Journal of Acoustical Engineering Society of Iran, Vol. 11, No. 2, 2024

(Research Article)
Identifying outlier data in the time series of the fluvial acoustic tomography system (FATS) for river flow measurement using modified singular spectrum analysis and Kernel density method

F. Hoveyzi[1], S.H. Momeni Masuleh[*1], M. Bahreinimotlagh[2]
Department of Mathematics, Shahed University
Department of Water Resources Studies and Research, Water Research Institute

Received: 2023/11/15, Accepted: 2024/03/17

Abstract

Continuous monitoring of river discharge is crucial for environmental, social, and economic evaluations in water resources management. River acoustic tomography is a system that uses the acoustic tomography method to improve the accuracy of flow measurements and solve the problems of discharge measurement methods. However, the collected data may contain outlier data, which can change the output results and measurement error if not identified and removed. In this research, modified singular spectrum analysis and hybrid kernel density method along with five replacement methods were used to identify and remove outliers. The replacement methods include linear interpolation, an average of outlier data, an average of unexamined outlier data, the replacement of outlier data with the last valid data before the outlier data, and the average of one point on each side of the outlier data. Out of 12232 data, 1076 data (about 9%) and 676 data (about 6%) of the total data were identified as outlier data using the modified singular spectrum analysis and kernel density method, respectively. According to the analyses and investigations, the modified singular spectrum analysis has performed better than the kernel density method.

Keywords: Fluvial acoustic tomography, Outlier detection, Singular spectrum analysis, Kernel density, River discharge.

pp. 33-40 (In Persian)

[*] Corresponding author E-mail: momeni@shahed.ac.ir

مقاله پژوهشی

مجله انجمن مهندسی صوتیّات ایران
سال ۱۴۰۲، دوره: ۱۱، شماره: ۲، صفحات: ۳۳-۴۰

شناسایی داده‌های پرت در سری‌زمانی تیکه‌نگاری صوتی رودخانه‌ای برای اندازه‌گیری دبی رودخانه با استفاده از تحلیل طیفی منفرد اصلاح‌شده و روش چگالی کرنل

فاطمه حویزی[۱]، سیّد حجت‌اله مؤمنی ماسوله*[۱]، مسعود بحرینی مطلق[۲]

۱. گروه ریاضی، دانشکده علوم پایه، دانشگاه شاهد

۲. پژوهشکده مطالعات و تحقیقات منابع آب، مؤسسه تحقیقات آب

دریافت: ۱۴۰۲/۰۸/۲۴، پذیرش: ۱۴۰۲/۱۲/۲۷

چکیده

دستیابی به داده‌های پیوسته‌ی دبی رودخانه به‌منظور ارزیابی‌های زیست‌محیطی، اجتماعی و اقتصادی، از چالش‌های مهم مدیریت منابع آب است. برای ارتقای دقت اندازه‌گیری‌های جریان و حل مشکلات روش‌های اندازه‌گیری دبی، تیکه‌نگاری صوتی رودخانه‌ای توسعه پیدا کرده‌اند. این سامانه از روش تیکه‌نگاری برای پایش پیوسته جریان استفاده می‌کند. داده‌های برداشت‌شده حاوی داده‌های پرتی است که عدم شناسایی و حذف آن‌ها موجب تغییر نتایج خروجی و خطای اندازه‌گیری می‌گردد. در این پژوهش از تحلیل طیفی منفرد اصلاح‌شده و روش ترکیبی چگالی کرنل به همراه پنج روش جایگزینی درون‌یابی خطی، میانگین داده‌های پرت‌کاوی‌شده، میانگین داده‌های پرت‌کاوی‌نشده، جایگزینی داده پرت با آخرین داده‌ی معتبر پیش از داده‌ی پرت و میانگین یک نقطه در هر طرف داده‌ی پرت استفاده شده است. در تحلیل طیفی منفرد اصلاح‌شده و روش چگالی کرنل از ۱۲۲۳۲ داده به ترتیب ۱۰۷۶ داده که حدود ۹ درصد و ۶۷۶ داده که حدود ۶ درصد از کل داده‌ها را شامل می‌شود به عنوان داده‌های پرت شناسایی شدند. با توجه به تحلیل‌ها و بررسی‌های انجام‌شده تحلیل طیفی منفرد اصلاح‌شده عملکرد بهتری نسبت به روش چگالی کرنل داشته است.

کلیدواژه‌ها: تیکه‌نگاری صوتی رودخانه‌ای، داده‌های پرت، تحلیل طیفی منفرد اصلاح‌شده، چگالی کرنل، دبی رودخانه.

۱. مقدمه

رودخانه‌ها از دیرباز تاکنون نقش مهمی در زندگی انسان و بوم‌سازگان داشته‌اند. آب‌های جاری این منابع طبیعی علاوه بر تأمین آب شرب و آبیاری، برای حمل و نقل، تجارت، صنعت، ورزش و تفریح نیز اهمیت دارند. از بنیادی‌ترین منابع تأمین آب، می‌توان به رودخانه‌ها اشاره نمود که نقش مهمی در چرخه‌ی آب جهان دارند. رودخانه‌ها، تأثیر غیرقابل انکاری در مدیریت منابع آب، کنترل سیلاب، بروز حوادث طبیعی در آب دارند.

از جمله ویژگی‌های مهم رودخانه که نقش مهمی در برنامه‌ریزی، کنترل منابع آب و حفاظت از محیط زیست دارند، دبی رودخانه، دما و شوری آب رودخانه هستند. لذا ایجاد یک روش قابل اعتماد و فن‌اوری مرتبط برای اندازه‌گیری، یکی از مؤلفه‌های اصلی و مسئله‌ای کلیدی و واجد اهمیت برای مدیریت منابع آب است. آگاه بودن از میزان دبی رودخانه چند ماه زودتر به خصوص در خشکسالی‌ها مهم است. اساسی‌ترین رابطه به‌منظور اندازه‌گیری دبی رودخانه، ضرب کردن میزان سرعت جریان آب عبوری از مقطع کال (کانال) در مساحت کال است. یکی از معمولی‌ترین وسیله‌ها برای اندازه‌گیری دقیق سرعت جریان در رودخانه استفاده از مولینه است. نظر به اینکه سرعت جریان آب در هر نقطه از مقطع متفاوت بوده به همین دلیل لازم است برای محاسبه‌ی دبی از سرعت متوسط استفاده گردد. رابطه‌ی دبی- اشل متداول‌ترین روش برای اندازه‌گیری دبی رودخانه است. در این روش سرعت جریان آب به صورت ماهانه و یا دو بار در ماه با

* نویسنده پاسخگو: momeni@shahed.ac.ir

فرض دائمی بودن جریان در رودخانه توسط سرعت‌سنج‌های مکانیکی اندازه‌گیری شده و با استفاده از یک رابطه‌ی یک به یک با ارتفاع آب، دبی رودخانه محاسبه می‌گردد. گاهی تخمین دبی رودخانه در زمان گذر موج یا سیلاب به‌دلیل غیردائمی بودن جریان رودخانه با استفاده از نمودار دبی- اشل دارای خطای زیادی است و تغییرات دبی نسبت به ارتفاع آب به‌صورت یک به یک نخواهد بود. بنابراین نه تنها اندازه‌گیری سرعت جریان با استفاده از مولینه و محاسبه دبی توسط دبی- اشل طاقت‌فرسا و زمان‌بر است، بلکه این روش در زمان جریان‌های غیردائمی دارای خطای بالایی است. برای غلبه بر محدودیت‌های روش دبی- اشل، سه نوع ابزار صوتی توسعه پیداکرده‌اند. اولین ابزار سرعت‌سنج‌های صوتی[1] هستند که از زمان جابجایی صوتی[2]، برای اندازه‌گیری سرعت متوسط در راستای مسیر صوتی استفاده می‌کنند. مشکل اصلی این روش این است که برای محاسبه دبی نیاز به چندین پس پردازش فرعی، مانند روش شاخص سرعت که سرعت اندازه‌گیری‌شده در راستای مسیر صوتی ارسالی را به سرعت متوسط مقطع عرضی تبدیل می‌کند، دارد [۱].

دومین ابزار، سرعت‌سنج‌های افقی داپلَر[3] است که به منظور پایش پیوسته جریان به‌کار می‌رود و از قانون داپلر، برای اندازه‌گیری سرعت جریان آب استفاده می‌کند، مشابه با سرعت‌سنج‌های صوتی، این سرعت‌سنج نیز امواج صوتی را در راستای یک خط مستقیم انتشار می‌دهد و قادر به پوشش تمام یک مقطع عرضی نیست. علاوه بر این، زمانی که مقدار رسوب معلق در آب افزایش می‌یابد، توانایی برداشت نیم‌رخ افقی سرعت توسط این سرعت‌سنج‌ها به سرعت کاهش می‌یابد؛ علاوه بر این مشکلات در صورت وجود جریان‌های جزر و مدی دقت اندازه‌گیری این دستگاه کاهش پیدا کرده و هم‌چنین برد اندازه‌گیری محدود نیز از دیگر محدودیت‌های استفاده از سرعت‌سنج‌های افقی است. برای ارتقای دقت اندازه‌گیری‌های جریان و حل مشکلات سرعت‌سنج‌های صوتی و سرعت‌سنج‌های افقی داپلر، تیکه‌نگاری صوتی رودخانه‌ای[4] توسعه پیدا کرده است، این سامانه از روش تیکه‌نگاری برای پایش پیوسته جریان استفاده می‌کند. فن‌اوری تیکه‌نگاری صوتی یکی از جدیدترین و دقیق‌ترین فن‌اوری‌های روز دنیا برای پایش پیوسته منابع آب در اقیانوس‌ها، دریاها و رودخانه‌ها و برای اندازه‌گیری ویژگی‌های مختلف آب است که در حال حاضر در کشورهای توسعه یافته مورد استفاده قرار می‌گیرد. فن‌اوری تیکه‌نگاری صوتی در مقایسه با سایر ابزارهای صوتی نظیر سرعت‌سنج‌های صوتی و سرعت‌سنج‌های داپلر، قابلیت انتشار امواج در فواصل طولانی‌تری را دارد. هم‌چنین برخلاف آن دو دستگاه که الگوی انتشار امواج صوتی در آن‌ها در راستای یک خط مستقیم است، در این روش امواج صوتی در کل مقطع محیط آب انتشار می‌یابند و نیازی به پس‌پردازش‌های پیچیده روی داده‌های خام نیست.

زنان‌ژو و همکاران، برای نظارت مداوم بر دبی رودخانه یانگتز، از ژانویه ۲۰۱۸ تا ۲۰۲۱ با داده‌های به‌دست‌آمده از نصب دستگاه‌های تیکه‌نگاری‌صوتی و داپلر در این رودخانه توانستند روشی ابتکاری برای دستیابی به دبی رودخانه حتی در زمان سیل فراهم کنند [۲].

حسن‌آبادی و همکاران، با استفاده از روش آستانه‌گذاری فضای فازی و روش سه‌بعدی روسیو- فضای فازی داده‌های پرت موجود در داده خام سامانه تیکه‌نگاری صوتی رودخانه‌ای را حذف و با مقایسه‌های انجام‌شده، روش سه‌بعدی روسیو- فضای فازی بهترین عملکرد را روی این داده‌ها داشته است[۳].

بحرینی‌مطلق و همکاران، با بررسی دقت اندازه‌گیری سرعت جریان با استفاده دستگاه تیکه‌نگاری صوتی رودالی ۵۰ کیلوهرتز برای رودخانه‌های با عرض کم و ۳۰ کیلوهرتز برای رودخانه‌های دارای پهنا نشان دادند که در رودخانه‌های دارای عرض بیش‌تر از ۴۰ متر استفاده از دستگاه تیکه‌نگاری صوتی رودالی ۵۰ کیلوهرتز توصیه می‌گردد [۴].

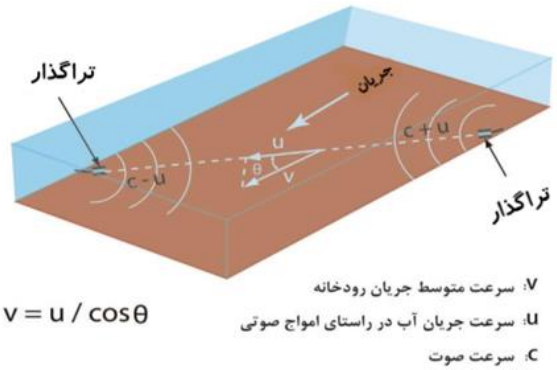
السواف و همکاران، با استفاده از تیکه‌نگاری صوتی

[1] AVMS; Acoustic Velocity Meter Systems
[2] Acoustic travel time
[3] H-ADCP; Horizontal Acoustic Doppler Current Profiler
[4] FATS; Fluvial Acoustic Tomography Systems

رودخانه‌ای براساس زمان جابجایی امواج، برای اندازه‌گیری سرعت متوسط مقطع عرضی، جهت جریان رودخانه و دبی رودخانه در دو رویداد سیل مورد بررسی قرار گرفت. در این پژوهش معادلات جدیدی برای ارزیابی جهت جریان و هم‌چنین دستورالعمل‌های جدیدی برای تخمین پیوسته جهت جریان با استفاده از توزیع مثلثی سامانه تیکه‌نگاری صوتی ارائه دادند. به‌طورکلی در این پژوهش روشی نوین برای پایش حرکت جریان رودخانه ارائه شده است [۵].

داده‌های برداشت‌شده از روش تیکه‌نگاری صوتی، سرعت‌های متوسط جریان در راستای امواج صوتی هستند که در فواصل زمانی منظم به ثبت رسیده‌اند.

دو مؤلفه مهم برای محاسبه سرعت صدا، دمای متوسط جریان و سرعت متوسط جریان آب، زمان رسیدن موج‌های صوتی در ایستگاه‌های پایین‌دست و بالادست و دانستن فاصله دو ایستگاه است. دقت اندازه‌گیری سرعت جریان متوسط در دستگاه‌های تیکه‌نگاری صوتی، ۱± سانتی‌متر بر ثانیه است و هم‌چنین این دستگاه‌ها توانایی اندازه‌گیری سرعت‌های بسیار کم را دارند. انتشار امواج صوتی در محیط آب تابع سرعت جریان آب u(x, y) و سرعت صدا C(x, y, z) در آن محیط است (شکل ۱).



شکل ۱ سرعت جریان آب در راستای امواج صوتی u سرعت در جهت جریان آب v و زاویه بین امواج صوتی جهت جریان آب θ.

با توجه به این‌که سرعت عمودی جریان ناچیز است، سرعت صدا c_m و سرعت متوسط جریان u_m در مقطع بین دو ایستگاه را می‌توان به‌صورت افقی در نظر گرفت.

سرعت صدا و سرعت متوسط جریان در مقطع بین دو ایستگاه به‌صورت رابطه (۱) تعریف می‌شوند [۶]:

$$\begin{cases} C_m = \dfrac{L}{t_m}, \\ u_m = \dfrac{C_m^{۲}}{۲L}\Delta t \end{cases} \qquad (۱)$$

که درآن L فاصله بین دو ایستگاه، t_m حاصل میانگین زمان رسیدن امواج به ایستگاه‌ها و Δt تفاضل زمان رسیدن امواج به ایستگاه‌ها هستند و به‌صورت زیر تعریف می‌شوند:

$$\begin{cases} t_m = \dfrac{t_{up} + t_{Down}}{۲}, \\ \Delta t = t_{up} - t_{Down} \end{cases} \qquad (۲)$$

که در آن t_{up} زمان رسیدن امواج صوتی به ایستگاه بالادست و t_{down} زمان رسیدن امواج صوتی به ایستگاه پایین‌دست هستند و به صورت رابطه (۳) تعریف می‌شوند:

$$\begin{cases} t_{up} = \dfrac{L}{c_m - u_m}, \\ t_{down} = \dfrac{L}{c_m + u_m} \end{cases} \qquad (۳)$$

در نهایت سرعت جریان رودخانه V_m با اندازه‌گیری سرعت متوسط در راستای امواج صوتی u_m به‌دست می‌آید:

$$V_m = \frac{u_m}{cos\Theta} \qquad (۴)$$

و با استفاده از رابطه پیوستگی، دبی جریان برابر است با حاصل‌ضرب سرعت متوسط جریان در مقطع عرضی رودخانه:

$$Q = A(H)V_m \sin(\Theta) \qquad (۵)$$

که‌درآن A سطح مقطع عرضی و تابعی از ارتفاع آب H و Θ زاویه بین دو فرستنده و جهت جریان رودخانه است.

داده‌های به‌دست‌آمده از روش تیکه‌نگاری صوتی، داده‌های سری‌زمانی هستند که به دلیل عوامل مختلف، هنگام برداشت از طریق حسگرها، داده‌های پرتی نیز به همراه آن‌ها ثبت شده‌اند. عدم شناسایی و حذف این داده‌ها موجب تغییر نتایج خروجی و خطا می‌گردد؛ لذا ضرورت انجام پرت‌کاوی پیش از اجرای الگوریتم‌های مختلف داده‌کاوی برای کسب اطلاعات مفید در مورد داده‌ها، مسئله‌ای الزامی و غیرقابل انکار است.

۲. مواد و روش‌ها

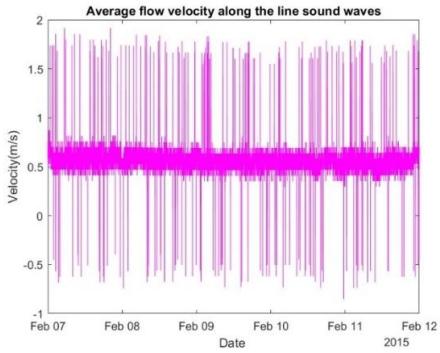
یک داده‌ی پرت مشاهده‌ای (یا مجموعه‌ای از مشاهدات) است که اختلاف معناداری با سایر داده‌ها دارد. با توجه به

اینکه سری‌زمانی به‌دست‌آمده از سامانه تیکه‌نگاری صوتی داده‌های سرعت متوسط جریان در راستای امواج صوتی هستند از این‌رو در این پژوهش پس از پرت‌کاوی با استفاده از تحلیل طیفی منفرد اصلاح‌شده و روش چگالی کرنل، سرعت جریان رودخانه در راستای اصلی جریان رودخانه محاسبه و سپس دبی رودخانه محاسبه گردیده است.

۱-۲. داده‌های منطقه مورد استفاده

داده‌های در دسترس قرارگرفته برای انجام این پژوهش در بازه زمانی ۷ فوریه تا ۱۲ فوریه ۲۰۱۵ در دانشگاه هیروشیما[1] ژاپن، به‌صورت پیوسته و هر ۳۰ ثانیه مورد اندازه‌گیری قرار گرفته است. محل برداشت این داده‌ها، رودخانه گونو واقع در حاشیه شهر میوشی استان هیروشیما ژاپن بوده است (شکل ۲).

شکل ۲ محدوده برداشت داده‌ها و نقشه شبکه رودخانه‌ها و مخازن سدهای موجود در بالای قسمت مورد مطالعه.



شکل ۳ نمودار داده‌های اولیه تیکه‌نگاری صوتی.

۲-۲. روش پژوهش

۲-۲-۱. تحلیل طیفی منفرد اصلاح‌شده

تحلیل طیفی منفرد یکی از روش‌های ناپارامتری برای تحلیل سری‌زمانی است. این روش شامل دو مرحله است: مرحله اول تجزیه و مرحله دوم بازسازی است. هر کدام از این مراحل خود دارای دو گام هستند. در مرحله تجزیه ابتدا تعبیه‌سازی[2] و سپس تجزیه[3] صورت می‌گیرد و در مرحله بازسازی[4] ابتدا گروه‌بندی[5] و سپس میانگین‌گیری قطری[6] صورت می‌گیرد.

الگوریتم روش

شناسایی داده‌های پرت با استفاده از تحلیل طیفی منفرد اصلاح‌شده به‌صورت زیر انجام می‌گردد [۷]:

$$f = \left[f\left(۱\right), f\left(۲\right), f\left(۳\right) \dots f(M)\right], \quad M > ۲$$

$f = a + b$ ترکیبی از دو علامت (سیگنال) است. با در نظر گرفتن علامت a به عنوان علامت نوفه و علامت b به عنوان علامت مطلوب، هدف استخراج علامت نوفه از علامت مطلوب است. الگوریتم تحلیل طیفی منفرد دارای مراحل زیر است:

۱. آستانه و طول پنجره براساس داده‌ها تعیین می‌شود.

۲. علامت یک متغیره f، در ماتریس چند متغیره F به‌صورت زیر جای‌گذاری می‌شود:

$$\mathrm{F} = \begin{pmatrix} f(۱) & f(۲) & \cdots & f(L) \\ f(۲) & f(۳) & \cdots & f(L+۱) \\ \vdots & \vdots & \vdots & \vdots \\ f(N) & f(N+۱) & \cdots & f(M) \end{pmatrix}, \qquad (۶)$$

که‌درآن $۱ < N < M$ طول ماتریس است که به آن طول پنجره گفته می‌شود و $L = M - N + ۱$ است. ماتریس مسیر F، ماتریس هنکل نامیده می‌شود (تعبیه‌سازی).

۳. تشکیل ماتریس کوواریانس و محاسبه مقادیر ویژه آن (تجزیه کردن).

$$C = FF^T = F^T F \qquad (۷)$$

۴. تحرک محلی[7] برای هر یک از بردارهای ویژه محاسبه می‌شود. بدین‌صورت که فرض کنید:

$$u = \left[u\left(۱\right), u\left(۲\right), \dots, u\left(N\right)\right]$$

u یک ماتریس است که هر کدام از ستون‌های آن یک

[1] Miyoshi City, Hiroshima, Japan

[2] Embedding
[3] Decomposition
[4] Reconstruction
[5] Grouping
[6] Diagonal averaging
[7] Local mobility

بردار است از این‌رو تحرک محلی آن به‌صورت زیر محاسبه می‌گردد:

$$m_u = \frac{m_1}{m_0} \tag{۸}$$

که در آن:

$$\begin{cases} m_0 = \sqrt{\frac{1}{N}}\,\|u(n)\|_2 \\ m_1 = \sqrt{\frac{1}{N-1}}\,\|s(n)\|_2 \end{cases}$$

و

$$s(n) = u(n) - u\left(n-1\right), \quad n = 1, 2, \dots, N$$

۵. برای هر بردار ویژه‌ای که تحرک محلی آن کم‌تر یا مساوی آستانه است یک ماتریس با استفاده از رابطه (۹) محاسبه می‌گردد و همه ماتریس‌های به‌دست‌آمده با یکدیگر جمع می‌شوند. (گروه‌بندی)

$$F_k = u_k u_k^T F, \qquad k = 1, 2, \dots, N, \tag{۹}$$

$$F = \sum_{j=1}^{d} F_i \tag{۱۰}$$

که در آن d تعداد ماتریس‌هایی است که با استفاده از رابطه‌ی ۹ به‌دست‌آمده‌اند.

۶. با استفاده از رابطه (۱۱) میانگین‌گیری قطری روی ماتریس به‌دست‌آمده از رابطه (۱۰)، برای بازسازی علامت تک متغیره صورت می‌گیرد:

$$\begin{cases} \frac{1}{t}\sum_{k=1}^{t} \hat{a}_{k,t-k+1}, & 1 \le t < N, \\ \frac{1}{N}\sum_{k=1}^{N} \hat{a}_{k,t-k+1}, & N \ge t \le L, \\ \frac{1}{M-t+1}\sum_{k=t-L+1}^{M-L+1} \hat{a}_{k,t-k+1}, & L < t \le M \end{cases} \tag{۱۱}$$

۲-۲-۲. تخمین چگالی کرنل دو متغیره

روش دوم، روش تخمین چگالی دو متغیره کرنل[1] برای شناسایی داده‌های پرت استفاده شده است. چگالی کرنل یک برآورد ناپارامتریک چگالی است. در این روش با استفاده از یک بیضی که توسط محاسباتی روی داده‌ها ترسیم می‌گردد داده‌های پرت شناسایی می‌شوند. اگر پهنای باند مناسب انتخاب شود، می‌تواند تخمین خوبی از چگالی را ارائه دهد و به اندازه شبکه وابسته نباشد. چگالی دو متغیره کرنل با کرنل گاوسی را می‌توان با استفاده از رابطه‌ی (۱۲) محاسبه کرد [۸]:

$$\hat{f}\,(x.y) = \frac{1}{2\pi N h_x h_y} \sum_{i=1}^{N} exp\left[-\frac{(x-x_i)^2}{2h_x^2} - \frac{(y-y_i)^2}{2h_y^2}\right] \tag{۱۲}$$

که‌درآن N تعداد داده‌ها، h_y و h_x پهنای باند در هر دو محور، x_i و y_i متغیرها، x و y مکان‌هایی که چگالی در آن‌ها برآورد می‌گردد و $\hat{f}\,(x,y)$ چگالی تخمین‌زده شده است. می‌توان (x,y) را $(u,\Delta u)$ در نظر گرفت. که در آن Δu مشتق مرتبه اول با دقت درجه دوم داده‌ها است (در مقاله مرجع از مشتق مرتبه اول استفاده شده است ولی در این مقاله برای افزایش دقت از مشتق مرتبه اول با دقت درجه دوم استفاده شده است).

الگوریتم روش

۱. محاسبه مشتق پیشرو (مرتبه‌ی اول) با دقت مرتبه‌ی دوم برای نقطه‌ی اول، مشتق پسرو با دقت مرتبه‌ی دوم برای نقطه پایانی و مشتق مرکزی برای نقاط میانی که به صورت زیر محاسبه می‌گردند:

$$\begin{cases} \Delta u_1 = \frac{1}{2h}\left[3u_1 - 4u_2 + u_3\right], \\ \Delta u_N = -\frac{1}{2h}\left[3u_N - 4u_{N-1} + u_{N-2}\right], \\ \Delta u_i = \frac{u_{i+1} - u_i}{2h} \end{cases} \tag{۱۳}$$

۲. برآورد کردن زاویه چرخش محورهای اصلی با استفاده از تقریب حداقل مربعات:

$$\theta = tan^{-1}\left(\frac{N\sum_{i=1}^{N} x_i y_i - \sum_{i=1}^{N} x_i \sum_{i=1}^{N} y_i}{N\sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i\right)^2}\right) \tag{۱۴}$$

۳. تبدیل داده‌ها با استفاده از رابطه‌ی زیر:

$$\begin{cases} x_t = x\cos\theta + y\sin\theta, \\ y_t = -x\sin\theta + y\cos\theta \end{cases} \tag{۱۵}$$

[1] Bivariate kernel density estimation

۴. مقیاس کردن مجدد داده‌ها در بازه [۰,۱] با رابطه زیر:

$$\begin{cases} x_s = \dfrac{x - min(x)}{max(x) - \min(x)}, \\ y_s = \dfrac{y - min(y)}{max(y) - \min(y)} \end{cases} \quad (۱۶)$$

۵. پس از مقیاس کردن داده‌ها در گام قبل، با استفاده از رابطه‌ی (۱۲) چگالی کرنل داده‌ها محاسبه می‌گردد.

۶. تخمین نقاط قطع برای تعیین اندازه بیضی و رسم بیضی.

برای شناسایی نقاط قطع از شیب‌های (C_x, C_y) چگالی نرمال‌شده در امتداد محورهای بیضی استفاده شده است و مقدار آن‌ها کم‌تر یا مساوی ۰/۴ انتخاب می‌شود، که به صورت زیر محاسبه می‌گردند:

$$\begin{cases} C_x = n_x \left|\Delta\hat{f}(x, y_p)\right| / \hat{f}(x_p, y_p) \\ C_y = n_y \left|\Delta\hat{f}(x_p, y)\right| / \hat{f}(x_p, y_p) \end{cases} \quad (۱۷)$$

که در آن n_x و n_y تعداد نقاط شبکه استفاده شده در ماتریس چگالی کرنل هستند. $\Delta\hat{f}(x_p, y)$ و $\Delta\hat{f}(x, y_p)$ به‌صورت زیر محاسبه می‌گردند:

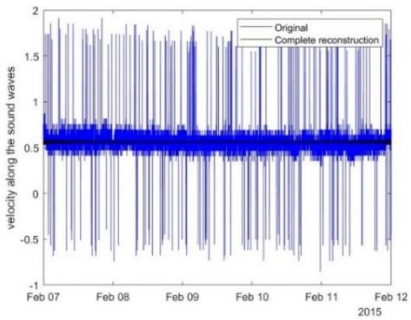
$$\begin{cases} \Delta\hat{f}(x, y_p) = \hat{f}(x_{i+1} - x_i,\ y_p) \\ \Delta\hat{f}(x_p, y) = \hat{f}(x_p,\ y_{i+1} - y_i) \end{cases} \quad (۱۸)$$

با استفاده از محورهای بزرگ و کوچک بیضی که توسط نقاط قطع، به دست می‌آیند و رسم بیضی، داده‌های خارج از بیضی، داده‌های پرت هستند. پس از شناسایی داده‌های پرت و حذف آن‌ها با استفاده از روش چگالی کرنل، با به‌کارگیری پنج روش جایگزینی:۱. درون یابی خطی، ۲. میانگین یک نقطه در هر طرف داده‌ی پرت جمعا دو نقطه، ۳. آخرین داده‌ی معتبر پیش از داده‌ی پرت، ۴. میانگین علامت داده‌های پرت‌کاوی‌شده و ۵. میانگین علامت داده‌های پرت‌کاوی‌نشده، داده‌های پرت حذف‌شده جایگزین می‌گردند.

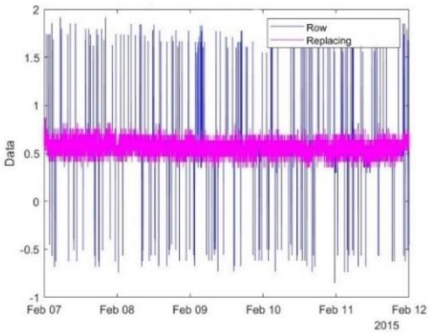
۳. نتایج

داده‌های مورد بررسی، داده‌های سرعت جریان رودخانه هستند که تعداد آن‌ها ۱۲,۲۳۲ است. تعداد داده‌های پرت شناسایی‌شده در تحلیل طیفی منفرد اصلاح‌شده ۱,۰۷۶ عدد است که حدود ۹ درصد از کل داده‌ها را در برمی‌گیرد. شکل ۴ خروجی این تحلیل را نشان می‌دهد.

در روش چگالی کرنل تعداد داده‌های پرت شناسایی‌شده ۶۷۶ عدد از کل داده‌هاست که حدود ۶ درصد از کل داده‌ها را در بر می‌گیرد. شکل ۵ خروجی پرت‌کاوی این روش را نشان می‌دهد.

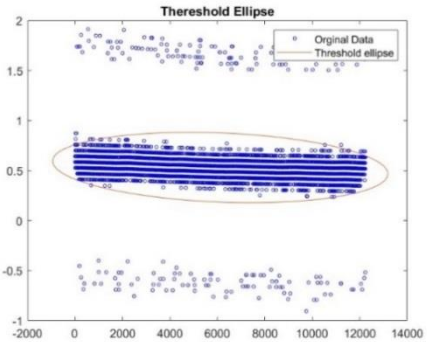


شکل ۴ پرت‌کاوی با استفاده از تحلیل طیفی منفرد اصلاح‌شده.



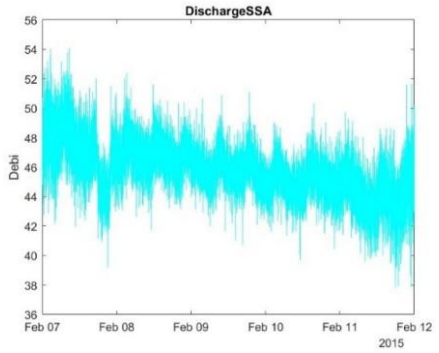
شکل ۵ پرت‌کاوی با استفاده از روش چگالی کرنل.

پس از پرت‌کاوی، سرعت متوسط جریان در راستای اصلی رودخانه با استفاده از رابطه (۴) محاسبه گردید و سپس با استفاده از رابطه (۵) دبی رودخانه به‌دست‌آمد (شکل ۶).

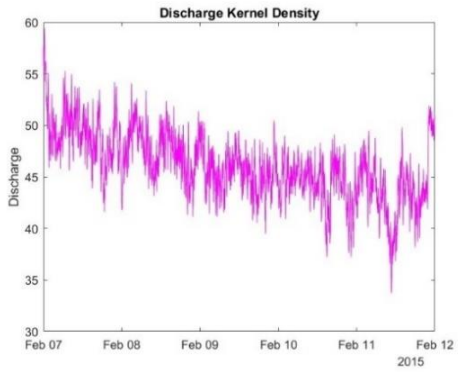


شکل ۶ بیضی به‌دست‌آمده از روش چگالی کرنل.

بیضی به‌دست‌آمده از روش چگالی کرنل با استفاده از داده‌ها را نشان می‌دهد. در شکل ۷ دبی رودخانه محاسبه‌شده روی داده‌های پرت‌کاوی‌شده توسط تحلیل طیفی منفرد اصلاح‌شده و در شکل ۸ دبی رودخانه



شکل ۷ دبی محاسبه‌شده برای داده‌های پرت‌کاوی‌شده با استفاده از تحلیل طیفی منفرد اصلاح‌شده.



شکل ۸ دبی محاسبه‌شده برای داده‌های پرت‌کاوی‌شده با استفاده از روش چگالی کرنل.

محاسبه‌شده روی داده‌های پرت‌کاوی‌شده توسط روش چگالی کرنل قابل مشاهده است. پس از پرت‌کاوی با استفاده از روش چگالی کرنل جایگزینی توسط روش‌های بیان‌شده انجام گردید.

۴. بحث

در تحلیل طیفی منفرد اصلاح‌شده آستانه با استفاده از تحرک محلی در دو مرحله تعیین گردیده است و در روش چگالی کرنل برای افزایش دقت محاسبات از مشتق مرتبه اول با دقت درجه دوم استفاده گردیده است.

با توجه به اینکه متداول‌ترین روش برای تخمین دبی، روش دبی- اشل است، با استفاده از چهار مؤلفه ضریب همبستگی پیرسون[1]، جذر خطای میانگین مربعات[2]، میانگین درصد خطای نسبی[3] و ضریب کارایی نش- ساکلیف[4]، دبی به‌دست‌آمده از داده‌های به‌دست‌آمده از روش تیکه‌نگاری صوتی رودخانه‌ای که توسط تحلیل طیفی منفرد اصلاح‌شده و روش چگالی کرنل پرت‌کاوی شده‌اند مورد ارزیابی و بررسی قرار گرفت. در جدول ۱ این محاسبات قابل مشاهده هستند.

روشی که دارای ضریب کارایی الگوی نش- ساکلیف

جدول ۱ ارزیابی دبی داده‌های پرت‌کاوی‌شده و دبی- اشل.

روش	همبستگی پیرسون	خطای جذر میانگین مربعات	میانگین خطای درصد مطلق	ضریب کارایی الگوی نش- ساکلیف
چگالی کرنل با جایگزینی یک نقطه در هر طرف داده پرت	۰٫۵۱	۲٫۸۴	۴٫۶۴	-۰٫۴۶
چگالی کرنل با جایگزینی آخرین داده معتبر پیش از داده پرت	۰٫۵۲	۲٫۸۲	۴٫۵۴	-۰٫۴۵
چگالی کرنل با جایگزینی میانگین علامت پرت‌کاوی‌شده	۰٫۵۳	۲٫۶۵	۴٫۴۰	-۰٫۲۸
چگالی کرنل با جایگزینی میانگین علامت پرت‌کاوی‌نشده	۰٫۵۴	۲٫۶۹	۴٫۳۰	-۰٫۳۲
چگالی کرنل با جایگزینی درون‌یابی خطی	۰٫۵۴	۲٫۷۴	۴٫۴۱	-۰٫۳۷
تحلیل طیفی منفرد اصلاح‌شده	۰٫۵۵	۱٫۹۵	۳٫۱۶	۰٫۳۰
روش سه‌بعدی روسیو فضای فازی	۰٫۳۱	۵٫۳۷	۹٫۲۷	-۴٫۲۴
روش انحراف معیار با ضریب ۲	۰٫۵۴	۵٫۴۸	۴٫۵۰	-۰٫۴۱
روش انحراف معیار با ضریب ۵	۰٫۵۴	۵٫۴۸	۴٫۵۰	-۰٫۴۱

1 Pearson Correlation
2 RMSE; Root Mean Squared Error
3 MRPE; Mean Relation Percentage Error
4 NSE; Nash-Sutcliffe Efficiency

نزدیک به عدد ۱، بیش‌ترین مقدار ضریب همبستگی پیرسون و کم‌ترین مقدار برای دیگر مؤلفه‌هاست، روش بهتری می‌باشد.

همان‌گونه که در جدول ۱، مشاهده می‌شود تحلیل طیفی منفرد اصلاح‌شده به نسبت روش ترکیبی تخمین چگالی کرنل و روش‌های جایگزینی آن، روش انحراف معیار با ضریب ۲ و ۵ و روش سه‌بعدی روسیو فضای فازی، تحلیل طیفی منفرد اصلاح‌شده، با توجه به محاسبات انجام‌شده عملکرد بهتری دارد از این‌رو این تحلیل، روش بهتری برای پرت‌کاوی است.

تقدیر و تشکر

نویسندگان این مقاله از داوران محترم مقاله کمال تشکر را دارند.

تضاد منافع

نویسندگان هیچ منافع مالی یا روابطی که بتواند بر انتشار این مقاله تاثیرگذار باشد، ندارند.

فهرست منابع

[1] P. Dobriyal, R. Badola, C. Tuboi, S.A. Hussain, “A review of methods for monitoring streamflow for sustainable water resource management,” Applied Water Science, vol. 7, no. 6, pp. 2617-2628, 2017.

[2] Z. Chen, H. Zheng, Y. Tang, C. Chen, “Measurement of yangtze river flow based on coastal acoustic tomography,” Journal of Physics: Conference Series, vol. 1739, no. 1, pp. 012032, 2021.

[3] A. Hasanabadi, M. Bahreinimotlagh, E. Jabbari, K. Kawanisi, H. Alizadeh, Y. Olfatmiri, “Improving the accuracy of streamflow data acquired from the acoustic tomography technique using data despiking algorithms,” Journal of Hydrology, vol. 608, no. 127587, 2022.

[4] M. Bahreinimotlagh, R. Roozbahani, Y. Zohrabi, H. Kardanmoghadam, H. Dehban, K. Mohtasham, “Feasibility study of real-time and automated monitoring of Iranian Rivers using 50-kHz fluvial acoustic tomography system,” Journal of Acoustical Society of Iran, vol. 8, no. 1, pp. 14-21, 2020.

[5] M.B. Al Sawaf, K. Kawanisi, G.N.N. Gusti, F. Khadami, C. Xiao, M. Bahreinimotlagh, “Continuous measurement of flow direction and streamflow based on travel time principles using a triangular distribution of acoustic tomography systems,” Journal of Hydrology, vol. 617, no. 128917, 2023.

[6] M. Bahreinimotlagh, R. Roozbahani, Y. Zohrabi, H. Kardanmoghadam, H. Dehban, K. Mohtasham, “Feasibility study of real-time and automated monitoring of Iranian Rivers using 50-kHz fluvial acoustic tomography system,” Journal of Acoustical Society of Iran, vol. 8, no. 1, pp. 14-21, 2020.

[7] A. Sharma, A.K. Maddirala, B. Kumar, “Modified singular spectrum analysis for despiking acoustic Doppler velocimeter (ADV) data,” Measurement, vol. 117, pp. 339-346, 2018.

[8] M.R. Islam, D.Z. Zhu, “Kernel density-based algorithm for despiking ADV data,” Journal of Hydraulic Engineering, vol. 139, no. 7, pp. 785-793, 2013.