

# Análisis de frecuencia de crecientes por medio de métodos no paramétricos

Daniel Francisco Campos Aranda

Universidad Autónoma de San Luis Potosí

*El objetivo fundamental de este trabajo consiste en describir con detalle dos métodos no paramétricos utilizados en el análisis de frecuencia de crecientes anuales. Este nuevo enfoque no selecciona a priori un modelo de función de densidad de probabilidad (fdp), sino que la calcula con base en los valores de la muestra o registro disponible, por ello es adecuado para fdp multimodales o de extremo derecho oscilante. Los procedimientos expuestos corresponden al método de las funciones núcleo (circular, rectangular y gaussiana) y al estimador parcial directo de la distribución de probabilidades acumuladas. Se comentan los programas de cómputo elaborados y sus aplicaciones de prueba; por último, se describen y analizan dos aplicaciones numéricas en dos estaciones hidrométricas de características muy diferentes.*

**Palabras clave:** métodos no paramétricos, frecuencia de crecientes, funciones núcleo, funciones de densidad de probabilidad.

## Introducción

Cuando en un sitio bajo estudio se dispone de información hidrométrica, la estimación de las avenidas de diseño puede ser realizada por medio de métodos paramétricos (más difundidos) y no paramétricos (más recientes). En términos concretos el enfoque paramétrico involucra el ajuste a la serie de gastos máximos anuales observados, de una función de densidad de probabilidades teórica, fdp, previamente adoptada. La fdp es empleada para describir la naturaleza aleatoria de los datos, cuya integral desde su inicio hasta un valor dado  $x$  conduce a la probabilidad de no excedencia  $F(x)$ , lo cual permite definir la función de distribución de probabilidades, FDP.

En hidrología superficial, diversas FDP son actualmente utilizadas, por ejemplo, las Log-normal de dos y tres parámetros (Burgess *et al.*, 1975; Stedinger, 1980), la Gumbel (Clarke, 1973; Kite, 1977), la General de Valores Extremos (Jenkinson, 1969; Raynal, 1984), la Pearson tipo III (Thom, 1971; Kottegoda, 1980), la Log-Pearson tipo III (Bobée, 1975; WRC, 1977) y recientemente la Wakeby (Houghton, 1978; Martínez y Raynal, 1988; Campos, 1995). Además de los modelos para poblaciones mezcladas (González, 1970; Rossi *et al.*,

1984; Gutierrez y Raynal, 1988; Campos, 1989). El enfoque paramétrico presenta tres severas desventajas:

- Emplear distribuciones unimodales
- Las crecientes muy grandes, o valores disparados (*outliers*) tienen gran influencia en el comportamiento del extremo derecho de la distribución
- La totalidad de datos determinan la forma de la distribución, incluso se ha demostrado que pequeñas diferencias en las crecientes de valor bajo afectan el extremo derecho, que es donde se estiman las grandes crecientes asociadas a periodos de retorno elevados (Gingras & Adamowski, 1993).

En los métodos no paramétricos no se asume una forma analítica de la fdp, sino que ésta es determinada con base en la muestra o serie anual de gastos máximos aleatorios ( $x_1, x_2, x_3, \dots, x_n$ ), por ello el enfoque no paramétrico es particularmente apropiado para fdp multimodales. La principal limitación de los métodos no paramétricos en el análisis de frecuencia de crecientes, corresponde a la escasa extrapolación que se puede realizar más allá del mayor dato observado. Sin embargo es posible modificar el enfoque no paramétrico para remediar este defecto, por ejemplo a través

del uso de mezclas de distribuciones paramétricas y no paramétricas (Schuster & Yakowitz, 1985), o bien incorporando la información histórica de crecientes (Bardsley, 1989a).

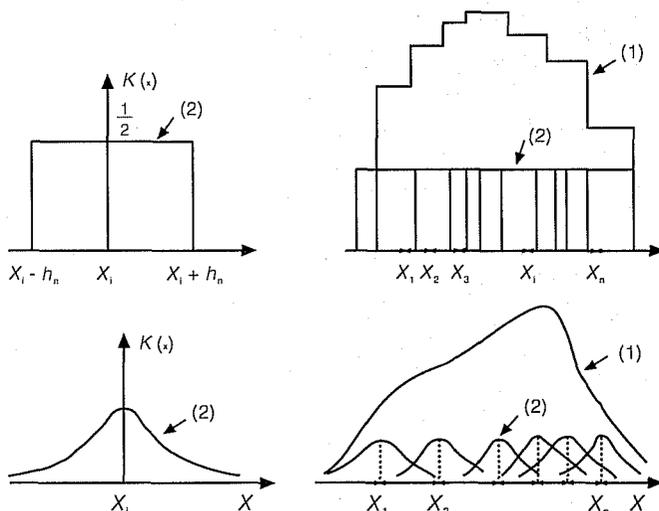
**Método de las funciones núcleo**

Los desarrollos recientes en teoría estadística han proporcionado una nueva técnica alternativa (métodos no paramétricos) para las estimaciones de fdp, los procedimientos disponibles son variados, sin embargo el más popular y mejor desarrollado teóricamente es el basado en las funciones núcleo (kernel). En este método la fdp  $f(x)$  está definida como (Adamowski, 1985):

$$f(x) = [1/(n h)] \sum_{i=1}^n K[(x - x_i)/h] \quad (1)$$

siendo  $n$  el número de datos u observaciones  $x_i$ ,  $K(\bullet)$  es una función núcleo seleccionada (la cual ella misma es una fdp simétrica) y  $h$  es un factor de suavizado o ancho de banda que debe ser estimado a partir de los datos de la muestra (registro de gastos máximos anuales). Por lo anterior, las condiciones impuestas a una función núcleo  $K(\bullet)$  son que sea no negativa, simétrica y que su integral sume uno (Adamowski, 1987). El procedimiento expresado por la ecuación 1, consiste en localizar en cada observación una función núcleo adoptada, cuya suma con un valor adecuado de  $h$  conduce a la fdp buscada. Esto último se muestra en la ilustración 1, para el caso de funciones núcleo rectangular y gaussiana.

**1. Aproximación de la función de densidad de probabilidad (1) por medio de la función núcleo (2) rectangular y gaussiana**



El proceso no paramétrico de construir una fdp es similar al de formar un histograma con ancho de clase  $h$ . El histograma se integra con rectángulos de altura  $z_i/(n h)$  y ancho  $h$ , siendo  $z_i$  el número de datos que corresponden a cada intervalo, por ello el área bajo el histograma es uno. La fdp no paramétrica se integra con la suma de las funciones núcleo de área  $1/n$ , cada una centrada en un dato de la muestra (Gingras y Adamowski, 1993).

La función núcleo puede tener una gran variedad de formas, las más difundidas son la circular, la rectangular y la gaussiana. La selección de la función núcleo no es crucial o crítica, ya que se ha demostrado que la mayoría conduce a resultados similares, en cambio, la determinación del factor  $h$  sí es crítica, pues un alto valor de  $h$  exagera el suavizado ocultando la verdadera variación y un valor bajo de  $h$  resulta en un insuficiente suavizado (Adamowski, 1989).

Uno de los métodos de selección de  $h$  consiste en minimizar, por medio de una técnica de validación cruzada, el error medio cuadrado integrado, EMCI, definido como (Adamowski, 1985):

$$EMCI = \int_{-\infty}^{\infty} E[f_n(x) - f(x)]^2 dx \quad (2)$$

en donde  $E$  significa la esperanza matemática o valor medio y  $f_n(x)$  es una estimación de la fdp desconocida  $f(x)$ . En general,  $h$  depende de la fdp verdadera, de la función núcleo y del tamaño de la muestra; la selección óptima de  $h$  conduce consistente y asintóticamente al menor error medio cuadrado integrado.

**FUNCION NUCLEO CIRCULAR.** Está definida por la expresión siguiente (Adamowski, 1989):

$$K(x) = (3/4) (1 - x^2) \quad \text{para } -1 \leq x \leq 1 \quad (3)$$

$$K(x) = 0 \quad \text{para cualquier otro valor de } x$$

en donde  $x$  será igual a  $(x - x_i)/h$ . Al usar el criterio de EMCI se ha encontrado que los diversos algoritmos de cómputo de  $h$ , se comportan de manera similar y conducen al valor óptimo teórico expresado como (Adamowski, 1989):

$$h = [ \sum_{i=2}^n \sum_{j=1}^{i-1} (x_i - x_j) ] / [5^{1/2} n(n - 10/3)] \quad (4)$$

siendo  $x_i$  y  $x_j$  son los valores observados ordenados en forma creciente.

Cuando la muestra es muy sesgada, o la fdp es multimodal, no es conveniente utilizar un valor constante de

$h$ , sino uno que varíe de manera inversa a la densidad local, esto se logra con una función núcleo variable, lo cual se puede consultar en Adamowski (1989).

**FUNCION NUCLEO RECTANGULAR.** Definida por las expresiones siguientes (Adamowski, 1985):

$$K(\mathbf{x}) = 0 \quad \text{para } \mathbf{x} \leq (x_i - h) \text{ y } \mathbf{x} \geq (x_i + h) \quad (5)$$

$$K(\mathbf{x}) = 1/2 \quad \text{para } (x_i - h) \leq \mathbf{x} \leq (x_i + h)$$

el factor de suavizado  $h$  deberá determinarse minimizando la siguiente expresión del error medio cuadrático:

$$\sum_{j=1}^n [p_j - p_n(x_j)]^2 \quad (6)$$

en la cual  $p_j$  es una probabilidad empírica y  $p_n(x_j)$  es una probabilidad de excedencia evaluada con la función núcleo; entonces (Adamowski, 1985):

$$p_j = (j - 0.25) / (n + 0.50) \quad (7)$$

$$p_n(x) = \int_x^{\infty} f(x) dx \quad (8)$$

Al sustituir la ecuación 1 en la 8 se obtiene:

$$p_n(x_i) = (1/nh) \sum_{i=1}^n \int_{x_i}^{\infty} K(x) dx = (1/nh) \sum_{i=1}^n C_i(x_i) \quad (9)$$

siendo:

$$C_i(x_i) = h \quad \text{para } x_i \leq (x_i - h) \quad (10)$$

$$C_i(x_i) = 0.50 (x_i + h - x_i) \quad \text{para } (x_i - h) \leq x_i \leq (x_i + h)$$

$$C_i(x_i) = 0 \quad \text{para } x_i \geq (x_i + h)$$

Al diferenciar la ecuación 6 e igualarla a cero se obtiene:

$$\sum_{j=1}^n [p_j - p_n(x_j)] [d p_n(x_j)/d h] = 0 \quad (11)$$

en donde:

$$d p_n(x_j)/d h = (1/nh) \sum_{i=1}^n [d C_i(x_i)/d h - C_i(x_i)/h] \quad (12)$$

al usar la función núcleo dada por las ecuaciones 5 se obtienen las expresiones siguientes:

$$d C_i(x_i)/d h = 1 \quad -\infty \leq x_i \leq (x_i - h)$$

$$d C_i(x_i)/d h = 1/2 \quad (x_i - h) \leq x_i \leq (x_i + h) \quad (13)$$

$$d C_i(x_i)/d h = 0 \quad (x_i + h) \geq x_i \geq \infty$$

El valor buscado de  $h$  se obtiene resolviendo la ecuación 11, por ejemplo a través de un método de aproximaciones sucesivas. Con base en la ecuación 9 se estima la probabilidad de excedencia ( $p$ ) asociada a cualquier valor de  $x$ , esto permite construir la FDP muestral. Por tanteos se puede estimar el valor de  $x$  correspondiente a un determinado periodo de retorno  $T_r = 1/p$  (Adamowski, 1985).

**FUNCION NUCLEO GAUSSIANA.** Esta función es actualmente la más utilizada (Gingras y Adamowski, 1992,1993,1994; Gingras *et al.*,1994) y está definida como:

$$K(\mathbf{x}) = [1/((2\pi)^{1/2})] [\exp(-\mathbf{x}^2/2)] \quad \text{para } -4.5 \leq \mathbf{x} \leq 4.5$$

$$K(\mathbf{x}) = 0 \quad \text{Para cualquier otro valor de } \mathbf{x} \quad (14)$$

La solución a través del EMCI conduce a encontrar la solución, por tanteos, de la ecuación siguiente (Gingras *et al.*,1994):

$$\{1/[(2\pi)^{1/2} nh]\} \left\{ \sum_{i=1}^n \sum_{j=1}^n \exp(\delta) [(1-4(2n)^{1/2} \exp(\delta)/(n-1)) (\gamma-1)(\gamma+1)-1] \right\} = 0 \quad (15)$$

en donde:

$$\delta = d_{ij}/4 \quad \text{con } d_{ij} = -[(x_i - x_j)/h]^2 \quad \gamma = (x_i - x_j)/h$$

la primera sumatoria se efectúa cuando  $i$  es diferente de  $j$ .

### Método del estimador no paramétrico parcial directo

Bardsley (1989 a,b) ha propuesto un estimador libre de parámetros de ajuste, que estima globalmente la función de FDP, con base en la probabilidad de excedencia  $p_n(x)$ , éste es:

$$p_n(x) = (1/n) \sum_{i=1}^n \exp\{-\exp[(x - r_i)/S_i]\} \quad (16)$$

$S_i = 0.50 (r_i - r_{i-1})$ , para  $i = 2$  y se acepta que  $S_1 = S_2$ , siendo  $r_i$  los gastos máximos anuales observados ordenados en forma creciente. Asignando por tanteos

valores a  $x$ , se puede estimar su magnitud correspondiente a los periodos de retorno de diseño [ $T_r = 1/p_n(x)$ ].

### Descripción de las aplicaciones numéricas

#### *Programas de computo y aplicaciones de prueba*

Los tres programas de cómputo desarrollados para la aplicación del método de las funciones núcleo, tienen una estructura similar, primero ordenan los datos, después calculan o estiman el parámetro  $h$ , en seguida presentan en pantalla la fdp y la FDP para un intervalo ( $\Delta x$ ) de la variable seleccionada y por último, imprimen, para otro intervalo  $\Delta x$  adoptado, lo anterior y el periodo de retorno correspondiente.

En el programa NOPACIR (ajuste NO PARAMétrico con función núcleo CIRCular), el parámetro  $h$  se calcula con la ecuación 4; en cambio, en el programa NOPAREC la estimación de  $h$  se realiza por tanteos, con base en las ecuaciones 6 a 12, buscando el mínimo de la ecuación 11; en forma similar en el programa NOPAGAUS el parámetro  $h$  se estima por tanteos, con la ecuación 15.

En el programa NOPACIR se utilizan las ecuaciones 1 y 3 para evaluar la fdp, cuya integración por medio de la regla del trapecio conduce a la FDP, caracterizada por la probabilidad de no excedencia hasta el valor de  $x$  asignado. En el programa NOPAREC se utiliza la ecuación 9 para estimar la probabilidad de excedencia de la FDP desde un cierto valor de  $x$  adoptado; en contraste en el programa NOPAGAUS se evalúa la fdp con la ecuación 1, cuya integración conduce al valor de probabilidad de no excedencia correspondiente al valor de  $x$  asignado. Por último, el programa NOPABAR emplea la ecuación 16 y por tanteos evalúa el valor de  $x$  correspondiente a un periodo de retorno de diseño.

Se realizaron dos aplicaciones numéricas de prueba a los programas citados, la primera para los 67 años de registro (1917-1983) del noreste del río Margaree en la estación Valle Margaree, Canadá (Adamowski, 1985, 1987, 1989) y la segunda para los 21 años de registro (1968-1988) del río Nackawic en la estación 01AK007, Canadá, citados por Gingras y Adamowski (1992).

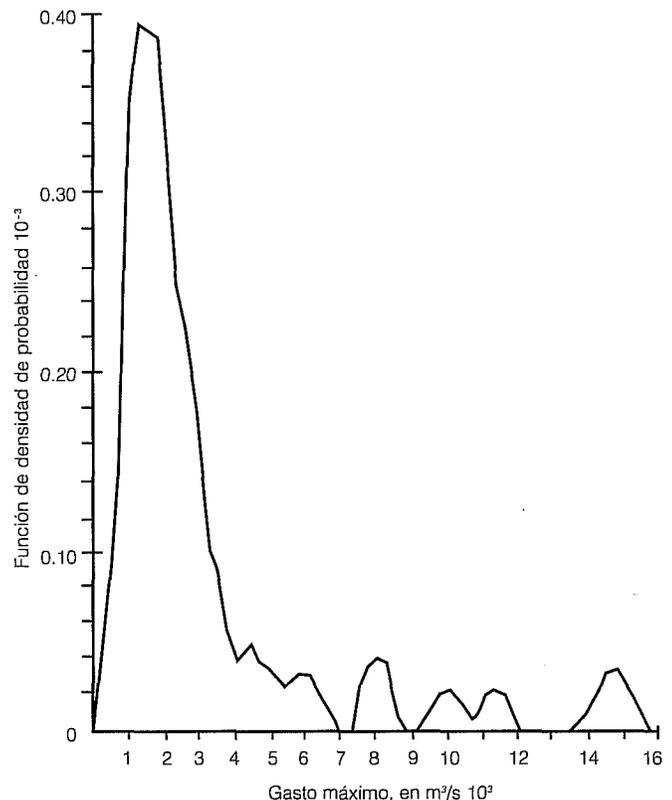
#### *Aplicación a los datos de la estación hidrométrica Huites*

La estación hidrométrica Huites se localiza en el río Fuerte, dentro de la Región Hidrológica No. 10 (Sinaloa), su área de cuenca es de 26 020 km<sup>2</sup> y sus cre-

cientes anuales presentan una marcada estacionalidad, siendo las más severas las que ocurren durante el invierno en diciembre y enero, y las más frecuentes las del mes de agosto con un 24% de ocurrencias en dicho mes. En general, el mecanismo que genera las crecientes de invierno es originado por lluvias de amplia extensión, conocidas como *equipatas* (SRH, 1975); en cambio, las crecientes del verano y otoño son generadas por tormentas convectivas locales. El registro disponible de gastos máximos anuales en la estación Huites comprende 53 años, cuyos valores y parámetros estadísticos correspondientes se presentan en el cuadro 1.

La aplicación de los cuatro métodos no paramétricos descritos condujo a los valores que se han tabulado en el cuadro 1, para los ocho periodos de retorno indicados. Los valores correspondientes al método de las funciones núcleo fueron obtenidos utilizando un intervalo  $\Delta x = 5 \text{ m}^3/\text{s}$ , en cambio, los del método de Bardsley se obtuvieron por tanteos. Es notable la consistencia en el orden de magnitud que muestran los resultados, aún para los altos periodos de retorno.

### 2. Función de densidad de probabilidad (fdp) en la estación hidrométrica Huites (Sinaloa), estimada por medio de la función núcleo circular para un intervalo $\Delta x = 250 \text{ m}^3/\text{s}$ y $h = 680.758$



1. Datos y resultados de los métodos no paramétricos y paramétricos en la estación hidrométrica Huites (Sinaloa)

Año	Q <sub>máx</sub> (m³/s)	Año	Q <sub>máx</sub> (m³/s)	Año	Q <sub>máx</sub> (m³/s)
1941	2 085	1959	1 908	1977	1 119
1942	2 531	1960	15 000	1978	6 178
1943	14 376	1961	1 396	1979	4 443
1944	2 580	1962	1 620	1980	1 474
1945	1 499	1963	2 702	1981	2 508
1946	1 165	1964	1 319	1982	1 530
1947	1 127	1965	1 944	1983	8 000
1948	3 215	1966	2 420	1984	5 496
1949	10 000	1967	2 506	1985	3 385
1950	3 229	1968	1 534	1986	1 374
1951	677	1969	1 508	1987	1 245
1952	1 266	1970	1 558	1988	2 299
1953	1 025	1971	2 200	1989	1 345
1954	955	1972	2 225	1990	11 350
1955	4 780	1973	7 960	1991	2 509
1956	696	1974	4 001	1992	2 006
1957	593	1975	1 067	1993	1 180
1958	3 010	1976	3 233		

Parámetros Estadísticos:

Media = 3 176.4 m³/s	Mediana = 2 085.0 m³/s	Moda = 1 260.9 m³/s
Desviación Estándar = 3 235.6 m³/s	Coeficiente de Curtosis = 8.730	
Coeficiente de Asimetría (Cs) = 2.426	Cs de los logaritmos = 0.777461	

Periodo de Retorno (Años)

	2	5	10	20	50	100	500	1 000
<i>Ajuste No Paramétrico:</i>								
FN <sup>1</sup> Circular (h = 680.8)	2 065	3 920	7 850	11 225	14 660	14 990	15 425	15 515
FN Rectangular (h = 140.0)	2 095	3 975	7 940	11 310	14 500	14 990	15 110	15 125
FN Gaussiana (h = 29.0)	2 085	3 995	7 965	11 340	14 420	15 000	15 035	15 050
Estimador Parcial Directo	2 065	3 730	7 820	10 815	14 612	15 042	15 605	16 006
<i>Ajuste Paramétrico:</i>								
FDP <sup>2</sup> Wakeby	2 125	4 770	7 307	10 519	16 276	22 286	45 570	62 051
FDP Log-Pearson tipo III	2 046	4 086	6 292	9 334	15 189	21 573	46 943	64 883
FDP Gral. de Val. Extremos	2 035	3 995	6 195	9 405	16 097	24 044	60 616	90 151
FDP Log-normal (2 parám.)	2 282	4 303	5 998	7 889	10 740	13 191	19 997	23 464
FDP Log. de Boughton	2 221	4 359	6 084	7 910	10 445	12 426	17 112	19 129
FDP Gumbel ó VEI	2 645	5 504	7 397	9 213	11 564	13 326	17 396	19 146

1 Función Núcleo

2 Función de Distribución de Probabilidades

En contraste, los resultados de los métodos paramétricos presentan dispersión en los resultados desde el periodo de retorno de cincuenta años. Lo anterior es probable que se origine por el uso inadecuado de una fdp unimodal, ya que el gráfico de la fdp en la estación Huites obtenido a través de la función núcleo circular, ilustración 2, muestra claramente una fdp con extremo derecho oscilante. En la ilustración 3 se muestra la comparación para las FDP empírica y la definida con la función núcleo circular, en papel de probabilidad Gumbel.

Con base en esto último y la dispersión de los gastos máximos del cuadro 1 correspondientes a periodos de retorno superiores a cincuenta años, se puede indicar que los resultados de los métodos paramétricos no son completamente confiables, y por ello se deben buscar otras técnicas de análisis de frecuencia de crecientes para los datos de la estación hidrométrica Huites, por ejemplo, en el campo de las distribuciones para poblaciones mezcladas o contaminadas (González, 1970; Rossi *et al.*, 1984; Gutierrez y Raynal, 1988; Campos, 1989), bajo el enfoque de segregación de poblaciones (Campos, 1987), transformando los datos originales (Bethlahmy, 1977 y Chander *et al.*, 1978), o bien en los análisis regionales de crecientes (Campos, 1994a,b).

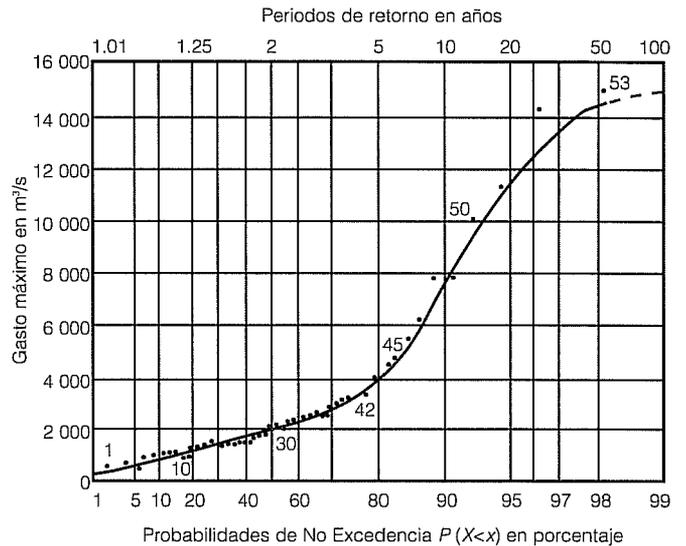
*Aplicación a los datos de la estación hidrométrica Santa Rosa*

En la estación hidrométrica Santa Rosa se miden los escurrimientos del río Valles, a unos 7 km aguas arriba de Ciudad Valles, San Luis Potosí, y dentro de la Región Hidrológica No. 26 correspondiente a la cuenca baja del río Pánuco (SRH, 1977). El área que drena hasta la estación Santa Rosa es de 3 521 km<sup>2</sup> y el registro disponible de gastos máximos anuales abarca 36 años, cuyos valores y parámetros estadísticos se aprecian en el cuadro 2.

Los resultados de los cuatro métodos no paramétricos se han concentrado en el cuadro 2, en este caso los valores correspondientes al método de las funciones núcleo fueron obtenidos utilizando un intervalo  $\Delta x = 1 \text{ m}^3/\text{s}$ , en cambio, los del método de Bardsley se estimaron por tanteos, apoyándose en los valores previamente obtenidos. Nuevamente es notable la regularidad que muestran los resultados, aún para los altos periodos de retorno.

También en el cuadro 2 se tienen los valores obtenidos con el ajuste paramétrico empleando siete FDP, los resultados se han tabulado en orden creciente de error estándar de ajuste (Kite, 1977), o desviación estándar entre los valores observados y los estimados

**3. Comparación entre las distribuciones de probabilidad empírica y la obtenida con la función núcleo circular en la estación hidrométrica Huites**



con el modelo de FDP adoptado para la misma probabilidad de posición gráfica (Benson, 1962). Los valores obtenidos con las FDP varían bastante a partir de los periodos de retorno que superan el número de datos, y tales valores son función de comportamiento de la cola derecha del modelo de FDP empleado.

En la ilustración 4 se muestra la fdp obtenida por medio del programa NOPAGAUS, para un intervalo  $\Delta x = 50 \text{ m}^3/\text{s}$ , tal gráfico muestra que la fdp es multimodal, lo cual probablemente se debe al hecho de tener dos modelos de generación de las crecientes máximas anuales, el primero asociado a la ocurrencia de tormentas convectivas y el segundo a la presencia de los ciclones; el primer mecanismo origina crecientes principalmente en junio y julio y el segundo en septiembre y octubre (SRH, 1977). La forma encontrada para la fdp destaca la inconveniencia del uso de los métodos paramétricos, cuya fdp en general es unimodal, y por lo tanto no pueden reproducir el comportamiento muestral en el extremo derecho de la fdp, conduciendo a estimaciones bastante diferentes en los periodos de retorno superiores al tamaño del registro. En la ilustración 5 se presenta el contraste de las FDP empírica y la obtenida con la función núcleo gaussiana, en papel de probabilidad log-normal.

**Conclusiones**

La ventaja fundamental de los métodos no paramétricos radica en permitir la obtención de una estimación

1. Datos y resultados de los métodos no paramétricos y paramétricos en la estación hidrométrica Santa Rosa (S.L.P.)

Año	Gasto máximo Anual (m³/s)	Año	Gasto máximo Anual (m³/s)
1958	1 176	1976	2 588
1959	505	1977	1 884
1960	342	1978	372
1961	728	1979	283
1962	324	1980	162
1963	973	1981	289
1964	145	1982	44
1965	360	1983	578
1966	1 692	1984	462
1967	912	1985	361
1968	770	1986	353
1969	1 244	1987	187
1970	803	1988	898
1971	636	1989	267
1972	394	1990	800
1973	951	1991	1 163
1974	1 480	1992	1 636
1975	622	1993	2 036

Parámetros Estadísticos:

Media = 798.4 m³/s	Mediana = 629.0 m³/s	Moda = 345.0 m³/s
Desviación Estándar = 601.9 m³/s	Coeficiente de Curtosis = 4.534	
Coeficiente de Asimetría (Cs) = 1.269	Cs de los logaritmos = -0.674603	

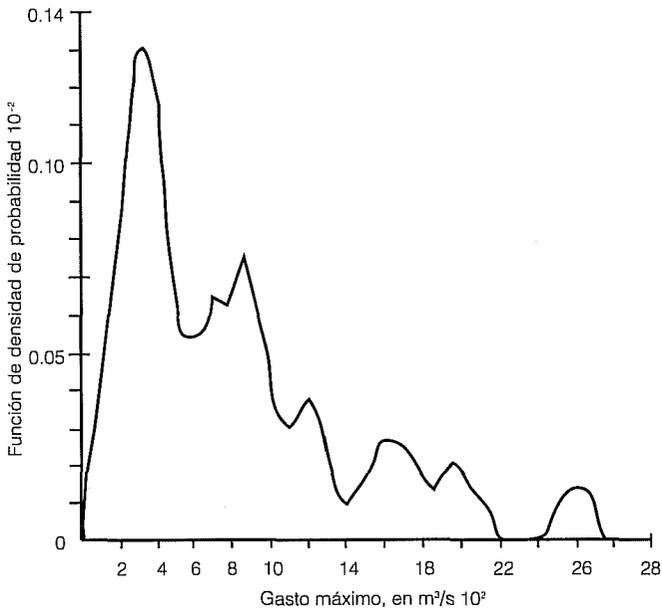
Periodo de Retorno (Años)

	2	5	10	20	50	100	500	1 000
<i>Ajuste No Paramétrico:</i>								
FN <sup>1</sup> Circular (h = 156.3)	654	1 239	1 746	2 605	2 605	2 698	---	---
FN Rectangular (h = 170.5)	637	1 228	1 726	1 994	2 513	2 636	2 734	15 125
FN Gaussiana (h = 56.0)	647	1 224	1 720	2 026	2 589	2 650	---	---
Estimador Parcial Directo	626	1 181	1 677	1 945	2 281	2 594	2 855	2 920
<i>Ajuste Paramétrico:</i>								
FDP <sup>2</sup> Wakeby	622	1 268	1 732	2 182	2 767	3 209	4 269	4 755
FDP Log. de Boughton	659	1 260	1 677	2 071	2 559	2 902	3 621	3 898
FDP Log-normal (2 parám.)	581	1 186	1 724	2 347	3 321	4 186	6 687	8 005
FDP Pearson tipo III	650	1 192	1 570	1 937	2 413	2 770	3 597	3 954
FDP Gumbel ó VEI	678	1 268	1 658	2 032	2 516	2 879	3 718	4 078
FDP Log-Pearson tipo III	650	1 209	1 582	1 921	2 326	2 602	3 154	3 357
FDP Gral. de Val. Extremos	610	1 092	1 501	1 978	2 747	3 463	5 725	7 039

1 Función Núcleo

2 Función de Distribución de Probabilidades

**4. Función de densidad de probabilidad (fdp) en la estación hidrométrica Santa Rosa (S.L.P.) estimada por medio de la función núcleo Gaussiana para un intervalo  $\Delta x = 50 \text{ m}^3/\text{s}$  y  $h = 56.0$**



gráfica de fdp muestral, la cual si no es unimodal, pone de manifiesto la inconveniencia de utilizar el enfoque paramétrico clásico en el análisis de frecuencia de crecientes, que consiste en emplear modelos de distribución de probabilidades de una sola población. En tales casos, el enfoque paramétrico se debe abordar a través del uso de distribuciones para poblaciones mezcladas o contaminadas, por medio de la segregación de las poblaciones, o bien, transformando los datos originales.

Con base en los diversos contrastes y aplicaciones realizados, se sugiere emplear un intervalo  $\Delta x$  para dibujar la fdp de  $1/50$ , como máximo, del valor del gasto máximo observado, en cambio, para el cálculo del gasto máximo asociado a un periodo de retorno es conveniente utilizar  $\Delta x$  igual a la unidad, en cuencas pequeñas; en las mediatas y grandes el valor  $\Delta x$  se puede variar de 5 a 10 metros cúbicos por segundo.

La forma o modelo que presenta la fdp muestral, obtenida por medio de los métodos no paramétricos permite agrupar regionalmente estaciones hidrométricas según un comportamiento similar, o bien, ayuda a identificar zonas de influencia de los diferentes mecanismos de generación de las crecientes.

La otra ventaja del análisis no paramétrico es la gran consistencia o similitud de resultados para los periodos de retorno que son estimables, en lo anterior se

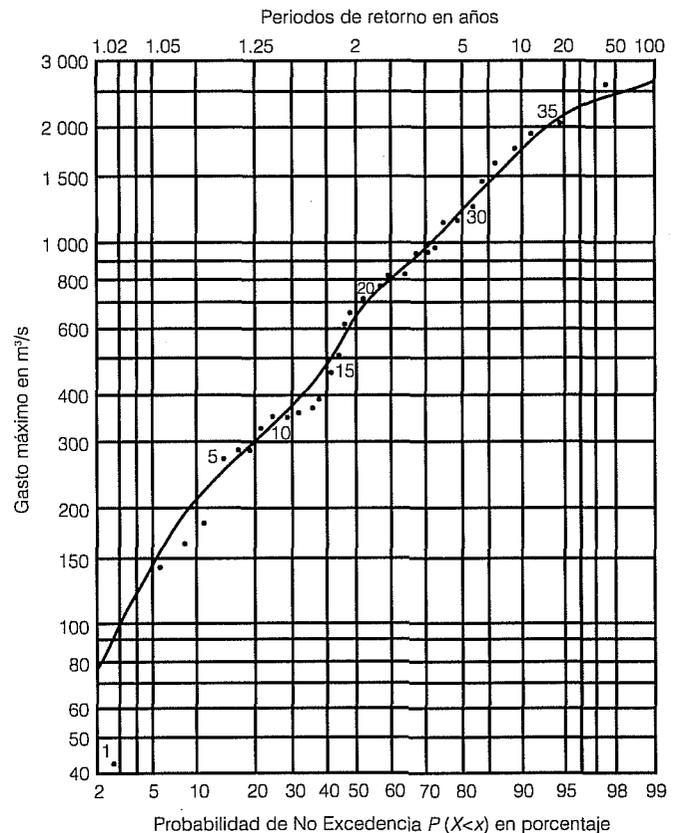
encuentra implícita la principal desventaja de estos procedimientos, que es la escasa extrapolación que permiten.

Con base en las observaciones anteriores, se sugiere utilizar el enfoque no paramétrico, por ejemplo a través del uso de la función núcleo circular que no implica tanteos y que es sumamente sencillo de programar o aplicar manualmente, para obtener una estimación gráfica de la fdp muestral y por otra parte, utilizar sus resultados asociados a los periodos de retorno estimables como apoyo cuantitativo para la selección de los resultados del enfoque paramétrico.

Conviene destacar la utilidad del enfoque no paramétrico, como una técnica muy confiable para estimar gastos máximos asociados a periodos de retorno que son interpolados dentro de los valores muestrales. Para fines de extrapolación se deben realizar mezclas óptimas de fdp no paramétricas y paramétricas.

Recibido: julio, 1995  
Aprobado: enero, 1996

**5. Comparación entre las distribuciones de probabilidad empírica y la obtenida con la función núcleo Gaussiana en la estación hidrométrica Santa Rosa**



## Agradecimientos

Al profesor Kaz Adamowski de la Universidad de Ottawa por haberme proporcionado los datos del río Margaree en Canadá, así como otros documentos y programas de cómputo relacionados con los ajustes no paramétricos, material que me permitió llevar este trabajo a su fase de aplicación numérica.

Al doctor Francisco Javier Aparicio Mijares, del Instituto Mexicano de Tecnología del Agua, IMTA, y al ingeniero Dionisio Castillo Rios, de la Gerencia Regional Noreste de la Comisión Nacional del Agua, CNA, por haberme proporcionado los datos actualizados de las estaciones Huites y Santa Rosa.

A los revisores anónimos del documento por sus comentarios y correcciones sugeridas, todas fueron realizadas, y sin duda contribuyeron a mejorar sustancialmente el trabajo.

## Referencias

- Adamowski, K. 1985. Nonparametric kernel estimation of floods frequencies. *Water Resources Research*, 21(11): 1585-1590.
- Adamowski, K. 1987. Nonparametric techniques for analysis of hydrological events. *Proceedings of the International Symposium on Water for the Future (Hydrology in Perspective)*, pp. 67-76. Edited by J.C. Rodda & N.C. Matalas. Rome, Italy. *Publication No. 164* of the International Association of Hydrological Sciences. Paris, France.
- Adamowski, K. 1989. A Monte Carlo comparison of parametric and nonparametric estimation of flood frequencies. *Journal of Hydrology*, 108:295-308.
- Bardsley, W. E. 1989a. Using historical data in nonparametric flood estimation. *Journal of Hydrology*, 108:249-255.
- Bardsley, W. E. 1989b. A simple parameter-free flood magnitude estimator. *Hydrological Sciences Journal*, 34(2):129-137.
- Benson, M. A. 1962. Plotting positions and economics of engineering planning. *Journal of the Hydraulics Division*, 88(HY6):57-71.
- Bethlahmy, N. 1977. Flood analysis by SEMAX transformation. *Journal of the Hydraulics Division*, 103(HY1):69-78.
- Bobée, B. 1975. The log Pearson type 3 distribution and its application in hydrology. *Water Resources Research*, 11(5):681-689.
- Burges, S. J.; D. P. Lettenmaier & C. L. Bates. Properties of the three-parameter log normal probability distribution. *Water Resources Research*, 11(2):229-235.
- Campos A., D. F. 1987. Análisis probabilístico de avenidas máximas con poblaciones mezcladas. *Ingeniería Hidráulica en México*, 1(3):9-18.
- Campos A., D. F. 1989. Estimación de los parámetros óptimos de la distribución Gumbel mixta por medio del algoritmo de Rosenbrock. *Ingeniería Hidráulica en México*, 1V(1):9-18.
- Campos A., D.F. 1994a. Ajuste regionalizado de las distribuciones Gumbel y GVE en la región hidrológica No. 10, Sinaloa. 2: Método basado en los MPP. *Memorias Técnicas del XIII Congreso Nacional de Hidráulica*, Tomo II, Tema V, ponencia No. 35. Puebla, Pue.
- Campos A., D. F. 1994b. Aplicación del método del índice de crecientes en la región hidrológica número 10, Sinaloa. *Ingeniería Hidráulica en México*, 1X(3):41-55.
- Campos A., D. F. 1995. Ajuste de la distribución de probabilidades Wakeby por medio de optimización. *Memoria del II Simposium Internacional de Computación*, Tema II: Algoritmos, 2a. ponencia. Centro Nacional de Cálculo del I.P.N. 8 al 10 de Noviembre. México, D.F.
- Chander, S.; S. K. Spolia y A. Kumar. 1978. Flood frequency analysis by power transformation. *Journal of the Hydraulics Division*, 104(HY11):1495-1504.
- Clarke, R. T. 1973. *Mathematical models in hydrology*. Chapter 5: The estimation of floods with given return period. Irrigation and Drainage Paper 19. FAO, Rome, Italy.
- Gingras, D. & K. Adamowski. 1992. Coupling of nonparametric frequency and L-moment analyses for mixed distribution identification. *Water Resources Bulletin*, 28(2):263-272.
- Gingras, D. & K. Adamowski. 1993. Homogeneous region delineation based on annual flood generation mechanisms. *Journal of Hydrological Sciences*, 38(2):103-121.
- Gingras, D. & K. Adamowski. 1994. Performance of L-moments and nonparametric flood frequency analysis. *Can. Journal of Civil Eng.*, 21:856-862.
- Gingras, D.; K. Adamowski & P. J. Pilon. 1994. Regional flood equations for the provinces of Ontario and Quebec. *Water Resources Bulletin*, 30(1):55-66.
- González V., F. J. 1970. Contribución al análisis de frecuencias de valores extremos de los gastos máximos de un río. *Publicación No. 277*, Instituto de Ingeniería de la UNAM. México, D.F. (24 páginas).
- Gutiérrez O., C. y J. A. Raynal V. 1988. Distribuciones mezcladas en el análisis de frecuencias de gastos máximos. *Memorias del 10° Congreso Nacional de Hidráulica*, Tomo II, Subtema 6, páginas 220-228. Octubre, Morelia, Michoacán.
- Houghton, J. C. 1978. Birth of a parent: the Wakeby distribution for modeling flood flows. *Water Resources Research*, 14(6):1105-1109.
- Jenkinson, A. F. 1969. Statistics of extremes. Chapter 5 en Technical Note No. 98, WMO-No. 233. TP: 126: *Estimation of maximum floods*. Secretariat of the World Meteorological Organization. Geneva, Switzerland.
- Kite, G. W. 1977. *Frequency and risk analyses in hydrology*. Chapters 8 & 12: Type I extremal distribution & Comparison of frequency distributions. Water Resources Publication. Fort Collins, CO.: U.S.A.
- Kottegoda, N. T. 1980. *Stochastic water resources technology*. Theme 6.6: Pearson type III function applied to extreme values. The MacMillan Press, Ltd. London, England.
- Martínez A., H. R. y J. A. Raynal V. 1988. Métodos de estimación de parámetros de la distribución Wakeby. *Memorias del 10° Congreso Nacional de Hidráulica*, Tomo II, Subtema 6, páginas 464-472. Octubre, Morelia, Michoacán.
- Raynal V., J. A. 1984. La Distribución general de valores extremos en hidrología: 2.- Estado actual y aplicaciones. *Memorias del 8° Congreso Nacional de Hidráulica*, Tomo

- I, Subtema 2, páginas B.9-B.19. Toluca, Estado de México.
- Rossi, F., M. Fiorentino & P. Versace. 1984. Two-component extreme value distribution for flood frequency analysis. *Water Resources Research*, 20(7):847-856.
- Secretaría de Recursos Hidráulicos (SRH). 1975. *Boletín Hidrológico No. 36*. Tomos I y VI. Región Hidrológica No. 10 (Sinaloa). Subsecretaría de Planeación. México, D.F.
- Secretaría de Recursos Hidráulicos (SRH). 1977. *Boletín hidrológico No. 32*. Tomos I y III. Región Hidrológica No. 26 (Pánuco). Subsecretaría de Planeación. México, D.F.
- Schuster, E. & S. Yakowitz. 1985. Parametric/nonparametric mixture density estimation with application to flood-frequency analysis. *Water Resources Bulletin*, 21(5):797-804.
- Stedinger, J. R. 1980. Fitting log normal distributions to hydrologic data. *Water Resources Research*, 16(3):481-490.
- Thom, H. C. S. 1971. *Some methods of climatological analysis*. Chapter 3: General statistical methods, pp. 17-53. Technical Note No. 81, WMO-No. 199, TP: 103. World Meteorological Organization, Geneva, Switzerland.
- United States Water Resources Council (WRC). 1977. *Guidelines for determining flood flow frequency*. Bulletin #17A of the Hydrology Committee. Washington, D.C., U.S.A. Revised edition.

#### Abstract

Campos-Aranda, D.F. "Flood Frequency Analysis Through Nonparametric Methods". *Hydraulic Engineering in Mexico (in Spanish)*. Vol XI. Num. 3, pages 25-34, September-December, 1996.

A detailed description is given of two nonparametric methods that are used in annual flood frequency analysis. This new approach does not select a priori a probability density function (pdf) model but based on the available register or sample values, computes one. It is appropriate for multimodal or heavy-tailed pdf. The procedures put forward correspond to kernel function methods (circular, rectangular and Gaussian) and the direct partial estimator of the accumulated probability distribution. The computer programs that were used and their test applications are discussed. Finally, two numerical applications are described and analysed for two hydrometric stations with very different characteristics.

**Key words:** Nonparametric methods, flood frequency, kernel functions, probability density functions.