

10.1 Introducción

De la referencia 9 extraemos los siguientes comentarios, con ciertos ajustes para su adaptación al presente texto.

El objetivo básico de la aplicación de la estadística en Hidrología es el análisis de la información hidrológica en forma de muestras, a fin de inferir las características con que debe ser esperado en el futuro el fenómeno que se estudia. El avance en el campo de las computadoras y el desarrollo creciente de métodos numéricos han dado una importancia particular al uso de la estadística en todas las ciencias naturales, especialmente en Hidrología.

Existe en muchos la idea de que la estadística es usada sólo cuando no es posible dar una solución exacta a un problema hidrológico. En esta interpretación la solución exacta es una solución determinística del problema. Sin embargo, se puede demostrar que la solución determinística constituye una solución particular de la solución estadística o probabilística.

En forma general, la mayoría de los problemas hidrológicos se pueden agrupar en tres categorías principales de acuerdo al objetivo principal del proyecto:

- a. Diseño de estructuras hidráulicas, siendo necesaria la evaluación y cuantificación de los valores extremos (máximos y mínimos) del escurrimiento superficial.
- b. Satisfacción de demandas, siendo necesario evaluar y cuantificar las descargas disponibles en el punto de interés.
- c. Diseño y operación de embalses, siendo necesario evaluar y cuantificar la variación del escurrimiento superficial en todas sus características estadísticas, como valores medios, máximos y mínimos.

En cada una de las tres categorías mencionadas se presentan diferentes tipos de problemas, dependiendo la simplicidad o complejidad de la solución del tipo, cantidad y calidad de la información disponible, así como de la magnitud del proyecto. Los casos más comunes que se presentan en cada una de las tres categorías mencionadas son:

1. Cuencas con suficiente información hidrológica. Este es el caso más optimista donde se pueden aplicar todo tipo de metodologías existentes.
2. Cuencas con escasa información hidrológica. En este caso se pueden desarrollar modelos que relacionen las precipitaciones con las descargas, mediante el uso de la regresión simple o múltiple, lineal o no lineal.
3. Cuencas sin información hidrológica. Este es el caso más crítico y el más común, el cual puede resolverse mediante un análisis regional.

Conviene recordar aquí que las técnicas probabilísticas se aplican a la información hidrológica sólo después que ésta ha sido sometida a un tratamiento previo en que se atiende: el relleno de la información, el análisis de consistencia y la extensión del registro. Para este trabajo previo resulta fundamental la referencia 8.

10.2 Uso de Modelos Probabilísticos

Los fenómenos que se presentan en la ingeniería pueden clasificarse, desde el punto de vista de la certeza de su ocurrencia, en determinísticos y probabilísticos. Si la ocurrencia de las variables en un proceso es cierta, es decir si las variables siguen una ley determinada se habla de un proceso determinístico. En cambio, si se toma en cuenta la probabilidad de ocurrencia y la falta de certeza existente, entonces se habla de un proceso de naturaleza probabilística. Es conveniente hacer notar que la gran mayoría de los procesos que interesan al ingeniero, en especial en el campo de la Hidrología, pertenecen a la categoría de fenómenos probabilísticos.

Entre los procesos probabilísticos es necesario distinguir los probabilísticos a secas de los probabilísticos estocásticos o simplemente estocásticos. Se denomina proceso estocástico a aquél en el que las características de las variables aleatorias varían con el tiempo. En un proceso probabilístico, independiente de la variable tiempo, la secuencia de las variables no interesa y se supone que ellas siguen un determinado comportamiento dado por el modelo probabilístico o distribución de frecuencias. En el presente texto sólo en el apartado 10.7 se hace una breve referencia a los procesos estocásticos.

Dada pues una variable aleatoria, interesará describir la probabilidad de ocurrencia de los distintos estados. Esto se consigue gracias a un modelo matemático de su comportamiento o modelo probabilístico. Esta distribución probabilística permite calcular:

1. Las probabilidades de los distintos estados o valores que puede tomar la variable aleatoria.
2. La probabilidad de tener valores mayores o menores que un determinado límite.
3. Los valores de probabilidad de ocurrencia asociados a cada valor de la variable aleatoria.

En resumen, puede decirse que el modelo probabilístico o distribución permite conocer y manejar fácilmente el comportamiento de la variable y sintetiza toda la información sobre probabilidades asociadas a cada estado.

Según se trate de variables discretas o continuas, se usarán modelos de distribución probabilísticos discretos o continuos. Serán modelos discretos aquéllos cuya función densidad de probabilidad y función de probabilidad acumulada se encuentran definidas para determinados valores que puede tomar la variable.

Las principales distribuciones discretas son:

1. Distribución binomial
2. Distribución de Poisson

Las principales distribuciones continuas son:

1. Distribución uniforme
2. Distribución normal
3. Distribución logarítmico-normal
4. Distribución Gamma
5. Distribuciones de valores extremos
 - a. Tipo I o tipo exponencial (ley de Gumbel)
 - b. Tipo II o tipo Cauchy
 - c. Tipo III o distribuciones truncadas
6. Distribución Chi cuadrado

Una vez que el ingeniero, en base a su experiencia, escoge el modelo probabilístico que va a usar debe proceder a calcular los parámetros de su modelo y después revisar si este modelo es consistente con la realidad. Ambas cosas las hace con los datos observados (registro o muestra). Para la estimación de los parámetros hay disponibles dos métodos:

- el método de los momentos
- el método de máxima verosimilitud

Y para el estudio de la consistencia dos grupos de métodos:

- * métodos gráficos
- * métodos cuantitativos:
 - test Chi cuadrado
 - test W
 - test Student
 - test de Kolmogoroff.

10.3 Análisis de Frecuencia de Valores Extremos

En este apartado se describe el análisis de frecuencia de valores extremos referido a caudales, es decir el análisis a que son sometidos los caudales máximos anuales. El objeto es calcular el caudal de diseño de estructuras como los aliviaderos de las presas de embalse.

Supongamos que se cuenta con el registro de caudales máximos mensuales de una serie de 30 años (caudales en m^3/seg):

	E	F	M	A	M	J	J	A	S	O	N	D
1959	102	110	205	40	20	15	10	8	5	5	4	2
1960	80	90	85	30	12	8	6	2	2	1	0	0
1961	16	96	74	42	36	25	14	6	8	8	6	16
1962	95	125	196	45	31	21	16	15	15	14	3	2
⋮												
1988	97	116	91	50	28	19	12	11	6	5	2	12

Como esta serie abarca toda la información disponible es denominada se

rie de duración completa. La serie anual máxima se obtiene eligiendo el valor máximo de cada año:

1959	205
1960	90
1961	96
⋮	
1988	116

10.3.1 Posiciones de Trazado

Una vez seleccionada la serie con la que se va a realizar el análisis de frecuencia se ordenan los valores de mayor a menor, prescindiendo del año de ocurrencia:

m	Q
1	220
2	196
3	116
⋮	⋮
n = 30	78

Luego es necesario asignarle a cada valor una probabilidad de excedencia. Esta probabilidad de excedencia o frecuencia (P) que se asigna a cada valor de la serie es lo que se conoce como posición de trazado. Su inversa es el período de retorno (T).

A través del tiempo diferentes autores han desarrollado fórmulas para determinar posiciones de trazado (tabla 10.1).

De todas las fórmulas propuestas la que mejor aceptación ha tenido hasta el momento es la de Weibull. En el ejemplo:

m	Q	P	T
1	220	0.032	31
2	196	0.064	15.5
3	116	0.097	10.3
⋮			
n = 30	78	0.968	1.03

Las cifras de la tabla final deben interpretarse así: existe una probabilidad del 6.4% de que el valor 196 m³/seg sea igualado o excedido ~~2 veces en 30 años~~; también: es probable que el caudal de 196 m³/seg se presente una vez cada 15.5 años en promedio
 → 1 vez en 30 años en promedio

TABLA 10.1 POSICIONES DE TRAZADO

N = 10

Nombre	Año	Probabilidad de excedencia (P)	m = 1		m = 5	
			P	T	P	T
California	1923	$\frac{m}{N}$	0.10	10	0.500	2
Hazen	1930	$\frac{2m - 1}{2N}$	0.05	20	0.450	2.2
Weibull	1939	$\frac{m}{N + 1}$	0.091	11	0.455	2.2
Chegodayev	1955	$\frac{m - 0.3}{N + 0.4}$	0.067	14.9	0.451	2.2
Blom	1958	$\frac{m - 3/8}{N + 1/4}$	0.061	16.4	0.451	2.2
Tukey	1962	$\frac{3m - 1}{3N + 1}$	0.065	15.4	0.451	2.2
Gringorten	1963	$\frac{m - 0.44}{N + 0.12}$	0.055	18.2	0.447	2.2

N = número total de valores de la muestra

m = número de orden de los valores ordenados de mayor a menor

10.3.2 Ley de Gumbel

De las varias distribuciones de valores extremos hay dos que tienen mayor aceptación, al haber demostrado que se ajustan bien al fenómeno de las crecidas de los ríos: la distribución de valores extremos tipo I o ley de Gumbel y la distribución log-Pearson tipo III.

Vente Chow ha encontrado que estas distribuciones pueden expresarse en la forma:

$$x = \bar{x} + K \sigma_x$$

x ... caudal con una probabilidad dada

\bar{x} ... media de la serie de caudales pico

σ_x ... desviación estándar de la serie

K ... un factor de frecuencias definido por cada distribución. Es una función del nivel de probabilidad asignado a x.

La ley de Gumbel está dada por la expresión:

$$P = 1 - e^{-e^{-y}} \quad (10.1)$$

P ... probabilidad de que un valor x sea igualado o excedido
 y ... variable reducida, dada por la expresión:

$$y = a(x - u)$$

u ... moda de la distribución

a ... parámetro de dispersión

Para una muestra de tamaño finito, Gumbel encontró que:

$$u = \bar{x} - \sigma_x \frac{\bar{y}_n}{\sigma_n}$$

$$a = \frac{\sigma_n}{\sigma_x}$$

\bar{y}_n ... valor medio esperado de la variable reducida

σ_n ... desviación estándar de la variable reducida.

Y también que \bar{y}_n , σ_n , son funciones sólo del tamaño de la muestra.

n	20	30	40	50	100	200
\bar{y}_n	0.52	0.54	0.54	0.55	0.56	0.57
σ_n	1.06	1.11	1.14	1.16	1.21	1.24

Reemplazando valores:

$$y = ax - au$$

$$= \frac{\sigma_n}{\sigma_x} x - \frac{\sigma_n}{\sigma_x} \left(\bar{x} - \sigma_x \frac{\bar{y}_n}{\sigma_n} \right)$$

$$= \frac{\sigma_n}{\sigma_x} x - \frac{\sigma_n}{\sigma_x} \bar{x} + \bar{y}_n$$

$$\frac{\sigma_n}{\sigma_x} x = \frac{\sigma_n}{\sigma_x} \bar{x} + y - \bar{y}_n$$

$$x = \bar{x} + \frac{y - \bar{y}_n}{\sigma_n} \sigma_x \quad (10.2)$$

$$x = \bar{x} + K \sigma_x \quad (10.3)$$

Con la ecuación 10.3 es posible hallar los caudales con largos períodos de recurrencia (avenida centenaria, avenida milenaria, avenida diezmilenaria). Esta ecuación es la ecuación de una línea recta en papel probabilístico de Gumbel (página 168). Precisamente la manera de comprobar que el modelo de Gumbel es el apropiado para el problema en estudio consiste en graficar la recta y plotear los puntos de la muestra; deberá cumplirse que todos los puntos caen alineados cerca de la recta.

Ejemplo 10.1

Hallar el valor de K para una probabilidad de 1% si se trabaja con una muestra de 40 años.

$$P = 0.01$$

De la ecuación 10.1: $y = 4.600$

$$\begin{aligned} \text{Para } n = 40 \rightarrow \bar{y}_n &= 0.54 \\ \sigma_n &= 1.14 \end{aligned}$$

De modo que:

$$K = \frac{y - \bar{y}_n}{\sigma_n} = 3.55$$

Como $T = \frac{1}{P} = 100$, se indica $K_{100} = 3.55$

Ejemplo 10.2

Usando el registro de caudales que se acompaña, encontrar las crecidas de los 10, 100 y 1,000 años, utilizando la distribución de Gumbel. Los caudales están en pie³/seg.

Año	Q _{máx}	Año	Q _{máx}	Año	Q _{máx}
1969	36,600	1976	53,100	1983	49,600
1970	69,900	1977	58,800	1984	58,600
1971	99,000	1978	64,100	1985	39,700
1972	76,200	1979	77,800	1986	38,200
1973	62,600	1980	71,200	1987	103,000
1974	44,200	1981	59,600	1988	47,900
1975	49,200	1982	55,100		

Ordenando los caudales de mayor a menor:

m	Q	T	m	Q	T
1	103,000	21.0	11	58,600	1.91
2	99,000	10.5	12	55,100	1.75
3	77,800	7.0	13	53,100	1.62
4	76,200	5.25	14	49,600	1.50
5	71,200	4.2	15	49,200	1.40
6	69,900	3.5	16	47,900	1.31
7	64,100	3.0	17	44,200	1.24
8	62,600	2.63	18	39,700	1.17
9	59,600	2.33	19	38,200	1.11
10	58,800	2.1	n = 20	36,600	1.05

Para $n = 20 \rightarrow \bar{y}_n = 0.52$
 $\sigma_n = 1.06$

Para	T = 10	T = 100	T = 1,000
	P = 0.1	P = 0.01	P = 0.001
	y = 2.25	y = 4.60	y = 6.91
	K ₁₀ = 1.63	K ₁₀₀ = 3.85	K ₁₀₀₀ = 6.03

Además: $\bar{x} = 60,720$
 $\sigma_x = 18,233.65$

De modo que:

$$Q_{10} = 90,441 \quad Q_{100} = 130,920 \quad Q_{1000} = 170,669 \text{ pie}^3/\text{seg}$$

Como comprobación, ploteando los tres puntos encontrados (Q,T) en papel probabilístico de Gumbel deben caer en una misma línea recta; además todos los puntos de la muestra deben caer en o cerca de esta recta.

10.3.3 Distribución Log-Pearson Tipo III

Para el uso de esta distribución se convierten los valores de la serie a sus logaritmos decimales y se hallan los siguientes parámetros:

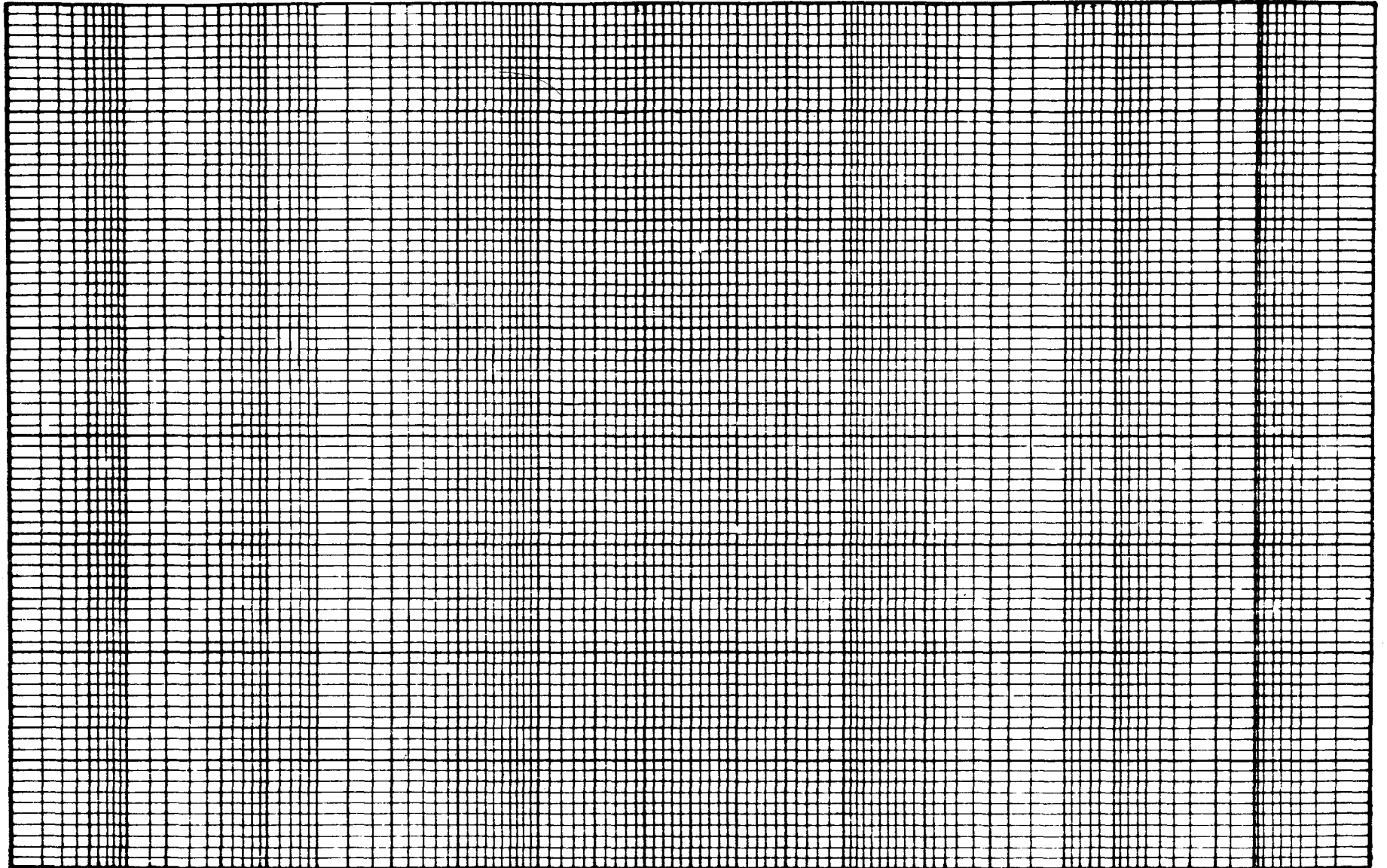
media $\overline{\log x} = \frac{\sum \log x}{n}$

desviación estándar $\sigma_{\log x} = \sqrt{\frac{\sum (\log x - \overline{\log x})^2}{n - 1}}$

Coefficiente de asimetría $Ag = \frac{n \sum (\log x - \overline{\log x})^3}{(n-1)(n-2)(\sigma_{\log x})^3}$

El valor de x para cualquier nivel de probabilidad se puede calcular a partir de:

99.99 99.9 99.8 99 98 95 90 80 70 60 50 40 30 20 10 5 2 1 0.5 0.2 0.1 0.05 0.01



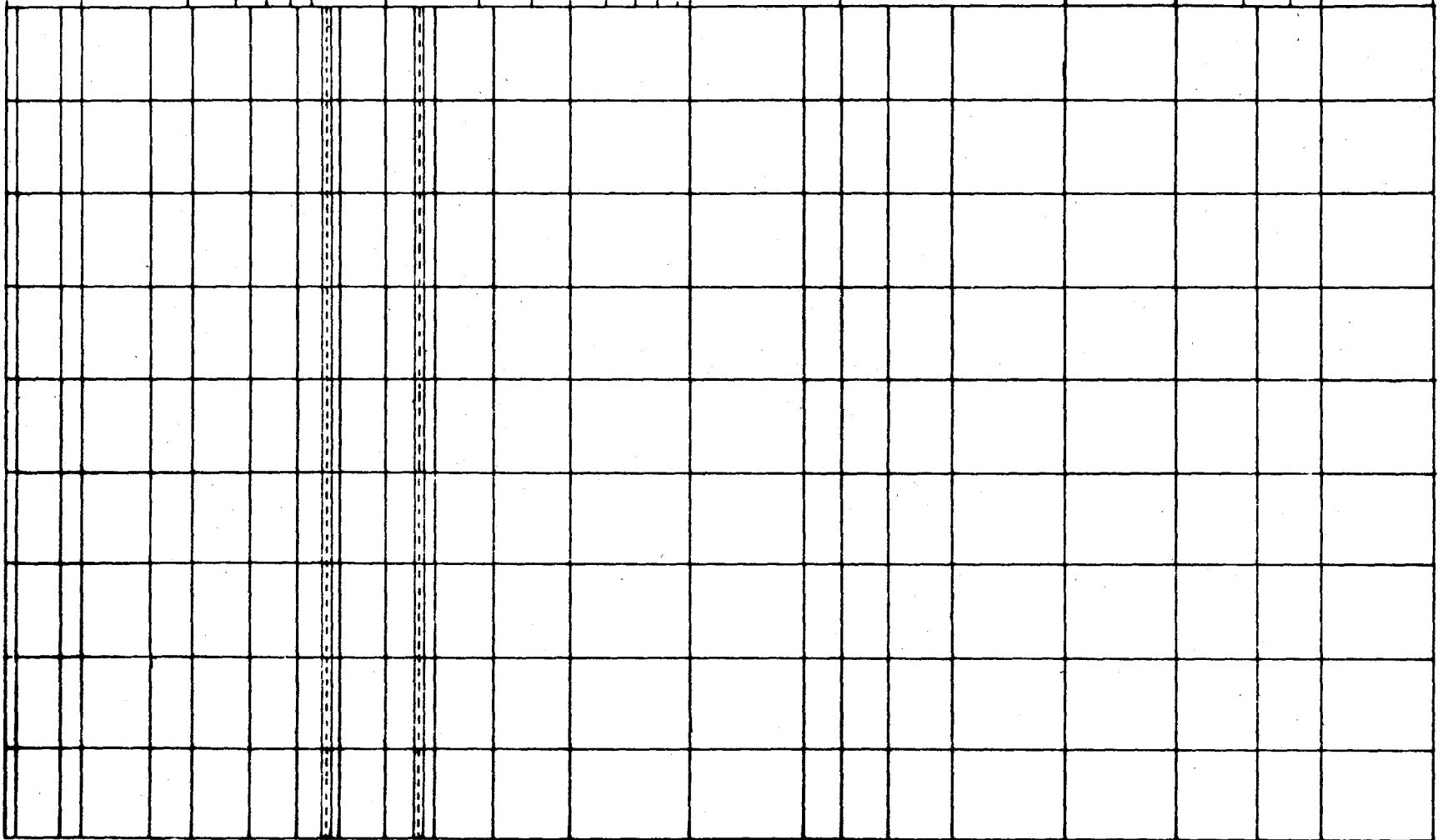
167

0.01 0.05 0.1 0.2 0.5 1 2 5 10 20 30 40 50 60 70 80 90 95 98 99 99.8 99.9 99.99

PAPEL DE PROBABILIDAD NORMAL

PERIODO DE RETORNO (años)

1.1 1.2 1.3 1.4 1.5 2 3 4 5 10 25 50 100 200 300 400 500 1000



168

.1 .5 1.0 5 10 20 30 40 50 60 70 80 90 95 96 97 98 99 99.5 99.7 99.8 99.9

PROBABILIDAD $\left[\frac{m}{n+1} \cdot 100 \right]$

-2.0 -1.5 -1.0 -0.5 0 0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5 5.0 5.5 6.0 6.5 7.0

VARIABLE REDUCIDA

PAPEL DE PROBABILIDAD DE GUMBEL

$$\log x = \overline{\log x} + K \sigma_{\log x} \quad (10.4)$$

Los valores de K se toman de la tabla 10.2

TABLA 10.2 VALORES DE K EN LA ECUACION 10.4

Coeficiente de asimetría	Período de retorno, años							
	1,0101	1,2500	2	5	10	25	50	100
	Nivel de probabilidad, porcentaje							
Ag	99	80	50	20	10	4	2	1
3,0	-0,667	-0,636	-0,396	0,420	1,180	2,278	3,152	4,051
2,8	-0,714	-0,666	-0,384	0,460	1,210	2,275	3,114	3,973
2,6	-0,769	-0,696	-0,368	0,499	1,238	2,267	3,071	3,889
2,4	-0,832	-0,725	-0,351	0,537	1,262	2,256	3,023	3,800
2,2	-0,905	-0,752	-0,330	0,574	1,284	2,240	2,970	3,705
2,0	-0,990	-0,777	-0,307	0,609	1,302	2,219	2,912	3,605
1,8	-1,087	-0,799	-0,282	0,643	1,318	2,193	2,848	3,499
1,6	-1,197	-0,817	-0,254	0,675	1,329	2,163	2,780	3,388
1,4	-1,318	-0,832	-0,225	0,705	1,337	2,128	2,706	3,271
1,2	-1,449	-0,844	-0,195	0,732	1,340	2,087	2,626	3,149
1,0	-1,588	-0,852	-0,164	0,758	1,340	2,043	2,542	3,022
0,8	-1,733	-0,856	-0,132	0,780	1,336	1,993	2,453	2,891
0,6	-1,880	-0,857	-0,099	0,800	1,328	1,939	2,359	2,755
0,4	-2,029	-0,855	-0,066	0,816	1,317	1,880	2,261	2,615
0,2	-2,178	-0,850	-0,033	0,830	1,301	1,818	2,159	2,472
0	-2,326	-0,842	0	0,842	1,282	1,751	2,054	2,326
-0,2	-2,472	-0,830	0,033	0,850	1,258	1,680	1,945	2,178
-0,4	-2,615	-0,816	0,066	0,855	1,231	1,606	1,834	2,029
-0,6	-2,755	-0,800	0,099	0,857	1,200	1,528	1,720	1,880
-0,8	-2,891	-0,780	0,132	0,856	1,166	1,448	1,606	1,733
-1,0	-3,022	-0,758	0,164	0,852	1,128	1,366	1,492	1,588
-1,2	-3,149	-0,732	0,195	0,844	1,086	1,282	1,379	1,449
-1,4	-3,271	-0,705	0,225	0,832	1,041	1,198	1,270	1,318
-1,6	-3,388	-0,675	0,254	0,817	0,994	1,116	1,166	1,197
-1,8	-3,499	-0,643	0,282	0,799	0,945	1,035	1,069	1,087
-2,0	-3,605	-0,609	0,307	0,777	0,895	0,959	0,980	0,990
-2,2	-3,705	-0,574	0,330	0,752	0,844	0,888	0,900	0,905
-2,4	-3,800	-0,537	0,351	0,725	0,795	0,823	0,830	0,832
-2,6	-3,889	-0,499	0,368	0,696	0,747	0,764	0,768	0,769
-2,8	-3,973	-0,460	0,384	0,666	0,702	0,712	0,714	0,714
-3,0	-4,051	-0,420	0,396	0,636	0,660	0,666	0,666	0,667

El método consiste en plotear 2 puntos (Q, T) y trazar la línea recta que los une, en papel de probabilidad log-normal. Se prolonga esta recta para encontrar el caudal con período de retorno grande.

10.3.4 Eventos Históricos

Algunas veces, las posiciones de trazado calculadas deben ser modificadas. Esto sucede cuando existen antecedentes históricos que permiten suponer que en el lugar en estudio han ocurrido crecidas que vale la pena tomar en cuenta. Veamos tres ejemplos de casos en que hay que considerar eventos históricos.

Primer Caso:

Se tiene el registro de los últimos 30 años, pero se sabe que la mayor crecida de este período es la mayor ocurrida en los últimos 80 años. A la mayor crecida del período habrá que asignarle una probabilidad de $1/81$. Para la segunda mayor crecida el cálculo se hace en forma normal ($2/31$, etc).

Segundo Caso:

Se tiene el registro de los últimos 30 años, pero se sabe que en 1909 hubo una crecida, de magnitud conocida, que es la mayor hasta la fecha (80 años). La probabilidad de esta crecida será $1/81$. Para las crecidas medidas el cálculo se hace en forma normal.

Tercer Caso:

Se tiene el registro de los últimos 30 años, pero se sabe que en 1909 ocurrió una crecida, de magnitud conocida, que es la segunda en orden de magnitud en relación con las medidas. Las probabilidades se calculan así:

mayor crecida de todas	$1/81$
segunda crecida de todas	$2/81$
segunda crecida del registro	$2/31$
tercera crecida del registro	$3/31$, etc

10.3.5 Longitud de Registro

Un problema que siempre ha preocupado a los hidrólogos es cuantificar el error cometido al trabajar con registros de pocos años en la estimación de caudales con altos períodos de retorno.

Benson, ingeniero del U.S. Geological Survey, partió de una lista hipotética de 1000 crecidas que definían una recta de frecuencia, de base o referencia, trazada en papel probabilístico de valores extremos. Este registro lo dividió en forma aleatoria en 100 registros de 10 años, 40 de 25 años, 20 de 50 y 10 de 100. Para cada uno de estos 170 registros se dibujó su recta de frecuencia y todas se compararon con la recta base.

Entre la crecida anual media calculada con diferentes longitudes de registro y la crecida anual media base se obtuvieron las siguientes diferencias: 28% para registros de 10 años, 14%.

para registros de 25 años, 12% para los de 50 años y 5% para los de 100 años. Esto da una idea de lo que puede esperarse en la estimación de la crecida anual media con registros cortos.

Se observó que la diferencia entre la crecida anual media base y las crecidas anuales medias determinadas gráficamente en la ordenada correspondiente a un período de retorno de 2.33 años, no diferían en más de 1.5%. Ello permite concluir que es posible calcular gráficamente el promedio sin error apreciable.

Se estudió también los valores de las crecidas de períodos de retorno de 10, 25, 50 y 100 años determinadas con los registros de distintas longitudes, mediante la ley de Gumbel. Los resultados se resumen en la tabla 10.3.

TABLA 10.3 LONGITUDES DE REGISTRO NECESARIAS

T	Precisión 25%		Precisión 10%	
	80% de las veces	95% de las veces	80% de las veces	95% de las veces
2.33	--	12	25	40
10	8	18	38	90
25	12	31	75	105
50	15	39	90	110
100	--	48	100	115

Estos resultados muestran que el 95% de las veces un registro de 39 años de longitud permite definir la crecida de los 50 años, con una precisión del 25%. Se tiene también que el 95% de las veces un registro de 12 años permite definir la crecida anual media con el 25% de error.

Las cifras de este estudio no pueden aplicarse a otras curvas de frecuencia, pero son bastante indicativas de la variación que puede esperarse al trabajar con registros de pequeña longitud. Visto de otro modo, se nota la conveniencia de extender un registro corto antes de realizar con él un análisis de frecuencia.

10.3.6 Probabilidad de Diseño

Frente a un caso concreto el ingeniero debe decidir el período de retorno de su crecida de diseño. Para ello debe precisar en primer término la vida útil de la obra; luego se preguntará sobre la probabilidad de ocurrencia de crecidas durante ese período. Escogiendo un porcentaje adecuado de riesgo determinará el período de retorno de la crecida de diseño y procederá a calcular la magnitud de ésta con la ley de Gumbel o la distri-

bucción log-Pearson tipo III; o mejor aún, con las dos a fin de tomar el promedio.

El período de retorno se ha venido manejando como un valor medio. Es decir, el suceso son período de retorno de T años será excedido una vez cada T años en promedio; pero también será excedido en más de la mitad de las veces en una serie de T años.

Aclaremos. Supongamos un evento de T = 10 años; la probabilidad de no excedencia en un año cualquiera es $(1-0.1) = 0.9$ y la probabilidad de no excedencia durante un período de 10 años es $0.9^{10} = 0.35$. O lo que es lo mismo, la probabilidad de excedencia de un suceso con un período de retorno de 10 años en un período de 10 años es 65%.

La fórmula a usar es pues:

$$(1 - P)^n = \text{probabilidad de no excedencia}$$

$$\left(1 - \frac{1}{T}\right)^n = \text{probabilidad de no excedencia} \quad (10.5)$$

n ... período de diseño o vida útil de la obra.

Con la 10.5 puede calcularse el período de retorno a usar asociado a un período de diseño o vida útil y a un nivel de probabilidad asignado. Así es como se ha construido la tabla 10.4

TABLA 10.4 PERIODOS DE RETORNO

Período de diseño (n)	Probabilidad de no exceder				
	0.01	0.25	0.50	0.75	0.99
2	1.1	2.0	3.4	7.5	200
5	1.7	4.1	7.7	17.9	498
10	2.7	7.7	14.9	35.3	996
20	4.9	14.9	29.4	70.0	1,990
30	7.0	22.2	43.0	105.0	3,330
50	11.4	36.6	72.0	175.0	5,000
100	22.2	72.5	145.0	345.0	10,000

La tabla 10.4 indica que existe una probabilidad igual a 1% de tener un evento con período de retorno de 996 años en un período de diseño de 10 años. Indica también que si se quiere correr un riesgo de 25% para una estructura de vida útil 50 años, debe ser diseñada con un caudal pico de período de retorno 175 años.

Toma de decisión

El estudio de las avenidas de largo período de retorno tiene una aplicación importante en el funcionamiento de los vertederos de demasías de las presas de embalse. Por esta razón el breve comentario que sigue se refiere a estas obras.

La selección de la magnitud de la avenida de proyecto es algo sumamente complejo y que debe ser encarado por ingenieros experimentados en estudios de avenidas. Básicamente se toman en consideración los efectos derivados de un eventual fracaso de la obra. La determinación de los daños probables no debe limitarse a las condiciones que existen en la época de la construcción. Deben, por el contrario, evaluarse los futuros aprovechamientos en la llanura de inundación aguas abajo, como viviendas, granjas, caminos, puentes y otros.

Las presas que forman grandes vasos, construidos en los ríos principales, con un elevado potencial de escurrimiento, sin duda pueden considerarse en la categoría de muy peligrosas. En estos casos deberá asumirse un criterio conservador, tomando como base que no se podría tolerar una falla debido a la probable pérdida de vidas y debido a los daños potenciales que pueden alcanzar las proporciones de una catástrofe. En cambio, las pequeñas presas construidas en corrientes aisladas en áreas rurales, donde la falla ni amenaza la vida humana ni produciría daños mayores, se pueden considerar en la categoría de poco riesgo. En este caso, el criterio de diseño se puede establecer en una base mucho menos conservadora.

En los Estados Unidos, los vertederos de demasías de las grandes presas, cuando existe un gran peligro de vidas humanas, son diseñados para la creciente más desfavorable posible llamada creciente máxima probable (CMP). La experiencia demuestra que la CMP es mayor que aquellas que pueden ser obtenidas utilizando el análisis de frecuencia. A la CMP no se le puede asignar una probabilidad realista ni es susceptible de un estudio económico; el análisis debe estar basado en la decisión de dar la máxima protección. La CMP es muy grande y casi siempre superior a las posibilidades de control; su consideración en el estudio sólo sirve para eliminar la posibilidad adicional de una falla repentina en la presa; aparte de la protección de vidas humanas puede ser útil, por ejemplo, colocar las plantas de agua potable, así como otros servicios públicos, por encima del nivel de la CMP.

La obtención de la CMP es un proceso largo, laborioso y costoso. Se determina a partir de la precipitación máxima probable (PMP). La PMP se puede obtener a partir del análisis de frecuencia de las precipitaciones máximas. A partir de la PMP puede obtenerse la CMP mediante técnicas hidrológicas consistentes en el uso de modelos lluvia-escorrentía e hidrogramas unitarios. Se requiere mucha experiencia. Las condiciones extremas serán condiciones antecedentes muy húmedas, hidrograma unitario con pico muy pronunciado y embalse lleno al presentar se la tormenta.

10.3.7 Método de Grádex

Los métodos basados en distribuciones teóricas para el estudio de frecuencia de crecidas (Gumbel y log-Pearson tipo III), se basan en la utilización de registros de caudales pico. Sin embargo, los datos de caudales son por lo general cortos y relativamente escasos. Guillot y Duband elaboraron una metodología que permite evaluar los gastos máximos de crecidas utilizando registros cortos de caudales conjuntamente con registros largos de lluvias. El método consiste en extrapolar la curva de frecuencia de gastos observados, a partir de cierto valor, mediante una recta paralela a la de precipitaciones máximas diarias.

Guillot y Duband desarrollaron el método del Grádex apoyándose en las siguientes hipótesis:

1. Las precipitaciones máximas diarias siguen la ley de Gumbel o valores extremos tipo I.
2. La lluvia media diaria sobre una cuenca tiene por grádex el promedio de los grádex de las subcuencas.

Se entiende por grádex a la pendiente de la recta precipitación-probabilidad en un papel probabilístico de Gumbel. Se demuestra que es igual a la razón σ_x/σ_n .

3. La retención, a medida que aumenta la lluvia, tiende hacia un límite superior, función de las características de la cuenca y de la humedad del suelo. A medida que se alcanza un equilibrio entre el volumen de agua que llueve y el escurrido todo incremento de precipitación tiende a producir un aumento igual de volumen escurrido.

El método del grádex es aplicable a cuencas de hasta el orden de los 1,000 km² de superficie. En su aplicación se procede analizando los gastos observados hasta definir la crecida decenal y luego se extrapola la recta con el grádex medio de las precipitaciones sobre la cuenca.

De la ecuación 10.2:

$$\begin{aligned}x &= \bar{x} + \frac{y - \bar{y}_n}{\sigma_n} \cdot \sigma_x = \bar{x} + \frac{\sigma_x}{\sigma_n} (y - \bar{y}_n) \\&= \bar{x} + \frac{\sigma_x}{\sigma_n} \cdot y - \frac{\sigma_x}{\sigma_n} \cdot \bar{y}_n \\&= \left(\bar{x} - \frac{\sigma_x}{\sigma_n} \cdot \bar{y}_n\right) + \frac{\sigma_x}{\sigma_n} \cdot y\end{aligned}$$

de modo que σ_x/σ_n , que se define como grádex, viene a ser la pendiente de la recta de frecuencia en papel probabilístico de Gumbel, tal como se había anticipado.

Este método fue aplicado en Chile para alrededor de 40 estaciones, alcanzando resultados bastante aceptables. Encontró que

los resultados no diferían en más de un 10% de las crecidas obtenidas siguiendo el método de Gumbel.

10.3.8 Análisis de Frecuencia Regionales

Los análisis de frecuencia regionales se refieren al establecimiento de relaciones válidas para toda la región. Para ello se requiere contar con registros en estaciones cercanas con características hidrológicas semejantes.

Método de la crecida índice

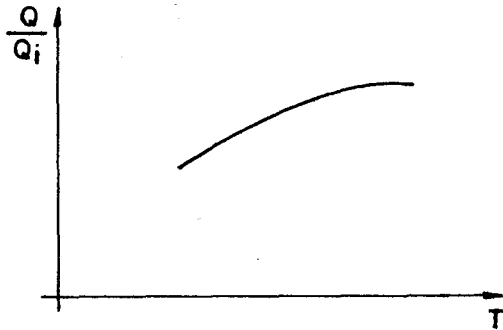
Es el utilizado por el U.S. Geological Survey en sus estudios regionales de frecuencia de crecidas. Consiste fundamentalmente en seleccionar una región de características meteorológicas y fisiográficas similares y efectuar un análisis de frecuencia para cada una de las estaciones en estudio. Estas curvas se adimensionalizan en base a una crecida índice (puede ser la crecida media), luego se superponen y se determina una sola distribución adimensional cuyo único parámetro es la crecida índice. Después se efectúa una correlación entre las crecidas índice con las características de sus cuencas y se obtiene la curva de regresión. La combinación de la distribución adimensional y la curva de regresión permite efectuar el análisis de frecuencia en cualquier cuenca de la región.

El método, paso a paso, es el siguiente:

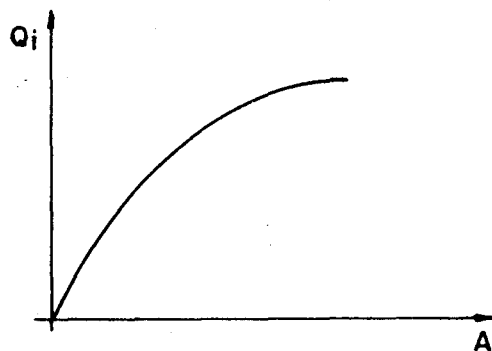
1. Se tabulan las crecidas máximas anuales de cada una de las estaciones existentes en la región, siempre que dicho registro tenga una longitud de 5 ó más años. No se deben utilizar estaciones que estén muy afectadas por obras de regulación artificial o cualquier otro factor artificial, salvo que se corrijan las inconsistencias.
2. Mediante un diagrama de barras se puede representar la longitud de cada registro. En base a este diagrama se elige la longitud del período para el cual se va a hacer el estudio (período común). Hay que tratar de elegir el período común más largo posible.
3. En base a correlaciones se extienden los registros de corta duración para que abarquen un período igual al período común elegido.
4. Se numeran las crecidas para cada estación según su magnitud, asignándole el número 1 a la mayor.
5. Se calcula el período de retorno o probabilidad de excedencia mediante la fórmula de Weibull, solamente para las crecidas medidas. A las crecidas estimadas mediante correlaciones no se les calcula período de retorno ya que no se recomienda graficarlas. Sólo se utilizan para darles un número de orden más realista a las crecidas medidas.
6. Para cada una de las estaciones, graficar las crecidas versus su probabilidad de excedencia en papel probabilístico de Gumbel y trazar las respectivas rectas de distribución.

Con cada una de estas rectas se determina gráficamente la crecida índice. Se recomienda que sea la crecida media y para su determinación se entra al gráfico con una probabilidad de excedencia de 42.9% ($T = 2.33$ años).

7. Tabular para cada una de las estaciones las razones entre el caudal para varios períodos de retorno y la crecida índice.
8. A partir de la tabla del punto 7, determinar la mediana de las razones para cada período de retorno. Graficar estos valores y luego pasar por ellos una curva que representará la curva de frecuencias regional y cuyo único parámetro es la crecida índice (Q_i).



9. En base a los datos de las crecidas índice para cada estación y a las características físicas y meteorológicas de cada cuenca determinar una ecuación de regresión con Q_i como variable dependiente. Las características físicas y meteorológicas de las cuencas son múltiples y de difícil evaluación. Afortunadamente se ha demostrado en la práctica que en la mayoría de las regiones se obtiene una buena correlación utilizando únicamente el área de la cuenca como variable independiente.



10. Utilizando las curvas determinadas en 8 y 9 se puede obtener una curva de frecuencia para cualquier sección dentro de la región en estudio.

10.3.9 Resumen del Estudio de Avenidas

A continuación se presenta un comentario de los aspectos más relevantes en el estudio de avenidas.

En cuanto a lo que se persigue.- Unas veces sólo se requiere

el caudal máximo, como en el diseño de alcantarillas, otras se requiere el hidrograma de la avenida, como en el estudio del tránsito.

En cuanto a las condiciones precedentes.- De acuerdo a la climatología del lugar, ha de precisarse si la avenida se presenta cuando en el río hay un flujo base apreciable o no. Si se están manejando precipitaciones la precisión debe ser en el sentido de que si la precipitación de diseño se presenta cuando el suelo de la cuenca está húmedo o no.

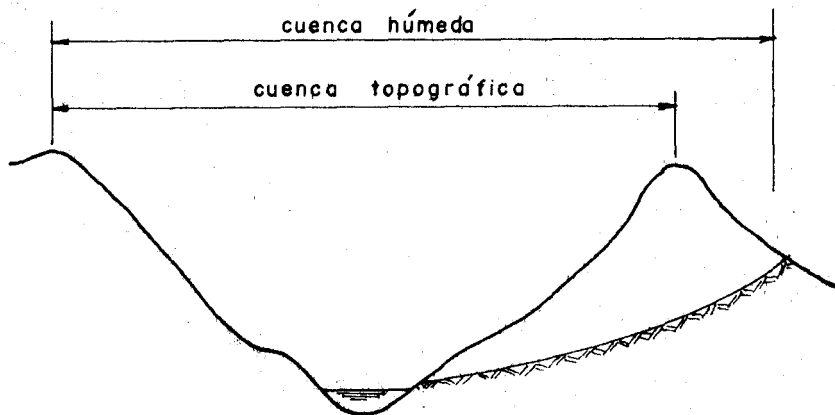
En cuanto a la información disponible.- Se presenta toda una gama de posibilidades.

- a) El caso ideal lo constituye la cuenca con estación de aforo en el punto de interés y estaciones pluviométricas repartidas, una de las cuales por lo menos es registradora. Se puede en tal caso aplicar el análisis de frecuencia a los caudales y volúmenes extremos y también se puede hacer el análisis de los hidrogramas de crecidas.
- b) Igual que el anterior, con la diferencia que es corto el período de registro de caudales. Se puede aplicar el método del Grádex.
- c) Igual que los casos a) y b), sólo que los caudales no son medidos sino inferidos a partir de valores medidos en el mismo río, aguas arriba y/o aguas abajo del punto de interés. Se aplican los mismos métodos. Se hace notar que deben evitarse los análisis de frecuencia de caudales con valores inferidos de otras cuencas.
- d) Sólo se conocen los caudales. Se aplica el análisis de frecuencia para obtener los valores de la descarga y el volumen de los T años. El hidrograma correspondiente puede obtenerse aplicando el hidrograma básico.
- e) La cuenca posee sólo las estaciones pluviométricas repartidas, una de las cuales por lo menos es registradora. Se aplica el análisis de frecuencia a las precipitaciones para obtener el valor de la precipitación de los T años y duración igual al tiempo de concentración (T_C) de la cuenca. Luego se estima la escorrentía directa correspondiente a esta precipitación aplicando el método del hidrograma unitario triangular.
- f) Igual que el anterior, pero sin ninguna estación registradora. Se puede estimar la escorrentía directa con los datos de suelos y cubierta vegetal. Es posible, además, utilizar el hidrograma básico para obtener el hidrograma aproximado de la avenida.
- g) La cuenca no posee ningún tipo de estación. Se puede utilizar el método de la sección y la pendiente para estimar el caudal máximo de la corriente. Se puede también realizar un análisis de frecuencia regional. Es posible además utilizar el hidrograma sintético para generar el hidrograma aproximado de la avenida.

En cuanto al trabajo de campo,.- El estudio de avenidas, como parte del estudio hidrológico de un proyecto, es necesario que esté a cargo de un ingeniero con mucha experiencia de campo. En el campo deben reunirse todos los datos que se refieren a las características hidrológicas de la cuenca. Debe prepararse un mapa a curvas de nivel mostrando la delimitación de la cuenca y la ubicación de cada estación hidrométrica y/o pluviométrica. El ingeniero que hace el estudio de las avenidas debe hacer un recorrido de inspección sobre la cuenca, a fin de verificar las divisorias del área de drenaje y los datos de cubierta vegetal. El ingeniero experimentado sabe que ciertos detalles topográficos ejercen una influencia importante en la magnitud de la precipitación, y que algunas formaciones geológicas tienen la característica de producir elevadas descargas en las avenidas mientras que otras tienden a reducir el potencial de las avenidas. El recorrido debe incluir visitas a las cuencas vecinas si es que se prevé el uso de los registros de las cuencas vecinas en el estudio.

Casos especiales

1. Las relaciones lluvia-escorrentimiento se establecen para la hipótesis general de que la cuenca húmeda coincide con la cuenca topográfica. En estas circunstancias el coeficiente de escorrentía C tiene valores menores que la unidad. Hay ocasiones sin embargo en que esto no es así. La figura adjunta muestra cómo la presencia de un estrato impermeable puede hacer que el valor del coeficiente de escorrentimiento resulte mayor que la unidad.



Este caso requiere un estudio especial para determinar la cantidad de agua de escorrentía de que se puede disponer.

2. Las descripciones hechas en este texto, acerca de las avenidas, se refieren al caso general de avenidas que resultan de las precipitaciones pluviales sobre una cuenca sin deshielos. Hay ocasiones, sin embargo, en que los escurrimientos producidos por los deshielos son una característica dominante. En estos casos el gasto de la corriente proveniente de los deshielos puede ser grande y debe añadirse al gasto calculado para la lluvia. Además, cuando se estima el escurrimiento producido por las lluvias en estas áreas, debe considerarse que parte de la cuenca está cu-

bierta por nieve. El escurrimiento de la lluvia que cae en la nieve puede llegar a ser igual a la precipitación. Por lo tanto, las pérdidas totales por retención para la lluvia en una cuenca parcialmente cubierta por la nieve serán menores que las de la misma cuenca cuando no está cubierta por la nieve.

3. Ciertas cuencas poseen accidentes o medios que retardan el escurrimiento, como lagunas, pantanos, hojarasca y tierra vegetal porosa. Estas cuencas requieren también un estudio especial que se encarga a ingenieros con mucha experiencia.

10.4 Análisis de Frecuencia de Valores Medios

El análisis de frecuencia de valores medios (por ejemplo, gastos medios mensuales, precipitaciones anuales, volúmenes de escorrentía, etc) se efectúa con distribuciones como la normal o log normal.

En esta ocasión sólo se verán dos aspectos importantes del análisis de frecuencia de valores medios: la curva de gastos clasificados o curva de duración y el tratamiento especial que conviene darle a las series que contienen valores ceros.

Curva de Duración

En el apartado 6.5 fue descrita esta curva y la forma cómo se construye. Ahora vamos a ver de qué manera se obtiene la curva de duración a partir de registros cortos.

Una curva de duración será más representativa de las condiciones de escurrimiento en una cuenca mientras mayor sea el período de tiempo que abarquen los datos. Si se cuenta con un registro corto, se puede dibujar una curva de duración representativa de un período más largo mediante el siguiente método recomendado por el U.S. Geological Survey.

1. Se dibuja una curva de duración para la estación en estudio (A) con los registros que ella posee.
2. Se elige otra estación B de régimen fluviométrico similar y con registro más extenso. Se construye una curva de duración de esta estación B para un período de registro concurrente con el de la estación A.
3. A partir de las curvas dibujadas en 1 y 2 se determinan los caudales para varias probabilidades de excedencia. Estos puntos se llevan a un gráfico y se determina una curva de relación entre ambas estaciones.
4. Se construye luego la curva de duración de B para todo el período que tiene registros.
5. En base a las curvas determinadas en 3 y 4 se construye la curva de duración de A representativa del período de larga duración. Para ello, se entra a la curva determinada en 4 con una probabilidad de excedencia y se obtiene un caudal para B. Con ese caudal se en

tra a la curva determinada en 3 y se tiene el caudal para A con la probabilidad de excedencia dada. Repitiendo este procedimiento para varias probabilidades se puede trazar la nueva curva de duración para la estación A.

La selección de la estación de registros más largo debe hacerse considerando que debe tener condiciones climáticas y fisiográficas similares a la de la cuenca en estudio. Es por ello que debe preferirse seleccionar una estación en el mismo río. Además, la estación seleccionada y la estación en estudio deben tener un período concurrente de registros suficiente como para poder establecer una curva de relación aceptable. Otro factor que es conveniente considerar al seleccionar la estación es su distancia a la estación en estudio. A igualdad de otros factores, estaciones cercanas darán mejores relaciones que estaciones más alejadas.

Estimación de curvas de duración.

En algunas oportunidades se requieren curvas de duración en ríos que no tienen datos fluviométricos. Cuando la parte baja de la curva (caudales pequeños) no es de gran interés, se puede recurrir a la trasposición desde cuencas vecinas de curvas de duración por unidad de área. En cambio, si interesa la distribución de los caudales chicos la trasposición puede no ser aceptable ya que ellos dependen en gran medida de las características geológicas de la cuenca.

Análisis de frecuencia de datos con ceros

Al intentar hacer un análisis de frecuencia de cierto tipo de datos como por ejemplo, datos de valores medios, crecidas en cuencas pequeñas de regiones áridas, etc, es fácil encontrar muestras que contengan valores cero. El problema se presenta porque las distribuciones comunes no se ajustan a una serie con esas características. Por esta razón, Jennings y Benson desarrollaron un método basado en probabilidades condicionadas que permite realizar en forma fácil el análisis de frecuencia de una serie del tipo señalado.

En la figura se indica el espacio muestral y los diferentes eventos que se van a considerar.

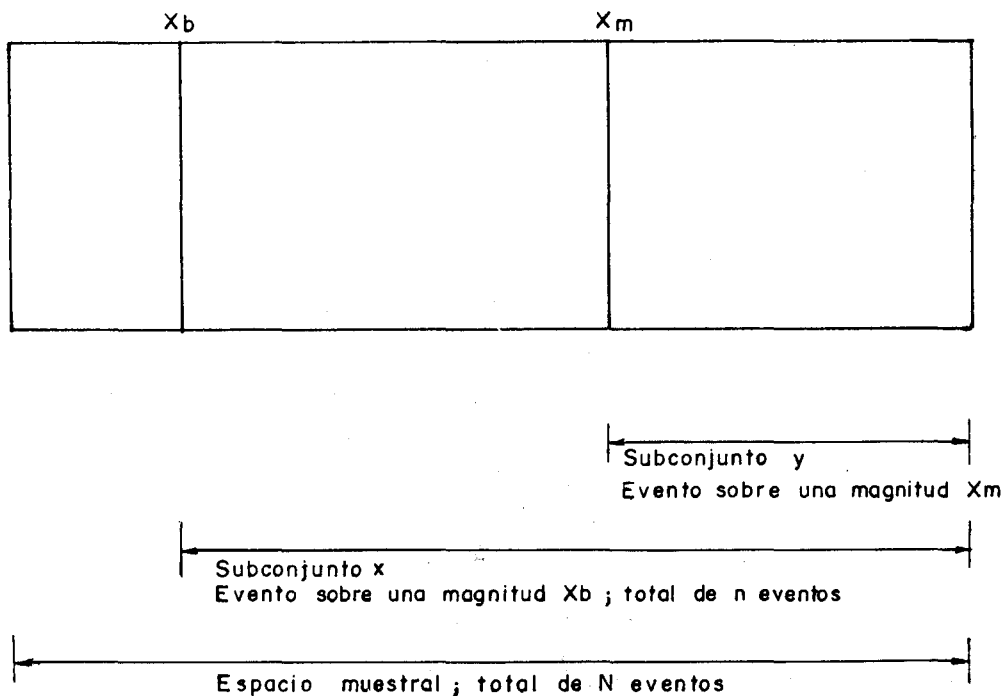
La fórmula para probabilidades condicionadas indica que:

$$P(y/x) = \frac{P(y \cdot x)}{P(x)}$$

De acuerdo a la figura la intersección de y,x es y de modo que:

$$P(y/x) = \frac{P(y)}{P(x)}$$

Luego: $P(y) = P(x) \cdot P(y/x)$



La probabilidad de que ocurra el evento x es igual a n/N , de manera que:

$$P(y) = \frac{n}{N} P(y/x)$$

La última ecuación expresada en términos hidrológicos y haciendo $x_b=0$, quedaría:

$$\text{Prob (evento} > x_m) = \frac{n}{N} \text{Prob (evento} > x_m / \text{evento} > 0) \quad (10.6)$$

De acuerdo con lo anterior, para obtener las probabilidades de excedencia de una serie con valores nulos, se efectúa un análisis de frecuencia de la serie que contiene todos los valores mayores que cero y luego se multiplican estas últimas probabilidades por la razón n/N .

10.5 Análisis de Frecuencia de Precipitaciones

Lo estudiado sobre análisis de frecuencia de crecientes se aplica también a las precipitaciones. No obstante los numerosos estudios comparativos realizados con las diferentes distribuciones no se ha llegado a una conclusión general definitiva.

Los valores de la precipitación máxima horaria o diaria se ajustan bien a distribuciones tales como la de valores extremos tipo I y log-Pearson tipo III.

En zonas húmedas el valor medio es alto y por lo tanto la precipitación mensual, estacional o anual se aproximan bien a una distribución normal. En zonas áridas los valores se ajustan mejor a distribuciones como la log-normal y las transformadas de la distribución normal.

Curvas intensidad-duración-frecuencia

En el apartado 2.7 se describió la forma de obtener las curvas intensidad-duración-frecuencia. En el manejo de la fórmula racional (apartado 7.3) se vió la aplicación de estas curvas para obtener el valor de i que entra en dicha fórmula.

Lamentablemente, los registros pluviográficos son escasos, de modo que el procedimiento descrito para obtener las curvas pocas veces es aplicable en la práctica. Esta situación ha llevado a algunos investigadores a buscar procedimientos que permitan superar tal deficiencia, aprovechando al máximo la información realmente existente. Frederick Bell (1969) publicó un trabajo en el cual generalizaba las curvas intensidad-duración-frecuencia, a partir de datos recogidos principalmente en Estados Unidos. El argumento físico en que se apoyó es el hecho de que las lluvias extremas de menos de dos horas de duración se deben a tormentas de tipo convectivo, las cuales poseen características similares en todas las regiones del mundo.

La expresión matemática propuesta por Bell es la siguiente:

$$p_t^T = (0.21 T + 0.52)(0.54 t^{0.25} - 0.50) p_{60}^{10} \quad (10.7)$$

t ... duración en minutos

T ... período de retorno en años

p_t^T ... precipitación caída en t minutos con período de retorno T años.

La fórmula es aplicable a lluvias de menos de dos horas de duración y con períodos de retorno comprendidos entre 2 y 100 años. Estudios hechos en diferentes partes del mundo han conducido a valores sensiblemente iguales para todos los lugares. Se observa que se requiere conocer la precipitación de una hora de duración y 10 años de período de retorno.

En realidad no siempre se cuenta con información de lluvias de una hora de duración. Pero Espíldora, obtuvo en Chile que la relación entre la lluvia máxima diaria y la lluvia de una hora es más o menos constante e igual a 4.04. Esto hace posible obtener la lluvia p_{60}^{10} que entra en la fórmula, a partir de las lluvias máximas diarias cuyos registros son más frecuentes.

El procedimiento completo entonces viene a ser el siguiente. A partir de los registros de precipitaciones máximas diarias obtener, mediante un análisis de frecuencia, la magnitud de la lluvia con período de retorno 10 años. Usando el coeficiente de Espíldora tener p_{60}^{10} y luego aplicar la fórmula de Bell. Por último, calcular, a partir de las magnitudes encontradas de lluvia, las intensidades correspondientes a fin de poder construir las curvas intensidad-duración-frecuencia.

Mapas de lluvia-duración-frecuencia

En vez de intensidades se grafican otras veces las cantidades de lluvia, siempre para diferentes duraciones y frecuencias, lográndose confeccionar mapas regionales e incluso para todo el país. Estados Unidos, por ejemplo, tiene mapas de lluvia para duraciones entre 30 minu-

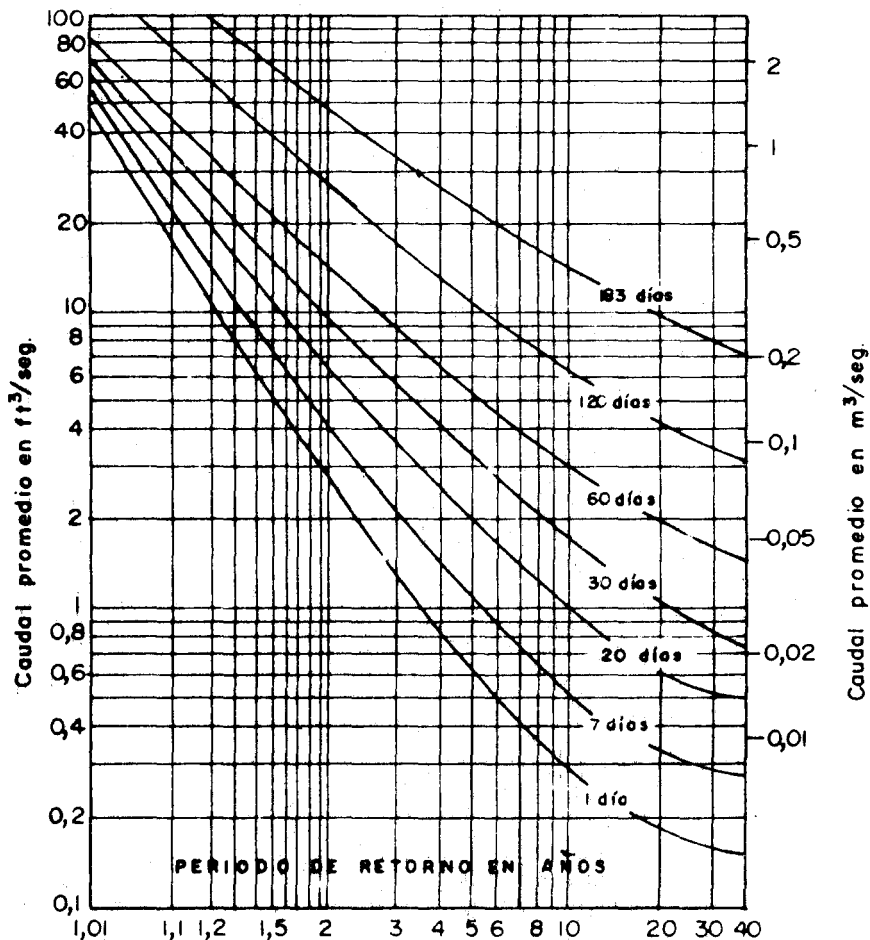
tos y 24 horas, para períodos de retorno entre 1 y 100 años. Estos mapas se elaboran a partir de precipitaciones en parte observadas y en parte estimadas a través de correlaciones múltiples usando parámetros como la pendiente, la elevación, las distancias, etc.

Todos los mapas de lluvia-duración-frecuencia muestran datos puntuales de lluvia, pero se supone que estos valores son aplicables para áreas de hasta 25 km². La precipitación media en una área debe ser menor que la precipitación máxima puntual en el área.

10.6 Análisis de Frecuencia de Sequías

Se acostumbra definir la sequía en términos de períodos fijos de tiempo con menos de una cantidad mínima de agua (caudal o lluvia).

Por lo general interesan los caudales bajos. Se puede en este caso seleccionar el caudal mínimo cada año durante períodos de diversas longitudes (1 día, 7 días, 30 días, etc) y aplicar a esta serie el análisis de frecuencia según la distribución log normal. También se puede utilizar el modelo de Gumbel, pero con escala logarítmica para los caudales debido a los amplios límites de los valores. Las líneas de frecuencia ya no resultan rectas como en el análisis de crecidas, sino curvas como muestra la figura.



La frecuencia de la sequía en términos de la lluvia puede presentarse en un gráfico similar, que mostraría la frecuencia de varias cantidades mínimas de lluvia en períodos de diversas longitudes.

10.7 Breve Mención de los Procesos Estocásticos

10.7.1 Introducción

Sea una variable x . Si el comportamiento de x puede predecirse con certeza, x es una variable determinística. Si no, x es una variable aleatoria o random, pudiendo decirse que x está gobernada por las leyes de la probabilidad.

Asumamos ahora que el comportamiento de x puede observarse de manera secuencial x_1, x_2, \dots , en que los subíndices representan intervalos de tiempo. Dicha secuencia se llama serie de tiempo. A estas series nos referimos en este apartado.

Las series de tiempo son series estocásticas. Y el juego de variables aleatorias x_1, x_2, \dots , asociadas según su distribución de probabilidad se llama proceso estocástico. El término estocástico tiene pues una significación más amplia que el término probabilístico.

Los procesos estocásticos en hidrología pueden representarse de dos maneras, en forma discreta (fig. a) o en forma continua (fig. b), siendo la segunda la más común.

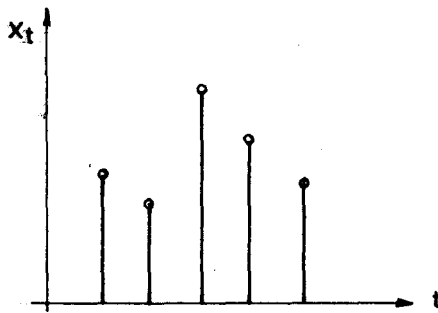


Fig. a

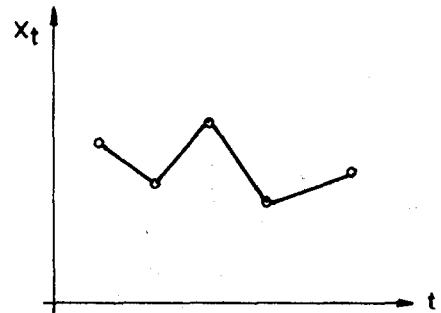


Fig. b

Aparte de la función de distribución $f(x_1, x_2, \dots)$, conviene indicar algunas propiedades, como valor esperado, varianza y covarianza. El valor esperado de un proceso estocástico x_1, x_2, \dots , se compone del juego de valores esperados de cada posición en el tiempo: $E(x_1), E(x_2), \dots$. El juego de varianzas es $\text{Var}(x_1), \text{Var}(x_2), \dots$. También se usa la notación:

$$u_t = E(x_t)$$

$$\sigma_t^2 = \text{Var}(x_t), t = 1, 2, \dots$$

Considerando dos posiciones cualesquiera t y $t-K$, la covarianza entre las variables x_t y x_{t-K} se representa por:

$$\text{Cov}(K) = \text{Cov}(x_t, x_{t-K})$$

La covarianza es la propiedad que describe la dependencia lineal del proceso estocástico.

Un proceso estocástico (serie de tiempo) es estacionario en el promedio si los valores esperados no varían con el tiempo:

$$E(x_1) = E(x_2) = \dots = E(x_t) = E(x) = u$$

Análogamente, cuando $\text{Var}(x_t) = \sigma^2$, $t = 1, 2, \dots$ es una constante, el proceso estocástico es estacionario en la varianza.

Un proceso estocástico es estacionario en la covarianza cuando la covarianza depende sólo del retardo K pero no depende de la posición t :

$$\text{Cov}(x_t, x_{t-K}) = \text{Cov}(K), \text{ sin importar } t$$

Un proceso estocástico es estacionario de primer orden cuando es estacionario en el promedio y estacionario de segundo orden cuando es estacionario en el promedio y en la covarianza (se hace notar que estacionario en la covarianza implica estacionario en la varianza).

En las definiciones que se han dado, en vez de usar el término proceso estocástico estacionario se puede decir también series de tiempo estacionarias o simplemente series estacionarias.

Modelos

Un modelo matemático representando un proceso estocástico se llama modelo estocástico o modelo de serie de tiempo. Tiene una cierta forma matemática o estructura y un juego de parámetros.

Un modelo de serie de tiempo simple podría estar representado por una simple función de distribución de probabilidad.

$$f(x; \theta)$$

con parámetros $\theta = \{\theta_1, \theta_2, \dots\}$ válidos para todas las posiciones $t = 1, 2, \dots$ y sin ninguna dependencia entre x_1, x_2, \dots

Por ejemplo, si x es normal con promedio u y varianza σ^2 , el modelo de la serie de tiempo puede escribirse:

$$x_t = u + \sigma \epsilon_t \quad t = 1, 2, \dots \quad (10.8)$$

donde ϵ_t es también normal con promedio cero, varianza uno y $\epsilon_1, \epsilon_2, \dots$ son independientes.

El modelo 10.8 tiene parámetros u, σ y desde que ellos son constantes (no varían con el tiempo) el modelo es estacionario.

Un modelo de serie de tiempo con estructura dependiente puede formarse como:

$$\epsilon_t = \phi \epsilon_{t-1} + \xi_t \quad (10.9)$$

donde ξ_t es una serie independiente con promedio cero y varianza $(1 - \phi^2)$. ε_t es la serie dependiente y ϕ es el parámetro del modelo.

Si ε_t de 10.8 fuera representada por el modelo dependiente 10.9 entonces x_t sería también un modelo dependiente. En este caso los parámetros del modelo de x_t serían u , σ y ϕ .

Desde que los parámetros de los modelos recién considerados son constantes los modelos son estacionarios, representando series de tiempo estacionarias o procesos estocásticos estacionarios.

Los modelos no estacionarios resultan cuando dichos parámetros varían con el tiempo.

Modelaje de series de tiempo

El modelaje de series de tiempo es un proceso que puede ser simple o complejo, dependiendo de las características de la serie que se tiene, del tipo de modelo a usar y de las técnicas de modelaje que se sigan.

Por ejemplo, series con características estadísticas que no varían con el tiempo generalmente conducen a modelos y técnicas de modelaje que son más simples que aquéllas de series con características que sí varían con el tiempo.

Hay varios modelos estocásticos que pueden usarse para representar una serie de tiempo, unos más complejos que otros.

Para cada modelo hay técnicas conocidas para estimar los parámetros del modelo y para probar cuán bueno es el modelo; aquí también unas técnicas son más complejas que otras.

En general, los pasos en el modelaje son:

1. Selección del tipo de modelo.
2. Identificación de la forma o el orden del modelo (número de parámetros).
3. Estimación de los parámetros del modelo.
4. Comprobación de la bondad del modelo.

El modelaje es por eso un proceso iterativo.

10.7.2 Modelos de Series de Tiempo

En 1914 Hazen mostró ya la posibilidad de usar la teoría de la estadística y la probabilidad en el análisis de la secuencia de los caudales en un río. Pero no fue sino por los años 60 que se inició el desarrollo formal del modelaje estocástico.

Los principales modelos estocásticos conocidos se señalan a continuación, haciéndose notar que no es única la manera de expresar cada uno de ellos.

1) Modelo de filtro lineal

La idea es que una serie de tiempo en la que los valores sucesivos tienen alta dependencia, puede ser considerada como generada por una serie de variables aleatorias a_t generadas por una distribución única de valor esperado cero y varianza conocida σ_a^2 . Usualmente estas variables son normales y en este caso se conocen como ruido blanco.

El proceso queda definido entonces por el modelo:

$$Z_t = u + a_t + \phi_1 a_{t-1} + \phi_2 a_{t-2} + \dots \quad (10.10)$$

donde u es un parámetro que determina el nivel del proceso y la secuencia ϕ_1, ϕ_2, \dots puede ser finita o infinita. Si la secuencia es finita o infinita pero convergente el proceso es estacionario.

2) Modelo de autorregresión

Este modelo de proceso estocástico expresa el valor del proceso en el instante t , como una combinación lineal de p valores anteriores y de una variable aleatoria a_t con valor esperado cero y varianza σ_a^2 conocida.

Sea $Z_t, Z_{t-1}, \dots, Z_{t-p}$

$$\bar{Z}_t = Z_t - u$$

Entonces el proceso

$$Z_t = \bar{Z}_t + u$$

$$\bar{Z}_t = \phi_1 \bar{Z}_{t-1} + \phi_2 \bar{Z}_{t-2} + \dots + \phi_p \bar{Z}_{t-p} + a_t \quad (10.11)$$

se denomina proceso autorregresivo de orden p . En forma abreviada AR(p). También se conoce con el nombre de proceso markoviano.

El modelo contiene $p+2$ parámetros por determinar $u, \phi_1, \phi_2, \dots, \phi_p, \sigma_a^2$ y es posible hacer ver que se trata de un caso particular del modelo de filtro lineal.

El modelo AR(1) resulta de considerar solamente un término autorregresivo en la ecuación 10.11:

$$Z_t = \bar{Z}_t + u$$

$$\bar{Z}_t = \phi_1 \bar{Z}_{t-1} + a_t$$

y se llama modelo autorregresivo de primer orden o modelo Markov de primer orden.

De manera similar, el modelo AR(2) se obtiene haciendo $p=2$ en la ecuación 10.11.

Estudiar las propiedades de los modelos AR es referirse al valor esperado, la varianza, la función de autocorrelación o correlograma y las condiciones que deben reunir los parámetros del modelo.

3) Modelo de promedios móviles

El proceso estocástico definido por un modelo de promedios móviles representa el valor de \bar{Z}_t como combinación lineal de q valores de una secuencia de variables aleatorias a_t , con valor esperado cero y varianza σ_a^2 .

Entonces:

$$Z_t = \bar{Z}_t + u$$

$$\bar{Z}_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \quad (10.12)$$

se denomina proceso de promedios móviles de orden q . En forma abreviada MA(q). El nombre de promedio móvil se conserva por tradición, pues los valores $1, \theta_1, -\theta_2, \dots, -\theta_q$, no deben ser necesariamente positivos ni sumar uno. Este proceso contiene $q+2$ parámetros por determinar a partir de los datos de las series.

4) Modelo autorregresivo de promedios móviles

Un tipo de proceso que permite una extraordinaria flexibilidad para ajustar series reales es uno que combina las características de los dos modelos anteriores. Es así como se puede definir un proceso:

$$Z_t = \bar{Z}_t + u$$

$$\bar{Z}_t = \phi_1 \bar{Z}_{t-1} + \dots + \phi_p \bar{Z}_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}$$

(10.13)

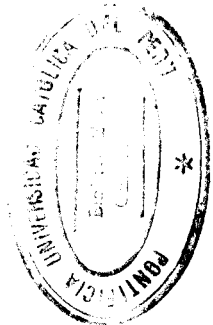
que necesita $p+q+2$ parámetros que deben ser estimados. El nombre abreviado de este modelo es ARMA.

5) Modelos no estacionarios

Se trata de una generalización de los modelos previamente mencionados. Se trata de representar series cuyo comportamiento no es estacionario y en particular que no varían en torno a un valor medio, pero que sin embargo tienen un cierto tipo de comportamiento homogéneo.

Un proceso de esta naturaleza puede ser representado porque sus diferencias de orden d conforman un proceso estacionario tipo ARMA, es decir si

$$W_t = \nabla^d Z_t$$



donde ∇^d representa las diferencias de orden d del proceso Z_t :

$$W_y = \phi_1 W_{t-1} + \dots + \phi_p W_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \quad (10.14)$$

El proceso así definido se denomina proceso autorregresivo de promedios móviles integrado, de orden (p, d, q) . En forma abreviada ARIMA.

Es interesante hacer notar dos elementos importantes de este tipo de procesos. El primero es que puede ser considerado como la aplicación en serie de tres modelos de filtro lineal con coeficientes especiales, y la segunda es que ha sido empleado con gran éxito en la representación de series de tiempo estacionales.

10.7.3 La Función de Autocorrelación

La función de autocorrelación mide la dependencia lineal que existe entre la serie dada y la misma serie desplazada en el tiempo.

Este concepto permite analizar una serie de tiempo e identificar parcialmente el proceso subyacente si la función de autocorrelación tiene la estructura esperada.

Para definir formalmente la función de autocorrelación o correlograma es necesario definir previamente la autocovarianza para un desplazamiento K :

$$\gamma_K = E |(Z_t - u)(Z_{t+K} - u)| \quad (10.15)$$

Para procesos estacionarios, la función de autocorrelación se puede definir entonces como:

$$r(K) = \frac{\gamma_K}{\gamma_0} \quad K = 0, 1, \dots \quad (10.16)$$

donde $\gamma_0 = E |(Z_t - u)^2|$ es la varianza constante del proceso

Estimación de correlogramas

Hasta aquí se ha considerado solamente la función teórica de autocorrelación que describe un proceso estocástico. En la práctica se tiene una serie de tiempo finita Z_1, \dots, Z_n , con N observaciones, de la cual sólo se puede obtener una estimación de la función.

Se han sugerido diversas estimaciones de la función, y cada su gerencia tiene características y propiedades particulares. Se ha concluido que la estimación más satisfactoria del valor de la función de autocorrelación para un desplazamiento K es:

$$r = \frac{C_K}{C_0} \quad (10.17)$$

siendo C_K la estimación de la autocovarianza, según:

$$C_K = \frac{1}{N} \sum_{t=1}^{N-K} (Z_t - \bar{Z}) \cdot (Z_{t+K} - \bar{Z}) \quad (10.18)$$

$$K = 0, 1, 2, \dots$$

$$\bar{Z} = \frac{1}{N} \sum_{t=1}^N Z_t$$

Técnicas de análisis de series de tiempo

Las técnicas principales de análisis de series de tiempo utilizan herramientas matemáticas bastante desarrolladas que no corresponde tratar en este texto. Sin embargo, hay dos formas que son usualmente utilizadas y que no requieren de un desarrollo teórico muy complicado. Una es mediante los correlogramas y la otra es mediante el desarrollo en series de Fourier de la serie de tiempo.

10.7.4 Aplicaciones del Modelaje en Hidrología

El modelaje de series de tiempo en Hidrología tiene dos aplicaciones globales:

1. Para la generación de series hidrológicas de tiempo sintéticas.
2. Para la predicción de series hidrológicas de tiempo futuras.

Se requiere la generación de series sintéticas en los siguientes casos:

- a) Para dimensionamiento de reservorios.
- b) Para planear expansiones de la capacidad de sistemas de suministro.
- c) Para determinar el riesgo de falla de suministro de agua para irrigaciones.
- d) Para determinar el riesgo de falla de capacidades confiables de centrales hidroeléctricas.
- e) En casos similares.

Se requiere la predicción de series de tiempo futuras en los siguientes casos:

- 1) Para planeamiento a corto plazo de operación de reservorios.

- 2) Para planeamiento de operación durante sequías.
- 3) En casos similares.

10.7.5 Reflexiones Acerca del Modelaje

Siempre existen diferencias entre los modelos reales y los estimados, y entre los parámetros de los modelos reales y los estimados. Dichas diferencias representan las incertidumbres del modelaje. Una forma de disminuir tales incertidumbres es mediante la selección del modelo que mejor represente la realidad física del sistema. Algunas veces puede ser posible usar leyes físicas para inferir cuáles serían las expresiones matemáticas del correspondiente modelo estocástico de serie hidrológica.

El modelaje de procesos de descargas ha seguido esencialmente dos caminos: el determinístico o simulación física del sistema hidrológico y el estadístico o simulación estocástica del sistema. En el primero hay siempre una correspondencia única entre la entrada, digamos precipitación, y la salida, digamos descarga. En el segundo los modelos más ampliamente seguidos son los modelos autorregresivos. Claro que también se han usado otros modelos determinísticos o estocásticos. Actualmente la tendencia es la de conciliar ambos puntos de vista. Por un lado el método determinístico trata la precipitación como una variable aleatoria y por otro se busca una justificación física de los modelos estocásticos.

Ejemplo 10.3

Si la recesión del flujo en el río es de la forma

$$Q_t = Q_0 e^{-Kt}$$

Q_0 ... caudal al inicio del año cronológico

K ... constante de recesión

Entonces la serie anual dependiente de descargas Z_i seguirá el modelo autorregresivo de primer orden, AR(1):

$$Z_i - \bar{Z} = \phi (Z_{i-1} - \bar{Z}) + \epsilon_i$$

\bar{Z} ... promedio de Z_i

ϕ ... coeficiente de autorregresión o coeficiente autorregresivo

ϵ_i ... componente estocástica independiente.

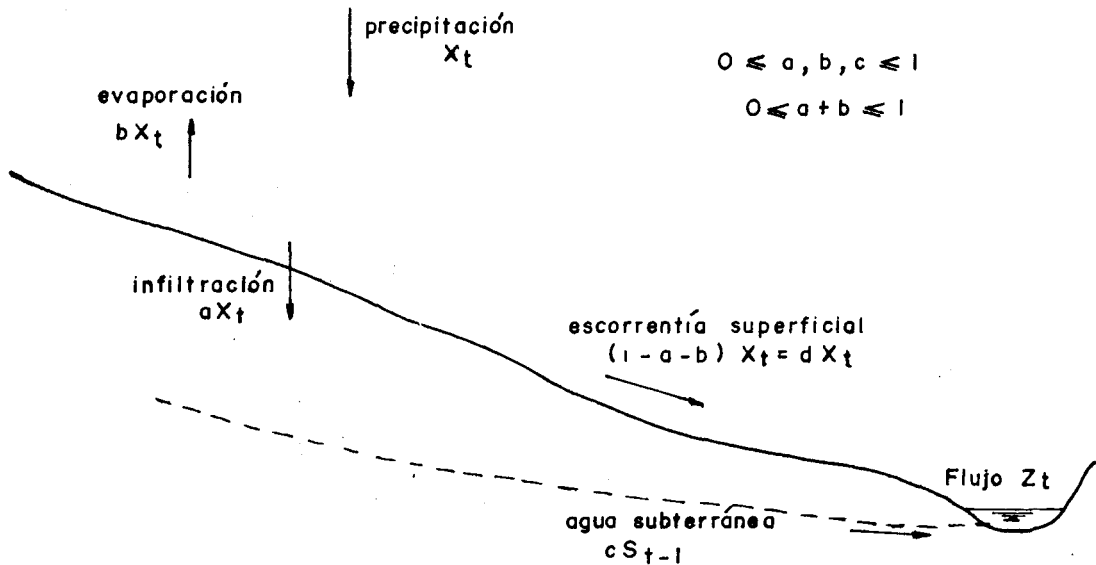
Se demuestra que:

$$\phi = e^{-K}$$

Sin embargo, qué tan bueno resulta este modelo AR(1), depende de qué tan buena es la suposición de recesión exponencial en el río.

Ejemplo 10.4

Justificación física de los modelos autorregresivos de promedios móviles (ARMA) para simulación de caudales.



flujo anual en el río = contribución de a. subterránea + escorrentía directa.

$$Z_t = c S_{t-1} + d X_t \quad (10.19)$$

La ecuación de continuidad para el almacenamiento subterráneo da:

$$S_t = S_{t-1} + a X_t - c S_{t-1}$$

$$\text{o} \quad S_t = (1 - c) S_{t-1} + a X_t \quad (10.20)$$

Combinando 10.19 y 10.20 Salas y Smith (referencia 10) encontraron que el modelo para el flujo anual Z_t puede escribirse como:

$$Z_t = (1 - c) Z_{t-1} + d X_t - |d(1 - c) - ac| X_{t-1} \quad (10.21)$$

que tiene la forma de un modelo ARMA (1, 1) cuando la precipitación anual es una serie independiente.