Iranian Journal of Irrigation and Drainage
No. 5, Vol. 10, Nov.-Dec. 2016, p. 556-569

نشریه آبیاری و زهکشی ایران
شماره 5، جلد 10، آذر - دی 1395، ص. 569-556

# تحلیل فراوانی دبی اوج سیلاب با پهنای باند متغیر و ثابت روش چگالی هسته‌ای مطالعه موردی: رودخانه دز

محمدعلی محمدجعفرشعرباف[1*] ، سید سعیدموسوی ندوشنی[2]

تاریخ دریافت: 1394/1/18 تاریخ پذیرش: 1395/7/28

## چکیده

روش متداول در تحلیل فراوانی سیلاب رویکرد پارامتری است، این روش در تحلیل چگالی‌های نامتقارن و دارای چند نقطه اوج توانمند نیست. به منظور رفع این مشکل می‌توان از مدل‌های ناپارامتری استفاده کرد. روش‌های تخمین چگالی هسته با پهنای باند ثابت و پهنای باند متغیر از جمله روش‌های ناپارامتری در تخمین تابع است. تابع چگالی احتمال در روش برآورد چگالی هسته با پهنای باند ثابت با انتخاب یک تابع هسته و یک پهنای باند بهینه برآورد می‌شود. تابع چگالی احتمال در روش پهنای باند متغیر با انتخاب یک تابع هسته و محاسبه پهنای باند در هر نقطه مشاهداتی تخمین زده می‌شود. روش‌های قاعده سرانگشتی و صحت‌سنجی مضاعف روش‌های رایج در محاسبه پهنای باند بهینه است. روش مرتبط با پهنای باند علاوه بر روش‌های محاسباتی پهنای باند مذکور در این تحقیق استفاده شده است. تحلیل فراوانی سیلاب با رویکرد ناپارامتری با توجه به دبی‌های حداکثر لحظه-ای سالانه رودخانه دز برای 38 سال آماری انجام و نتایج با روش پارامتری مقایسه شد. با توجه به معیار کارایی جذر میانگین مربع خطا، نتایج نشان داد پهنای باند بهینه محاسبه شده بر اساس روش مرتبط با پهنای باند دقیق‌ترین روش محاسبه پهنای باند تابع هسته نسبت به روش‌های قاعده سرانگشتی و صحت‌سنجی مضاعف است. همچنین روش چگالی هسته با پهنای باند متغیر دقیق‌تر از روش چگالی هسته با پهنای باند ثابت است، هر دو روش ناپارامتری مذکور نیز از توزیع پارامتری لوگ‌پیرسون نوع 3 دقیق‌تر هستند.

**واژه‌های کلیدی:** پارامتری، پهنای باند، تابع هسته ثابت، تابع هسته متغیر، تحلیل فراوانی، ناپارامتری

## مقدمه [1]

الگوهای هیدرولوژیکی تحلیل فراوانی به طور رایج با روش‌های پارامتری تحلیل شده‌اند، یعنی یک توزیع احتمالی برای داده‌ها فرض می‌شود. انتخاب این توزیع مهمترین مساله در تحلیل فراوانی سیلاب است. توزیع واقعی در عمل غالبا ناشناخته است. انتخاب توزیع‌های مناسب دقت تخمین را کاهش می‌دهد(Karmakar et al., 2007). همچنین اغلب توابع پارامتری تک اوجه هستند و تنها تعداد کمی از توابع پارامتری نظیر توزیع‌های آمیخته پدیده‌های چنداوجی را تحلیل می‌کنند در حالی که تعداد زیادی از مسایل در عمل متضمن چگالی-های چند اوجه است.

بنابراین تحلیل فراوانی با رویکرد ناپارامتری تا حدودی محدودیت‌های روش پارامتری را رفع می‌کند. در این روش، توزیع مشخصی برای داده‌ها همچون روش پارامتری در نظر گرفته نمی‌شود و شکل‌های توابع چگالی ناپارامتری به طور مستقیم به واسطه داده‌ها تعیین می‌شود (Polanski et al., 2000; Shabri., 2002).

یکی از ساده‌ترین روش‌های تخمین تابع ناپارامتری روش هیستوگرام است. این روش به طور گسترده‌ای استفاده می‌شود (Polanski et al., 2000)

آداموفسکی روش تخمین هسته را برای تحلیل فراوانی سیلاب در روش ناپارامتری پیشنهاد داد. وی چگالی هسته با پهنای باند ثابت را با دو توزیع لوگ پیرسون نوع 3 و مقادیر حدی تعمیم یافته برای داده‌های سیلاب رودخانه مارگاری مقایسه نمود. وی با محاسبه پهنای باند به روش قاعده سرانگشتی نشان داد که روش ناپارامتری برآورد چگالی هسته با پهنای باند ثابت، دقت بهتری نسبت به روش پارامتری دارد (Adamowski., 1987) .

اسکات روش‌های ناپارامتری را برای تحلیل داده‌ها در سه و چهار بعد بررسی کرد و نتایج نشان داد که استفاده از این روش مفید و دارای پتانسیل خوبی در آینده است(Scott., 1985).

شابری روش تابع چگالی هسته با پهنای باند ثابت را در تحلیل

1- دانشجوی دکتری مهندسی عمران، مهندسی و مدیریت منابع آب، دانشگاه شیراز
2- استادیار دانشکده مهندسی عمران، آب و محیط زیست، دانشگاه شهید بهشتی
(*- نویسنده مسئول: Email: mohammad.sharbaf@gmail.com)

فراوانی سیلاب به کار برد. او از صحت سنجی مضاعف در محاسبه پهنای باند تابع هسته بهره گرفت. لذا این روش را رویکرد مناسبی معرفی کرد (Shabri., 2002).

کیم و همکاران نشان دادند که استفاده از روش ناپارامتری با چگالی هسته با پهنای باند ثابت دارای نتایج پایداری در تخمین تابع چگالی احتمال نسبت به روش پارامتری است(Kim et al., 2003) .

حقیقت‌جو با استفاده از چگالی هسته با پهنای باند ثابت به تخمین بارش سالانه در سرتاسر ایران پرداختند و نتایج را با روش پارامتری مقایسه نمودند. آن‌ها نشان دادند تابع ناپارامتری برازش بهتری به داده‌های بارش حداکثر سالانه نسبت به روش پارامتری دارند (Haghighatjou., 2013).

هدف و نوآوری این تحقیق، تحلیل فراوانی سیلاب به روش تابع چگالی هسته با پهنای باند متغیر و پهنای باند ثابت بر اساس محاسبه پهنای باند بهینه بر مبنای روش مرتبط با پهنای باند[1] است. همچنین از روش‌های متداول نظیر قاعده سرانگشتی[2] و صحت‌سنجی مضاعف[3] با توجه به چهار نوع از توابع هسته نظیر نرمال[4]، دو وزنی[5]، سه وزنی[6]، اپینیچنیکوو[7] به منظور سنجش میزان دقت روش مرتبط با پهنای باند نسبت به رویکردهای مذکور استفاده گردید که در بخش بعدی معرفی می‌گردد.

## مواد و روش‌ها

### تعریف و روش برآورد تابع چگالی هسته[8]

گسترده‌ترین نظریه و رویکرد در محاسبه تابع چگالی احتمال به روش ناپارامتری استفاده از توابع هسته است. این روش شامل میانگین وزنی محرک توزیع فراوانی تجربی است. تابع چگالی احتمال در این رویکرد به صورت رابطه 1 تعریف می‌شود (Adamowski., 1987).

$$\hat{f}(x)=\frac{1}{2nh}\sum_{i=1}^{n} I(|x-x_i|\le h) \quad (1)$$

$x_i$ داده‌های مشاهداتی و I(.) تابع نشان‌گر است.

$$I(x-x_i)=\begin{cases}0 & x-x_i<0\\ 1 & x-x_i\ge 0\end{cases} \quad (2)$$

### انواع روش‌های تخمین چگالی هسته

تخمین چگالی هسته به دو روش بر حسب ثابت[9] یا متغیر[10] بودن پهنای باند تقسیم می‌شود (Bowman and Azzalini., 1997) .

### تخمین چگالی هسته ثابت[11] (چگالی هسته با پهنای باند ثابت)

این روش تابع چگالی احتمال را با فرض وجود یک تابع هسته در هر نقطه مشاهداتی و تجمیع آن‌ها به همراه مقدار مناسب پهنای باند به روش ناپارامتری محاسبه می‌کند (Adamowski., 1987).

یک تخمین چگالی احتمال هسته برای داده‌های مستقل $x_i$ مطابق با رابطه 3 است.

(Scott et al.,1981; Silverman., 1986).

$$f(x)=\frac{1}{n}\sum_{i=1}^{n}k_h(x-x_i)=\frac{1}{nh}\sum_{i=1}^{n}k_h\left(\frac{x-x_i}{h}\right) \quad (3)$$

n تعداد داده‌های مشاهداتی، h پهنای باند یا ضریب هموارسازی که برای همه داده‌های مشاهداتی ثابت است، $k_h(\ )$ تابع هسته که خود یک تابع چگالی احتمال است، x متغیر تصادفی، $x_i$ داده مشاهداتی.

با توجه به مقادیر تقریبی محاسبه شده اریبی[12] و واریانس، میانگین انتگرال مربع خطا تابع تخمین زده شده به روش ناپارامتری بر اساس رابطه 4 محاسبه می‌شود (Tsybakov., 2009).

$$MISE\approx\frac{1}{4}h^4k_2^2\beta(f)+\frac{1}{nh}j_2 \quad (4)$$

$$k_2=\int z^2k(z)dz,\ j_2=\int k(z)^2dz,\beta(f)=\int f''(x)^2dx$$

n تعداد داده‌های مشاهداتی، h پهنای باند، $k_2$ واریانس تابع هسته، f تابع چگالی احتمال، β(f) انتگرال مربع مشتق دوم تابع چگالی احتمال و $j_2$ انتگرال مربع تابع هسته می‌باشند.

انتخاب انواع تابع هسته به طور وسیعی توسط تعداد زیادی از محققین نظیر آداموسکی و رائو مورد مطالعه قرار گرفت. توابع مثلثی[13]، نرمال، دو وزنی، سه وزنی، اپینیچنیکوو و غیره تعدادی از توابع هسته مطرح هستند. اثبات نظری نشان می‌دهد که نوع تابع هسته انتخاب شده نقش تعیین کننده‌ای در عملکرد روش ندارد. اما محاسبه پهنای باند بسیار مهم است (Adamowski., 2000 ;Tsybakov., 2009)

ضابطه چهار تابع هسته مورد استفاده در این مقاله در جدول 1 درج شده است.

1- Plug in bandwidth
2- Rule of Thumb
3- Cross Validation
4- Gaussian
5- Biweight
6- Triweight
7- Epanechnikov
8- Kernel Density Function
9- Fixed Model
10- Variable Model
11 - Fixed Kernel Density
12- Bias
13- Triangle

جدول 1- ضابطه توابع هسته مورد استفاده
(Shabri, 2002; Tsybako, 2009)

| Range | k(u) | Kernel |
|---|---|---|
| $\|u\| \le 1$ | $\frac{3}{4}(1-u^2)$ | اپینیچنیکوو |
| $\|u\| \le 1$ | $\frac{15}{16}(1-u^2)^2$ | دو وزنی |
| $\|u\| \le 1$ | $\frac{35}{32}(1-u^2)^3$ | سه وزنی |
| به ازای کلیه مقادیر u | $\frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}u^2)$ | نرمال |

## انتخاب پهنای باند

انتخاب پهنای باند یک مرحله مهم در روش برآورد هسته است. یک تغییر در پهنای باند ممکن است به طور چشم‌گیری باعث تغییر در شکل تابع چگالی احتمال شود. به ازای مقادیر بسیار کوچک پهنای باند تابع چگالی احتمال بسیار ناهموار با اوج‌های تیز و فواصل بسیار کم خواهد شد و عبارت دوم (واریانس) در رابطه 4 بزرگ می‌شود و به ازای مقادیر بزرگ پهنای باند تابع چگالی احتمال بسیار هموار می‌شود و عبارت اول (اریبی) در همان رابطه افزایش پیدا می‌کند. در نتیجه انتخاب مقدار بهینه برای پهنای باند لازم و ضروری است (Simonovic., 2007 ; Karmakar and Tsybakov., 2009).

چندین روش تخمین پهنای باند بر اساس حداقل نمودن تخمین تابع میانگین انتگرال مربع خطا وجود دارد. با مشتق‌گیری از رابطه 4 بر حسب پهنای باند (h) می‌توان مقدار بهینه پهنای باند را محاسبه کرد (Wasserman., 2006).

$$h_{opt} = (\frac{1}{n}\frac{j_2}{\beta(f)k_2^2})^{\frac{1}{5}} = (\frac{1}{n}\frac{\gamma(k)}{\beta(f)})^{\frac{1}{5}} \quad (5)$$

پارامترهای موجود در رابطه 5 همان پارامترهای رابطه 4 می‌باشد و $\gamma(k)$ نسبت پارامتر $j_2$ به $\beta(f)$ است.

در این تحقیق محاسبه پهنای باند بهینه با توجه به سه رویکرد تقسیم بندی می‌شود. یک رویکرد بر اساس قاعده سرانگشتی سیلورمن است. دومین رویکرد بر اساس صحت‌سنجی مضاعف و سومین روش، روش مرتبط با پهنای باند است.

## قاعده سرانگشتی

استفاده از رابطه بهینه‌سازی جهت محاسبه پهنای باند به دلیل مجهول بودن تابع صحیح و مشتق دوم آن سخت است، بنابراین با استفاده از قاعده سرانگشتی می‌توان بر مشکلات روش بهینه‌سازی غلبه نمود.

روش قاعده سرانگشتی در محاسبه پهنای باند نتایج معقولی را برای تمامی توزیع‌هایی که نسبتا متقارن و تک اوجه هستند، دارد. در عمل مشکل اساسی این قاعده در محاسبه پهنای باند حساسیتش به داده‌های پرت است؛ زیرا یک داده پرت می‌تواند باعث تخمین بسیار بزرگی از پهنای باند شود. با توجه به استفاده از چارک پایین و بالا داده‌ها می‌توان تخمین مناسب‌تر و پایداری را در محاسبه پهنای باند به دست آورد (Lee., 2009). بنابراین محاسبه پهنای باند با توجه نوع تابع هسته از رابطه 8 محاسبه می‌شود (Silverman., 1986).

همچنین از دیگر مشکلات این روش این است که اگر توزیع داده‌ها نرمال نباشد محاسبه پهنای باند در این حالت از دقت کافی برخوردار نخواهد بود (Kim and Heo., 2002).

$$h_{opt,f} = C_f \frac{\hat{\sigma}}{n^{1/5}} \quad (6)$$

$$h_{opt,s} = C_s \frac{IQR}{n^{1/5}} \quad (7)$$

$$h_{opt} = \min(h_{opt.f}, h_{opt,s}) \quad (8)$$

$h_{opt,f}$: اولین پهنای باند با توجه به قاعده سرانگشتی،

$h_{opt,s}$: دومین پهنای باند با توجه به قاعده سرانگشتی،

$C_f$: ضریب اولین پهنای باند با توجه به قاعده سرانگشتی،

$C_s$: ضریب دومین پهنای باند با توجه به قاعده سرانگشتی،

ضرایب $C_S$ و $C_f$ بر اساس مقادیر مندرج در جدول 2 محاسبه می شود.

n: تعداد داده‌ها، $\sigma$ انحراف معیار داده‌ها، IQR: تفاضل چارک سوم و اول داده‌ها.

## روش صحت‌سنجی مضاعف

روش صحت‌سنجی مضاعف یکی از بهترین روش‌های محاسبه پهنای باند بهینه و همچنین یک روش عمومی و انعطاف پذیر است. دو روش صحت‌سنجی مضاعف از قبیل صحت‌سنجی مضاعف حداکثر درست نمایی و روش صحت سنجی مضاعف حداقل مربعات وجود دارد (Tsybako., 2009; Wesserman., 2006). در این تحقیق از

روش صحت‌سنجی مضاعف حداقل مربعات استفاده شده است.

**جدول2-مقادیر ضرایب $C_S$ و $C_f$ بر اساس قاعده سرانگشتی**
**(Henderson and Parmeter., 2012; Lee., 2009)**

| تابع هسته | $C_S$ | $C_f$ |
|---|---|---|
| اپینیچنیکوو | 1/75 | 2/35 |
| دو وزنی | 2/07 | 2/78 |
| سه وزنی | 2/35 | 3/16 |
| نرمال | 0/79 | 1/06 |

## صحت سنجی مضاعف حداقل مربعات[1]

در این روش پهنای باند بـا حـداقل نمـودن انتگـرال مربـع خطـا محاسبه می‌شود (Rudemo.,1982; Stone., 1984; Bowman., 1984).

$$ISE(\hat{f}(x)) = \int (\hat{f} - f)^2 (x)dx = \int \hat{f}^2(x)dx - 2\int \{\hat{f} * f\}(x)dx + \int f^2(x)dx \quad (9)$$

ISE انتگرال مربع خطا، f مقـدار صـحیح و واقعـی تـابع چگـالی احتمال و $\hat{f}$ مقدار تخمین زده شده تابع چگالی احتمال

در رابطه 9 $\int f^2(x)dx$ مستقل از پهنای باند است. در نهایت بـا به کارگیری جمع به جای انتگـرال در رابطـه 9 تـابع صـحت سـنجی مضاعف مطابق با رابطه 10 باز نویسی می‌شود (Bowman., 1984).

$$CV(h) = \frac{1}{n^2 h}\sum_{i=1}^{n}\sum_{j=1}^{n}\bar{k}\left(\frac{X_i - X_j}{h}\right) - \frac{2}{n(n-1)}\sum_{i=1}^{n}\sum_{i\neq j}^{n}k_h(X_i - X_j) \quad (10)$$

$k_h$ تابع هسته، n تعداد داده‌های مشـاهداتی، h پهنـای بانـد، X داده مشاهداتی و $\bar{k}$ از رابطه 11 محاسبه می‌شود

$$\bar{k}(x) = \int k(u)k(x-u)du \quad (11)$$

پهنای باند از رابطـه 12 بـا توجـه بـه حـداقل نمـودن رابطـه 10 محاسبه می‌شود.

$$h_{CV} = \arg\min CV(h) \quad (12)$$

برای دریافت نتایج دقیق از رابطه 12 نیاز است که تعداد داده‌های مشاهداتی زیـادتری در دسـترس باشـد روش فـوق بـرای داده‌هـای مشاهداتی با طول کم جواب دقیقی را نخواهد داد. بنابراین بـومن نیـز رابطه‌ای را به منظور اصلاح تابع صحت سنجی مضاعف ارایه کرد که پهنـای بانـد بهینـه بـا حـداقل کـردن ایـن تـابع محاسـبه مـی‌شـود (Bowman et al., 1998).

$$CV(h) = \frac{1}{n}\sum_{i=1}^{n}\int_{-\infty}^{+\infty}(I(x - x_i) - F_{-i}(x))^2 dx \quad (13)$$

$$\hat{F}_{-i}(x) = \frac{1}{n-1}\sum_{j\neq i} H\left(\frac{x - x_i}{h}\right) \quad (14)$$

$$H(x) = \int_{-\infty}^{x} k(t)dt \quad (15)$$

## روش مرتبط با پهنای باند

## روش مرتبط با پهنای باند آلتمن و لیگر[2]

این روش بر اساس لحاظ نمودن مربع خطـا بـین تـابع صـحیح و تابع تخمین زده شده نظیر انتگرال مربع خطا اسـت. متـدولوژی ارایـه شده توسط آلتمن و لیگر برای محاسبه تابع ناپارامتری مطابق با رابطه 16 است (Altman and Leger., 1995).

$$ISE(\hat{F}(h)) = \int (\hat{F}_h(x) - F(x))^2 (x)dx \quad (16)$$

$$(x) = \int_{-\infty}^{x} \hat{f}_h(t)dt \quad (17)$$

$\hat{F}_h(x)$ و $\hat{f}_h(t)$ به ترتیـب تـابع توزیـع (تـابع تجمعـی) و تـابع چگالی تخمین زده شده و F(x) تابع توزیع واقعی داده‌ها.

در نتیجه تخمین پهنای باند بهینه با اسـتفاده از روش مـرتبط بـا پهنای باند به روش آلتمن و لیگر مطابق با رابطه 18 است (Altman and Leger., 1995).

$$h_{AISE}(\hat{F}_h) = \left(\frac{0.25V_2}{B_3}\right)^{1/3} n^{-1/3} \quad (18)$$

n: تعداد داده‌ها و تخمین مقادیر پارامترهای $V_2$ و $B_3$ بـه ترتیـب مطابق با رابطه 19 و20 است (Altman and Leger., 1995).

$$V_2 = \rho(k)\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j=1, j\neq 1}^{n}\frac{1}{\alpha}k\left(\frac{x_i - x_j}{\alpha}\right) \quad (19)$$

$$B_3 = 0.25\hat{D}_3(F)(\mu_2(k))^2 \quad (20)$$

در رابطه 19 و20، $x_i$ $x_j$ $x_k$ داده‌های مشـاهداتی، k تـابع چگـالی هسته، $\mu_2(k)$ که همان واریانس تابع هسته که نحوه محاسـبه آن در رابطـه 21 در ذیـل نشـان داده شـده اسـت، همچنـین $\rho(k)$ و

1- Least Square Cross Validation

2- Altman and Leger

$\widehat{D}_3(F)$ پارامترهایی هستند که با استفاده از رابطه 22 و 23 تخمین زده می‌شوند.

(21)
$$\mu_2(k)=\int_{-\infty}^{+\infty}x^2k(x)dx$$

(22)
$$\rho(k)=2\int_{-\infty}^{+\infty}xk(x)H(x)dx$$

(23)
$$\widehat{D}_3(F)=\frac{1}{n^3\alpha^4}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}k'(\frac{x_i-x_j}{\alpha})k'(\frac{x_i-x_k}{\alpha})$$

در رابطه‌های فوق، H تابع توزیع چگالی هسته که از رابطه 15 محاسبه می‌شود، $k'$ مشتق تابع هسته است و α از رابطه 24 برای تابع هسته اپینچنیکوو برآورد می‌شود (Altman and Leger., 1995):

(24)
$$\alpha = n^{-0.3}\widehat{\sigma}(x_i)=n^{-0.3}\min\left\{s',\frac{Q_3-Q_1}{1.349}\right\}$$

$Q_3$ چارک سوم، $Q_1$ چارک اول و $s'$ انحراف معیار داده‌ها.

## روش مرتبط با پهنای باند پلانسکی و بیکر[1]

پلانسکی و بیکر یک روش تکراری را به منظور محاسبه پهنای باند بهینه ارایه کردند که در ذیل به طور خلاصه شرح داده شده است (Polansky and Baker., 2000):

محاسبه $\widehat{\omega}_{2j}(\widehat{g}_{2j})$ با توجه به رابطه 25:

(25)
$$\widehat{\omega}_r=\frac{(-1)^{\frac{r}{2}}r!}{(2\widehat{\sigma}(x_i))^{r+1}(\frac{r}{2}!)\pi^{\frac{1}{2}}}$$

r مرتبه مشتق تابع چگالی احتمال، $\widehat{\sigma}$ انحراف معیار داده‌ها است. برای محاسبه مقدار فاکتوریل در رابطه 25 از تابع گاما استفاده می‌شود.

ابتدا از j=۱ تا j=b مقدار پارامتر g از رابطه 26 سپس مقدار $\widehat{\omega}_{2j}(\widehat{g}_{2j})$ با توجه به رابطه 27 محاسبه و در نهایت پهنای باند بهینه از رابطه 28 تخمین زده می‌شود (Polansky and Baker., 2000).

(26)
$$\widehat{g}_{2j}=(\frac{2L^{(2j)}(0)}{-n\mu_2(L)\widehat{\omega}_{2j+2}})^{1/(2j+3)}$$

(27)
$$\widehat{\omega}_{2j+2}=\begin{cases}\widehat{\omega}_{2b+2} & if \quad j=b\\ \widehat{\omega}_{2j+2}(\widehat{g}_{2j+2}) & if \quad j<b\end{cases}$$

(28)
$$h_{PB}=(\frac{\rho(k)}{-n\mu_2^2(k)\widehat{\omega}_{2j}(\widehat{g}_{2j})})^{1/3}$$

1- Polansky and Baker

در رابطه 26 تا 28: g یک پارامتر مثبت است که به طور معمول به عنوان یک پارامتر هموارکننده یا پهنای باند می‌تواند لحاظ شود. L یک تابع هسته می‌باشد که لزوماً برابر با با تابع هسته k نیست. b یک عدد صحیح مثبت که نشان‌دهنده تعداد مراحل محاسبات در روش مذکور است. H تابع توزیع (تابع تجمعی) چگالی هسته، $\rho(k)$ و $\mu_2(k)$ پارامترهایی هستند که با استفاده از رابطه 21 و 22 محاسبه می‌شوند و در نهایت پارامتر $h_{PB}$ مقدار پهنای باند بهینه محاسبه شده به روش پلانسکی و بیکر است.

روش برآورد هسته با پهنای باند ثابت از لحاظ محاسباتی ساده است اما صد در صد روش مطلوبی نیست در نتیجه باعث کاهش اثر تابع هسته‌ای می‌شود، زیرا:

یک نمونه مشاهداتی با تعداد داده‌های زیاد می‌تواند همیشه تخمین بهتری از برآورد چگالی احتمال هسته ارایه دهد بنابراین چنانچه اگر داده‌های مشاهداتی کم‌تر باشد روش برآورد هسته با پهنای باند ثابت دارای بازده خوبی برای تخمین نخواهد بود (Karmakar and Simonovic., 2007).

انتخاب یک مقدار ثابت برای پهنای باند برای توزیع یک متغیره در توزیع‌هایی که دارای چولگی هستند غیر منطقی است (Karmakar and Simonovic., 2007).

این روش در برون‌یابی، مقدار تابع چگالی احتمال بسیار کوچکی را برای بزرگ‌ترین داده مشاهده شده در نمونه سبب می‌شود. برون-یابی بر اساس نوع تابع چگالی هسته و پهنای باند است، بنابراین تعداد داده‌های کمی در فاصله پهنای باند مورد نظر قرار می‌گیرد که روی برون‌یابی تاثیر می‌گذارد (Karmakar and Simonovic., 2007).

روش برآورد چگالی هسته برای توابع چگالی احتمال ناهموار مانند توابع چگالی یکنواخت روش مطلوبی نیست (Markovich, 2008).

## تخمین چگالی هسته متغیر[2] (چگالی هسته با پهنای باند متغیر)

یکی از ویژگی‌های داده‌ها ممکن است وجود برجستگی‌ها[3] در انتهای تابع چگالی احتمال و یا ابتدای آن به دلیل تعداد داده‌های کم مشاهداتی در آن بازه باشد که در این صورت داده‌ها در آن بازه تنک[4] می‌باشند. ممکن است مناسب‌ترین روش استفاده از پهنای باند بزرگ برای از بین بردن این برجستگی‌ها باشد. از طرف دیگر ممکن است در مقادیر کم داده‌ها مثل یک خوشه به یکدیگر نزدیک باشند که در این صورت استفاده از پهنای باند کوچک مناسب است. با توجه به این امر میزان هموارسازی در ناحیه‌های متفاوت را بایستی با درنظرگرفتن عرض باندهای متفاوت تعدیل نمود. این ایده محرکی برای استفاده از پهنای باند متفاوت در هر نقطه مشاهداتی برای محاسبه تابع چگالی

2- Variable Kernel Density
3- Bumps
4- Sparse

احتمـال بـه روش ناپـارامتری اسـت ( Bowman and Azzalini., 1997).

تابع چگالی احتمال بر مبنای چگالی هسته با پهنـای بانـد متغیـر توسط رابطه 29 تعریف می‌شود (Bowman and Azzalini., 1997، Silverman., 1986 ).

$$f(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_i}k_h\left(\frac{x - X_i}{h_i}\right) \quad (29)$$

$$h_i = hd_k(x_i) \quad (30)$$

$h_i$ پهنای باند متغیر، h پهنای باند بهینه و $d_k(x_i)$ فاصله از $x_i$ تا نزدیک ترین همسایه میان Kامین داده و $k_h(\,)$ تابع هسته.

همچنین پیشنهاد مفیدی در جهت معرفی عنصر متغیر به عنـوان یک اصلاح برای تمامی مقادیر ثابت هموار مطابق با رابطـه 31 ارایـه داده شد (Bowman and Azzalini., 1997).

$$h_i = hd_k(x_i)/\bar{d} \quad (31)$$

$\bar{d}$ : میانگین هندسی $d_k(y_i)$

اصلی‌ترین تاثیر ویژگی‌های پهنای باند متغیر برآورد تـابع چگـالی احتمال با اوج‌های تیز نسبت به پهنای باند ثابـت اسـت ( Bowman and Azzalini, 1997).

اصلی‌ترین تاثیر ویژگی‌های پهنای باند متغیر برآورد تـابع چگـالی احتمال با اوج‌های تیز نسبت به پهنای باند ثابـت اسـت ( Bowman and Azzalini, 1997).

در هر دو مورد تخمین چگالی متغیر و تخمین چگالی ثابـت بایـد رویکرد بهینه برای تابع هسته و پهنای باند در نظرگرفته شود.

### محاسـبه دبـی بـا دوره بازگشــت‌هـای مـورد نظـر بـه روش ناپارامتری

مقادیر دبی با دوره بازگشت‌های مورد نظر به روش ناپارامتری بـا حل عددی معادله 32 محاسبه می‌شود.

$$x_T = F^{-1}\left(1 - \frac{1}{T}\right) \quad (32)$$

$x_T$ دبی با دوره بازگشت (T) مورد نظر، $F^{-1}$ معکـوس تـابع توزیـع چگالی. تابع توزیع چگالی نیز از رابطه 33 محاسبه می‌شود:

$$F(x) = \int_{-\infty}^{x} f(x) \quad (33)$$

### زبان برنامه‌نویسی R

برای اجـرای محاسـبات و عملیـات مربـوط بـه تحلیـل فراوانـی ناپارامتری از بسته‌های[1] sm و kerdiest در محیط برنامه نویسی R و همچنین برنامه‌نویسی در این محیط، استفاده شد.

## نتایج و بحث

### منطقه مورد مطالعه

تحلیل فراوانی به روش ناپارامتری بر مبنای روش‌های ارایه شـده در فوق و همچنین با رویکرد پارامتری برای دبی‌های حداکثر لحظـه-ای سالانه رودخانه دز انجام شد. این رودخانه که از زردکـوه بختیـاری سرچشمه می‌گیرد پس از گذشتن از شمال خوزستان و شهر دزفول در منطقه‌ای به نام گرگر در شرق شوشتر به رود کارون می‌پیوندد.

طول آمار مورد استفاده در این تحقیق 38 سال آماری است که در این فاصله زمانی مقادیر حداکثر لحظه‌ای دبی سالانه به عنوان سـری مورد نظر انتخاب گردید.

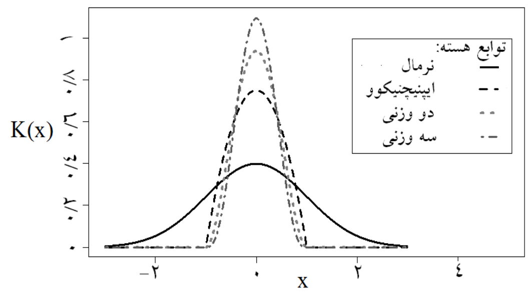
### محاسبات پارامتری با توجه دبی‌های حداکثر لحظه‌ای سـالانه رودخانه دز

با توجه به آزمون نکویی برازش کولموگروف-اسمیرنوف، مطـابق با جدول3، توزیع لوگ پیرسون نوع 3 مناسـب‌تـرین توزیـع پـارامتری برای دبی‌های حداکثر لحظه‌ای سالانه رودخانه دز است.

### محاسبات ناپارامتری

### محاسبات ناپارامتری بر اساس روش برآورد چگالی هســته بـا پهنای باند ثابت

در این تحقیق از چهار تابع هسته اپینچنیکوو، دو وزنی، سه وزنی و نرمال برای محاسبه پهنای باند بهینه در روش برآورد چگالی هسـته (با توجه به دبی‌های حداکثر لحظه‌ای سالانه رودخانه دز) استفاده شده است تا به طور تجربی در این مقاله تاثیر نوع تابع هسته در شکل تابع چگالی احتمال بررسی شود. همچنین مقادیر سیلاب تا دوره بازگشـت 1000 سال نیز با توجه به توابـع چگـالی نـاپرامتری بـه دسـت آمـده تخمین زده شد و نمودارهای مربوطه در شکل‌هـای 2 و 3 نشـان داده شده است.




**شکل 1- نمودار چهار تابع هسته مورد استفاده**

1- Package

جدول 3- مقادیر ریشه دوم مربع خطا و ضریب همبستگی با توجه به سه توزیع پارامتری

| | توزیع حدی تعمیم یافته | لوگ‌پیرسون نوع 3 | لوگ‌نرمال دو پارامتره |
|---|---|---|---|
| RMSE | 0/3206 | 0/1752 | 0/20164 |
| $r(Q_{cal},Q_{obs})$ | 0/9393 | 0/9884 | 0/9806 |

به منظور محاسبه پهنای باند به روش‌های صحت‌سنجی مضاعف حداقل مربعات، روش مرتبط با پهنای باند آلتمن - لیگر و روش مرتبط با پهنای باند پلانسکی - بیکر از بسته نرم‌افزاری kerdiest استفاده شده است (Rio and Perez., 2012).

مقادیر محاسبه شده پهنای باند بر اساس روش‌های مذکور در جدول 4 درج شده است.

پارامترهای مورد استفاده در جدول 4 به شرح ذیل هستند:

$h_{ROT}$: پهنای باند محاسبه شده بر اساس روش قاعده سر انگشتی.

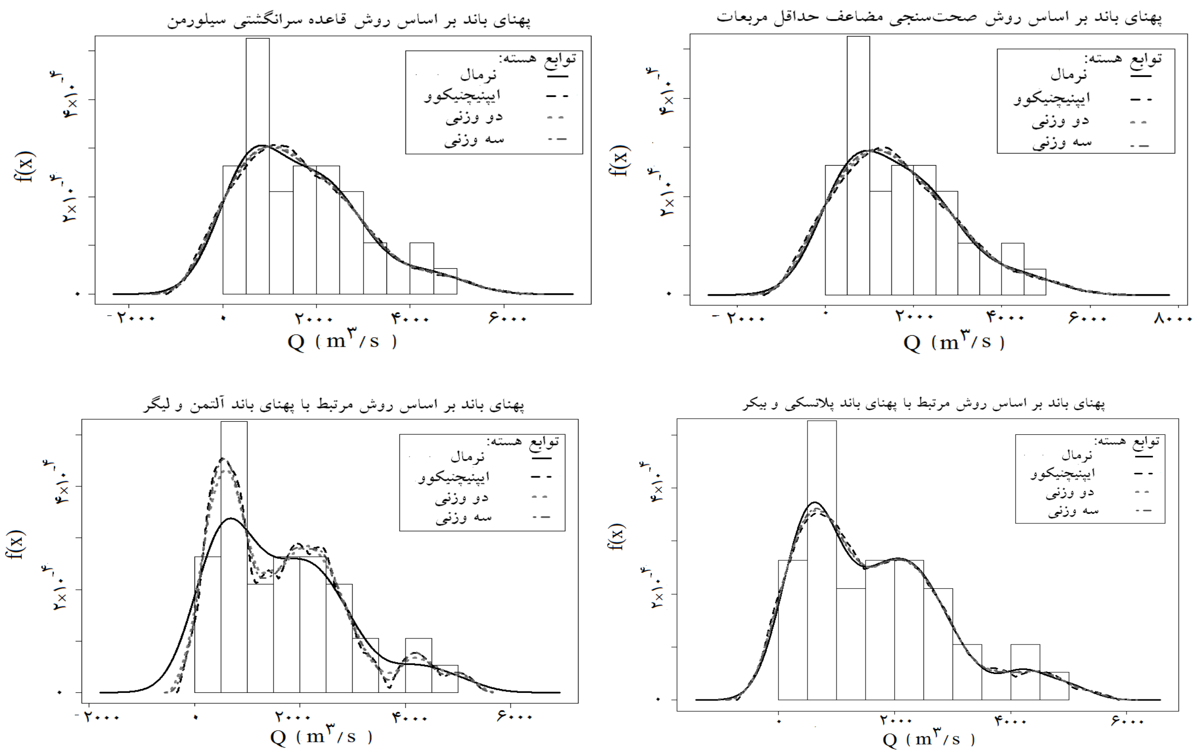
$h_{CV}$: پهنای باند محاسبه شده بر اساس روش صحت‌سنجی مضاعف.

$h_{PB}$: پهنای باند محاسبه شده بر اساس روش مرتبط با پهنای باند پلانسکی و بیکر.

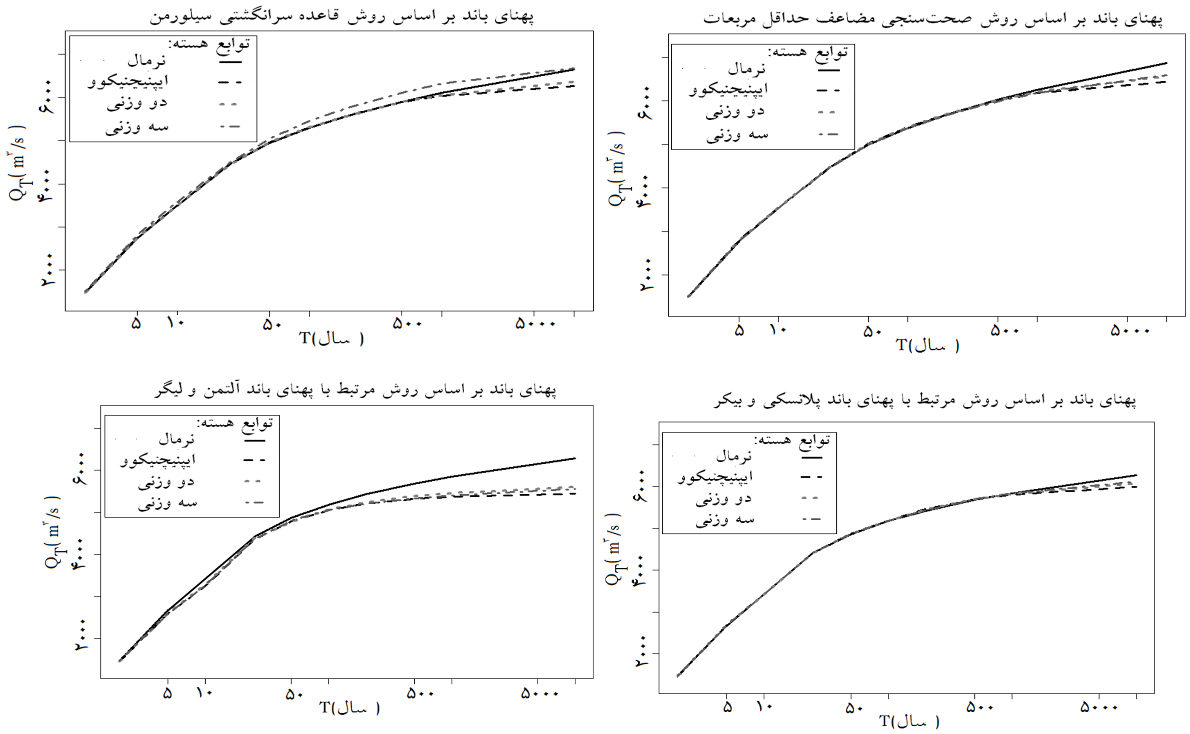
$h_{AL}$: پهنای باند محاسبه شده بر اساس روش مرتبط با پهنای باند آلتمن و لیگر.

جدول4- مقادیر محاسبه شده پهنای باند بهینه با توجه به روش‌های قاعده سرانگشتی، روش اصلاح شده صحت سنجی مضاعف حداقل مربعات، روش مرتبط با پهنای باند آلتمن - لیگر و روش مرتبط با پهنای باند پلانسکی- بیکر با توجه به چهار نوع تابع هسته

| تابع چگالی | $h_{ROT}$ | | $h_{ROT}$= min($h_{opt,f}$ , $h_{opt,s}$) | $h_{CV}$ | $h_{AL}$ | $h_{PB}$ | | |
|---|---|---|---|---|---|---|---|---|
| | $h_{opt,f}$ | $h_{opt,s}$ | | | | n_stage=1 | n_stage=3 | n_stage=3 |
| نرمال | 624/24 | 679/92 | 624/24 | 704/03 | 490/35 | 485/16 | 441/69 | 400/08 |
| اپینچنیکوو | 1383/93 | 1506/15 | 1383/93 | 1578/3 | 510/18 | 1091/73 | 993/90 | 900/28 |
| دو وزنی | 1637/16 | 1781/56 | 1637/16 | 1869/72 | 730/68 | 1290/04 | 1174/45 | 1063/82 |
| سه وزنی | 1860/94 | 2022/54 | 1860/94 | 2064/01 | 737/27 | 1461/4 | 1330/45 | 1205/14 |




شکل 2- توابع چگالی ناپارامتری با پهنای باند ثابت براساس چهار نوع تابع هسته




**شکل 3- مقادیر دبی با دوره بازگشت‌های متفاوت با توجه به توابع چگالی احتمال ناپارامتری با پهنای باند ثابت براساس چهار نوع تابع هسته**

**جدول5- ریشه دوم میانگین مربع خطا و ضریب همبستگی برای روش‌های محاسبه پهنای باند ثابت با توجه به دبی‌های حداکثر لحظه‌ای سالانه رودخانه دز**

| | | | نرمال | سه وزنی | دو وزنی | اپینچنیکوو |
|---|---|---|---|---|---|---|
| قاعده سرانگشتی | | RMSE | 0/781 | 0/795 | 0/797 | 0/806 |
| | | $r(Q_{cal},Q_{obs})$ | 0/9880 | 0/9873 | 0/9871 | 0/9867 |
| | | | نرمال | سه وزنی | دو وزنی | اپینچنیکوو |
| صحت‌سنجی مضاعف حداقل مربعات | | RMSE | 0/933 | 0/935 | 0/967 | 0/974 |
| | | $r(Q_{cal},Q_{obs})$ | 0/9880 | 0/9854 | 0/9846 | 0/9841 |
| | | | نرمال | سه وزنی | دو وزنی | اپینچنیکوو |
| روش مرتبط با پهنای باند آلتمن و لیگر | | RMSE | 0/540 | 0/183 | 0/219 | 0/170 |
| | | $r(Q_{cal},Q_{obs})$ | 0/9914 | 0/9963 | 0/9957 | 0/9964 |
| | | | نرمال | سه وزنی | دو وزنی | اپینچنیکوو |
| روش مرتبط با پهنای باند پلانسکی و بیکر | b=2 | RMSE | 0/531 | 0/551 | 0/556 | 0/562 |
| | | $r(Q_{cal},Q_{obs})$ | 0/9916 | 0/9910 | 0/9909 | 0/9905 |
| | b=3 | RMSE | 0/457 | 0/475 | 0/479 | 0/484 |
| | | $r(Q_{cal},Q_{obs})$ | 0/9927 | 0/9922 | 0/9920 | 0/9918 |
| | b=4 | RMSE | 0/390 | 0/405 | 0/408 | 0/411 |
| | | $r(Q_{cal},Q_{obs})$ | 0/9936 | 0/9932 | 0/9931 | 0/9930 |

به منظور مقایسه روش‌های محاسبه پهنای باند ثابت با توجه به چهار نوع تابع هسته مورد استفاده از ریشه دوم میانگین مربع خطا و ضریب

همبستگی میان داده‌های مشاهداتی و محاسباتی استفاده شده است این مقادیر در جدول 5 درج شده است.

**محاسبات ناپارامتری بر اساس روش برآورد چگالی هسته با پهنای باند متغیر**

با توجه به اینکه در رویکرد پهنای باند ثابت روش مرتبط با پهنای باند آلتمن و لیگر کم‌ترین خطا را بر اساس مقادیر مندرج در جدول 5 در محاسبه پهنای باند برای داده‌های مورد نظر دارد، بنابراین در رویکرد پهنای باند متغیر نیز از این روش استفاده شده است.

در این روش با توجه به رابطه 32 فاصله از $x_i$ تا نزدیک‌ترین همسایه میان kامین داده توسط بسته نرم‌افزاری

sm (Bowman and Prvan., 1998) محاسبه شده است و با کدنویسی در محیط برنامه‌نویسی R مقادیر پهنای باند و درنهایت تابع چگالی احتمال ناپارامتری در هر نقطه دلخواه به ترتیب مطابق با شکل‌های 4 و 5 برای چهار تابع هسته مورد نظر تخمین زده شد.

مقادیر خطا و ضرایب همبستگی با توجه به مقدار K امین داده (در $d_K(x_i)$) برای چهار تابع هسته مطابق با شکل 6 و 7 محاسبه شده است.

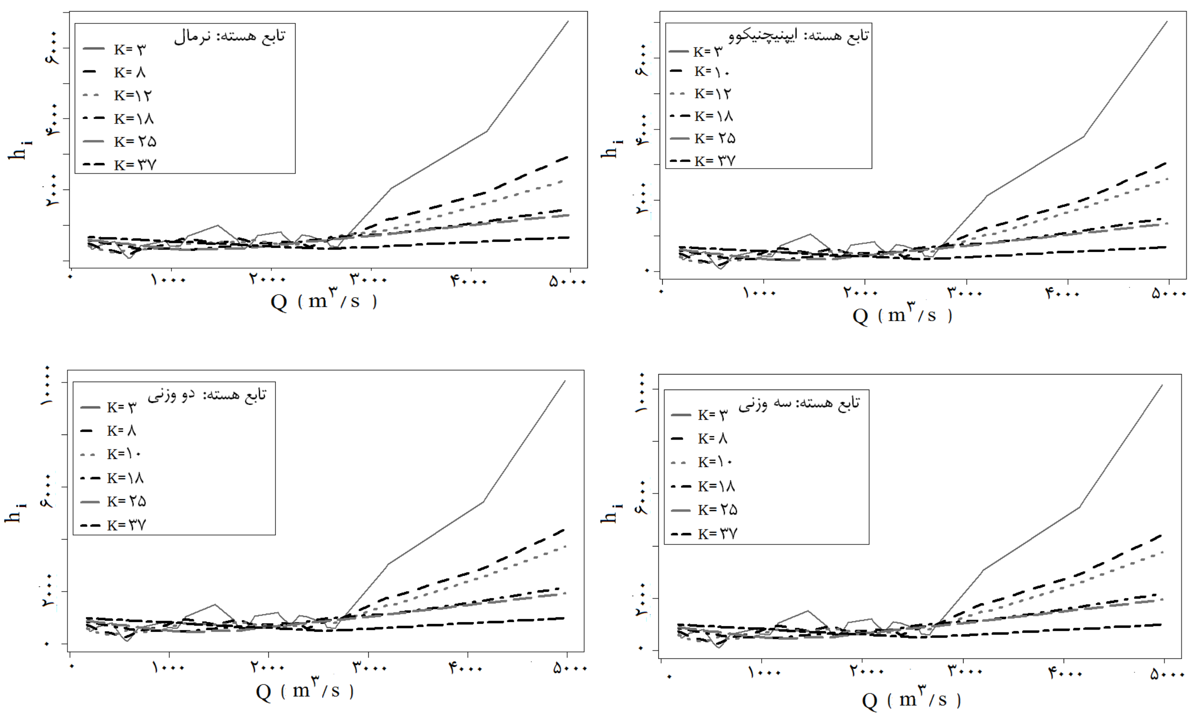
به منظور مقایسه روش‌های مورد نظر، نمودارهای بهترین توابع چگالی ناپارامتری بر اساس پهنای باند ثابت، متغیر و بهترین توزیع پارامتری و همچنین نمودار چندک تجربی و نظری مطابق با شکل‌های 8، 9 و 10 رسم شده است.

**تحلیل فراوانی ناپارامتری**

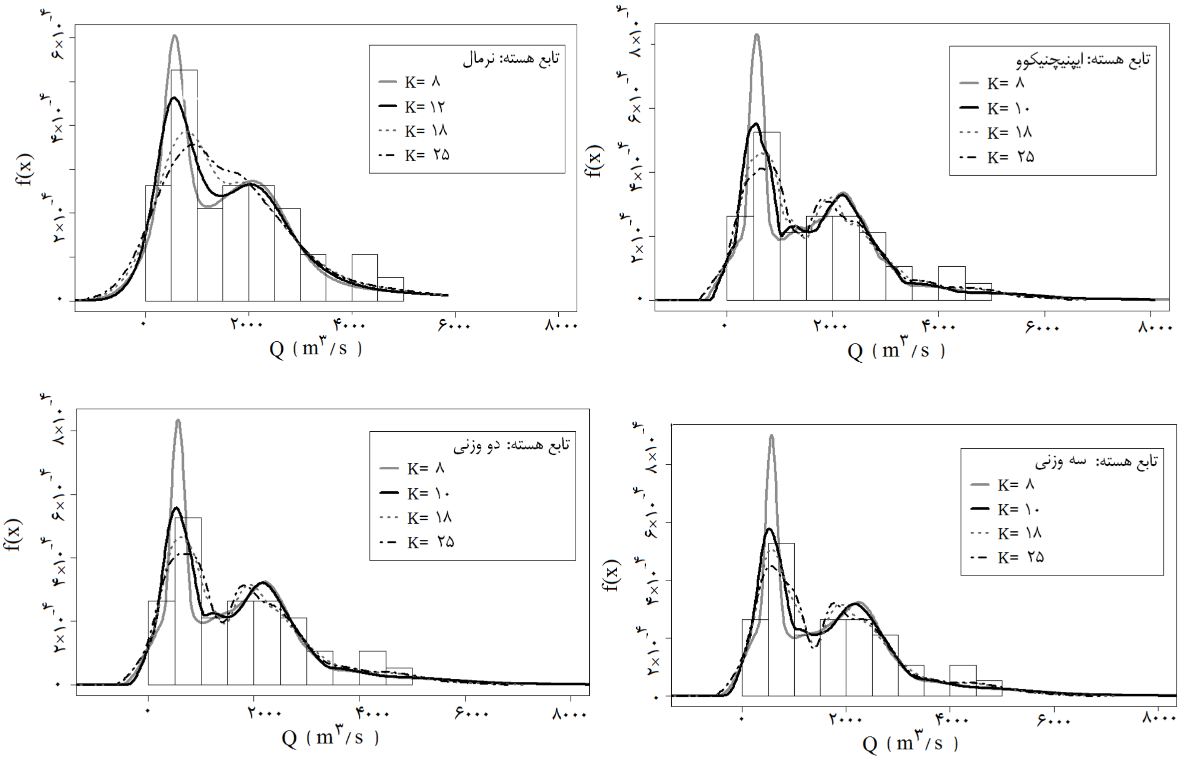
به منظور تحلیل فراوانی به روش‌های ناپارامتری مذکور، مقادیر سیلاب با دوره بازگشت‌های مورد نظر با حل عددی معادله 32 محاسبه و این مقادیر در مقابل دوره بازگشت‌های مورد نظر مطابق با شکل 11 رسم شده است.

## نتیجه‌گیری

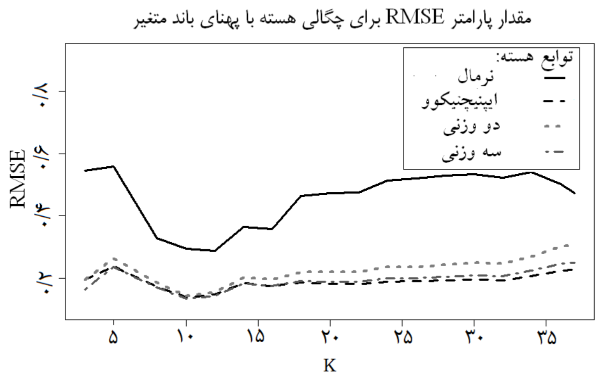
با توجه به مقادیر ریشه دوم مربع خطا مندرج در جدول 5 روش مرتبط با پهنای باند دقیق‌ترین روش محاسبه پهنای باند بهینه نسبت به دو روش قاعده سرانگشتی و صحت‌سنجی مضاعف است. که در این میان روش مرتبط با پهنای باند آلتمن و لیگر از روش مرتبط با پهنای باند پلانسکی و بیکر دقیق‌تر است.



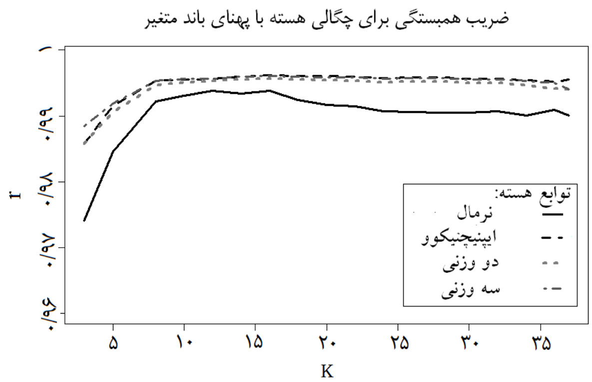

**شکل4- نمودار پهنای باند متغیر – دبی با توجه به چهار نوع تابع هسته**



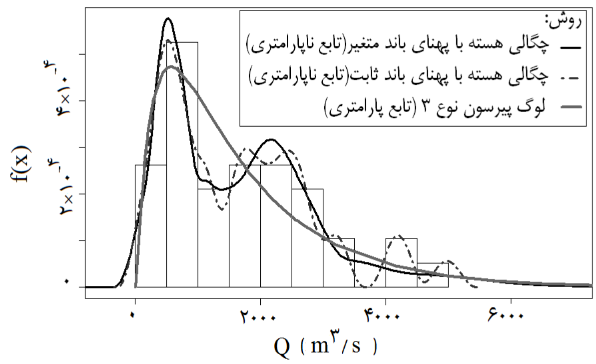

**شکل 5-** توابع چگالی احتمال ناپارامتری بر اساس محاسبات پهنای باند متغیر و با توجه به چهارتابع هسته



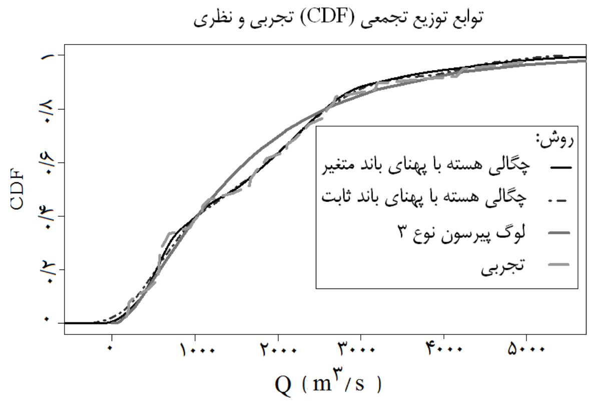

**شکل 6-** خطاهای محاسبه شده بر اساس مقادیر k امین داده (در $d_K(x_i)$) با توجه به چهار تابع هسته مورداستفاده



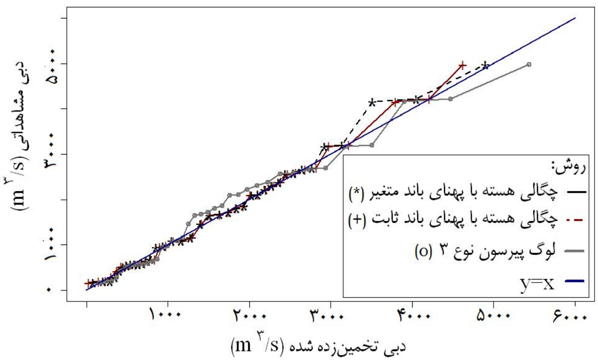

**شکل 7-**ضرایب همبستگی محاسبه شده بر اساس مقادیر k امین داده (در $d_K(x_i)$) با توجه به چهار تابع هسته مورداستفاده



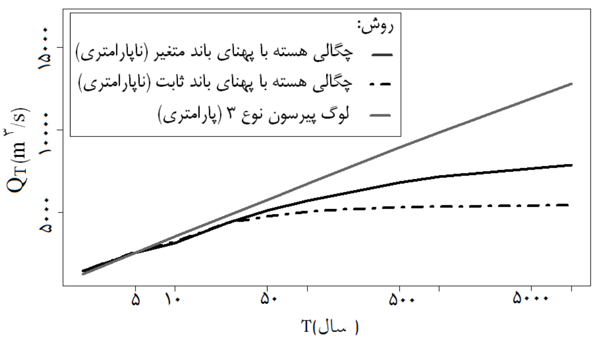

**شکل8- نمودار تابع چگالی احتمال پارامتری و ناپارامتری با پهنای باند ثابت و متغیر**




**شکل9- نمودار توابع توزیع نظری و تجربی بر اساس محاسبات پهنای باند ثابت و متغیر**




**شکل10- نمودار چندک تجربی و نظری**




**شکل 11- نمودار دبی-دوره بازگشت برای توزیع‌های پارامتری و ناپارامتری با پهنای باند متغیر و ثابت**

همچنین مطابق با جدول 5 استفاده از روش مرتبط با پهنای باند آلتمن و لیگر با تابع هسته اپینیچنیکوو به منظور محاسبه تابع چگالی احتمال ناپارامتری در محاسبه چندک‌های سیلاب این رودخانه بهترین دقت را دارد.

بر اساس نمودارهای نشان داده شده در شکل 2 استفاده از سایر توابع هسته (به استثنای تابع هسته نرمال در روش مرتبط با پهنای باند آلتمن و لیگر) تفاوت عمده‌ای در تخمین توابع ناپارامتری و در نهایت تخمین چندک‌های سیلاب (نشان داده شده در شکل 3) ایجاد نخواهد کرد.

نتایج حاصل از محاسبات ریشه دوم مربع خطا در شکل 6 حاکی از آن است که تابع هسته سه وزنی در محاسبه تابع ناپارامتری با رویکرد پهنای باند متغیر با مقدار K=10 (Kامین داده) دارای دقت بیش‌تری نسبت به سایر توابع هسته مورد استفاده است.

نمودارهای نشان داده شده در شکل 5 با مقادیر بهینه K=10 برای سه تابع هسته اپینیچنیکوو، دو وزنی و سه وزنی، بیانگر آن است که نوع تابع هسته در محاسبه توابع ناپارامتری با پهنای باند متغیر در شکل تابع چگالی احتمال تفاوت قابل ملاحظه‌ای را ایجاد نخواهد کرد. اما این تفاوت تا حدی برای تابع هسته نرمال وجود دارد.

با بررسی نتایج مطرح شده می‌توان به این نتیجه رسید که در تحلیل فراوانی به روش ناپارامتری با پهنای باند متغیر و ثابت، مطابق با تئوری مذکور، توابع هسته مختلف منجر به تخمین‌های قابل مقایسه می‌شود و انتخاب نوع تابع هسته خیلی مهم نخواهد بود.

مطابق با نمودار نشان داده شده در شکل 1 و مقادیر پهنای باند مندرج در جدول 4 هرچه تابع چگالی هسته مقدار اوج بیش‌تری داشته باشد در نتیجه پهنای باند محاسبه شده متناسب با آن تابع هسته برای داده‌های مورد نظر بزرگ‌تر است. این بزرگی پهنای باند باعث کاهش ناهمواری تابع چگالی احتمال ناپارامتری می‌شود و این امر علت برابری تقریبی توابع چگالی احتمال ناپارامتری محاسبه شده با انواع توابع هسته است.

با توجه به مقادیر ریشه دوم میانگین مربع خطا مندرج در جداول3، 5 و نمودار رسم شده در شکل 6 روش چگالی هسته با پهنای باند متغیر با هسته سه وزنی و مقدار K=10 (RMSE=0.1336) دقیق‌تر از روش چگالی هسته با پهنای باند ثابت محاسبه شده به روش مرتبط با پهنای باند با هسته اپینیچنیکوو (RMSE=0.1699) است. هم‌چنین دو روش ناپارامتری مورد استفاده ذکر شده از توزیع پارامتری لوگ پیرسون نوع 3 (RMSE=0.1752) دقیق‌تر می‌باشند. همچنین با توجه به اینکه مقادیر ضرایب همبستگی برای روش‌های ناپارامتری مندرج در جدول 5 و شکل 7 نسبت به روش پارامتری مطابق با جدول 3 بیش‌تر و نزدیک به یک می‌باشند می‌توان برتری و اعتماد پذیری نتایج حاصل از روش ناپارامتری را نسبت به روش پارامتری نشان داد. همچنین مطابق با شکل‌های 2 و 5 روش‌های ناپارامتری مذکور بر مبنای پهنای باند محاسبه شده به روش مرتبط با پهنای باند، توانایی توصیف دو یا چند اوجه بودن داده‌های مورد مطالعه را دارند.

در تحلیل فراوانی سیلاب به روش‌های ناپارامتری مطابق با شکل 1 شیب نمودار دبی-دوره بازگشت از یک دوره بازگشت به بعد کم می‌شود. زیرا در این روش به منظور تخمین دبی برای دوره بازگشت مورد نظر از داده‌های مشاهداتی و تعداد کمی از داده‌هایی که در همسایگی حداقل و حداکثر داده‌های مشاهداتی می‌باشند استفاده می‌شود.

به عبارت دیگر روش ناپارامتری تابع داده‌های مشاهده شده می‌باشد. بنابراین روش ناپارامتری در برون‌یابی دارای محدودیت است. ولی در روش پارامتری در برون‌یابی به منظور تخمین دبی چندان محدودیتی مشابه با روش ناپارامتری وجود ندارد.

## منابع


Adamowski,K. 1987. Nonparametric techniques for analysis of hydrological events. Water for future: Hydrol in Perspect, Proc. the Rome Symposium. IAHS Publ 164:67–76

Adamowski,K. 2000. Regional analysis of annual maximum and partial duration flood data by nonparametric and L-moment methods. Journal of Hydrology. 229.3: 219-231.

Altman,N and Leger,C. 1995. Bandwidth selection for kernel distribution function estimation. Journal of Statistical Planning and Inference. 46.2: 195-214.

Bowman,A.W. 1984. An alternative method of cross-validation for the smoothing of density estimates. Biometrika. 71.2: 353-360.

Bowman,A.W and Azzalini,A. 1997. Applied Smoothing Techniques for Data Analysis: The Kernel Approach with S-Plus Illustrations: The Kernel Approach with S-Plus Illustrations. Oxford University Press.

Bowman,A., Hall,P and Prvan,T. 1998. Bandwidth selection for the smoothing of distribution functions. Biometrika. 85.4: 799-808.

del Rio,A.Q and Perez,G.E. 2012. Nonparametric Kernel Distribution Function Estimation with kerdiest: An R Package for Bandwidth Choice and Applications. Journal of Statistical Software. 50.8: 1-21.

Henderson,D.J and Parmeter,C.F. 2012. Normal reference bandwidths for the general order, 193-200.

Rudemo,M. 1982. Empirical choice of histograms and kernel density estimators. Scandinavian Journal of Statistics. 9: 65-78.

Shabri,A. 2002. Nonparametric Kernel estimation of annual maximum stream flow quantiles. Matematika, 18.2: 99-107.

Silverman, B.W., 1986. Density estimation for statistics and data analysis (Vol. 26). Chapman and Hall CRC press. London, England.

Stone,C.J. 1984. An asymptotically optimal window selection rule for kernel density estimates. The Annals of Statistics. 12: 1285-1297.

Scott,D.W. 1985. Averaged shifted histogram: effective nonparametric density estimators in several dimensions. The Annals of Statistics.13.3: 1024-1040.

Scott,D.W and Factor,L.E. 1981. Monte Carlo study of three data-based nonparametric probability density estimators. Journal of the American Statistical Association. 76.373: 9-15.

Tsybakov,A.B. 2009. Introduction to nonparametric estimation. Revised and extended from the 2004 French original. Translated by Vladimir Zaiats. New York, Springer .

Wasserman,L. 2006. All of nonparametric statistics. Springer Science and Business Media.

multivariate kernel density derivative estimator. Statistics and Probability Letters. 82.12: 2198-2205.

Karmakar,S and Simonovic,S.P. 2007. Flood Frequency Analysis Using Copula with Mixed Marginal Distributions. Department of Civil and Environmental Engineering. The University of Western Ontario.

Kim,T.W., Valdés,J.B and Yoo,C. 2003. Nonparametric approach for estimating return periods of droughts in arid regions, Journal of Hydrologic Engineering. 8.5: 237–246.

Kim,K.D and Heo,J.H. 2002. Comparative study of flood quantiles estimation by nonparametric models. Journal of hydrology. 260.1: 176-193.

Lee,S.B. 2004. A Comparative Study on Parametric and Nonparametric Methods of Rainfall Frequency Analyses. MSc Thesis, Yonsei University, 155 pages.

Markovich,N. 2008. Nonparametric analysis of univariate heavy-tailed data: research and practice (Vol. 753). John Wiley & Sons.

Polansky,A.M and Baker,E.R. 2000. Multistage plug—in bandwidth selection for kernel distribution function estimates. Journal of Statistical Computation and Simulation. 65.1-4: 63-80.

Haghighatjou,p and Akhoond-Ali,A.M and Nazemosadat,M.J. 2013. Nonparametric kernel estimation of annual precipitation over Iran. Theoretical and applied climatology.112.1-2:

Iranian Journal of Irrigation and Drainage
No. 5, Vol. 10, Nov.-Dec. 2016, p. 556-569

نشریه آبیاری و زهکشی ایران
شماره 5، جلد 10، آذر - دی 1395، ص. 569-556

# Flood Frequency Analysis Using Variable And Fixed Kernel Density Case study: Dez river

**M.A Mohammad Jafar Sharbaf[1*], S. Mousavi Nadoushani[2]**
Recived: Oct.19, 2016 Accepted: Apr.07, 2016

## Abstract

Traditional method in flood frequency analysis is parametric approach. This method lacks the ability to describe multimodal and Asymmetric densities. In order to overcome this problem, the nonparametric models can be used. Two methods of nonparametric approach are: fixed and variable kernel density. In fixed kernel density method, the probability density function can be estimated by selecting a kernel function and optimal bandwidth and in variable kernel density method the probability density function can be estimated by selecting a kernel function and bandwidth at each observation point. Cross validation and Rule of thumb are common methods for estimating the optimum bandwidth. In this paper, besides mentioned methods Plug in bandwidth method is used and nonparametric flood frequency analysis is performed using annual maximum flood data of the Dez river. Finally results were compared with parametric method. According to RMSE, it is concluded that plug in bandwidth is the most accurate method for estimating optimum bandwidth. As well as Nonparametric method based on variable kernel density is more accurate than fixed kernel density and both types of these models are more accurate than LP3 distribution.

**Key words**: Bandwidth, Flood frequency, Kernel function, Nonparametric, Parametric.

1 -PhD Student of Shiraz University

2- Assistant professor, Water Engineering Department, Abbaspour College of Technology, Shahid Beheshti University
(*- Corresponding Author Email: mohammad.sharbaf@gmail.com)