Gráficos de probabilidad

Muchas veces en gráficosque representan precipitaciones o caudales en función de su probabilidad, los valores de ésta (0,1; 0,2; 0,3; etc.) aparecen extrañamente distribuidos de acuerdo con una escala de probabilidad o probabilística... Vamos a comprender la naturaleza de esta escala deformada.

Frecuencias acumuladas con datos individuales

A	B	C	D	E
n	Q medios (m^3/s) ordenados	n/N	n/(N+1) [Weibull]	(n-0,5)/N [Hazen]
1	5,17	0,022	0,022	0,011
2	8,46	0,044	0,043	0,033
3	8,75	0,067	0,065	0,056
4	10,52	0,089	0,087	0,078
5	10,68	0,111	0,109	0,100
6	11,28	0,133	0,130	0,122
7	14,32	0,156	0,152	0,144
8	14,70	0,178	0,174	0,167
9	15,08	0,200	0,196	0,189
10	15,59	0,222	0,217	0,211
11	16,00	0,244	0,239	0,233
12	16,29	0,267	0,261	0,256
13	16,45	0,289	0,283	0,278
14	16,51	0,311	0,304	0,300
15	16,70	0,333	0,326	0,322
16	17,02	0,356	0,348	0,344
17	17,17	0,378	0,370	0,367
18	17,46	0,400	0,391	0,389
19	17,46	0,422	0,413	0,411
20	17,59	0,444	0,435	0,433
21	18,35	0,467	0,457	0,456
22	19,17	0,489	0,478	0,478
23	19,81	0,511	0,500	0,500
24	20,22	0,533	0,522	0,522
25	20,38	0,556	0,543	0,544
26	20,63	0,578	0,565	0,567
27	21,74	0,600	0,587	0,589
28	21,86	0,622	0,609	0,611
29	22,09	0,644	0,630	0,633
30	22,40	0,667	0,652	0,656
31	22,47	0,689	0,674	0,678
32	22,66	0,711	0,696	0,700
33	25,41	0,733	0,717	0,722
34	26,97	0,756	0,739	0,744
35	27,09	0,778	0,761	0,767
36	27,85	0,800	0,783	0,789
37	28,74	0,822	0,804	0,811
38	29,85	0,844	0,826	0,833
39	30,74	0,867	0,848	0,856
40	31,56	0,889	0,870	0,878
41	31,91	0,911	0,891	0,900
42	32,92	0,933	0,913	0,922
43	36,57	0,956	0,935	0,944
44	37,55	0,978	0,957	0,967
45	41,83	1,000	0,978	0,989

En el tema *Cálculos estadísticos en Hidrología* hemos visto los gráficos de valores acumulados (*función de distribución*), que nos permiten evaluar la probabilidad de que se supere un valor dado (o viceversa). En Hidrología normalmente disponemos de pocos datos (típicamente de 30 a 60 valores), que no son suficientes para clasificarlos en varios intervalos. Pero la curva de valores acumulados puede conseguirse trabajando con datos individuales, sin agrupar en intervalos.

Veámoslo con un ejemplo: disponemos de 45 caudales medios anuales del río Tormes[1] y vamos a construir un gráfico de frecuencias acumuladas.

El proceso de trabajo es el siguiente:

1) Ordenamos los caudales de menor a mayor (Recogimos los datos en orden cronológico y en la columna B ya aparecen ordenados).

2) Cálculo de frecuencias:

En la columna A hemos indicado el número de orden, que es también la **frecuencia absoluta acumulada**: por ejemplo, existen 6 valores inferiores o iguales a 11,28 m^3/s.

La columna C es la **frecuencia relativa acumulada**: Como hay 45 datos (N) y entre ellos 6 caudales tienen un valor igual o menor a 11,28 m^3/s, calculamos: 6/45 = 0,133; o sea que el 13,3% de los datos disponibles son iguales o menores que 11,28 m^3/s.

Este cálculo intuitivo (6/45, es decir: n/N) presenta un problema: al llegar al valor más alto obtenemos 45/45=1,00 (columna C, abajo). La frecuencia del 100% nos induce a suponer (erróneamente) que todos los caudales en este punto de aforo son siempre iguales o menores que 41,83 m^3/s (nuestro mayor valor disponible), lo cual es una suposición obviamente falsa: si conseguimos medidas de más años, seguro que se producirá algún caudal mayor que 41,83 m^3/s

Para solucionar este problema se han ideado varios procedimientos de calcular las frecuencias relativas: en lugar de n/N pueden utilizarse (columnas D y E):

n/(N+1) [Weibull] ; En este ejemplo: 1/46 ; 2/46, etc

(n–0,5)/N [Hazen][2] ; En este ejemplo: 0,5/45 ; 1,5/ 45; etc.

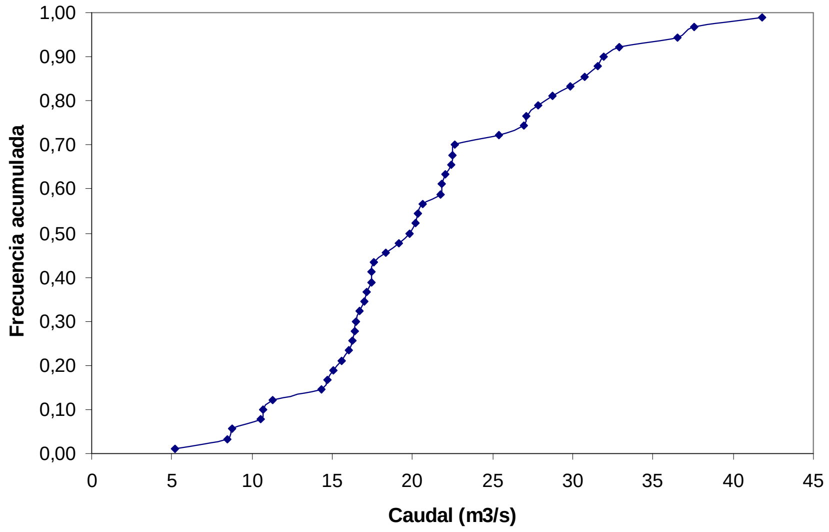
[1] Río Tormes en Barco de Ávila (parte alta de la cuenca, 900 km^2), años hidrológicos 1940-41 a 1984-85

[2] Según Ragunath (2012), p.222 es *n/(N-0,5)*. Todos los textos (por ej., Chow et al. (1987)) indican *(n-0,5)/N*

3) Representamos gráficamente los caudales en función de su frecuencia acumulada (columna B en un eje, columna E en el otro —fórmula de Hazen—).

Con esta curva (de la figura 1) ya podemos realizar evaluaciones aproximadas: por ejemplo, qué caudal no será superado el 20% de años, o inversamente, qué probabilidad existe de que no se alcance un caudal de 30 m^3/s.

Fig. 1.- Valores de 45 caudales en función de su frecuencia acumulada (proporción de casos menores o iguales a él)



Frecuencia se refiere a los datos medidos de la muestra.

Probabilidad se refiere a toda la población, normalmente basándonos en la ley estadística que rige la distribución de los datos (ya que en Hidrología nunca disponemos de datos de toda la población).

Por **ejemplo** (con referencia a la Tabla anterior): El caudal 27,09 m^3/s **no** es superado el 76,7% de los años (**frecuencia** acumulada 0,767, columna E)

Como esta muestra parece ajustarse a la Ley de Gauss, se calcula que la **probabilidad** de que **no** se supere ese caudal de 27,09 m^3/s es de 0,774 (~77,4%)

Gráficos de probabilidad

En el tema *Cálculos estadísticos en Hidrología* vimos (figuras 4 y 6 de aquel tema) que si los datos se ajustan a la Ley Normal o de Gauss, la curva de frecuencias acumuladas adquiere forma de ***S*** alargada. Pero si distribuimos las marcas del eje vertical de acuerdo con la ecuación de Gauss, podremos conseguir que esa línea en ***S*** se convierta en una línea recta:

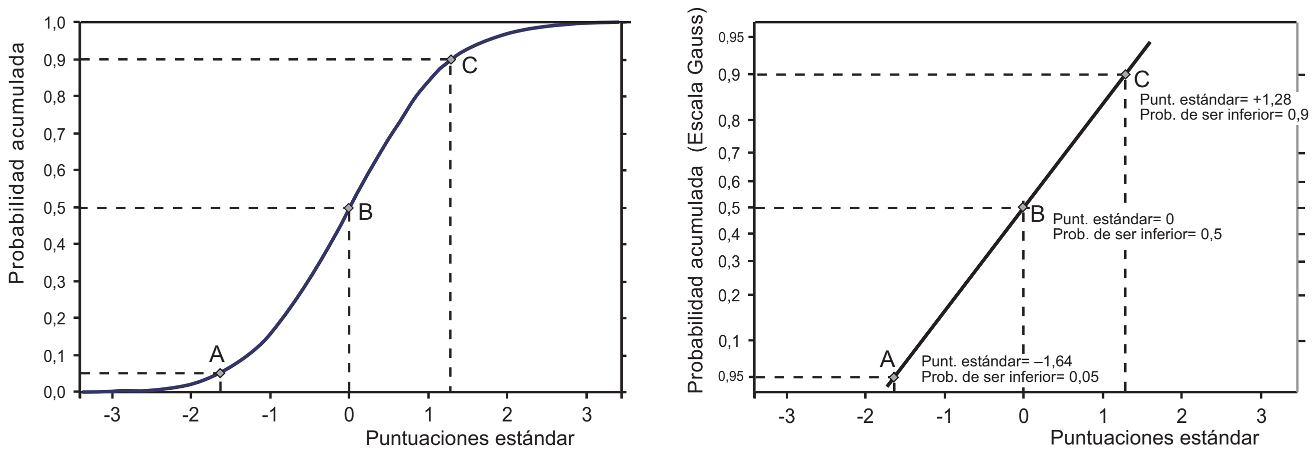


Fig. 2.- Los puntos A, B y C del gráfico izquierdo y derecho son equivalentes (ver su significado en la figura derecha).
Izquierda: Si la probabilidad acumulada está distribuida en el eje vertical aritméticamente, los puntos A, B, C siguen la curva de Gauss acumulada (forma de *S*). Derecha: Si el eje vertical está preparado de acuerdo con la ley de Gauss, los tres puntos se ajustan a una recta

En esta escala de Gauss (fig. 2, derecha, eje vertical) los extremos (0 y 1,00) se situarían en –∞ y +∞

Podemos dibujar los puntos manualmente si disponemos de un papel especial gauss-aritmético o *papel probabilístico* (incluido al final de ese documento).

En un eje se sitúan los valores brutos (en nuestro ejemplo, caudales), y en el otro eje la frecuencia acumulada (frecuencias distribuidas en el eje de acuerdo a la ley de Gauss).

En estas figuras se representan en el **eje vertical** las frecuencias o probabilidad, por coherencia con la explicaciones previas, pero en la práctica es más frecuente que la probabilidad o frecuencias se representen en el **eje horizontal** y los valores brutos (caudales, precipitaciones) en el eje vertical.

Si los puntos se alinean aproximadamente en línea recta, parece que la serie de datos sí se ajusta a la ley normal o de Gauss[3]. La bibliografía francesa[4] la denomina *recta de Henry*[5]

En la figura 3 hemos representado los mismos caudales del ejemplo utilizado en el apartado anterior, que se representaron en la figura 1, pero la escala de probabilidad está 'deformada' de acuerdo con la ley de Gauss y por ello se ajustan aproximadamente a una recta.

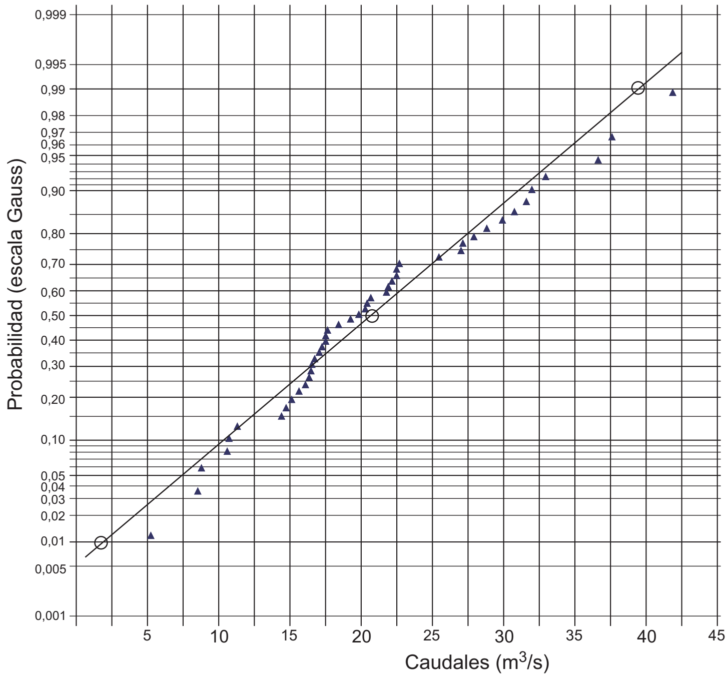


Fig. 3.- Representación de 45 caudales medios anuales en papel probabilísitco de Gauss (mismos datos de la fig. 1). Los tres pequeños círculos son los puntos calculados para trazar la recta.

Para **trazar la recta de ajuste** necesitamos al menos dos puntos:

Calculamos la media y la desviación estándar de la muestra:

Media = 20,98 m^3/s
Desviación estándar = 8,13 m^3/s

La recta debe pasar por el punto: Probabilidad= 0,50 ; Caudal= 20,98 (media aritmética)

A la probabilidad 0,99 (de ser menor) equivale una probabilidad 0,01 de que ese valor sea superado. Según la tabla de la función de Gauss, a la probabilidad 0,01 corresponde una puntuación estandarizada de +2,33. El valor bruto correspondiente será:

$$2{,}33 = \frac{x - 20{,}98}{8{,}13} \quad ; \quad x = 39{,}92 \text{ m}^3/\text{s}$$

Análogamente, a la probabilidad 0,01 (de ser menor) equivale una probabilidad 0,99 de ser superado. Según la tabla de Gauss, a la probabilidad 0,99 corresponde una puntuación estandarizada de –2,33. El valor bruto correspondiente será:

$$-2{,}33 = \frac{x - 20{,}98}{8{,}13} \quad ; \quad x = 2{,}04 \text{ m}^3/\text{s}$$

Por tanto el trazado de la recta que refleja el ajuste a la ley de Gauss se hace con estos puntos (marcados con círculos en la figura 3):

---

3 Para cuantificar la *bondad del ajuste* se utiliza el test χ^2 (chi-cuadrado)

4 Por ejemplo, Dubreuil, P (1974).- *Initiation a l'analyse hydrologuique*. Ed. Masson, 216 pp.

5 En referencia a Pierre Jean Paul Henri (1848-1907), oficial de artillería francés que consideró este tratamiento gráfico al estudiar la dispersión de los impactos de tiro alrededor del centro. (http://serge.mehl.free.fr/chrono/Henry.html)

$$Q = 20{,}98 \;\; ; \;\; Probabilidad = 0{,}50$$
$$Q = \;\; 2{,}04 \;\; ; \;\; Probabilidad = 0{,}01$$
$$Q = 39{,}92 \;\; ; \;\; Probabilidad = 0{,}99$$

Si el ajuste parece aceptable, la recta representa a la población, y a partir de ese momento cualquier estimación (qué caudal corresponde a cierta probabilidad o viceversa) se hará sobre la precta y no sobre la línea de puntos.

Confección de una escala de Gauss

Aunque no es usual que uno mismo deba preparar su propia escala de Gauss, es instructivo saber hacerlo.

Por ejemplo, para situar la marca de **0,95** (probabilidad de ser superado): Consultamos en la tabla de Gauss, obtenemos: –1,64 (la puntuacion estándar **–1,64** es superada por el 95% de los casos).

Debemos repetir la operación para todas las marcas que queramos representar. Por ejemplo, para situar la marca **0,005**: de nuevo consultamos la tabla de Gauss y obtenemos un valor estandarizado de **+2,58** (la puntuacion estándar +2,58 es superada por el 0,5% de los casos).

Con la Hoja de Cálculo se consigue lo mismo mediante las fórmulas:
=DISTR.NORM.ESTAND.INV(0,95) y =DISTR.NORM.ESTAND.INV(0,005).

Para representar estos dos puntos (–1,64 y +2,58), trazamos una escala aritmética de –3 a +3 así:

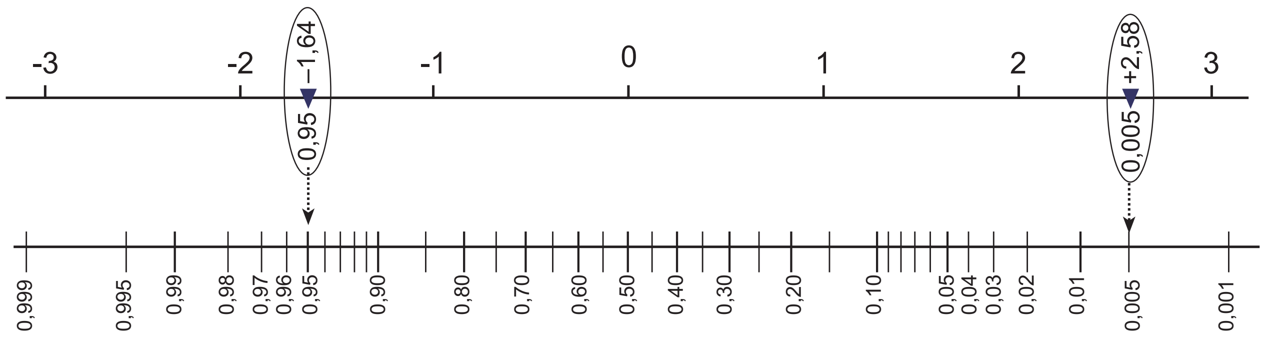


Fig. 4.- Realización de una escala de probabilidades según la ley de Gauss.

Se realiza la misma operación para el resto de las marcas del eje.

Finalmente, para rotular el eje, no escribimos los valores que nos han servido para ubicarlos (–1,64 ; +2,58) sino las correspondientes probabilidades (0,95 ; 0,005)

Al final de este docmento se incluyen varios gráficos con la escala Gauss en uno de sus ejes.

Realización del gráfico con Hoja de Cálculo

Una representación como la figura 3 puede realizarse a mano sobre una hoja de papel de probabilidad Gauss. Es más cómodo hacerlo mediante la Hoja de Cálculo.

Los pasos iniciales para ordenar los datos y calcular frecuencias se realizan fácilmente con la Hoja de Cálculo. Para representar los puntos en sus posiciones correspondientes a la escala Gauss, es necesario añadir una o dos nuevas columnas en la tabla de datos de nuestro ejemplo:

	A	B	C	D	E	F	G
1	n	Q medios (m^3/s) ordenados			(n-0,5)/N [Hazen]	Valores estándar segun Gauss	Valores brutos segun Gauss
2	1	5,17			0,011	-2,287	2.39
3	2	8,46			0,033	-1,834	6.07
4	3	8,75			0,056	-1,593	8.02
5	4	10,52			0,078	-1,420	9.43
...	. . .	. . .			. . .	. . .	. . .
22	21	18,35			0,456	-0,112	20.07
23	22	19,17			0,478	-0,056	20.52
24	23	19,81			0,500	0,000	20.98
25	24	20,22			0,522	0,056	21.43
26	25	20,38			0,544	0,112	21.88
...	. . .	. . .			. . .	. . .	. . .
43	42	32,92			0,922	1,420	32.52
44	43	36,57			0,944	1,593	33.93
45	44	37,55			0,967	1,834	35.89
46	45	41,83			0,989	2,287	39.57

Para conseguir la columna F: en la celda F2 se utliza la fórmula:

=DISTR.NORM.ESTAND.INV(E2).

Con esta fórmula, la Hoja de Cálculo aplica la fórmula de Gauss a la inversa, es decir: qué caudal (en valor estandarizado) corresponde a cada frecuencia de la celda E2.

Ahora sólo es necesario realizar un gráfico con las columnas B y F. Este gráfico (figura 5) será idéntico al que hemos trazado a mano sobre papel de probabilidad (figura 3). Pero aunque sea el mismo gráfico, en el eje vertical no aparecen rotuladas las probabilidades, sino las puntuaciones estandarizadas correspondientes a cada probabilidad. La correspondencia entre ambas series de valores es la que se mostraba en la figura 4.

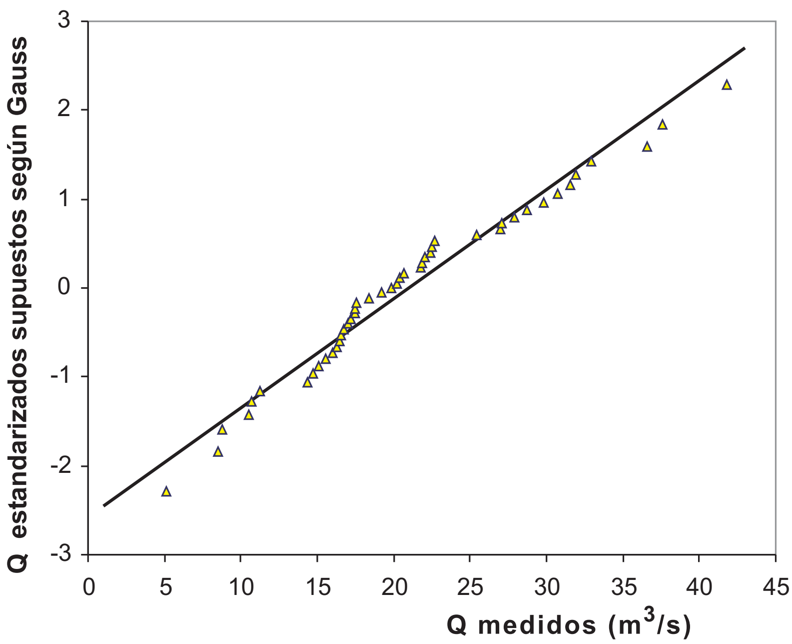
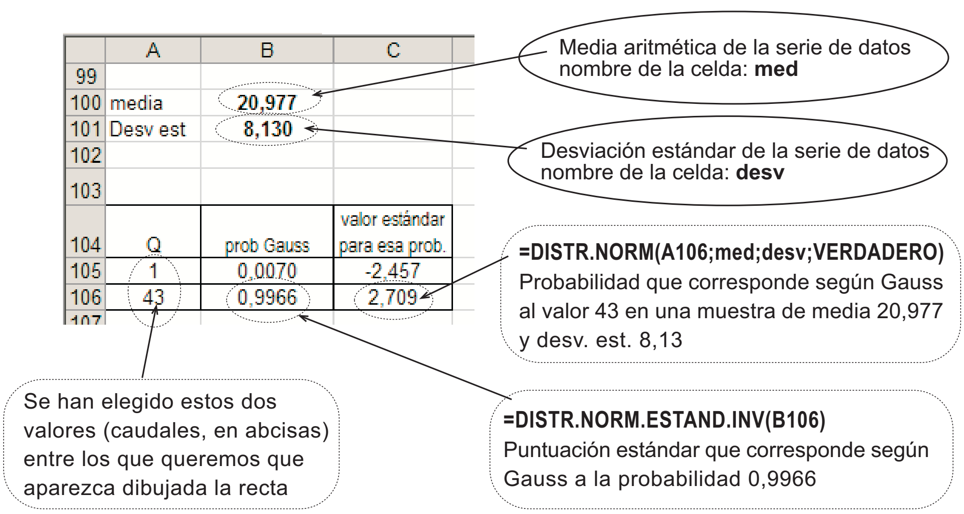


Fig. 5.- Mismo gráfico de la figura 3 dibujado mediante la Hoja de Cálculo

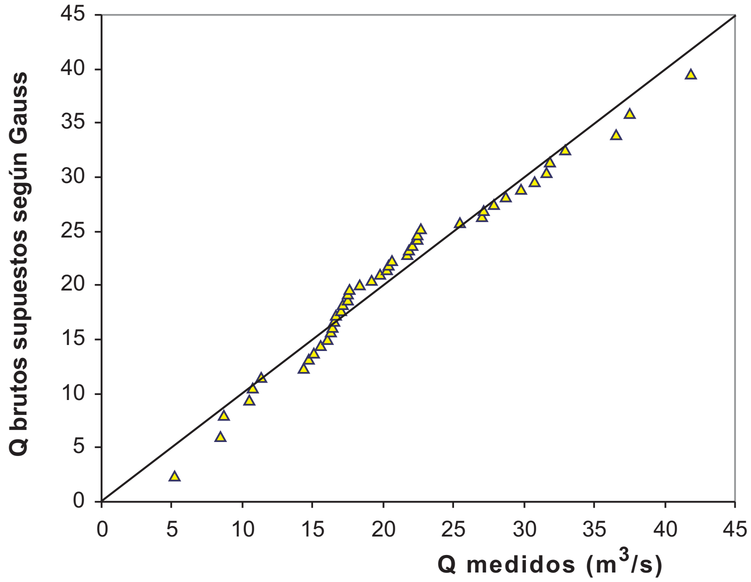


El **trazado de la recta de ajuste** (fig.5) **no** puede hacerse con la "línea de tendencia" facilitada por Excel.

Se hará añadiendo otra **serie de datos nueva** que generará un segundo gráfico sobre el anterior. Este segundo gráfico se hace con dos puntos, abcisas A105:A106, y ordenadas C105:C106, eligiendo el tipo de gráfico en que no se muestran los puntos, sólo la línea que los une.

El mismo gráfico podemos obtenerlo utilizando las columnas B y G de la tabla de la página anterior:

Fig. 6.- Mismo gráfico de la figura 5 , pero ahora en el eje vertical: caudales brutos (m^3/s) que deberían haberse producido para cada frecuencia, según la ley de Gauss



Esto pone de manifiesto el **fundamento de los gráficos de probabilidad: comparar los valores reales medidos con los que se deberían haber producido de acuerdo con la Ley de Gauss**; obviamente si hubiera un ajuste perfecto a Gauss, ambas series de valores serían idénticas y los puntos formarían una recta perfecta. En este caso, no es necesario ningún cálculo para dibujar la recta de ajuste a la ley de Gauss: basta unir los puntos (0,0) y (45,45).

El gráfico de la figura 6 adolece del mismo problema que el de la figura 5: no es visualmente significativo, porque en el eje vertical no aparecen las probabilidades.

Gráfico de probabilidad Gumbel

Todo el proceso explicado es similar para una distribución que se ajuste a la ley de Gumbel.

En la tabla adjunta se presenta los caudales máximos (día de mayor caudal de cada año) para la misma estación de aforos y para los mismos años[6] que habíamos visto en el ajuste a la Ley de Gauss. En la segunda columna los caudales, que hemos ordenado en orden decreciente, y en la tercera columna se ha calculado su frecuencia por el método de Hazen (0,5/45; 1,5/45, etc.). La última columna (retorno) es el inverso de la anterior.

n	Q max m^3/s)	frec Hazen	Retorno (1/frec)	n	Q max m^3/s)	frec Hazen	Retorno (1/frec)	n	Q max m^3/s)	frec Hazen	Retorno (1/frec)
1	487	0,011	**90,00**	16	310	0,344	**2,90**	31	177	0,678	**1,48**
2	465	0,033	**30,00**	17	310	0,367	**2,73**	32	171	0,700	**1,43**
3	465	0,056	**18,00**	18	310	0,389	**2,57**	33	166	0,722	**1,38**
4	440	0,078	**12,86**	19	295	0,411	**2,43**	34	166	0,744	**1,34**
5	403	0,100	**10,00**	20	279	0,433	**2,31**	35	166	0,767	**1,30**
6	391	0,122	**8,18**	21	279	0,456	**2,20**	36	156	0,789	**1,27**
7	387	0,144	**6,92**	22	275	0,478	**2,09**	37	156	0,811	**1,23**
8	387	0,167	**6,00**	23	259	0,500	**2,00**	38	150	0,833	**1,20**
9	387	0,189	**5,29**	24	257	0,522	**1,91**	39	144	0,856	**1,17**
10	387	0,211	**4,74**	25	233	0,544	**1,84**	40	144	0,878	**1,14**
11	346	0,233	**4,29**	26	233	0,567	**1,76**	41	132	0,900	**1,11**
12	342	0,256	**3,91**	27	229	0,589	**1,70**	42	125	0,922	**1,08**
13	341	0,278	**3,60**	28	223	0,611	**1,64**	43	120	0,944	**1,06**
14	338	0,300	**3,33**	29	207	0,633	**1,58**	44	112	0,967	**1,03**
15	328	0,322	**3,10**	30	194	0,656	**1,53**	45	78	0,989	**1,01**

[6] Río Tormes en Barco de Ávila, años 1940-41 a 1984-85

Representamos estos puntos en un gráfico Gumbel-aritmético (los tres valores menores caen fuera del gráfico, por la izquierda), utilizando el modelo incluído al final de este documento:

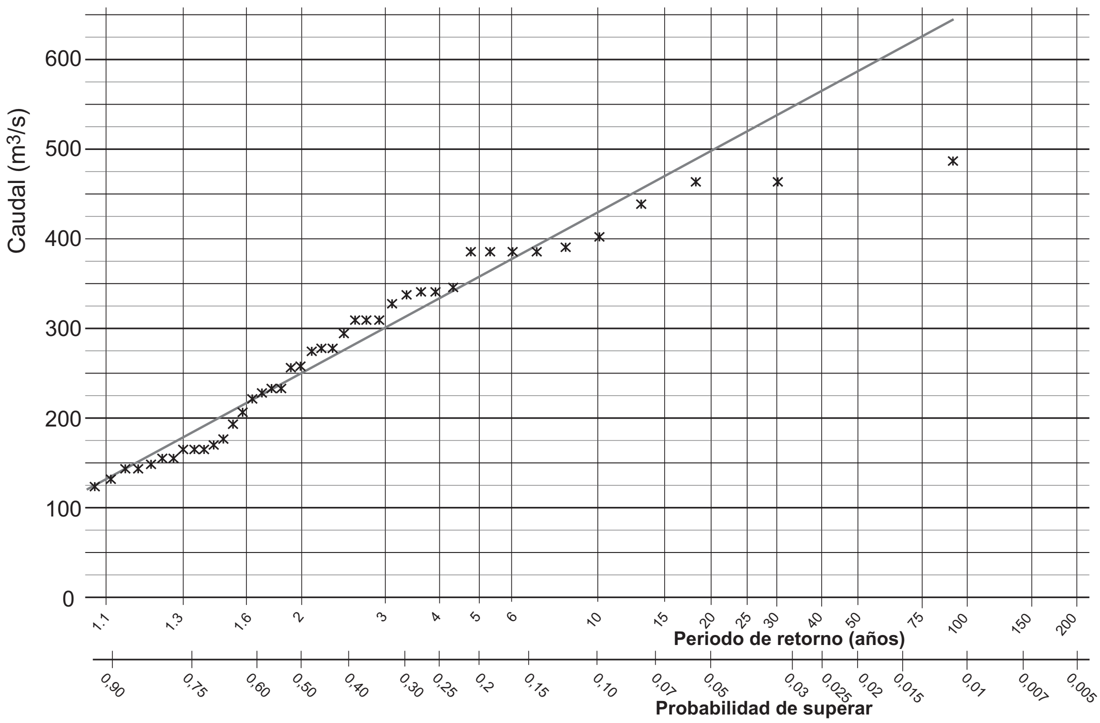


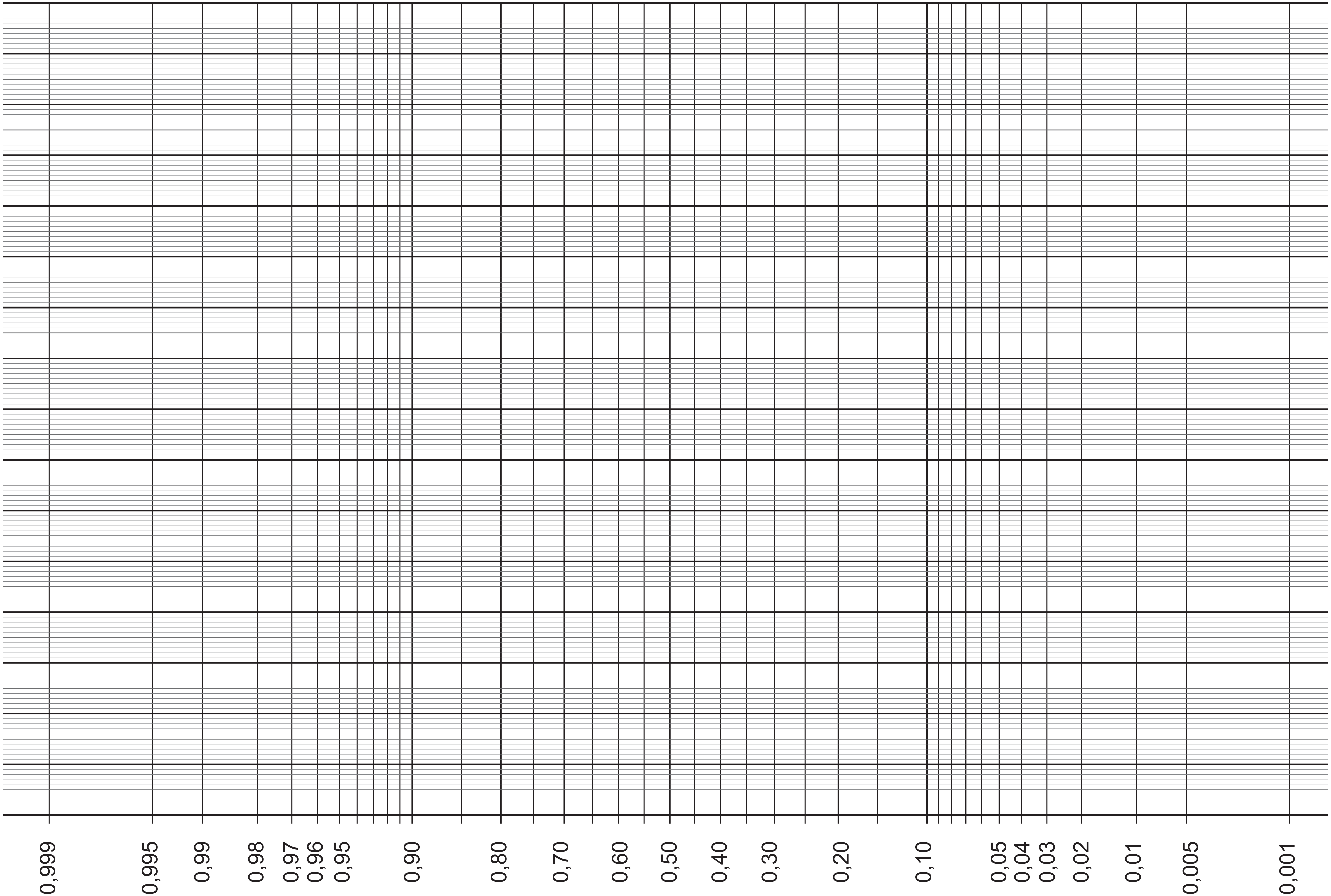
Fig. 7.- Caudales máximos representados en un gráfico de probabilidad Gumbel.

Si una serie de datos se ajustara a la ley de Gumbel, al ser representados en un papel de probabilidad en el que uno de los ejes ha sido 'deformado' de acuerdo con dicha ley, los puntos deberían alinearse, y la recta de ajuste nos permitiría realizar estimaciones.

En este caso, vemos que el ajuste puede ser aceptable para valores bajos, pero para retornos mayores de 20 años los datos empíricos se alejan de la recta de ajuste. Par trazar la recta de ajuste, se opera como en el caso de Gauss: para la media y la desviación estándar del grupo de datos considerado, se calcula el caudal que debería producirse según Gumbel para retornos de 2 y 50 años; esto nos proporcionará dos puntos en el gráfico para trazar la recta.

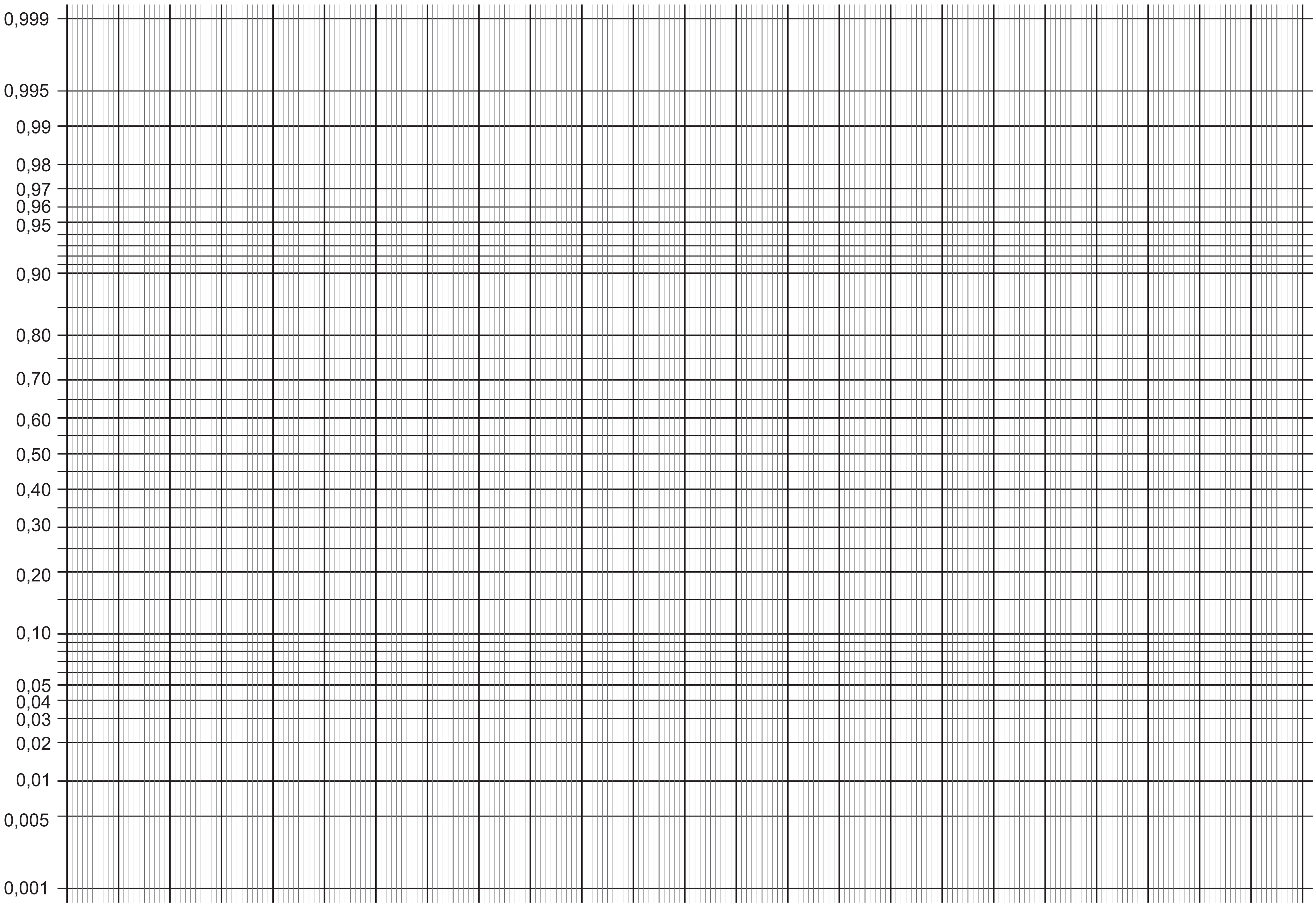
Análogamente a como vimos en los gráficos Gauss, se puede hacer el gráfico con la Hoja de Cálculo situando en un eje los caudales medidos y en el otro los caudales que deberían haberse producido de acuerdo con Gumbel (ese será el eje del retorno o probabilidad).

Escala Gauss



http://hidrologia.usal.es

F. Javier Sánchez San Román--Dpto. Geología--Univ. Salamanca (España)

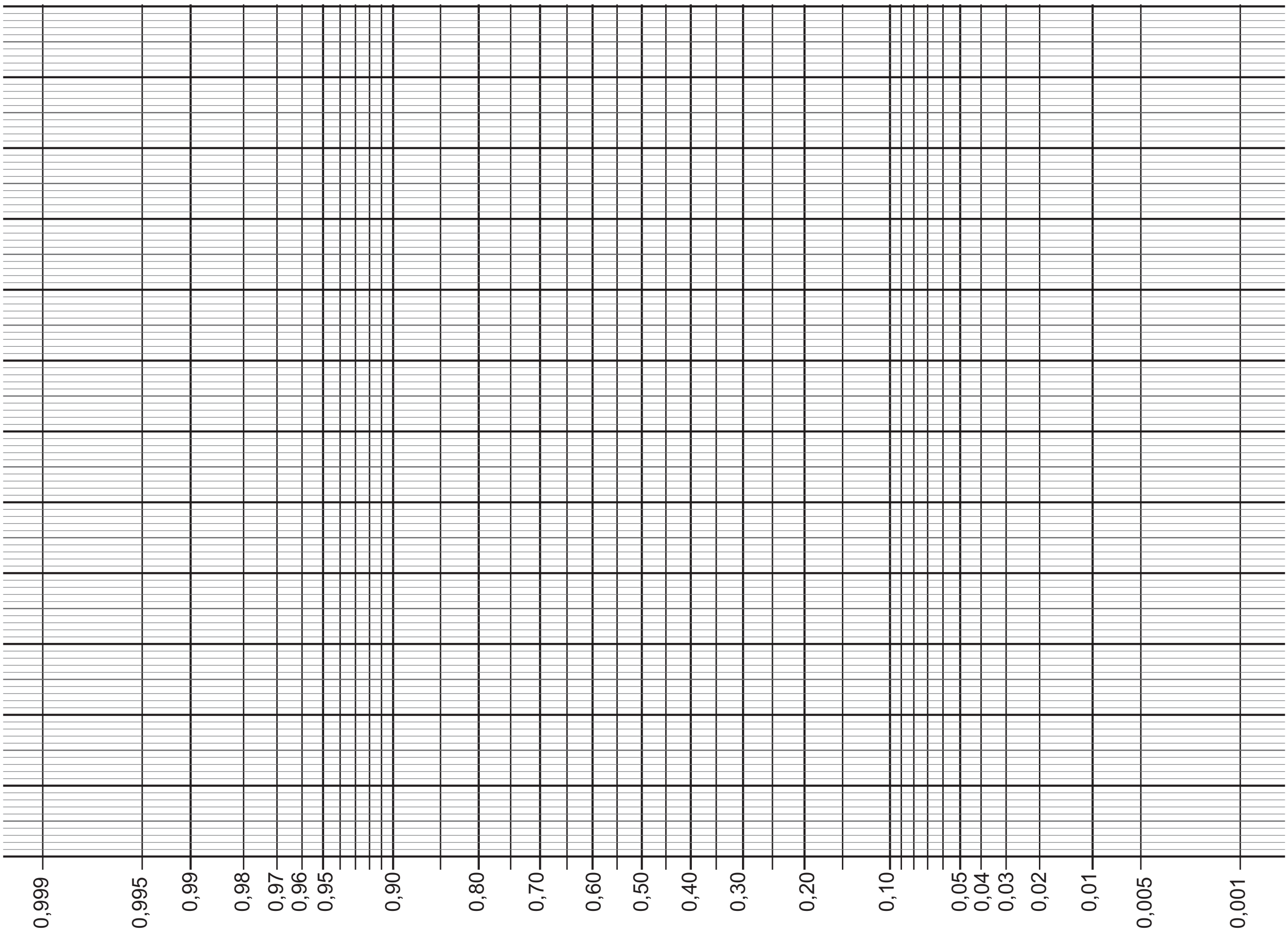


Escala Gauss

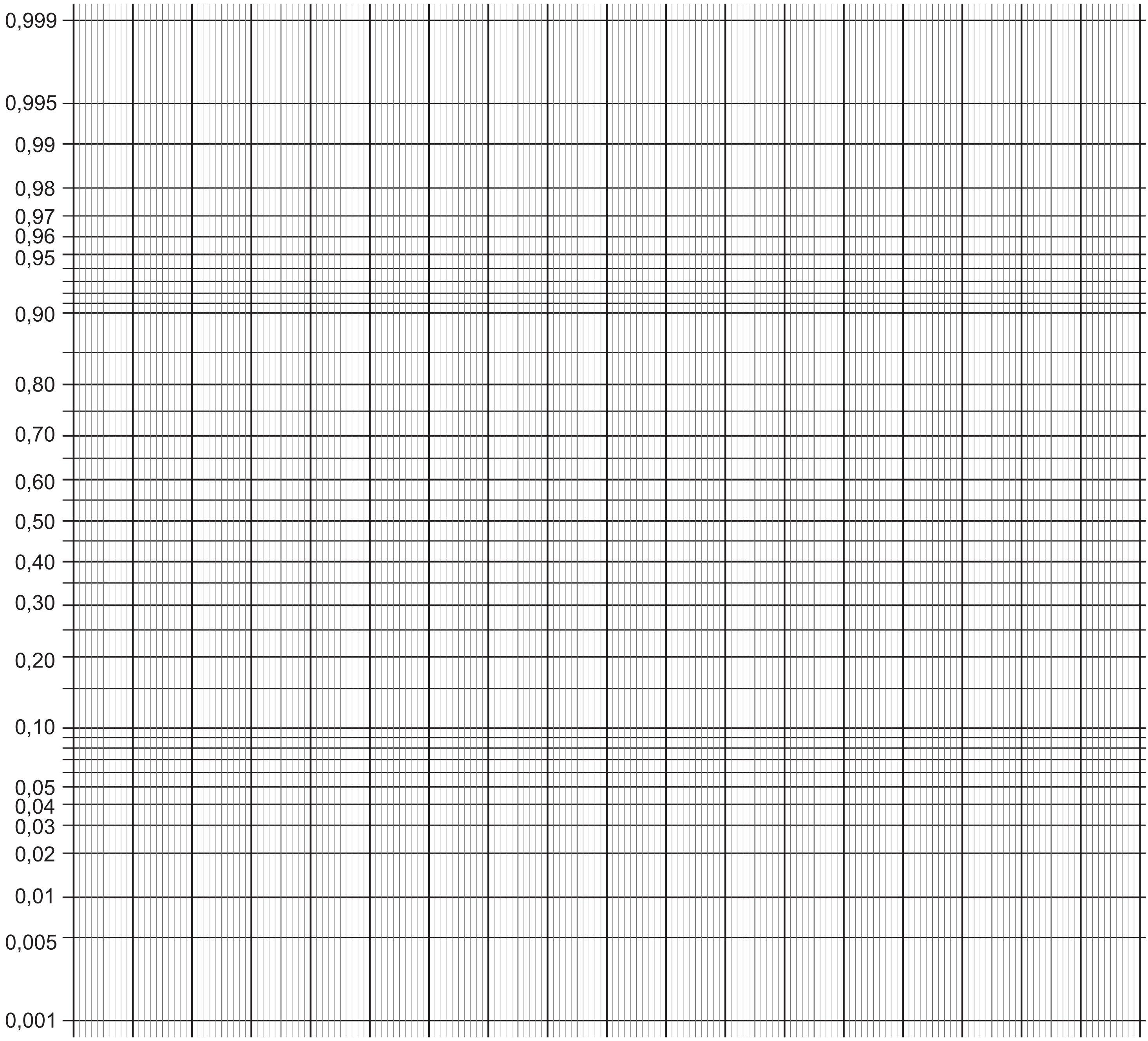
F. Javier Sánchez San Román--Dpto. Geología--Univ. Salamanca (España)

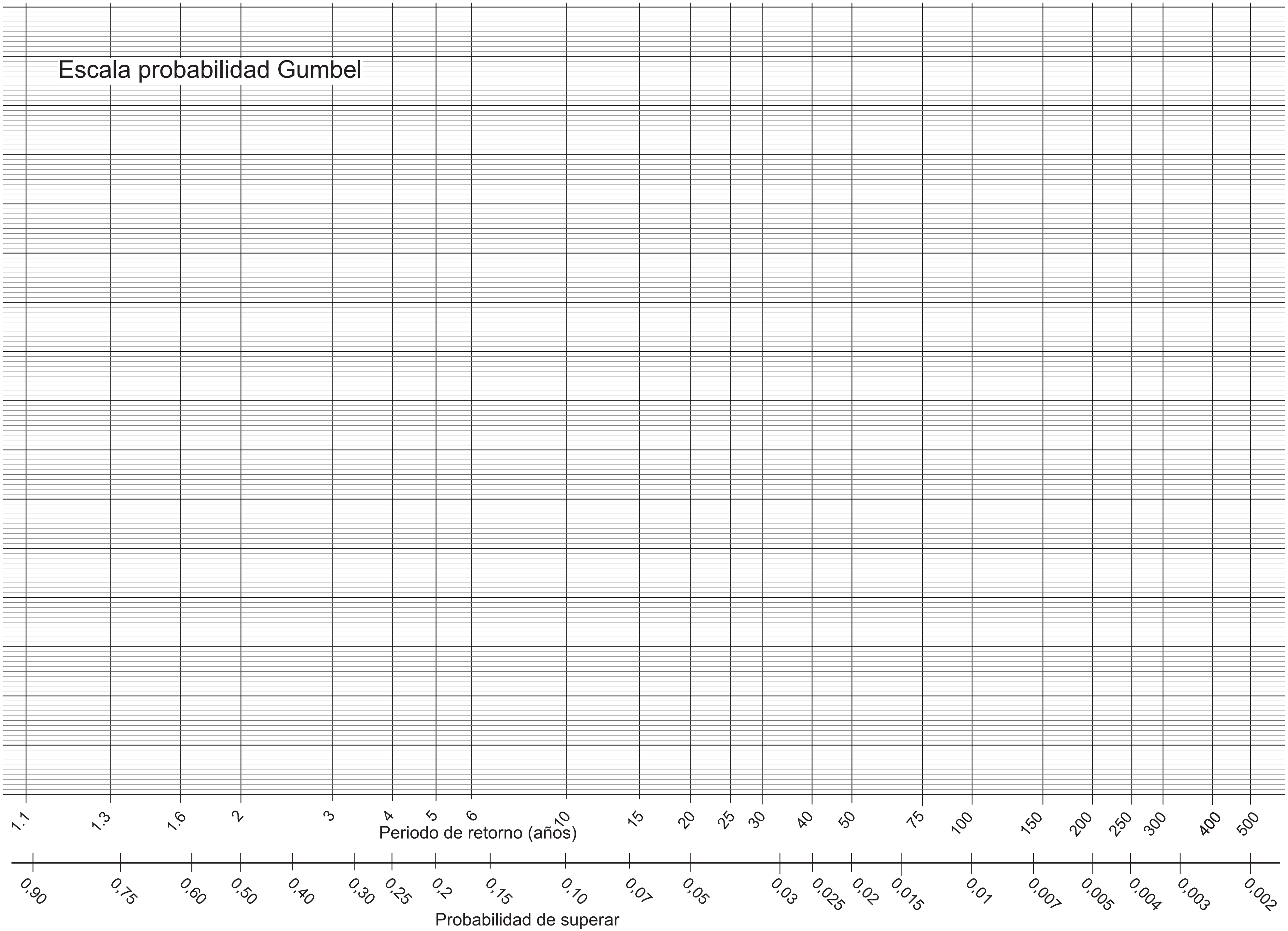
http://hidrologia.usal.es

Escala Gauss



Escala Gauss





F.J. Sánchez San Román
Dpto. Geología Univ. Salamanca (España)
http://hidrologia.usal.es/