

بهبود دقت داده‌های فناوری تیکه‌نگاری صوتی

با استفاده از روش آستانه‌گذاری فضای فازی

امیرحسین حسن‌آبادی^۱، ابراهیم جباری^۲، مسعود بحرینیموتلاهی^{۳*}، حسین علیزاده^۴ و یوسف الفت‌میری^۵
^۱ دانشجوی کارشناسی ارشد مهندسی عمران-آب و سازه‌های هیدرولیکی دانشگاه علم و صنعت ایران
^۲ عضو هیئت علمی دانشکده عمران دانشگاه علم و صنعت ایران
^۳ عضو هیئت علمی موسسه تحقیقات آب، تهران، ایران

چکیده

فناوری تیکه‌نگاری صوتی، یکی از شاخه‌های دانش سنجش‌از‌دور جهت پایش منابع آب سطحی است. داده‌های پرت در این فناوری تا به حال با استفاده از روش انحراف معیار تشخیص داده‌شده و حذف می‌شده‌اند. در این پژوهش، از روش آستانه‌گذاری فضای فازی به منظور تشخیص داده‌های پرت و از روش میانگین چهار نقطه در طرفین هر داده پرت جهت جایگزینی داده‌های پرت تشخیص داده‌شده استفاده شده است. داده‌های مورد استفاده قرار گرفته در این پژوهش شامل مجموعه داده ۱۲ روزه برداشت شده از رودخانه گونو واقع در شهر میوشی، استان هیروشیما ژاپن بوده است. مجموعاً ۸۰۱۷ داده، معادل با ۳۲٪ از ۲۵۰۳۱ داده اولیه، به عنوان نقاط پرت شناسایی و جایگزین شدند. همچنین مقدار انحراف معیار داده‌ها پیش از انجام فرایند پرت‌کاوی ۰/۲۰۶ و پس از روش آستانه‌گذاری فضای فازی به ۰/۱۱۹ رسید. نتایج نشان داد که روش آستانه‌گذاری فضای فازی نسبت به روش انحراف معیار از دقت و عملکرد بالاتری در شناسایی نقاط پرت برخوردار است؛ در نهایت مشاهده شد که مقایسه خطای نسبی اندازه‌گیری دبی بین روش آستانه‌گذاری فضای فازی و دبی-اشل (به عنوان مرجع) در اکثر نقاط کمتر از بیست درصد است. در حالی که این مقدار برای روش انحراف معیار و دبی-اشل به بیش از پنجاه درصد می‌رسد.

واژگان کلیدی: پرت‌کاوی، تیکه‌نگاری صوتی، داده‌کاوی، آستانه‌گذاری فضای فازی

Improving the accuracy of Acoustic Tomographic data using Phase-Space Thresholding method

AmirHosein Hasanabadi¹, Ebrahim Jabbari², Masoud Bahreinimotlagh^{3*}, Hossein Alizadeh⁴ and Yousef Olfatmiri⁵

^{1,5} MSc student, Department of Civil Engineering, Iran University of Science & Technology, Tehran, Iran

² Professor, Department of Civil Engineering, Iran University of Science & Technology, Tehran, Iran

³ Assistant Professor, Department of Water Resources and Studies, Water Research Institute, Tehran, Iran

⁴ Assistant Professor, Department of Civil Engineering, Iran University of Science & Technology, Tehran, Iran

Abstract

Acquiring reliable flow velocity and discharge estimates is critical in aquatic and water management studies. Acoustic Tomography is a Cutting-edge River monitoring technology that has been developed innovatively for continuous monitoring of surface water currents in oceans, seas, and, in recent years, rivers. It is a promising method to measure flow characteristics such as velocity and discharge precisely and continuously in time. Unlike classic point/transect flow velocity measurements, the acoustic tomography system can measure depth-averaged and range-averaged flow velocity along the ray path in a fraction of a second. The FAT system's backbone concept is the "time-of-travel" principle, which is analogous to acoustic velocity meters to some extent. The reciprocal sound transmission between two acoustic stations that are deployed diagonally on both sides of the river occurs in this system, and the cross-sectional averaged velocity is measured using the travel time of the transmitted signals between the two stations.

* Corresponding author

* نویسنده عهده‌دار مکاتبات



The FAT system velocity data, similar to all acoustic instruments measurements, contains outliers that were previously cleaned using the simple standard deviation (STD) method. The problem of this method is that it not only fails to replace identified spikes with the appropriate values, resulting in signal discontinuity, but it also employs engineering judgement in its process, increasing the method's error.

The Phase-Space Thresholding (PST) method, which was developed for despiking output of Acoustic Doppler Velocimeter (ADV) data, is proposed in this study to implement on the FAT system output and to detect its spikes. This method combines three concepts: 1) Differentiation enhances a signal's high-frequency portion, 2) the universal threshold parameter gives the expected maximum of a normal, random series, and 3) good data cluster in a dense cloud in phase space. These concepts are employed to construct an ellipsoid in three-dimensional phase space, and points that are outside the threshold ellipsoid are designated as spikes. An important advantage of this method in comparison with various other methods is that it requires no parameters tuning. Another advantage of this method over the standard deviation method is that identified spikes are replaced with a reliable result. We used the mean value of two adjacent data points on either side of the identified spike to replace the values of the detected spike.

The data used in this research included a 12-day dataset collected from the Gono River in Miyoshi City, Hiroshima Prefecture, Japan. After 6 iterations of applying the PST approach to the input dataset, a total of 8017 data were detected as spikes and replaced with the proper value, accounting for 32% of the 25031 data. Furthermore, before despiking, the standard deviation was 0/206, but improved to 0/119 after employing the PST technique. The decrease in standard deviation implies that as a result of despiking process, the data dispersion around the signal mean has decreased. According to the results the PST method is more accurate in despiking FAT system data compared to the standard deviation method.

Finally, it was discovered that the relative discharge error between the output of the PST method and the rating curve data (as a reference) is less than 20%, but this value surpasses 50% for the standard deviation and rating curve method data.

Key Words: Despiking, Acoustic Tomography, Phase-Space Thresholding

ارزیابی‌های زیست‌محیطی، اجتماعی و اقتصادی است. از این رو، کشورهای توسعه‌یافته در حال ارتقای فناوری‌های نوین به منظور پایش پیوسته و به‌هنگام منابع آب سطحی هستند [1].

پیش‌بینی فرایندهای هیدرولوژیکی و رفتار رودخانه‌ها به‌صورت دقیق و مناسب، می‌تواند کمک مؤثر و شایانی در حوزه طراحی بهینه پروژه‌های آبی و سازه‌های هیدرولیکی و همچنین جلوگیری از خطرات احتمالی ناشی از آن‌ها و افزایش آمادگی در مقابل این خطرات احتمالی به‌منظور کاهش خسارات و تلفات و تسهیل مدیریت چنین پدیده‌هایی را داشته باشد. در این راستا، ایجاد یک فناوری یا سامانه به‌منظور فراهم‌آوردن داده‌های به‌هنگام و پیوسته با قابلیت اطمینان و دقت بالا، به‌منظور استفاده برای پیش‌بینی وقایع پیش‌رو، از پراهمیت‌ترین و چالش‌برانگیزترین مسائل پایش روی متخصصین حوزه مدیریت منابع آب است [2].

پیش از انجام فرایندهای داده‌کاوی، ضرورتاً نیاز به پیش‌پردازش داده‌های خام و آماده‌سازی آن‌ها احساس می‌شود؛ در مرحله پیش‌پردازش داده سلسله‌عملیاتی صورت می‌پذیرد که باعث برطرف‌شدن مشکلات مختلف داده مسئله مورد بررسی خواهد شد؛ به‌این‌ترتیب داده برای انجام فرایند یادگیری مدل، آماده می‌شود. در

۱- مقدمه

رشد فناوری اطلاعات و روش‌ها و ابزارهای نوین تولید و جمع‌آوری داده‌ها در سالیان اخیر سبب شده که پژوهش‌گران به حجم انبوهی از داده‌ها در زمینه‌های گوناگون به سرعت و سهولت دسترسی پیدا کنند. موضوع مهم در این زمینه توانایی هرچه بیشتر در برداشت اطلاعات مفید و دانش نهفته کاربردی از این داده‌ها به‌منظور تجزیه و تحلیل آن‌ها و کسب دانش سودمند در مورد پدیده‌های گوناگون دنیای اطراف و همچنین بررسی امکان پیش‌بینی آینده از اطلاعات به‌دست‌آمده است. عوامل گوناگونی سبب می‌شوند که در هنگام برداشت داده‌ها، تعدادی از داده‌های ثبت‌شده وضعیت واقعی سامانه را نمایش ندهد و باعث اشتباه در اندازه‌گیری‌ها و درنهایت نتیجه‌گیری‌های اشتباه از آن‌ها شوند؛ لذا ایجاد فرایندهایی به‌منظور پالایش داده‌های خام و بررسی اعتبار آن‌ها ضرورتی غیرقابل‌انکار است؛ البته تعریف مرزی دقیق بین داده‌های معتبر و پرت، بحثی دشوار و مبهم است که دیدگاه محقق به داده‌ها و ماهیت آن‌ها و تسلط بر فرایند برداشت آن‌ها می‌تواند تا حدی ابهام موجود در این زمینه را کاهش دهد.

دستیابی به داده‌های پیوسته دبی رودخانه‌ها، از مهم‌ترین چالش‌های مدیریت منابع آب به‌منظور

تقسیم بر مجموع سرعت صوت و سرعت جریان آب در آن مقطع، و زمان رسیدن امواج صوت در ایستگاه بالادست (خلاف جهت جریان آب) عبارت است از فاصله دو ایستگاه تقسیم بر تفاضل سرعت صوت و سرعت جریان آب [8]:

$$t_{up} = \frac{L}{c_m - u_m} \quad (1)$$

$$t_{down} = \frac{L}{c_m + u_m} \quad (2)$$

در این روابط t_{up} و t_{down} به ترتیب زمان رسیدن علائم به ایستگاه‌های بالادست و پایین‌دست، L فاصله افقی بین دو ایستگاه صوتی، C_m سرعت صدا در آب، و u_m سرعت متوسط جریان در راستای امواج صوتی است. با حل معادلات بالا سرعت متوسط جریان و سرعت صدا محاسبه می‌شود [9]:

$$C_m = \frac{L}{t_m} \quad (3)$$

$$u_m = \frac{C_m^2}{2L} \Delta t \quad (4)$$

در روابط بالا $t_m = \frac{t_{up} + t_{Down}}{2}$ و $\Delta t = t_{up} - t_{Down}$ به ترتیب میانگین زمان رسیدن علائم صدا به ایستگاه‌ها و اختلاف زمانی رسیدن علائم صدا به ایستگاه‌ها است. با اندازه‌گیری سرعت متوسط در راستای امواج صوتی (U_m)، سرعت متوسط جریان رودخانه (V_m) به دست می‌آید:

$$V_m = \frac{u_m}{\cos\theta} \quad (5)$$

در نهایت با استفاده از رابطه پیوستگی، دبی جریان برابر است با، حاصل ضرب سرعت متوسط جریان در مساحت مقطع عرضی رودخانه:

$$Q = A(H)V_m \sin(\theta) \quad (6)$$

A مساحت مقطع عرضی بوده و تابعی از ارتفاع آب (H) است و θ زاویه بین راستای دو فرستنده و جهت جریان رودخانه است.

سرعت جریان اندازه‌گیری شده با روش تیکه‌نگاری صوتی همانند سایر دستگاه‌های صوتی، دارای داده‌های پرت است؛ که عدم توجه به موضوع حذف و جایگزینی این داده‌ها سبب کاهش دقت اندازه‌گیری‌های جریان می‌شوند.

عملیات پاک‌سازی داده، مشکل کیفیت و اعتبار داده‌ها برطرف می‌شود. از مشکلاتی که کیفیت داده‌ها را تحت تأثیر قرار می‌دهند، داده‌های پرت^۱ هستند. حضور داده‌های پرت سبب می‌شود که الگوریتم‌های یادگیری در داده‌کاوی نتوانند مدل دقیقی برای تبیین دانش نهفته در داده بیابند؛ به عبارت دیگر تجزیه و تحلیل سری‌های زمانی با حضور نقاط دورافتاده، هرگونه استنباط صحیح آماری را خدشه‌دار کرده و پیش‌بینی‌های به دست آمده را متأثر از خطا می‌سازد؛ لذا به دنبال روش‌هایی هستیم که تأثیر این نقاط ناخواسته را از بین برده و یا حداقل کاهش دهد [3]. یکی از فناوری‌های پایش پیوسته و خودکار منابع آب سطحی از جمله رودخانه‌ها [4] و دریاها [5]، فناوری تیکه‌نگاری صوتی است. دستگاه‌های تیکه‌نگاری صوتی با اتصال به ماهواره‌های سامانه موقعیت‌یاب جهانی^۲، با دقت نانوثانیه هم‌زمان‌سازی می‌شوند و در فواصل زمانی منظم و در یک‌زمان واحد، امواج صدا را به سمت یکدیگر ارسال می‌کنند. با تجزیه و تحلیل زمان رسیدن امواج صدا به دستگاه‌ها، خصوصیات جریان آب محاسبه می‌شوند [6]. این سامانه از دو پردازنده که هرکدام از آن‌ها توسط یک کابل به یک تراگذار^۳ که در درون آب قرار گرفته و امواج صوتی را ارسال و دریافت می‌کند، تشکیل شده است. با معلوم بودن فاصله دو ایستگاه از یکدیگر و ثبت زمان رسیدن موج‌های صدا در ایستگاه‌های بالادست و پایین‌دست با دقت نانوثانیه، سرعت صدا، دمای متوسط جریان و سرعت متوسط جریان آب با دقت بالا اندازه‌گیری می‌شود. شکل (۱) الگوی سامانه تیکه‌نگاری صوتی رودخانه‌ای به منظور پایش آن را نشان می‌دهد.



(شکل ۱- الگوی مفهومی سامانه تیکه‌نگاری صوتی [7])
(figure-1) Conceptual Pattern of Fluvial Acoustic Tomography System [7]

زمان رسیدن امواج صوتی در ایستگاه پایین‌دست (در جهت جریان آب) عبارت است از فاصله بین دو ایستگاه

¹ outlier

² GPS

³ Transducer

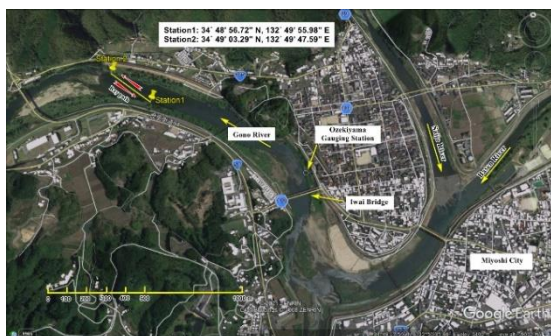
ایستگاه‌های اندازه‌گیری بالادست^۴ و پایین‌دست^۵ مشاهده می‌شود. زاویه بین راستای دو تراگذار ارسال‌کننده امواج صدا و جهت جریان اصلی رودخانه ۲۲ درجه اندازه‌گیری شده و در محاسبات، به منظور تبدیل سرعت جریان در راستای دو تراگذار به سرعت در راستای محور اصلی جریان لحاظ شده است.

در شکل (۳)، علامت سرعت جریان رودخانه در راستای ارسال امواج بین دو تراگذار قابل مشاهده است که پس از انجام پردازش‌های اولیه بر روی داده‌های خام خروجی این سامانه و حذف داده‌هایی که نسبت علامت به نوفه^۶ آن‌ها پایین‌تر از ۱۰ دسی‌بل هست، به دست آمده است. همان‌طور که مشاهده می‌شود، داده‌های خام خروجی برداشت‌شده از سامانه تیکه‌نگاری دارای نقاط پرت پرشماری بوده که عدم توجه به حذف و جایگزینی آن‌ها با مقادیر مناسب، باعث ایجاد خطا در محاسبات سرعت و دبی خواهد شد.

۲-۲- روش پژوهش

روش آستانه‌گذاری فضای فازی^۷ در سال ۲۰۰۲ توسط گورینگ و نیکورا به منظور پرت‌کاوی سری زمانی خروجی سرعت‌سنج صوتی داپلر^۸ توسعه پیدا کرد [10]. این روش سه مفهوم کلی را ترکیب می‌کند:

- ۱- مشتق‌گیری از یک علامت سبب تقویت بخش بسامد بالای یک علامت صدا می‌شود [11].
- ۲- مقدار بیشینه مطلق مورد انتظار یک سری زمانی تصادفی به کمک پارامتر آستانه جهانی^۹ قابل تعیین است [12].
- ۳- نقاط داده معتبر در فضای فازی در یک ابر متراکم تشکیل خوشه می‌دهند [13].



(شکل ۲-۲) سایت برداشت داده‌های تیکه‌نگاری، رودخانه گونو (figure-2) FATS data Collection site, Gono River, Japan

- ⁴ Station 1
- ⁵ Station 2
- ⁶ Signal to Noise Ratio (SNR)
- ⁷ Phase-Space Thresholding (PST)
- ⁸ Acoustic Doppler Velocimeter (ADV)
- ⁹ Universal Threshold

روشی که تا به حال به منظور پرت‌کاوی داده‌های این سامانه استفاده می‌شده است، روش ساده استفاده از پارامتر انحراف معیار بوده است؛ به این صورت که داده‌هایی که با فاصله مشخصی از روند اصلی علامت صدا قرار دارند و نوسانات آن‌ها حول روند از ضریب مشخصی از انحراف معیار کوچک‌تر است، به عنوان داده‌های نرمال در نظر گرفته می‌شوند و سایر نقاط خارج از این ناحیه پرت در نظر گرفته می‌شوند؛ و پس از حذف داده‌های پرت با استفاده از این روش، با میانگین‌گیری متحرک^۱ از سری زمانی، مسطح‌سازی آن صورت می‌پذیرفته است. در این روش برخلاف روشی که در این پژوهش معرفی و استفاده شده است، داده‌های پرت پس از تشخیص فقط حذف شده و جایگزین نمی‌شوند؛ علاوه بر این از دیگر مشکلات این روش می‌توان به عدم استفاده از روشی اشاره کرد که بتواند بر روی داده‌های گوناگون بدون دخالت و بدون نیاز به پارامترهای مورد نیاز بر اساس قضاوت انسانی اعمال شود، بلکه نیاز به مشخص کردن یک ضریب خاص از انحراف معیار داده‌ها توسط کاربری که روش را اعمال می‌کند وجود دارد.

هدف این پژوهش بررسی یکی از روش‌های توسعه یافته به منظور پرت‌کاوی خروجی سرعت‌سنج صوتی داپلر است که به علت هم‌جنس و هم‌نوع بودن این داده‌ها با خروجی سامانه تیکه‌نگاری صوتی، عملکرد این روش روی این داده‌ها مورد بررسی قرار گرفته است؛ علاوه بر این در این پژوهش از روش میانگین‌گیری دونقطه در هر طرف داده‌های پرت شناسایی شده به عنوان مقدار جایگزین پس از شناسایی و حذف نقاط پرت استفاده شده است.

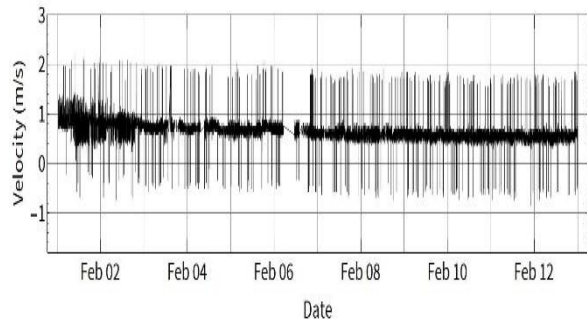
۲- مواد و روش‌ها:

۲-۱- منطقه مورد مطالعه و داده‌های

مورد استفاده:

داده‌های مورد استفاده در این پژوهش شامل علامت خروجی سامانه تیکه‌نگاری صوتی از نوع سرعت جریان رودخانه است که از رودخانه گونو^۲ واقع در شهر میوشی استان هیروشیما ژاپن^۳ برداشت شده است. این داده‌ها با طول زمانی ۱۲ روز و در بازه زمانی ۱ فوریه ۲۰۱۶ تا ۱۲ فوریه ۲۰۱۶ و هر ۳۰ ثانیه یک‌بار برداشت شده است. در شکل ۲ محدوده برداشت داده‌ها و محل قرارگیری

- ¹ Moving Average
- ² Gono River
- ³ Miyoshi City, Hiroshima, Japan



(شکل-۳) علامت سرعت جریان در راستای امواج ارسالی سامانه تیکه‌نگاری رودخانه‌ای
(figure-3) Flow Velocity Signal in the direction of transmitted FATS Waves

۱- محاسبه مشتقات اول و دوم علامت سرعت ورودی:

$$\Delta u_i = (u_{i+1} - u_{i-1}) / 2 \quad (7)$$

$$\Delta^2 u_i = (\Delta u_{i+1} - \Delta u_{i-1}) / 2 \quad (8)$$

۲- محاسبه انحراف معیار استاندارد و آستانه جهانی:

برای هر یک از متغیرهای u ، Δu ، $\Delta^2 u$ انحراف معیار را محاسبه می‌شود و محاسبه مقدار بیشینه مورد انتظار که برای یک متغیر تصادفی و نرمال که انحراف معیار استاندارد آن با σ تخمین زده شده و میانگین آن صفر است به کمک پارامتر آستانه جهانی:

$$\lambda_U = \sqrt{2 \ln(n)} \quad (9)$$

که n تعداد کل نقاط داده‌های سری زمانی است.

۳- محاسبه زاویه چرخش محور اصلی بیضوی آستانه در نمودار $u-\Delta^2 u$:

که به کمک رابطه زیر انجام می‌شود:

$$\theta = \tan^{-1} \left(\frac{\sum u_i \Delta^2 u_i}{\sum u_i^2} \right) \quad (10)$$

۴- محاسبه مقادیر اندازه محورهای بزرگ و کوچک بیضوی‌های آستانه:

این مقادیر به صورت زیر محاسبه می‌شوند:

برای نمودار $u-\Delta u$ ، اندازه محور بزرگ و کوچک بیضوی به ترتیب به صورت: $\lambda_U \sigma_{\Delta u}$ و $\lambda_U \sigma_U$

برای نمودار $u-\Delta^2 u$ ، اندازه محور بزرگ و کوچک بیضوی به ترتیب به صورت: $\lambda_U \sigma_{\Delta^2 u}$ و $\lambda_U \sigma_{\Delta u}$

و برای نمودار $u-\Delta^2 u$ ، اندازه محور بزرگ و کوچک بیضوی به کمک حل دستگاه دو معادله دو مجهول زیر حاصل می‌شود که a محور بزرگ و b محور کوچک است:

$$(\lambda_U \sigma_u)^2 = a^2 \cos^2(\theta) + b^2 \sin^2(\theta) \quad (11)$$

$$(\lambda_U \sigma_{\Delta^2 u})^2 = a^2 \sin^2(\theta) + b^2 \cos^2(\theta) \quad (12)$$

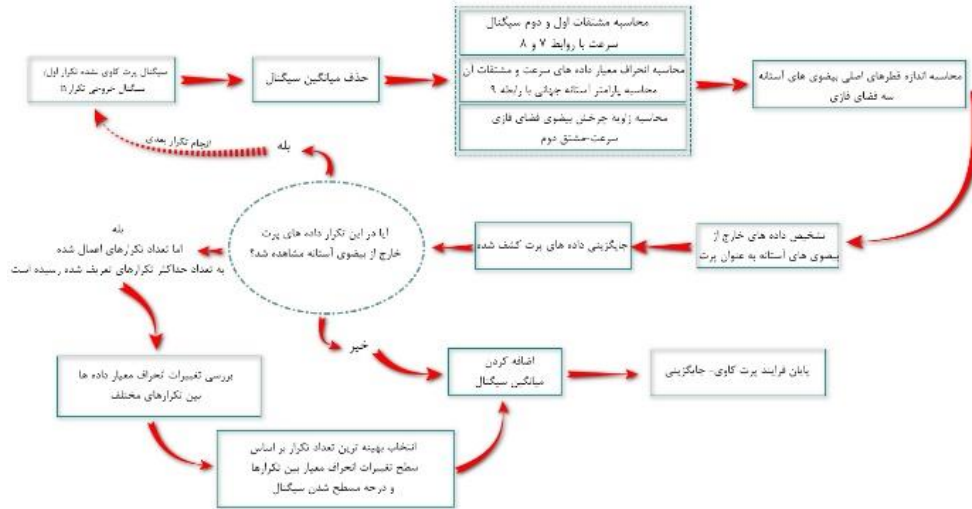
از این مفاهیم به منظور ساخت یک بیضوی در یک فضای فازی سه‌بعدی استفاده گردیده و سپس نقاط داده‌ای که خارج از این فضای بیضوی قرار می‌گیرند، به عنوان نقاط پرت شناسایی می‌شوند. برتری این روش نسبت به سایر روش‌های پیش از آن در این نکته نهفته است که در فرایند پرت‌کاوی از هیچ پارامتری که نیاز باشد توسط محقق و به صورت تجربی و تقریبی تعیین شود استفاده نمی‌کند. همچنین این روش یک روش تکرارشونده بوده و در هر گام پس از شناسایی و حذف داده‌های پرت بایستی آن‌ها را با مقادیر مناسبی جایگزین نمود؛ به این منظور در این پژوهش از روش میانگین مقادیر دونقطه داده در هر طرف نقطه پرت کشف شده، به عنوان مقدار جایگزین استفاده شده است.

نحوه انجام این فن به این صورت است که در ابتدا مقدار میانگین علامت را از داده‌ها کسر کرده و با مقدار نوسانات داده‌ها حول مقدار میانگین، کار پرت‌کاوی را انجام داده و مجدداً مقدار میانگین علامت صدا را به داده‌ها اضافه کرده و سپس مقدار داده‌هایی که به عنوان داده‌های پرت شناسایی شده‌اند را با مقادیر مناسبی جایگزین می‌شود. پس از حذف میانگین، مقادیر نوسانات سرعت در مقابل مشتق‌های اول و دوم آن به صورت دوبه‌دو ترسیم گردیده که باعث ایجاد سه نمودار دوبعدی $u-\Delta u$ ، $u-\Delta^2 u$ و $\Delta u-\Delta^2 u$ شده و سپس با محاسبه اندازه محورهای بزرگ و کوچک بیضوی‌های آستانه به کمک پارامترهای موجود اقدام به شناسایی نقاط خارج از این محدوده آستانه کرده و نقاط داده شناسایی شده را با داده‌هایی با مقادیر مناسب جایگزین می‌شود. این مراحل آن قدر تکرار می‌شود که تعداد نقاط معتبر ثابت گشته و یا به عبارت دیگر نقاط پرت جدیدی کشف نشوند؛ بنابراین می‌توان این روش را به این صورت گام بندی نمود:

۵- شناسایی و جایگزینی نقاط پرت:

برای هر یک از سه نمودار بیضوی‌های آستانه را ترسیم کرده، نقاط داده خارج از آن‌ها را به‌عنوان داده‌ی پرت شناسایی کرده و با مقادیر مناسب جایگزین می‌کنیم و به سراغ تکرار بعدی روش می‌رویم؛ در هر تکرار این روش، جایگزینی داده‌های پرت با مقادیر مناسب سبب کاهش مقدار انحراف معیار استاندارد شده و بنابراین اندازه بیضوی

آستانه در هر مرحله کاهش می‌یابد تا جایی که جایگزینی نقاط پرت دیگر تأثیری در کشف نقاط پرت جدید نداشته باشد که در این مرحله تکرارها به پایان می‌رسند. در شکل (۴) روندنمای نحوه انجام روش پرت‌کاوی آستانه‌گذاری فضای فازی قابل مشاهده می‌باشد.



(شکل-۴) روندنمای نحوه اعمال روش آستانه‌گذاری فضای فازی

(Figure-4) The PST Method Flowchart

در مجموعه داده‌های^۲ استفاده‌شده در این پژوهش مجموعاً کل نقاط داده برداشت‌شده در بازه ۱۲ روزه مذکور، ۲۷۰۷۵ داده بوده است که از میان آن‌ها تعداد ۲۵۰۳۱ داده (۹۲/۴۵٪ کل داده‌ها)، نقطه‌ای با نسبت علامت به نوفه بالای ۱۰ دسی‌بل بوده‌اند و الگوریتم پرت-کاوی روی این داده‌ها اعمال گردیده است.

در مرحله بعد و پس از حذف میانگین از علامت صدا ورودی، روش آستانه‌گذاری فضای فازی تا ۱۵ تکرار و روی نوسانات داده‌ها حول میانگین اعمال شد و در انتهای هر تکرار، مجدداً مقدار میانگین خروجی تکرار قبل به علامت پرت‌کاوی‌شده در آن مرحله اضافه شد. با توجه به شرط اتمام تکرارها که عدم شناسایی نقاط جدید در تکرار آخر به‌عنوان نقطه پرت جدید بود و با توجه به اینکه این مسئله در چندین تکرار انجام‌شده، حاصل نشد به‌عنوان یک معیار، از مقدار انحراف معیار داده‌ها در تشخیص اتمام تکرارها استفاده گردید به این صورت که از تکرار ششم به بعد تغییرات انحراف معیار داده‌ها بین دو تکرار، بسیار ناچیز و کمتر از یک‌هزارم محاسبه شد و آخرین مرحله این روش همان تکرار ششم لحاظ شد.

² Data Set

در پژوهشی که در سال ۲۰۱۶ توسط کاوانیسی و همکاران [14] از روش انحراف معیار برای پرت‌کاوی و بر روی گروه داده استفاده‌شده در این پژوهش استفاده‌شده است، به علت وجود درصد بالای نقاط پرت در این گروه داده، روش انحراف معیار عملکرد نامطلوبی از خود نشان داده بود به‌گونه‌ای که پژوهشگران این پژوهش به کمک قضاوت مهندسی خود یک محدوده سرعت معتبر برای داده‌ها در بازه نیم تا یک و نیم متر بر ثانیه برای داده‌ها در نظر گرفتند و داده‌های خارج از این محدوده را به‌عنوان نقاط پرت حذف کردند. نتایج خروجی روش استفاده‌شده در این پژوهش با خروجی روش قبل مقایسه می‌شود.

۳- نتایج و بحث

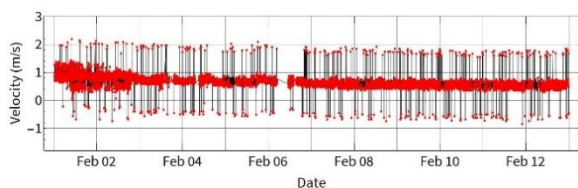
در مرحله اولیه پردازش داده‌های سامانه تیکه‌نگاری صوتی که داده‌های زمان رسیدن علائم به تراگذارهای ایستگاه‌های بالادست و پایین‌دست به داده‌های سرعت جریان تبدیل می‌شوند، با حذف داده‌هایی که نسبت علامت به نوفه^۱ آن‌ها کمتر از ۱۰ دسی‌بل است، داده‌هایی را که اعتبار آن‌ها کافی نبوده است، حذف می‌شوند.

¹ SNR

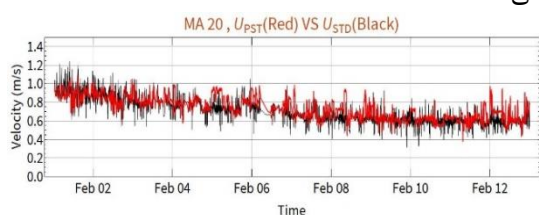
بوده که ۳۲/۰۲ درصد از کل ۲۵۰۳۱ داده را شامل می‌شوند.

(شکل-۷) نقاط شناسایی شده در شش تکرار روش فضای فازی به‌عنوان داده پرت (قرمز) و جایگاه آن‌ها در علامت اولیه (مشکی)

(Figure-7) Detected Data points as Spikes in PST method (Red Points) & their position in input signal (Black)



شکل (۸) مقایسه بین روش انحراف معیار و روش فضای فازی پس از پرت‌کاوی سرعت جریان ورودی اولیه و مسطح‌سازی به کمک میانگین متحرک بیست را نشان می‌دهد.



(شکل-۸) مقایسه سرعت جریان پس از پرت‌کاوی به روش فضای فازی (قرمز) و انحراف معیار (مشکی) و مسطح‌سازی با میانگین متحرک ۲۰

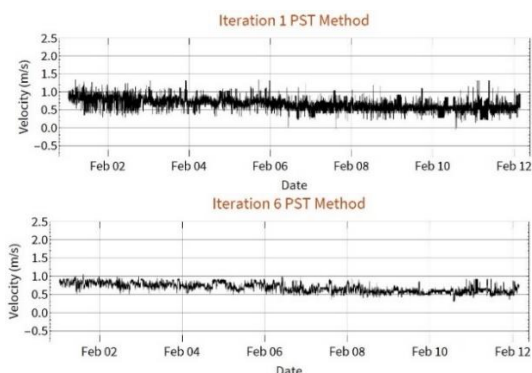
(Figure-8) Comparison of flow velocity after despiking with PST (Red) & STD (Black) & flattening with moving average 20

همان‌طور که مشاهده می‌شود هر دو روش به‌خوبی توانسته‌اند نقاط پرت آشکار را حذف نموده و علامت را مسطح نمایند اگرچه نوسانات سرعت در خروجی روش انحراف معیار بیشتر است. لازم به ذکر است که استفاده از میانگین متحرک بیست باعث تغییر تفکیک‌پذیری زمانی سری زمانی از سی ثانیه به ده دقیقه می‌شود ولی باعث کاهش هزینه محاسباتی الگوریتم‌های داده‌کاوی در حالتی که داده ورودی طول زیادی داشته و تعداد داده‌ها بسیار بالا است، خواهد شد.

مقدار اختلاف نسبی سرعت^۱ بین این دو روش پس از مسطح‌سازی به کمک میانگین متحرک بیست نیز در شکل (۹) نشان داده شده‌است. همان‌طور که مشخص است مقدار اختلاف نسبی سرعت برای اکثریت نقاط در بازه منفی تا مثبت تسی درصد است و در برخی نقاط خاص این اختلاف به بالای سی درصد نیز می‌رسد که نشان از اهمیت انتخاب روش پرت‌کاوی در نتایج نهایی و همچنین انطباق مناسب روش فضای فازی اعمال‌شده با روشی که بر اساس قضاوت مهندسی اعمال‌شده دارد.

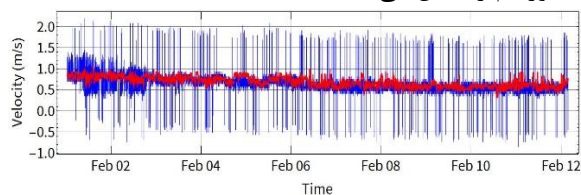
¹ Relative Velocity Difference

علت حاصل‌نشدن شرط اتمام تکرارها، برخلاف اینکه انحراف معیار به‌طور تقریبی ثابت بوده است و در نتیجه اندازه محورهای بزرگ و کوچک بیضوی‌های آستانه ثابت بوده است، این است که این روش پرت‌کاوی در گروه داده‌های با درصد بالای نقاط پرت، دارای محدودیت بوده و عملکرد مناسبی ندارد؛ همچنین ممکن است به محدودیت روش جایگزینی مورد استفاده نیز مرتبط باشد که به‌جای جایگزینی نقاطی با مقدار مناسب، مقداری را جایگزین کنند که خود همان مقدار در تکرار بعد به‌عنوان داده‌ی پرت شناخته می‌شود. در شکل (۵) نتیجه خروجی روش فضای فازی برای تکرار نخست و ششم، قابل مشاهده است:



(شکل-۵) نتایج تکرارهای اول و ششم روش فضای فازی (Figure-5) 1st & 6th Iteration Results of PST Method

همان‌طور که مشاهده می‌شود، این روش به‌خوبی توانسته نقاط پرت آشکار را که با مشاهده علامت صدا هم به‌سادگی قابل‌ردیابی هستند به‌خوبی حذف و جایگزین کند. در شکل (۶)، خروجی روش فضای فازی پس از شش تکرار با علامت صدا ورودی اولیه به این الگوریتم مقایسه شده است که دوباره عملکرد مناسب روش آستانه‌گذاری فضای فازی در پرت‌کاوی علامت صدا ورودی به این الگوریتم را نشان می‌دهد.

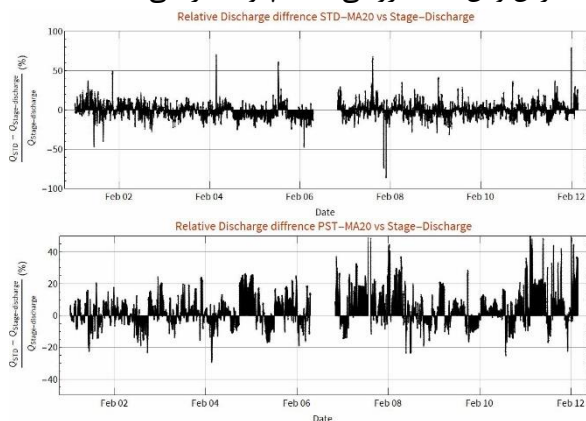


(شکل-۶) مقایسه خروجی روش آستانه‌گذاری فضای فازی با علامت صدا اولیه ورودی (Figure-6) Comparison between Input (Blue) & Output (Red) Signal of PST method

در شکل (۷) مجموعه داده‌هایی که در شش تکرار انجام‌شده به‌عنوان نقاط پرت شناسایی و جایگزین شده‌اند نشان داده شده‌است. تعداد کل این داده‌ها ۸۰۱۷ عدد

نسبی دبی در روش آستانه‌گذاری فازی و همچنین روش انحراف معیار با روش دبی اشل، بین منفی بیست تا مثبت سی درصد در اکثر نقاط متغیر است و در نتیجه می‌توان انطباق قابل قبول با نتایج روش دبی-اشل و اعتبار مناسب نتایج خروجی سامانه تیکه‌نگاری صوتی پس از انجام پرت‌کاوی را نتیجه گرفت. برای روش انحراف معیار، در تعداد قابل توجهی از نقاط، اختلاف بالای پنجاه درصد در نتایج مشاهده می‌شود؛ درحالی‌که در اکثریت قریب به‌اتفاق نتایج خروجی الگوریتم آستانه‌گذاری فازی روش این اختلاف در محدوده منفی تا مثبت بیست درصد قرار دارد که عملکرد مناسب‌تر روش آستانه‌گذاری فازی نسبت به روش انحراف معیار را نشان داده و همچنین حاکی از این حقیقت است که اگرچه روش آستانه‌گذاری فازی برای گروه داده‌های با درصد بالای حضور نقاط پرت در آن‌ها (مشابه گروه داده‌ای که در این پژوهش از آن استفاده شد)، عملکرد بهینه و کاملاً دقیقی ندارد اما توانسته از روشی که در گذشته بر اساس قضاوت مهندسی به منظور تعیین محدوده نرمال داده‌ها، اعمال می‌شده است، عملکرد بهتری داشته باشد.

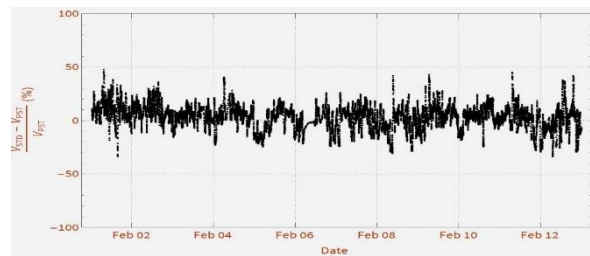
روش آستانه‌گذاری فازی برخلاف روش انحراف معیار به بحث جایگزینی نقاط پرت کشف‌شده نیز می‌پردازد که از برتری‌های آن نسبت به روش انحراف معیار است که سبب جلوگیری از کاهش طول سری زمانی و در نتیجه در دسترس بودن یک خروجی کاملاً پیوسته زمانی شده است.



(شکل-۱۱) اختلاف دبی نسبی روش‌های انحراف معیار و دبی-اشل (بالا) و روش‌های آستانه‌گذاری فازی و دبی-اشل (پایین)
(figure-11) Relative Discharge Difference between STD & Stage-Discharge Methods (up) and PST & Stage-Discharge Methods (down)

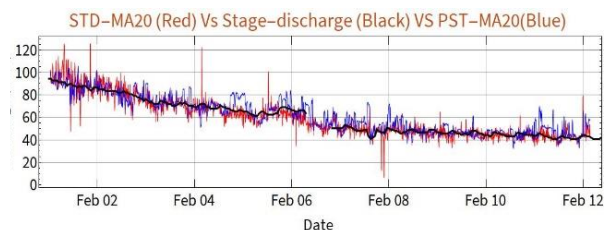
۴- نتیجه‌گیری و پیشنهادها:

خروجی سامانه تیکه‌نگاری صوتی رودخانه‌ای به عنوان یک ابزار مدرن و پیشرفته در مهندسی آب مانند سایر ابزارهای صوتی، تحت تأثیر عوامل ناشناخته‌ای، در برخی از نقاط



(شکل-۹) اختلاف نسبی سرعت بین دو روش فضای فازی و انحراف معیار
(Figure-9) Relative Velocity Difference between PST & STD Methods

به‌منظور صحت‌سنجی روش پیشنهادی این پژوهش، علامت سرعت خروجی سامانه تیکه‌نگاری پس از پرت‌کاوی با روش آستانه‌گذاری فازی و همچنین با دسترس بودن اطلاعات مربوط به مقطع عرضی مایل رودخانه بین دو تراگذار، به داده‌های دبی تبدیل شدند، و سپس با داده‌های خروجی روش دبی اشل^۱ مقایسه گردیدند. در این پژوهش به این منظور، از داده‌های دبی به‌دست‌آمده از روش دبی اشل که از یک ایستگاه هیدرومتری^۲ (که موقعیت آن در شکل ۲ نشان داده شده است) در بالادست محل داده‌برداری قرار داشت و در همین بازه زمانی برداشت‌شده بود، استفاده شد. در شکل ۱۰، خروجی دبی به‌دست‌آمده از سه روش انحراف معیار استاندارد، آستانه‌گذاری فازی و دبی اشل نشان داده شده است.



(شکل-۱۰) مقایسه دبی خروجی روش‌های انحراف -

معیار (قرمز)، آستانه‌گذاری فازی (آبی) و دبی-اشل (مشکی)
(Figure-10) Comparison of Output Discharge of STD (Red), PST (Blue) & Stage-Discharge Methods
 این شکل انطباق بسیار مناسب خروجی روش تیکه-نگاری صوتی با روش دبی اشل را نشان داده و همچنین عملکرد مناسب‌تر دبی به‌دست‌آمده از علامت سرعت پرت‌کاوی شده به روش فضای فازی نسبت به روش انحراف معیار همراه با اعمال شرط محدوده معتبر سرعت نیم تا یک و نیم متر بر ثانیه را نشان می‌دهد.

در شکل (۱۱) اختلاف نسبی دبی^۳ برای روش‌های انحراف معیار و فضای فازی در مقابل روش دبی اشل نشان داده شده است؛ همان‌طور که مشاهده می‌شود، اختلاف

¹ Stage Discharge Method

² Ozekiyama gauge station

³ Relative Discharge Difference

سیلاب با استفاده از دستگاه تیکه‌نگاری صوتی رودخانه‌ای و تعیین دقت اندازه‌گیری، حداق و حداکثر برد اندازه‌گیری"، آکوئیدرولولوژی، دوره ۶، شماره ۳، پاییز ۱۳۹۸، ص ۵۹۲-۵۸۵ [۳] حداد، پریسا، پایان‌نامه کارشناسی ارشد (۱۳۹۶)، عنوان: تشخیص داده پرت در داده‌های سری زمانی با الگوریتم داده-کاو، دانشگاه آزاد اسلامی، واحد مرودشت، دانشکده مهندسی، گروه کامپیوتر و فناوری اطلاعات.

[۴] مسعود بحرینی مطلق و همکاران، (۱۳۹۹)، "امکان سنجی پایش خودکار رودخانه‌های ایران با استفاده از فناوری تیکه‌نگاری صوتی رودخانه‌ای ۵۰ کیلوهرتز"، مجله انجمن مهندسی صوتیات ایران، سال هشتم، شماره ۱، pp. 14-21.

[۵] مسعود بحرینی مطلق و همکاران، (۱۳۹۸)، "امکان سنجی پایش جریان‌های خلیج فارس با استفاده از فناوری تیکه‌نگاری صوتی دریایی ۱۰ کیلوهرتز"، نشریه مهندسی دریا، سال پانزدهم، شماره ۳۰، پاییز و زمستان ۱۳۹۸، (۱۳۱-۱۳۸) .

[۶] مسعود بحرینی مطلق و همکاران، (۱۳۹۹)، "پایش پیوسته دمای آب با استفاده از فناوری تیکه‌نگاری صوتی"، نشریه مهندسی عمران امیرکبیر، دوره ۵۱، شماره ۵، سال ۱۳۹۹.

[۷] مسعود بحرینی مطلق و همکاران، (۱۳۹۷)، طراحی، ساخت و ارزیابی دستگاه تیکه‌نگاری صوتی رودالی، مجله انجمن مهندسی صوتیات ایران، سال ششم، شماره ۱، بهار و تابستان ۱۳۹۷.

[۸] مسعود بحرینی مطلق و همکاران، (۱۳۹۸)، "اولین تجربه سامانه تیکه‌نگاری صوتی برای پایش سرعت جریان رودخانه در ایران"، پژوهشات آب و خاک ایران، دوره ۵۰، شماره ۷، آذر ۱۳۹۸.

[۹] مسعود بحرینی مطلق و همکاران، (۱۳۹۷)، "ابرسی وضعیت جریان آب در دریاچه هفت برم با استفاده از فناوری تکه‌نگاری صوتی"، نشریه آب و خاک (علوم و صنایع کشاورزی)، جلد ۳۳، شماره ۱، فروردین-اردیبهشت، ص. ۳۵-۲۳.

[10] D. G. Goring and V. I. Nikora, "Despiking acoustic doppler velocimeter data," *J. Hydraul. Eng.*, vol. 128, no. 1, pp. 117-126, 2002, doi:10.1061/(ASCE)07339429(2002)128:1(117).

[11] M. Roy, V. R. Kumar, B. D. Kulkarni, J. Sanderson, M. Rhodes, and M. Vander Stappen, "Simple denoising algorithm using wavelet transform," *AICHE J.*, vol. 45, no. 11, pp. 2461-2466, 1999, doi:10.1002/aic.690451120.

[12] N. Carolina and N. Carolina, "Identification of Low-Dimensional Energy Containing / Flux Transporting Eddy Motion in the Atmospheric Surface Layer Using Wavelet Thresholding Methods," no. 1941, pp. 377-389, 1998.

[13] H. D. I. Abarbanel, J. P. Gollub, F. Combes, A. Mazure, and A. Blanchard, "Analysis of Observed Chaotic Data," vol. 49, no. 11, pp. 10-13, 1996, doi: 10.1063/1.881528.

[14] K. Kawanisi, M. Razaz, J. Yano, and K.

دارای داده‌های پرت است که حالت واقعی سامانه را نشان نداده و عدم پرت‌کاو، شناسایی و حذف و جایگزینی آن‌ها با مقادیر مناسب باعث افزایش خطای اندازه‌گیری‌ها می‌شود.

در این پژوهش سعی شد که به کمک یکی از روش‌های پرکاربرد در پرت‌کاو، عمل پرت‌کاو روی داده‌های سامانه تیکه‌نگاری انجام شود و عملکرد این روش بر روی این نوع داده‌ها پژوهش شود. اگرچه به دلیل ماهیت غیرقطعی کلیت بحث پرت‌کاو و پیچیدگی تعریف یک ناحیه مشخص به‌عنوان ناحیه معتبر، نمی‌توان به دقت و به‌صورت عددی روش‌های مختلف و عملکرد آن‌ها را با یکدیگر بررسی و مقایسه کرد، اما عدم توجه به موضوع پرت‌کاو به این علت نیز سبب می‌شود که نتوان به‌صورت کاملاً بهینه از ابزارهای نوین امروزی تولید داده‌ها بهره برد؛ لذا بایستی با تعریف برخی قوانین و آستانه‌ها به صورت قراردادی، کار پرت‌کاو را انجام داده شود؛ درنهایت با مقایسه نتایج این پژوهش با پژوهشات مشابه دیگری که انجام گرفته است، مشاهده شد که با افزایش درصد داده‌های پرت موجود در گروه داده ورودی، عملکرد روش آستانه‌گذاری فضای فازی مختل شده و نمی‌تواند به‌صورت بهینه عمل کند؛ اگرچه این روش در مقایسه با روش سنتی انحراف معیار عملکرد مناسبی از خود به نمایش گذاشت و اختلاف دبی نسبی بین این روش و روش دبی اشل نسبت به اختلاف دبی نسبی بین روش انحراف معیار و دبی اشل، در بیش‌تر نقاط داده کمتر بوده است. در این پژوهش در مجموع و در طول انجام شش تکرار روش آستانه‌گذاری فازی ۸۰۱۷ عدد داده به عنوان نقاط پرت شناسایی و با مقدار میانگین چهار نقطه داده در دو طرف مجاور خود (هر طرف دو داده) جایگزین شدند. این مقدار معادل با ۳۲ درصد از کل داده‌های ورودی بوده است؛ همچنین مقدار انحراف معیار داده‌ها پیش و پس از پرت‌کاو به روش فازی به ترتیب ۰/۲۰۶ و ۰/۱۱۹ محاسبه شد که کاهش پراکندگی نقاط حول میانگین علامت صدا را نشان می‌دهد.

۵- مراجع

K. Kawanisi, A. Kaneko, S. Nigo, and M. [1] Soltaniasl, "New acoustic system for continuous measurement of river discharge and water temperature," *Water Sci. Eng.*, vol. 3, no. 1, pp. 47-55, 2010, doi: 10.3882/j.issn.1674-2370.2010.01.005.

[۲] مسعود بحرینی مطلق و همکاران، (۱۳۹۸)، "امکان سنجی پایش

پژوهشی مورد علاقه ایشان شامل مدیریت و برنامه‌ریزی منابع آب در سطح حوضه آبریز، مدیریت تقاضا (مدیریت برنامه‌ریزی آبیاری، اکوهیدرولوژی، فرایندهای تصادفی و بهینه‌سازی تحت شرایط عدم قطعیت است.

نشانی رایانامه ایشان عبارت است از:

alizadeh@iust.ac.ir



یوسف الفت‌میری تحصیلات خود را

در مقطع کارشناسی در رشته مهندسی عمران در سال ۱۳۹۶ و در دانشگاه رازی کرمانشاه به پایان رساند و در حال حاضر دانشجوی کارشناسی ارشد

عمران گرایش آب و سازه‌های هیدرولیکی دانشگاه علم و صنعت ایران است. زمینه‌های پژوهشی مورد علاقه وی، مباحث مرتبط با سامانه تیکه‌نگاری صوتی و استفاده از الگوریتم‌های هوش مصنوعی به‌منظور پیش‌بینی سری‌های زمانی و مباحث مرتبط با مهندسی رودخانه و سواحل و مدیریت منابع آب است.

نشانی رایانامه ایشان عبارت است از:

yousefolfatmiri@gmail.com



امیرحسین حسن‌آبادی تحصیلات خود

را در مقطع کارشناسی و در رشته مهندسی عمران در سال ۱۳۹۷ و در دانشگاه صنعتی شاهرود به پایان رسانده و در حال حاضر دانشجوی کارشناسی ارشد

عمران گرایش آب و سازه‌های هیدرولیکی دانشگاه علم و صنعت ایران است. زمینه‌های پژوهشی مورد علاقه وی، مباحث مربوط به پردازش علائم صدا، ابزارهای نوین پایش آب و هیدرومتری و داده‌کاوی، مهندسی رودخانه و سواحل و مدیریت منابع آب است.

نشانی رایانامه ایشان عبارت است از:

AmirhosseinHsnabadi@gmail.com



دکتر ابراهیم جبّاری دوره کارشناسی

خود را در دانشگاه علم و صنعت ایران گذرانده و سپس مقاطع ارشد و دکترا را در دانشگاه لوون کشور بلژیک

گذرانده‌اند. ایشان از سال ۱۳۷۳ تا اکنون عضو هیئت علمی دانشکده عمران دانشگاه علم و صنعت ایران است.

نشانی رایانامه ایشان عبارت است از:

jabbari@iust.ac.ir



دکتر مسعود بحرینی مطلق مقطع دکترا

را در دانشگاه هیروشیما ژاپن در سال‌های ۲۰۱۴ تا ۲۰۱۷ گذرانده‌اند. ایشان هم

اکنون عضو هیئت علمی مؤسسه

پژوهش‌های آب وزارت نیرو هستند. زمینه‌های پژوهشی مورد علاقه ایشان شامل مباحث مربوط به روش تیکه‌نگاری صوتی و مهندسی سواحل و رودخانه است.

نشانی رایانامه ایشان عبارت است از:

m.bahreini@wri.ac.ir



دکتر حسین علیزاده هم اکنون

عضو هیئت علمی دانشکده عمران دانشگاه علم و صنعت ایران هستند.

ایشان در مقطع کارشناسی و کارشناسی ارشد در دانشگاه صنعتی شریف و در مقطع دکترا در دانشگاه امیرکبیر تحصیلات خود را گذرانده‌اند. زمینه‌های