

N_{SKL} : تعداد رکوردها در خوشه S که متغیر طبقه K ام که L دسته می‌گیرد.

σ_K^2 : واریانس برآورد شده از متغیر پیوسته K ام برای مجموع داده‌های ورودی

σ_{KS}^2 : واریانس برآورد شده از متغیر پیوسته K ام در خوشه‌های Z

بعد از مشخص شدن تعداد خوشه‌ها، میانگین و انحراف استاندارد هر خوشه محاسبه شد و به عنوان ورودی در مدل‌های رگرسیونی و برنامه‌ریزی بیان ژن بکار برده شد.

مدل برنامه‌ریزی بیان ژن:

برنامه‌ریزی بیان ژن تعمیم یافته روش الگوریتم ژنتیک می‌باشد که برای اولین بار توسط Koza (1992) بر اساس تئوری داروین ارائه شد. همچنین این مدل جزء روش‌های الگوریتم گردشی محسوب می‌ود. الگوریتم‌های یاد شده اقدام به تعریف یک تابع هدف در قالب معیارهای کیفی نموده و سپس تابع یاد شده را برای مقایسه جواب‌های مختلف حل مسئله در یک فرآیند گام به گام تصحیح ساختار داده‌ها به کار می‌گیرند و در نهایت، جواب مناسب را ارائه می‌نمایند (Alvisi et al., 2005). تفاوت اساسی موجود بین الگوریتم ژنتیک و برنامه‌ریزی بیان ژن به طبیعت هر یک از افراد برمی‌گردد، به نحوی که افراد در الگوریتم ژنتیک، ردیف‌های خطی با طول ثابت می‌باشند (کروموزوم‌ها) ولی در برنامه‌ریزی ژنتیک، همان شاخه‌های مجزا می‌باشند. همچنین در برنامه‌ریزی بیان ژن بر ساختار درختی مجموعه‌ها تأکید می‌شود ولی الگوریتم ژنتیک، بر اساس سیستم ارقام دودویی عمل می‌نماید. در این روش کروموزوم‌های خطی و ساده با طول ثابت، مشابه با آنچه که در الگوریتم ژنتیک استفاده می‌شود و ساختارهای شاخه‌ای با اندازه‌ها و اشکال متفاوت، مشابه با درختان تجزیه در برنامه‌ریزی بیان ژن ترکیب می‌شوند. اولین مرحله در الگوریتم مدل، تولید جمعیت اولیه از راه حل‌هاست. این امر می‌تواند به وسیله فرآیند تصادفی و یا در نظر گرفتن اطلاعات ورودی درباره مسئله انجام شود. سپس کروموزوم‌ها به صورت بیان درختی اظهار شده و توسط تابع برازش ارزیابی می‌گردند. در صورت دستیابی به راه حل مطلوب و یا رسیدن نسل‌ها به تعداد معین، تکامل متوقف شده و بهترین راه حل ارائه می‌شود. اگر شرایط توقف یافت نشود، نخبه‌گزینی انجام می‌شود و باقی راه‌حل‌ها به فرآیندی گزینشی واگذار می‌شوند. این فرآیند برای چندین نسل تکرار می‌شود و با پیش رفتن نسل به جلو کیفیت جمعیت نیز به طور نسبی بهبود می‌یابد (Ferreira, 2006). در برنامه‌ریزی بیان ژن از عملگرهای مختلفی مانند جهش و ترکیب استفاده می‌شود. هدف عملگر جهش، بهسازی تصادفی، داخل کروموزوم‌های معین است. خصوصیت این عملگر آن است که برای جلوگیری از ایجاد افراد معیوب از نظر قواعد، برخی عملیات بدون نقص را اجرا می‌نماید. در این مدل از ترکیب تک نقطه‌ای، دو نقطه‌ای و تقاطع ژنی استفاده می‌شود. از دیگر عملگرهای مورد استفاده می‌توان عملگر ترانهشت را نام برد. در این روش پدیده‌های مختلف با استفاده از مجموعه‌ای از توابع و ترمینال‌ها، مدل‌سازی می‌شوند. مجموعه توابع، معمولاً شامل توابع حسابی، مثلثاتی و یا توابع تعریف شده توسط کاربر هستند که معتقد است، می‌توانند برای تفسیر مدل مناسب باشند. مجموعه ترمینال‌ها، از مقادیر ثابت و متغیرهای مستقل مسئله تشکیل شده‌اند (Ferreira, 2006). در این پژوهش برای اجرای مدل برنامه‌ریزی بیان ژن از برنامه GeneXpro Tools (4.0) استفاده شده است.

مدل رگرسیونی

در حالت کلی معادله رگرسیون کلاسیک به صورت زیر است که در آن Y متغیر خروجی و X متغیر ورودی، a شیب خط رگرسیون و b نیز عرض از مبدا می‌باشد.

$$Y = aX + b \quad \text{رابطه (۱۲)}$$

آزمون آماری داده‌ها با استفاده از نرم‌افزار SPSS انجام گرفت. با استفاده از داده‌های میانگین و انحراف معیار هر خوشه، ابتدا همبستگی متغیرها محاسبه شده و سپس مناسب‌ترین معادله رگرسیونی بدست آمد.

ارزیابی مدل‌ها

برای ارزیابی دقت و اعتبار مدل‌ها از شاخص‌های مختلفی از جمله ضریب تبیین^۱ (R^2) و جذر میانگین مربعات خطا^۲ (RMSE)، استفاده شد (روابط ۱۳ و ۱۴)، به این ترتیب هر چه R^2 مدل بیشتر و RMSE کمتر باشد مدل‌های ایجاد شده دقیق‌تر و معتبرتر هستند.

$$R^2 = 1 - \frac{\sum_1^N (Y_i - \hat{Y}_i)^2}{\sum_1^N (Y_i - \bar{Y})^2} \quad \text{رابطه (۱۳)}$$

1 Coefficient of Determination

2 Root Mean Square Error

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (Y_i - \hat{Y}_i)^2}{N}} \quad \text{رابطه ۱۴}$$

که در آن‌ها Y_i ، \bar{Y} و \hat{Y}_i به ترتیب متغیر وابسته اندازه‌گیری شده، میانگین آن و متغیر وابسته برآورد شده می‌باشد. N نیز تعداد مشاهدات است.

معیارهای دیگری که برای ارزیابی دقت مدل‌ها به کار گرفته شد شامل؛ میانگین هندسی نسبت خطا^۱ (GMER) و انحراف استاندارد هندسی نسبت خطا^۲ (GSDER) بودند:

$$GMER = \exp\left[\frac{1}{N} \sum_{i=1}^N \ln\left(\frac{\hat{Y}_i}{Y_i}\right)\right] \quad \text{رابطه ۱۵}$$

$$GSDER = \exp\left[\frac{1}{N-1} \sum_{i=1}^N \left(\ln\left(\frac{\hat{Y}_i}{Y_i}\right) - \ln(GMER)\right)^2\right]^{\frac{1}{2}} \quad \text{رابطه ۱۶}$$

میانگین هندسی نسبت خطا (GMER) نشان دهنده میزان انطباق بین مقادیر اندازه‌گیری شده و برآورد شده است. اگر GMER برابر یک باشد، نشان دهنده انطباق کامل بین مقادیر اندازه‌گیری شده و پیش‌بینی شده است. اگر GMER بزرگ‌تر از یک باشد، نشان دهنده بیش‌برآوردی مدل و GMER کمتر از یک نشان دهنده کم‌برآوردی مدل است.

انحراف استاندارد هندسی نسبت خطا (GSDER) معیاری از پخشیدگی داده‌هاست. اگر نزدیک به یک باشد نشان دهنده پخشیدگی کمتر و هر چه اختلاف بیشتری از یک داشته باشد نشان دهنده انحراف بیشتر تخمین‌ها از داده‌های اندازه‌گیری شده است.

نتایج و بحث

بعد از قرار دادن مقادیر بدست آمده در رابطه ۷ شدت‌های مختلف بارش بر حسب میلی‌متر بر ساعت بدست آمد (جدول ۱) و جهت سهولت محاسبات و حذف اعداد اعشاری این شدت‌ها به سمت عدد بالاتر روند شد. طبق این جدول حداقل شدت بارندگی ۱ میلی‌متر بر ساعت و حداکثر شدت بارندگی ۱۵ میلی‌متر بر ساعت می‌باشد.

جدول ۱. مقادیر شدت باران

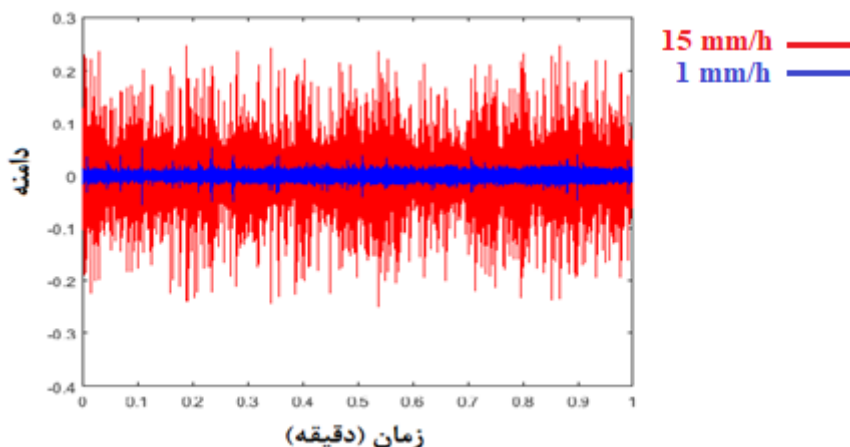
شماره ظرف	تکرار	حجم تولید شده (cm ³)	زمان (min)	سطح مقطع ظرف (cm ²) $A = \pi r^2$	شدت بارش	
					(mm.h ⁻¹)	(cm.min ⁻¹)
	۱	۰/۵	۱	۳۱۴	۰/۰۰۱۵۹	۰/۹۵۴
۱	۲	۰/۴۵	۱	۳۱۴	۰/۰۰۱۴۳	۰/۸۵۸
	۳	۰/۵	۱	۳۱۴	۰/۰۰۱۵۹	۰/۹۵۴
	۱	۰/۸	۱	۳۱۴	۰/۰۰۲۵۴	۱/۵۲۴
۲	۲	۰/۸	۱	۳۱۴	۰/۰۰۲۵۴	۱/۵۲۴
	۳	۰/۸	۱	۳۱۴	۰/۰۰۲۵۴	۱/۵۲۴
...
	۱	۷/۵	۱	۳۱۴	۰/۰۲۳۹	۱۴/۳۴۰
۴۰	۲	۷/۴	۱	۳۱۴	۰/۰۲۳۶	۱۴/۱۶۰
	۳	۷/۶	۱	۳۱۴	۰/۰۲۴۲	۱۴/۵۲۰

شکل (۲) دامنه صوتی حداقل و حداکثر شدت باران ایجاد شده در این پژوهش (۱ و ۱۵ میلی‌متر بر ساعت) را در بعد زمانی نشان می‌دهد. طبق این شکل بارندگی با شدت بیشتر دارای دامنه بیش‌تری نسبت به بارندگی با شدت کمتر می‌باشد. دامنه صوتی هر شدت بارندگی به انرژی جنبشی آن بارش بستگی دارد. انرژی جنبشی باران را معمولاً بر حسب شدت باران محاسبه می‌کنند زیرا شدت باران تابعی از قطر قطرات باران یا در واقع تابعی از جرم قطرات باران و سرعت نهایی آن‌هاست و بنابراین متناسب با انرژی جنبشی باران خواهد

1 Geometric Mean of Error Ratio

2 Geometric Standard Deviation of Error Ratio

بود (Wichmeier & Smith, 1978). بنابراین قطرات باران در شدت ۱۵ میلی‌متر بر ساعت دارای جرم و سرعت نهایی بیشتری بوده در نتیجه با انرژی بیشتری به سطح سینی برخورد کرده و دامنه صوتی بیشتری را نسبت به قطرات باران با ۱ میلی‌متر بر ساعت که دارای جرم و سرعت نهایی کمتر می‌باشد، ایجاد خواهد کرد.

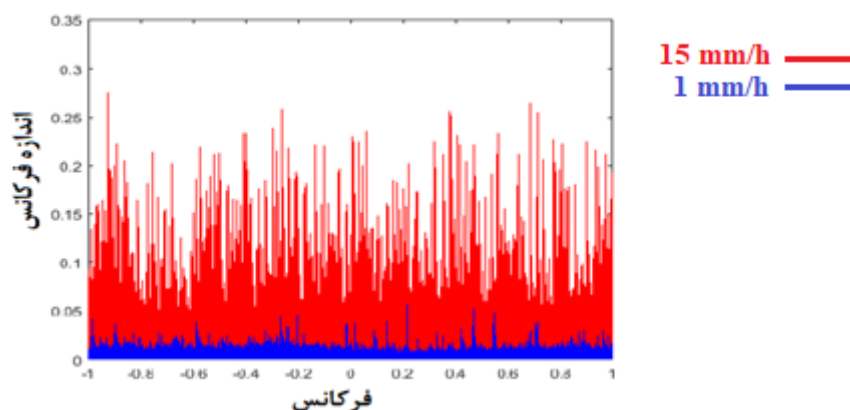


شکل ۲. نمودار دامنه صوتی شدت‌های باران ۱ و ۱۵ میلی‌متر بر ساعت در بعد زمانی

بارندگی‌هایی که دارای دامنه صوتی بیشتری هستند طبق رابطه (۹) دارای اندازه فرکانسی بیشتری نیز هستند بنابراین با کاهش شدت بارندگی اندازه فرکانس سیگنال صوتی کاهش و با افزایش شدت بارندگی اندازه فرکانس سیگنال صوتی افزایش می‌یابد. شکل (۳) اندازه فرکانسی شدت‌های بارش ۱ و ۱۵ میلی‌متر بر ساعت را در بازه فرکانسی π و $-\pi$ نشان می‌دهد. طبق این شکل بارش با شدت ۱۵ میلی‌متر بر ساعت دارای اندازه فرکانسی بیشتری نسبت به بارش با شدت ۱ میلی‌متر بر ساعت می‌باشد. هر چه شدت بارندگی بیش‌تر باشد انرژی جنبشی و در نتیجه قدرت فرساینده آن بیش‌تر است. طبق معادله Wichmeier & Smith (رابطه ۱۷) و معادله Hudson (رابطه ۱۸) انرژی جنبشی رابطه مستقیمی با شدت بارش دارد. Nakazato و همکاران (2018) برای تخمین شدت بارندگی، مشخصات فرکانسی فایل صوتی هر شدت بارندگی را بوسیله تبدیل فوریه بدست آورده و مشاهده کردند که مشخصات فرکانسی فایل صوتی شدت‌های مختلف بارندگی متفاوت است. آن‌ها مشاهده کردند که بین مشخصات فرکانسی و شدت بارندگی ارتباط وجود دارد. به طوری که با افزایش طیف توان فرکانسی صدای باران، شدت بارندگی نیز افزایش می‌یابد.

$$KE = 210.2 + 89 \log I \quad (\text{Wichmeier \& Smith, 1978}) \quad (\text{رابطه ۱۷})$$

$$KE = 29.8 - \frac{127.5}{I} \quad (\text{Hudson, 1981}) \quad (\text{رابطه ۱۸})$$



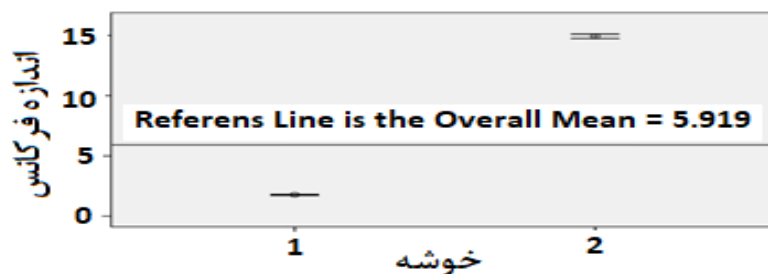
شکل ۳. اندازه فرکانسی شدت‌های بارش ۱ و ۱۵ میلی‌متر بر ساعت در بازه فرکانسی π و $-\pi$

بر اساس خوشه‌بندی دو مرحله‌ای، اندازه فرکانس‌های به دست آمده برای شدت‌های مختلف بارندگی در دو خوشه قرار گرفتند و میانگین و انحراف معیار هر خوشه مشخص شد و به عنوان ورودی در مدل‌های رگرسیون و برنامه‌ریزی بیان ژن مورد استفاده قرار گرفت

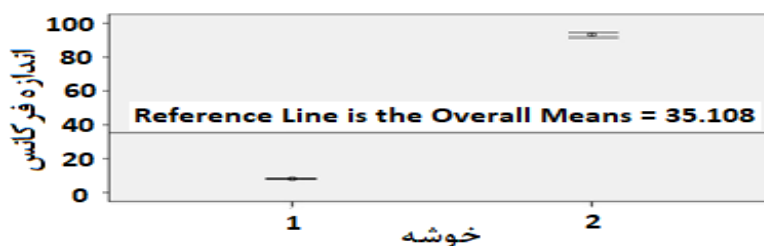
جدول ۲). شکل (۴) خوشه‌بندی شدت‌های بارش ۱ و ۱۵ میلی‌متر بر ساعت را با سطح اطمینان ۹۵ درصد نشان می‌دهد. طبق این شکل میانگین، حدود بالا و پایین هر خوشه در شدت بارش‌های مذکور نشان داده شده است. طبق جدول (۲) و شکل (۴)، شدت بارش ۱۵ میلی‌متر بر ساعت دارای میانگین بیشتری در هر دو خوشه نسبت به شدت بارش ۱ میلی‌متر بر ساعت می‌باشد. طبق رابطه‌های ۱۷ و ۱۸ هر چه شدت بارش افزایش یابد، انرژی جنبشی نیز افزایش می‌یابد در نتیجه دامنه سیگنال صوتی ناشی از برخورد قطرات باران با سطح افزایش می‌یابد و به تبع آن اندازه فرکانسی نیز افزایش می‌یابد.

جدول ۲. مقادیر میانگین و انحراف معیار اندازه فرکانس سیگنال برای شدت‌های مختلف بارندگی در هر خوشه

شماره ظرف	تکرار	خوشه	شدت بارش (mm/h)	میانگین	انحراف معیار
	۱	۱	۱	۱/۸۶۴	۴/۱۱۸
	۱	۲	۱	۱۵/۲۶۸	۱۱/۰۷۳
۱	۲	۱	۱	۱/۷۷	۳/۲۶۹
	۲	۲	۱	۱۴/۹۳۳	۹/۷۷۴
	۳	۱	۱	۱/۹۸۵	۴/۳۹۶
	۳	۲	۱	۱۵/۳۴۸	۱۱/۵۸۴
...
	۱	۱	۱۵	۷/۸۳۷	۱۳/۹۰۵
	۱	۲	۱۵	۸۴/۴۵۰	۵۹/۴۵۶
	۲	۱	۱۵	۷/۴۷۷	۱۳/۰۸۲
۴۰	۲	۲	۱۵	۸۱/۵۷۶	۵۷/۲۰۰
	۳	۱	۱۵	۸/۴۶۰	۱۳/۹۹۶
	۳	۲	۱۵	۹۳/۰۱۵	۶۷/۵۰۵



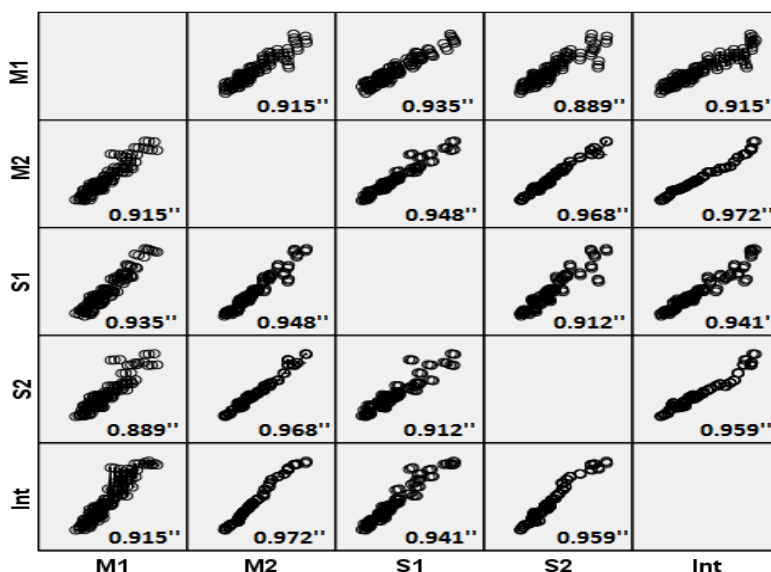
الف. شدت بارندگی ۱ میلی‌متر بر ساعت



ب. شدت بارندگی ۱۵ میلی‌متر بر ساعت

شکل ۴. خوشه بندی دو مرحله‌ای شدت‌های بارش ۱ و ۱۵ میلی‌متر بر ساعت

با توجه به شکل (۵) که ماتریس همبستگی پیرسون را نشان می‌دهد همه متغیرهای مستقل شامل میانگین خوشه ۱ (M1)، میانگین خوشه ۲ (M2)، انحراف معیار خوشه ۱ (S1) و انحراف معیار خوشه ۲ (S2) همبستگی معنی‌داری در سطح احتمال ۱ و ۵ درصد با شدت بارندگی (Int) نشان دادند. به ترتیب M1، S2، S1 و M2 همبستگی بالایی را با شدت بارندگی داشتند. همبستگی بالای این متغیرها نشان دهنده تاثیر شدت بارندگی بر اندازه فرکانس سیگنال‌های صوتی می‌باشد.



شکل ۵. ماتریس همبستگی پیرسون متغیرهای میانگین خوشه ۱ (M1)، میانگین خوشه ۲ (M2)، انحراف معیار خوشه ۱ (S1)، انحراف معیار خوشه ۲ (S2) و شدت بارندگی (Int)

با توجه به همبستگی بالای M1، M2، S1 و S2 با هم‌دیگر به عنوان متغیرهای مستقل، برای جلوگیری از پدیده چندهم‌خطی شدن، متغیر M2 به عنوان ورودی مدل‌ها مورد استفاده قرار گرفت چون در بین سایر متغیرها از همبستگی بالایی با شدت بارندگی برخوردار بود. رابطه (۱۹) معادله بدست آمده از مدل برنامه‌ریزی بیان ژن را نشان می‌دهد. در این رابطه Int شدت باران و M2 میانگین اندازه فرکانس خوشه دوم در هر شدت بارندگی را نشان می‌دهد.

$$\text{Int} = \sqrt{\left((M2) + (M2 + 2.473) \right)^{\frac{1}{3}} \times M2} + \left(\exp((-1.462) \times (-1.462 \times -1.462)) \right) \times M2^{\frac{1}{3}} \times (M2 \times 0.658) + \exp\left(\cos(-1.542)^{\frac{1}{3}} \right) - ((-1.861 + M2) + M2) \quad (\text{رابطه ۱۹})$$

شکل (۶) درخت توصیفی که از مدل تکامل یافته در پایان مراحل بهینه‌سازی بدست آمده را نشان می‌دهد که شامل ۳ زیر شاخه بوده و توسط عملگر جمع متصل شده‌اند. در این شکل do میانگین اندازه فرکانسی در خوشه دوم در هر شدت بارندگی (M2)، C0 و C1 ضرایب ثابت هستند.

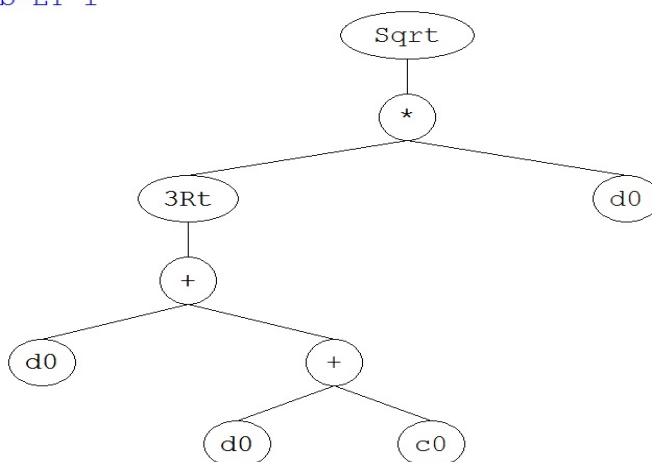
معادله رگرسیونی شدت بارندگی برای بارش‌های ایجاد شده طبق رابطه (۲۰) بدست آمد:

$$\text{Int} = 1.957 \times M2 - 21.051 \quad (\text{رابطه ۲۰})$$

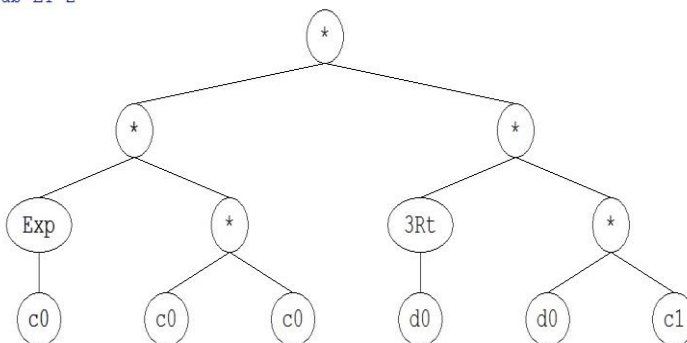
جدول (۳) مقادیر آماره‌های مورد استفاده برای ارزیابی مدل‌ها را نشان می‌دهد. همان‌طور که مشاهده می‌شود، مقادیر بیش‌تر R^2 و کمتر RMSE در مدل برنامه‌ریزی بیان ژن نسبت به مدل رگرسیونی، حاکی از دقت بالای مدل برنامه‌ریزی بیان ژن نسبت به مدل رگرسیونی می‌باشد. آماره GMER برای داده‌های آموزشی و صحت‌سنجی مدل برنامه‌ریزی بیان ژن به ترتیب برابر ۱/۱۱ و ۱/۱۴ و برای داده‌های آموزشی و صحت‌سنجی مدل رگرسیون به ترتیب برابر ۱/۲۵ و ۱/۲۸ بدست آمد، که نشان دهنده بیش‌برآوردی مدل رگرسیونی نسبت به مدل برنامه‌ریزی بیان ژن است. مقدار GSDER برای داده‌های آموزشی و صحت‌سنجی مدل برنامه‌ریزی بیان ژن به ترتیب برابر ۱/۰۹ و ۱/۱۲ و برای داده‌های آموزشی و صحت‌سنجی مدل رگرسیون به ترتیب برابر ۱/۳۴ و ۱/۳۷ بدست آمد، که آن نیز نشان دهنده عدم انطباق و پخشیدگی نسبتاً زیادتر مقادیر تخمینی مدل رگرسیونی در مقایسه با مدل برنامه‌ریزی بیان ژن است. بنابراین با توجه به مطالب فوق تمامی آماره‌های ذکر شده نشان دهنده مزیت و کارایی نسبتاً بیش‌تر مدل برنامه‌ریزی بیان ژن نسبت به مدل رگرسیونی می‌باشد. Ahmadi و همکاران (2017) برای برآورد هدایت هیدرولیکی اشباع خاک از مدل‌های برنامه‌ریزی بیان ژن و رگرسیون ریح استفاده کرده و دقت و صحت این مدل‌ها را بررسی کردند. همچنین نشان دادند که مدل برنامه‌ریزی بیان ژن نسبت به مدل رگرسیون ریح از دقت و صحت بالایی برخوردار است.



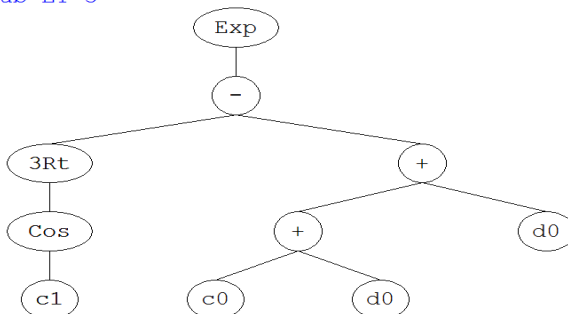
Sub-ET 1



Sub-ET 2



Sub-ET 3



شکل ۶. درخت توصیفی برای فرمول GEP که در آن میانگین اندازه فرکانس خوشه دوم (M2)، C0 و C1 ضرایب ثابت هستند.

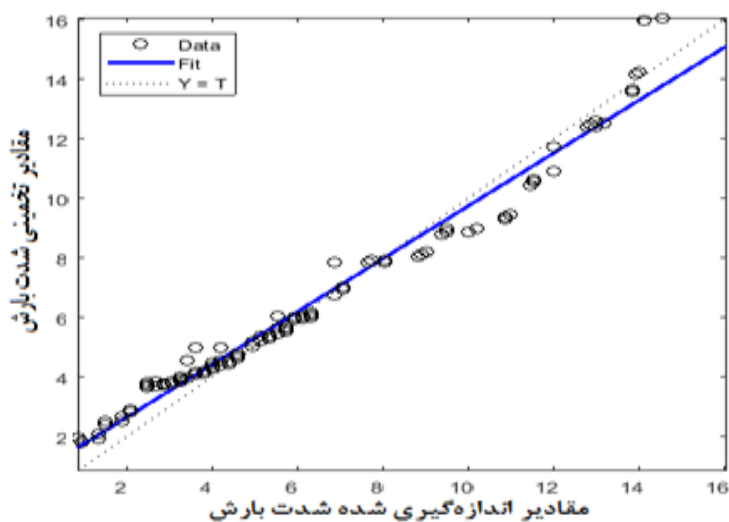
جدول ۳. آماره‌های ارزیابی مدل‌های برنامه‌ریزی بیان ژن و رگرسیون

داده‌های صحت‌سنجی				داده‌های آموزشی				مدل
GSDER	GMER	RMSE	R ²	GSDER	GMER	RMSE	R ²	
۱/۱۲	۱/۱۴	۲/۰۵	-۰/۹۶	۱/۰۹	۱/۱۱	۱/۸۵	-۰/۹۷	برنامه‌ریزی بیان ژن
۱/۳۷	۱/۲۸	۲/۹۱	-۰/۹۲	۱/۳۴	۱/۲۵	۲/۷۴	-۰/۹۴	رگرسیون

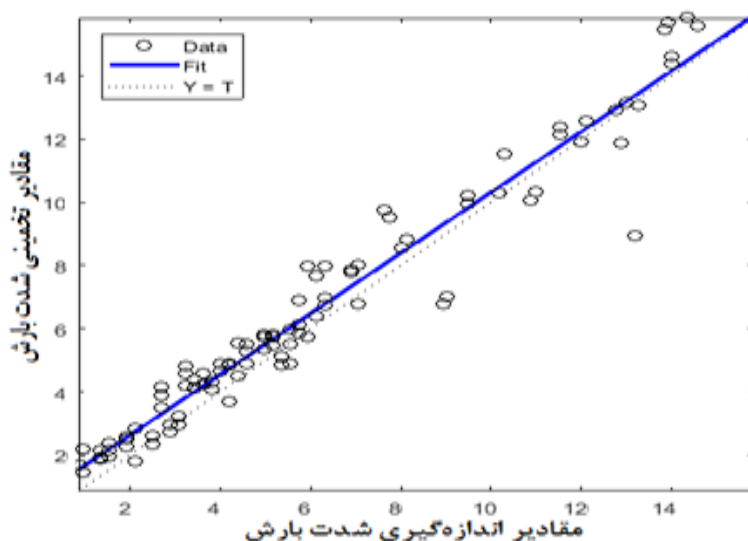
واحدهای RMSE، GMER و GSDER همگی بر حسب (mm/h) می‌باشد.

شکل (۷) خط برازش یافته و خط یک به یک بین مقادیر اندازه‌گیری شده و تخمینی شدت بارندگی توسط مدل برنامه‌ریزی بیان ژنی و مدل رگرسیون خطی را نشان می‌دهد. هر چه پراکندگی داده‌ها حول خط یک به یک کم‌تر باشد مقادیر برآوردی به مقادیر اندازه‌گیری شده نزدیک‌تر خواهد بود. در مدل برنامه‌ریزی بیان ژن داده‌ها به خط یک به یک نزدیک‌تر هستند که نشان دهنده دقت بیشتر مدل

برنامه‌ریزی بیان ژن است. Matar (2018) با استفاده از برنامه‌ریزی بیان ژن به مدل‌سازی تبخیر-تعرق مرجع ماهانه در مصر پرداخت. نتایج نشان داد که برنامه‌ریزی بیان ژن برای مدل‌سازی تبخیر-تعرق مرجع دقیق‌تر از مدل‌های تجربی است.

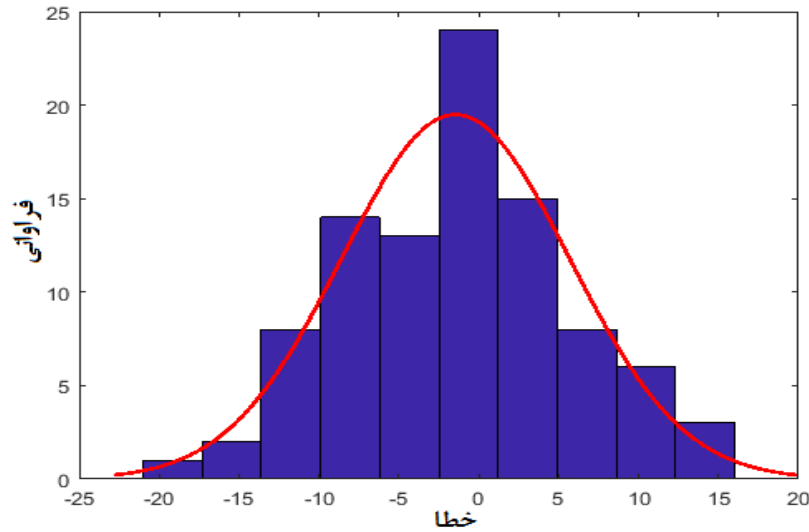


الف. مقادیر اندازه‌گیری شده و تخمینی شدت بارش (mm/h) توسط مدل برنامه‌ریزی بیان ژن

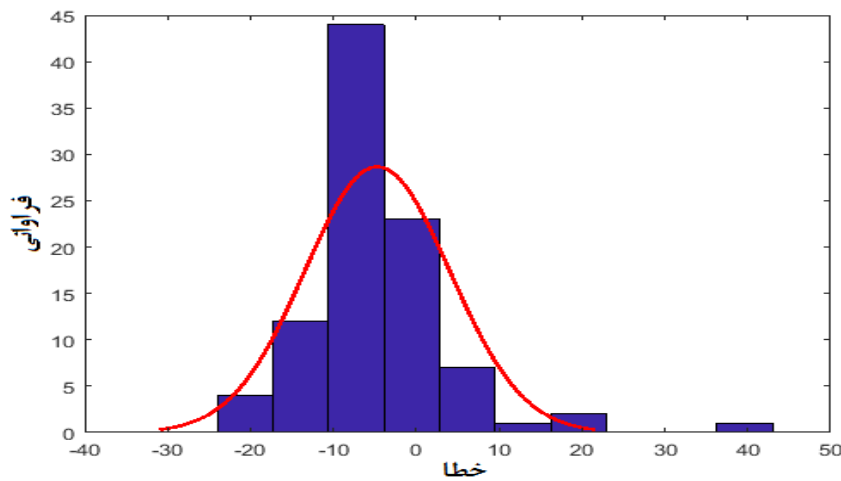


ب. مقادیر اندازه‌گیری شده و تخمینی شدت بارش (mm/h) توسط مدل رگرسیون خطی
شکل ۷. پراکنش مقادیر اندازه‌گیری شده و تخمینی شدت بارش

شکل (۸) که نشان دهنده نحوه توزیع خطا (اختلاف مقادیر بارش اندازه‌گیری شده و بارش تخمینی توسط مدل‌های برنامه‌ریزی بیان ژن و رگرسیون خطی) را نشان می‌دهد. طبق این شکل هر چه مدل‌ها در تخمین بارش دارای فراوانی بیش‌تری در محدوده خطای صفر یا نزدیک به صفر باشد دارای دقت بیش‌تری در تخمین شدت بارش خواهد داشت. با توجه به این مطلب مدل برنامه‌ریزی بیان ژن که بیش‌ترین فراوانی در محدوده خطای نزدیک به صفر دارد دقت بیش‌تری نسبت به مدل رگرسیونی دارد. Alinezhadi et al. (2021) دبی میانگین روزانه رودخانه کارون را با مدل‌های برنامه‌ریزی بیان ژن و رگرسیون‌های پارامتریک و ناپارامتریک تخمین زدند. نتایج نشان داد که مدل برنامه‌ریزی ژنی دارای دقت بیش‌تری نسبت به مدل‌های رگرسیون پارامتریک و ناپارامتریک می‌باشد.



الف. مدل برنامه‌ریزی بیان ژن



ب. مدل رگرسیون خطی

شکل ۸. نحوه توزیع خطا در مدل‌های برنامه‌ریزی بیان ژن و رگرسیون خطی

نتیجه‌گیری

بارندگی‌هایی که دامنه صوتی بیشتری داشتند، دارای اندازه فرکانسی بیشتری نیز بودند. بنابراین با کاهش شدت بارندگی اندازه فرکانس سیگنال صوتی کاهش و با افزایش شدت بارندگی اندازه فرکانس سیگنال صوتی افزایش یافت. بر اساس خوشه‌بندی دو مرحله‌ای، اندازه فرکانس‌های به دست آمده برای شدت‌های مختلف بارندگی به‌طور خودکار در دو خوشه قرار گرفته و میانگین و انحراف معیار هر خوشه مشخص شد. با توجه به همبستگی بین میانگین و انحراف معیار هر خوشه با یکدیگر و با شدت بارندگی و جلوگیری از پدیده هم‌خطی شدن، میانگین خوشه دوم به عنوان ورودی مدل‌های برنامه‌ریزی بیان ژن و رگرسیون خطی استفاده شد و توانست شدت بارندگی را در هر دو مدل برنامه‌ریزی بیان ژن و رگرسیونی با دقت مناسبی تخمین بزند. همچنین با توجه به معیارهای ارزیابی مشخص شد که مدل برنامه‌ریزی بیان ژن بهتر از مدل رگرسیون خطی توانست شدت بارندگی را تخمین بزند. بنابراین استفاده از این روش در تخمین شدت بارش توصیه می‌شود.

"هیچ‌گونه تعارض منافع بین نویسندگان وجود ندارد"

منابع

- احمدی، عباس؛ پالیزوان زند، پیمان و پالیزوان زند حبیب (۱۳۹۶). برآورد هدایت هیدرولیکی اشباع خاک با استفاده از برنامه‌ریزی بیان ژنی و رگرسیون ریب (مطالعه موردی در استان آذربایجان شرقی). *تحقیقات آب و خاک ایران*، (۵)، ۴۸، ۱۰۹۵-۱۰۸۷.
- علی‌نژادی، مهدی؛ موسوی، سید فرهاد و حسینی، خسرو (۱۴۰۰). مقایسه روش‌های برنامه‌ریزی بیان ژن و رگرسیون‌های پارامتریک و ناپارامتریک

در پیش‌بینی دبی میانگین روزانه رودخانه کارون (مطالعه موردی: ایستگاه هیدرومتری ملاثانی). نشریه علوم آب و خاک، ۲۵(۱)، ۶۲-۴۳.

REFERENCES

- Ahmadi, A., Palizvan Zand, P. & Palizvan Zand, H. (2017). Estimation of saturated hydraulic conductivity by using gene expression programming and ridge regression (A case study in East Azerbaijan province). *Iranian Journal of soil and water research*, 48(5), 1087-1095. (In Persian)
- Al-Amri, N.S. & Subyani, A.M. (2017). Generation of rainfall intensity duration frequency (IDF) curves for ungauged sites in arid region. *Earth Systems and Environment*, 1(1), 1-12.
- Alinezhadi, M., Mousavi, S.F. & Hosseini, Kh. (2021). Comparison of Gene Expression Programming (GEP) and Parametric and Non-parametric Regression Methods in the Prediction of the Mean Daily Discharge of Karun River (A case Study: Mollasani Hydrometric Station). *Journal of Water and Soil Science*, 25 (1), 43-62. (In Persian)
- Alvisi, S., Mascellani, G., Franchini, M. & Bardossy, A. (2005). Water level forecasting through fuzzy logic and artificial neural network approaches. *Hydrology and Earth System Sciences Discussions*, 2(3), 1107-1145.
- Beritelli, F., Capizzi, G., Sciuto, G.L., Napoli, C. & Scaglione, F. (2018). Rainfall estimation based on the intensity of the received signal in a LTE/4G mobile terminal by using a probabilistic neural network. *IEEE Access*, 6, 30865-30873.
- Chiu, T., Fang, D., Chen, J., Wang, Y., & Jeris, C. (2001). A Robust and Scalable Clustering Algorithm for Mixed Type Attributes in Large Database Environment. *In Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 263-268.
- Cristiano, E., Veldhuis, M., Wright, D.B., Smith, J.A. & van de Giesen, N. (2019). The Influence of Rainfall and Catchment Critical Scales on Urban Hydrological Response Sensitivity. *Water Resources Research*, 55(4), 3375-3390.
- Dang, T.A. (2020). Simulating Rainfall IDF Curve for Flood Warnings in the Ca Mau Coastal Area under the Impacts of Climate Change. *International Journal of Climate Change Strategies and Management*, 12, 705-715
- Ferreira, C. (2006). Gene expression programming: mathematical modeling by an artificial intelligence (Vol. 21). Springer.
- Hong, Y., Hsu, K.L., Sorooshian, S. & Gao, X. (2004). Precipitation estimation from remotely sensed imagery using an artificial neural network cloud classification system. *Journal of Applied Meteorology*, 43(12), 1834-1853
- Hu, Q., Li, Z., Wang, L., Huang, Y., Wang, Y. & Li, L. (2019). Rainfall Spatial Estimations: a review from spatial interpolation to multi-source data merging. *Water*, 11(3), 579.
- Hudson, N.W. (1981). Soil conservation. Batsford. London, England.
- Jaleel, L.A. & Farawn, M.A. (2013). Developing rainfall intensity-duration-frequency relationship for Basrah city. *Kufa Journal of Engineering*, 5(1), 105-112.
- Joyce, R., Janowiak, J., Arkin, Ph. & Xie, P. (2004). CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. *Journal of Hydrology*, 5(3), 487-503.
- Kizza, M., Westerberg, I., Rodhe, A. & Ntale, H.K. (2012). Estimating areal rainfall over Lake Victoria and its basin using ground-based and satellite data. *Journal of Hydrology*, 464, 401-411.
- Koza, J. (1992). Genetic Programming: on the Programming of Computers by Means of Natural Selection. MIT Press.
- Kyaw, A.K., Shahid, S. & Wang, X. (2022). Remote Sensing for Development of Rainfall Intensity-Duration-Frequency Curves at Ungauged Locations of Yangon, Myanmar. *Water*, 14(11), 1699.
- Liang, S., Li, X. & Wang, J. (2019). Advanced remote sensing: terrestrial information extraction and applications. Academic Press.
- Mattar, M.A., (2018). Using gene expression programming in monthly reference evapotranspiration modeling: a case study in Egypt. *Agricultural Water Management*, 198, 28-38.
- Mélèse, V., Blanchet, J., & Molinié, G. (2018). Uncertainty estimation of Intensity-Duration-Frequency relationships: a regional analysis. *Journal of Hydrology*, 558, 579-591.
- Nakazato, R., Funakoshi, H., Ishikawa, T., Kameda, Y., Matsuda, I. & Itoh, S. (2018). January. Rainfall intensity estimation from sound for generating CG of rainfall scenes. *In: Proceedings of 2018*



- International Workshop on Advanced Image Technology (IWAIT)*, 7-9 Jan., Institute of Electrical and Electronics Engineers Inc, Chiang Mai, Thailand, pp. 1-4.
- Neal, W. D., & Wurst, J. (2001). Advances in Market Segmentation. *Marketing research*, 13(1).
- Oppenheim, A.V., Willsky, A.S. & Hamid, S. (2006). Signals and Systems Second Edition. China, Publishing House of Electronics Industry.
- Prateek, G. (2017). Target detection using weather radars and electromagnetic vector sensors. *Signal Processing*, 137, 387-397.
- Rasel, M.M. & Islam, M.M. (2015). Generation of rainfall intensity-duration-frequency relationship for north-western region in Bangladesh. *IOSR Journal of Environmental Science, Toxicology and Food Technology*, 9 (9), 41-47.
- Şchiopu, D. (2010). Applying TwoStep cluster analysis for identifying bank customers' profile. *Buletinul*, 62(3), 66-75.
- Tramblay, Y., Thiemig, V., Dezetter, A. & Hanich, L. (2016). Evaluation of satellite-based rainfall products for hydrological modelling in Morocco. *Hydrological Sciences Journal*, 14(61), 2509-2519.
- Uijlenhoet, R. (2001). Raindrop size distributions and radar reflectivity–rain rate relationships for radar hydrology. *Hydrology and Earth System Sciences*, 5(4), 615–628.
- Weeks, M. (2010). Digital signal processing using MATLAB and wavelets. Publisher: Jones and Bartlett Learning.
- Wichmeier, W.H., & Smith, D.D. (1978). Predicting rainfall losses: a guide to conservation planning. *Agriculture Handbook* No. 537, US Department of Agriculture, Washington, DC.

Determining the intensity of rainfall using the analysis of sound frequencies resulting from the impact of raindrops

EXTENDED ABSTRACT

Introduction:

The application of sound data in many topics related to water and soil resources has not been used seriously yet. Especially in Iran, sound wave research in natural resources and environment sciences is considered as a new research. Therefore, it is necessary to conduct more and more diverse research in connection with the use of this method in various branches of comprehensive management of water and soil resources. Therefore, less time and money and more accurate and correct solutions can be obtained in related issues which increased the accuracy of predictions and modeling. In this research, a new and innovative method for estimating rainfall intensity based on audio data collection and audio frequency analysis is presented.

Materials and Methods:

In the laboratory of the Department of Soil Science, Faculty of Agriculture, University of Tabriz in 2022, 40 intensities of rainfall were created using designed rain simulators. The audio signals generated in different intensities of rainfall were recorded for 1 minute in 3 repetitions by REMAX model RP1 recorder in wav format and transferred to the computer for processing and the frequency size of audio files was extracted in MATLAB software. Then, the frequency measurements were automatically placed in two clusters in SPSS software using the two-stage clustering method. Then, the mean and standard deviation of each cluster were calculated and according to the correlation of each with each other and with rainfall intensity, and in order to avoid the phenomenon of multi-collinearity, only the mean of the second cluster was used as the input of the gene expression programming and linear regression models. To test the accuracy of the results obtained from the models, the coefficient of explanation (R^2), root mean square error (RMSE), geometric mean error ratio (GMER) and geometric standard deviation of error ratio (GSDER) statistics were determined.

Results Discussion:

Different intensities of rain were obtained using equation 7, which is the minimum rainfall intensity of 8 mm/h and the maximum rainfall intensity is 145 mm/h (Table 1). The greater the intensity of the rainfall, the greater the kinetic energy and, as a result, its erosive power. The sound amplitude of any rainfall intensity depends on the kinetic energy of that precipitation, as the intensity of the rainfall increases, the sound amplitude will also increase accordingly. According to equation (3), rains that have a larger sound amplitude also have a larger frequency size. Based on two-stage clustering, the obtained frequency sizes for different rainfall intensities were automatically placed into two clusters and the average and standard deviation of each cluster were determined. Considering the correlation between the mean and standard deviation of each cluster with each other and with the intensity of rainfall and avoiding the phenomenon of collinearity, the mean of the second cluster was used as an input for gene expression programming and linear regression models. The values of R^2 , RMSE (mm/h), GMER (mm/h) and GSDER (mm/h) for the gene expression programming model in the training series data were 0.97, 1.85, 1.11 and 1.09 respectively and for the validation series data were 0.96, 2.05, 1.14 and 1.12 respectively. While the values of the above criteria in the regression model were 0.94, 2.74, 1.25 and 1.34 respectively for the training series data and 0.92, 2.91, 1.28 and 1.37 respectively for the validation series data. The results of the above statistics indicate that the gene expression programming model is relatively more accurate than the regression and overestimation model, and the estimated data of the regression model is relatively more spread than the gene expression programming model.

Conclusion:

The kinetic energy of the rain is usually calculated according to the intensity of the rain, because the intensity of the rain is a function of the diameter of the raindrops, or actually a function of the mass of the raindrops and their final speed, and therefore it will be proportional to the kinetic energy of the rain. The greater the intensity of the rainfall, the greater the kinetic energy and, as a result, its erosive power. The sound amplitude of any rainfall intensity depends on the kinetic energy of that rainfall, as the intensity of the rainfall increases, the sound amplitude will increase accordingly, and as the intensity of the rainfall decreases, the sound amplitude will also decrease. Rainfalls that have a larger sound range also have a larger frequency range.

Keywords: Audio Signals, Clustering, Rain Simulator, Size Frequency.