

Investigación

Análisis de regresión lineal multivariable para la obtención del caudal pico de descarga en rotura de presas

Multivariate linear regression analysis to obtain dam breach peak outflow

José Manuel Sánchez Muñoz

Revista de Investigación



Volumen VI, Número 1, pp. 065-092, ISSN 2174-0410
Recepción: 26 Oct'15; Aceptación: 1 Mar'16

1 de abril de 2016

Resumen

En este artículo se hace un estudio del comportamiento dinámico e hidráulico del terreno en caso de avenida por rotura de presa, así como una presentación de una metodología de regresión lineal multivariable a partir del análisis de los datos históricos de rotura de presas para la obtención del caudal pico de descarga.

Palabras Clave: Rotura de presas, caudal pico de descarga, hidrograma, regresión lineal multivariable.

Abstract

This article presents a study of dynamic and hydraulic behavior of soil in case of dam breach flood, and a presentation of a multivariate linear regression methodology from the analysis of historical data for dam breach to obtain the peak outflow.

Keywords: Dam breach, peak outflow, hydrogram, multivariate linear regression.

1. Introducción a la modelización de rotura de presas

En un mundo en el que la gestión eficiente de los recursos hídricos resulta cada vez más importante, es fundamental entender el comportamiento dinámico e hidráulico del terreno durante el transcurso de una avenida, con el fin de optimizar el diseño de infraestructuras hidráulicas, encauzamientos fluviales, estabilizaciones de márgenes, estudios de zonas inundables, roturas de presas, etc.

Muchas son las metodologías instauradas en el contexto de la hidráulica y la hidrología geomorfológica que estudian desde un punto de vista descriptivo y analítico la predicción de los hidrogramas generados por fenómenos de rotura de presas.

Desde el punto de vista cronológico, podemos establecer el estudio y análisis de la onda de crecida y el daño potencial generado por la rotura de una presa en tres fases bien diferenciadas:

1. Predicción del hidrograma de rotura.
2. Progresión del hidrograma a través del cauce fluvial aguas abajo, haciendo uso de la correspondiente modelización.
3. Análisis de riesgos y predicción de daños.

Numerosos son los parámetros que influyen de manera determinante en el hidrograma de rotura de una presa, aunque las características geométricas como su altura y el volumen del embalse se han de considerar fundamentales para la formación de la brecha. También son importantes el modo de rotura así como el tiempo de formación de la brecha, que dependerá en gran medida de su tipología constructiva.

Con el fin de estimar estos parámetros de formación de la brecha de rotura, se ha consultado una amplia bibliografía, recurriendo en la mayoría de los casos a expresiones empíricas deducidas generalmente mediante un análisis de regresión sobre los datos obtenidos de experimentos bien en campo, bien en laboratorio simulando las condiciones reales mediante semejanza hidráulica. En estas expresiones han influido de forma especial también las dimensiones y los tiempos de desarrollo de las brechas de rotura correspondientes a casos históricos de colapsos de presas.

En este artículo pretendemos ofrecer una visión global descriptiva de la metodología utilizada a través del análisis de los modelos disponibles en la bibliografía para la predicción del hidrograma de rotura que resultará determinante para nuestra posterior modelización.

2. Formación de la brecha de rotura

Consideramos como *brecha* la apertura que se forma en la presa en su proceso de colapso. Existe una gran incertidumbre a la hora de obtener los parámetros de formación de dicha brecha, dimensiones y el tiempo de formación de la misma. Además de la necesidad de entender el mecanismo real por el que la presa ha colapsado (p.ej. tubificación¹ o sobrevertido²), resulta imprescindible realizar una diferenciación entre las distintas tipologías constructivas de una presa, ya que en los primeros intentos por predecir su modo de colapso, era usual considerar que éste se producía de forma completa e instantánea, ya que de este modo se simplificaba en gran medida el aparato de cálculo matemático, lo cual resultaba en cierto modo apropiado para las presas de arco, pero en ningún caso era aceptable para las presas de gravedad y de materiales sueltos.

A pesar de que en las últimas décadas se han dedicado muchos esfuerzos a la investigación con el objetivo fundamental de comprender estos mecanismos de rotura que nos permitan obtener una parametrización del proceso de colapso de una presa de un modo fiable, el desarrollo

¹ Los colapsos por tubificación ocurren cuando la formación inicial de la brecha tiene lugar en algún punto por debajo de la cota de coronación de la presa debido a la erosión interna que forma un canal a través del cual escapa el agua. A medida que la erosión avanza, la apertura va aumentando progresivamente su tamaño, lo que provoca el colapso de la parte superior de la presa.

² En este tipo de colapsos, el nivel del embalse de la presa sube por encima de la cota de coronación de la misma, produciéndose un vertido que puede erosionar el trasdós o paramento aguas abajo.

de una metodología fidedigna se encuentra aún en una fase de asimilación y comprensión de los fenómenos que provocan la formación de dichos mecanismos.

Resulta por todo ello de vital importancia la obtención de los parámetros del caudal pico y la forma del hidrograma de rotura, para lo que han de tomarse en cuenta:

- ✓ Las dimensiones y la geometría de la brecha de rotura.
- ✓ El tiempo de formación de la misma.
- ✓ La profundidad y el volumen de agua almacenada en el embalse.
- ✓ El caudal entrante en el embalse.

Tanto las dimensiones y la geometría de la brecha de rotura como el tiempo de formación de la misma dependen en gran medida de la tipología constructiva de la presa. Es por ello que atendiendo a la diversa bibliografía existente, el Ministerio de Medio Ambiente de España (2001, págs. 20–21) estableció un análisis entre las diferentes tipologías constructivas de presas consideradas. Se puede realizar pues un análisis diferenciando las presas en dos grandes grupos: presas de materiales sueltos y presas rígidas (hormigón, mampostería).

2.1. Presas de materiales sueltos

Durante las últimas décadas las presas de materiales sueltos se han postulado como la principal tipología a la hora de construir nuevas infraestructuras de almacenamiento hídrico y de explotación hidroeléctrica, debido fundamentalmente a motivos económicos (encarecimiento del hormigón en el caso de presas de gravedad o bóveda, o de la mano de obra en el caso de presas de contrafuertes) y de integración medioambiental. Todo ello se ha traducido en la existencia de un mayor número de este tipo de presas con respecto al resto.

Diferentes autores como Fread (1984, 1988), establecen que las presas de materiales sueltos suelen presentar roturas progresivas en el tiempo, es decir no instantáneas, de manera que dichas roturas evolucionan desde unos estados iniciales hasta la totalidad de la presa. La geometría que la brecha de rotura suele presentar es prácticamente trapezoidal. Una vez que la brecha comienza a formarse, el agua descargada origina una erosión de la presa lo que desencadena un aumento de las dimensiones de dicha brecha, de tal forma que se producirá la descarga hasta el momento en que se consuma todo el volumen de agua almacenada en el embalse, o bien hasta que dicha brecha sea capaz de resistir la erosión.

Una brecha completamente desarrollada en presas de tierra tiende a tener en promedio un ancho (b) en el rango de $h_p < b < 3h_p$ donde h_p representa la altura de la presa. Las longitudes de las brechas para presas de tierra resultan normalmente menores que la longitud total del cierre. La brecha requiere también un intervalo de tiempo para su formación. El tiempo total de colapso tiene un rango de duración de unos pocos minutos a pocas horas, dependiendo de la altura de la presa, el tipo de material usado en su construcción y la extensión de la compactación de los mismos, y la magnitud y la duración de la descarga del agua.

2.2. Presas rígidas

Estas presas tienden a presentar brechas parciales de rotura en forma monolítica rectangular, debido a que su construcción se hace en forma de bloques. Estas brechas por lo tanto guardan una correlación directa con su metodología constructiva, originadas normalmente por un sellado pobre de los diferentes bloques hormigonados, ya sea por falta de limpieza de las juntas en el momento de hormigonar, o debido a la utilización de una pobre dosificación del mortero

de retoma que actúa de “sellante”. El tiempo de formación de la brecha previo al colapso de la presa suele ser de unos pocos minutos.

3. Definición de los parámetros de la brecha de rotura

En la definición de los parámetros de la brecha de rotura toman especial protagonismo sus características físicas y geométricas, como la altura, anchura o el ángulo de los taludes laterales, así como las variables que definirán el tiempo medido desde el inicio de su formación y su posterior desarrollo.

Wahl (1998) estableció los parámetros físicos fundamentales para realizar el posterior análisis, que se muestran en la Figura 1.

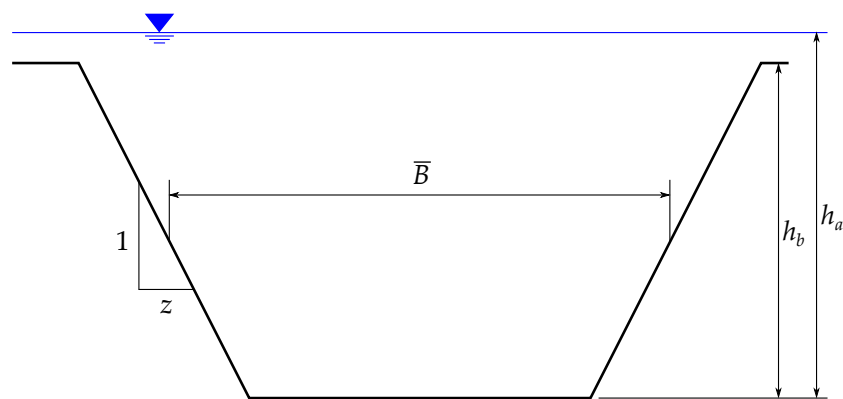


Figura 1. Parámetros de la brecha en una presa idealizada.

- ✓ **Profundidad de la brecha (h_b):** es la altura de la brecha, medida desde la coronación de la presa hasta el fondo de la misma. En algunas publicaciones también se habla de *Carga sobre la brecha* (h_a), que se refiere a la distancia medida desde la lámina de agua en el embalse hasta el fondo de la brecha.
- ✓ **Ancho de la brecha (B):** tanto el ancho final de la brecha como su tasa de expansión pueden afectar de forma dramática al aumento del caudal pico de descarga y el nivel de inundación aguas abajo de la presa. La bibliografía aceptada analiza diferentes casos de estudio estableciendo expresiones en función bien del ancho medio de la brecha (\bar{B}), o del ancho de la brecha en su parte superior o inferior.
- ✓ **Pendientes laterales de la brecha:** El valor de las pendientes laterales define la geometría trapezoidal o rectangular de la brecha de rotura. Este factor generalmente tiene una influencia pequeña.

Desde el punto de vista cronológico, se pueden destacar:

- ✓ **Tiempo de inicio de la brecha:** este instante es considerado en cuanto se presenta la descarga de los primeros caudales de sobrevertido o a través de la presa, lo cual desencadena el comienzo del aviso de alerta o evacuación por potencial colapso de la misma. Esta fase finaliza en el momento que comienza la fase de formación de la brecha de rotura. En la fase de inicio, la presa aún no colapsa y el caudal de descarga de la presa es pequeño. Durante la fase de iniciación es posible que la presa sea capaz de evitar el colapso si el sobrevertido o la erosión se detienen de manera inminente.

El tiempo de inicio de la brecha es un parámetro fundamental debido a que su influencia repercute de forma directa en el margen de tiempo de aviso disponible para evacuar las poblaciones aguas abajo. En programas como Dambrk o Fldwav no se trata de un parámetro de entrada. Actualmente existen pocas guías disponibles para la selección de los tiempos de inicio de brecha.

- ✓ **Tiempo de formación de la brecha:** también mencionado como *Tiempo de desarrollo de la brecha*; es el periodo transcurrido desde el momento de la aparición de la primera brecha en la cara aguas arriba de la presa hasta que la brecha está completamente desarrollada. En el caso de colapso por sobrevertido, se considera desde el instante en el que la presa comienza a erosionarse como resultado de la descarga. En la fase de inicio, la presa aún no colapsa, y la descarga es pequeña. La descarga puede ser considerada como un sobrevertido de apenas unos centímetros sobre la coronación de la presa, o el desarrollo de un canal de infiltración a través de la misma. Durante la fase de iniciación es posible que la presa sea capaz de evitar el colapso si el sobrevertido o la erosión se detienen de manera inminente.

4. Importancia de los parámetros de la brecha de rotura

Singh y Snorrason (1982) realizaron un estudio comparativo de la variación de los parámetros de la brecha de rotura con fin de predecir el caudal pico de descarga en el proceso de colapso de presas. Los estudios consideraron la influencia en la variación de: el ancho de la brecha, la profundidad, el tiempo de colapso y caudal de sobrevertido con rangos de amplitud similares a casos históricos observados.

Los resultados que estos investigadores obtuvieron, mostraron que:

1. Cambios en los parámetros de la brecha de rotura causaban cambios en el caudal pico de descarga.
2. La variación del tiempo de colapso causaba grandes cambios en el caudal pico de descarga.
3. Una reducción en el tiempo de colapso del 50% provocaba un incremento en el caudal pico de descarga de 13–83% en presas con volúmenes de embalse relativamente modestos pero únicamente de 1–5% para las de grandes volúmenes.
4. El incremento del caudal pico de descarga debido a grandes brechas resultaba de 6–50% para presas con volúmenes de embalse pequeños, mientras que era de 35–87% para las de grandes volúmenes.
5. La sensibilidad de la profundidad de la brecha fue relativamente pequeña (20%) y el cambio no demostró estar relacionado con el volumen de almacenamiento.
6. El caudal pico de descarga se correlaciona con el volumen de almacenamiento mejor que con la altura de la presa (un coeficiente de correlación del 96% frente al 70%).
7. Los resultados demostraron que las presas de tamaño modesto, con pequeñas capacidades de almacenamiento, son potencialmente más “peligrosas” que las grandes, ya que avenidas con periodos de retorno relativamente pequeños pueden hacerlas colapsar, produciendo caudales pico de descarga mucho mayores que los caudales de avenida que provocaron su colapso.

Petrascheck y Sydler (1984) demostraron la sensibilidad de la variación del caudal pico de descarga, los calados de las zonas de inundación, y el tiempo de arribo, a los cambios en el ancho de la brecha y en su tiempo de formación. Además concluyeron:

1. En localidades cercanas a la presa, la influencia de ambos parámetros (ancho de brecha y tiempo de formación de ésta) podía resultar dramática, ya que aguas abajo del pie de presa, el tiempo transcurrido hasta alcanzar el caudal pico de descarga puede alterarse de forma significativa debido a cambios en el tiempo de formación de la brecha.
2. El caudal pico de descarga y el calado de inundación resultaron independientes a los cambios en los parámetros de la brecha.

Estos resultados establecieron la necesidad de predecir con la máxima exactitud posible los parámetros de la brecha, para estimar de la manera más realista posible el caudal pico de descarga y la inundación resultante en las proximidades aguas abajo del pie de presa.

Wurbs (1987) concluyó que la simulación de la brecha es la principal fuente de incertidumbre en la modelización de una onda de crecida por rotura de presa. La importancia de los distintos parámetros depende del volumen del embalse. Sus resultados demostraron que en grandes volúmenes de embalses, el caudal pico de descarga sucederá cuando la brecha alcance su máxima anchura y profundidad. En estos casos, los cambios que se producen en la altura de descarga sobre la brecha resultan relativamente suaves durante el periodo de formación de la brecha. La exactitud de la estimación dependerá pues de la geometría de la brecha. Por otra parte, en pequeños embalses, existirá un cambio significativo en la cota del mismo durante la formación de la brecha. Para estos casos, el tiempo de formación de la brecha resulta ser un parámetro fundamental.

Wahl (1997) especificó la importancia de la precisión en la obtención de estos parámetros con el fin de evitar pérdidas materiales y humanas durante la avenida. Los casos de rotura de presas estudiados mostraron que la pérdida de vidas podría variar del 0,02 % de la población en riesgo con tiempos de alerta de 90 minutos, al 50 % de dicha población con tiempos de alerta menores de 15 minutos. El Tiempo de alerta, resultará directamente de la suma del tiempo de inicio de la brecha, el tiempo de formación de la brecha y el tiempo de la onda de avenida desde la presa hasta la población en riesgo.



Figura 2. Simulación artística de la rotura de la presa de South Fork (Thorton y otros, 2010).

La Figura 2 representa una simulación artística, del efecto devastador que el colapso de una presa, en este caso de materiales sueltos, puede llegar a provocar, lo cual nos hace recapacitar

sobre la importancia vital de estimar todos los parámetros de la brecha de un modo óptimo con el fin de minimizar daños materiales pero sobre todo pérdidas humanas. La Inundación de Johnstown (o la *Gran Inundación* como se conoce localmente) tuvo lugar el 31 de mayo de 1889, cuando la presa de South Fork construida a mediados del siglo XIX y situada 14 millas aguas arriba de la ciudad de Johnstown en Pensylvania (EE.UU), colapsó tras varios días de lluvias torrenciales (estimadas entre 150 y 250 mm en las últimas 24 horas), lo que desató un torrente de 20 millones de toneladas de agua y sedimentos (se estima un volumen de agua de 4800 millones de galones estadounidenses o 18,2 millones de m³). La inundación provocó más de 2200 pérdidas humanas y 17 millones de dólares en daños materiales. Se estima que la fuerza del muro de agua que golpeó la ciudad tuvo una fuerza similar al de las Cataratas del Niágara.

5. Fórmulas empíricas para la estimación de parámetros de rotura de presas

Desde finales de la década de los 70, muchos han sido los autores que mediante análisis correlativo de datos empíricos obtenidos en presas que tuvieron un colapso bien parcial o total, obtuvieron fórmulas empíricas capaces de estimar el caudal pico descarga producido por un colapso gradual de la presa. La Tabla 1, muestra algunas de las más importantes recogidas por Wahl (2004) y Thornton y otros (2010), clasificadas por tipo de parámetros de estudios y por diferentes autores, en unidades métricas (m, m³, m³/s), con el tiempo de colapso expresado en horas. Donde se muestran expresiones diferentes en idénticos autores, se hace para poner de manifiesto diferentes tipologías de presas (p.ej. presas de materiales sueltos de tierra frente presas de enrocamiento).

Como notación de las expresiones mostradas en la Tabla 1, consideramos que:

- V_a hace referencia al volumen de agua almacenada en el embalse sobre la base de la brecha en el momento de colapso de la presa (m³).
- V_{er} se refiere al volumen de la brecha erosionada (m³).
- C_b es un factor que varía dependiendo del volumen de embalse almacenado.
- K_0 se trata de un factor multiplicador que vale 1,4 para rotura por sobrevertido y 1,0 para rotura por tubificación.
- S se refiere al volumen de agua almacenada (m³).
- h_p es la altura de la presa (m).
- Se consideran también los parámetros ya especificados en la Figura 1.

6. Conjunto de datos a analizar

Se realizó un pequeño filtrado de la base de datos histórica de rotura de presas con el fin de poder idealizar el modelo de análisis multivariante a llevar a cabo, y se hicieron dos grupos de datos con los que aplicar la metodología fundamentados en las expresiones aparecidas en la Tabla 1.

Los datos principales utilizados hacen referencia a los siguientes parámetros:

- Q_p : Caudal pico de descarga en el proceso del colapso de la presa (medido en m³/s).

Tabla 1. Fórmulas empíricas para la estimación de parámetros de rotura de presas.

Referencia	Fórmula
Fórmulas de anchura de brecha	
Bureau of Reclamation (1988)	$\bar{B} = 3 h_a$
MacDonald y Landgridge-Monopolis (1984)	$V_{er} = 0,0261 (V_a \cdot h_a)^{0,769}$ (presas de mat. suelt.)
MacDonald y Landgridge-Monopolis (1984)	$V_{er} = 0,00348 (V_a \cdot h_a)^{0,852}$ (resto de presas)
Von Thun y Gillette (1990)	$\bar{B} = 2,5 h_a + C_b$
Froehlich (1995a)	$\bar{B} = 0,1803 K_0 (V_a)^{0,32} (h_b)^{0,19}$
Fórmulas para tiempos de colapso	
MacDonald y Landgridge-Monopolis (1984)	$t_f = 0,0179 V_{er}^{0,364}$
Von Thun y Gillette (1990)	$t_f = 0,015 h_a$ (presas altamente erosionables)
Von Thun y Gillette (1990)	$t_f = 0,020 h_a + 0,25$ (presas resistentes a erosión)
Von Thun y Gillette (1990)	$t_f = \bar{B}/(4 h_a)$ (presas resistentes a erosión)
Von Thun y Gillette (1990)	$t_f = \bar{B}/(4 h_a + 61)$ (presas altamente erosionables)
Froehlich (1995a)	$t_f = 0,00254 (V_a)^{0,53} (h_b)^{-0,9}$
Bureau of Reclamation (1988)	$t_f = 0,011 \bar{B}$
Fórmulas para el caudal pico de descarga	
Kirkpatrick (1977)	$Q_p = 1,268 (h_a + 0,3)^{2,5}$
Soil Conservation Service (1981)	$Q_p = 16,6 (h_a)^{1,85}$ (alt. presa > 31,4 m)
Hagen (1982)	$Q_p = 0,54 (S \cdot h_p)^{0,5}$
Hagen (1982)	$Q_p = 1,205 (V_a \cdot h_a)^{0,48}$
Bureau of Reclamation (1982)	$Q_p = 19,1 (h_a)^{1,85}$
Singh y Snorrason (1982)	$Q_p = 13,4 (h_p)^{1,89}$
Singh y Snorrason (1982)	$Q_p = 1,776 (S)^{0,47}$
MacDonald y Landgridge-Monopolis (1984)	$Q_p = 1,154 (V_a \cdot h_a)^{0,412}$
MacDonald y Landgridge-Monopolis (1984)	$Q_p = 3,85 (V_a \cdot h_a)^{0,411}$
Costa (1985)	$Q_p = 1,122 (S)^{0,57}$
Costa (1985)	$Q_p = 0,981 (S \cdot h_p)^{0,42}$
Costa (1985)	$Q_p = 2,634 (S \cdot h_p)^{0,42}$
Costa (1985)	$Q_p = 0,763 (V_a \cdot h_a)^{0,42}$
Evans (1986)	$Q_p = 0,72 (V_a)^{0,53}$
Froehlich (1995b)	$Q_p = 0,607 (V_a)^{0,295} (h_a)^{1,24}$
Walder y O'Connor (1997)	Q_p es estimado por métodos computacionales

- Shp: Factor multiplicativo del volumen acumulado de agua en el embalse (medido en m³) por la altura de la presa medida desde el lecho del cauce de la misma hasta su coronación (medido en m). Este factor por lo tanto viene medido en m⁴.
- Wavg: Ancho medio de la presa en sección transversal (medido en m).
- Vaha: Factor multiplicativo del volumen de agua acumulado en embalse desde la cota del fondo de la brecha de rotura hasta la cota de la lámina de agua de descarga, por dicha cota. Este factor por lo tanto viene medido en m⁴.
- hb: Altura de brecha de rotura medida desde la cota del fondo de dicha brecha hasta la coronación de la presa (medida en m).

- Bavg: Anchura media de la brecha de rotura (medida en m).
- Ver: Volumen de presa erosionado durante la formación de la brecha de rotura de la misma (medido en m³).

Con todo ello los dos conjuntos de datos seleccionados fueron los aparecidos en la Tabla 2.

Tabla 2. Conjunto de Datos.

Presa/Lugar	Qp	Wavg	Vaha	hb	Bavg	Ver
Apishapa, Colorado	6850	82,4	622000000	31,1	93	238000
Baldwin Hills, California	1130	59,6	11100000	21,3	25	31700
Buffalo Creek, West Virginia	1420	128	6780000	14	125	319000
Butler, Arizona	810	9,63	17000000	7,16	62,5	4310
Castlewood, Colorado	3570	47,4	133000000	21,3	44,2	55700
French Landing, Michigan	929	34,3	33000000	14,2	27,4	13800
Frenchman Creek, Montana	1420	37,3	173000000	12,5	54,6	28400
Hell Hole, California	7360	103,2	1070000000	56,4	121	555000
Ireland No. 5, Colorado	110	18	610000	5,18	13,5	1260
Johnstown (South Fork Dam, Penn.)	8500	64	465000000	24,4	94,5	68800
Kelly Barnes, Georgia	680	19,4	8780000	12,8	27,3	9940
Lake Avalon, New Mexico	2320	42,7	432000000	14,6	130	81000
Lake Genevieve, Kentucky	290	19,8	4560000	7,92	16,8	2630
Lambert Lake, Tennessee	1050	53,9	3790000	14,3	7,62	5870
Laurel Run, Pennsylvania	510	40,5	7830000	13,7	35,1	19500
Lawn Lake, Colorado	71	14,2	5350000	7,62	22,2	2400
Long Branch Canyon, California	340	11,3	900000	3,66	9,14	378
Lower Otay, California	1800	53,3	1950000000	39,6	133	107000
Merimac (Upper) Lake Dam, Georgia	1645	17,5	239000	3,05	14,2	758
Mossy Lake Dam, Georgia	9700	14,3	18200000	3,44	41,5	2040
Otter Lake, Tennessee	60	20,6	545000	6,1	9,3	1170
Potato Hill Lake, North Carolina	116	23,5	816000	7,77	16,5	3010
Prospect, Colorado	480	13,1	5950000	4,42	88,4	5120
Río Manzanares, New Mexico	7200	13,3	113000	7,32	13,3	1290
Schaeffer, Colorado	4500	80,8	135000000	30,5	137	227000
Teton, Idaho	65120	250	24000000000	86,9	151	3060000

7. Análisis

7.1. Propósito

Para construir el modelo, vamos a considerar que la variable Qp (Caudal pico de descarga) será nuestra *variable respuesta*, mientras que el resto serán nuestras *variables explicativas*.

7.2. Descripción de los datos analizados

Se ha utilizado una base de datos histórica de más de 100 presas con información de ciertos parámetros sobre el colapso de las mismas, con las que varios autores realizaron análisis

estadísticos con el fin de ofrecer varias expresiones empíricas con el fin de modelizar ciertos parámetros físicos como el caudal punta de descarga, o las dimensiones de la brecha de rotura en el proceso de colapso de las presas de materiales sueltos.

7.3. Metodología

Todo el estudio de la metodología utilizada se apoya en el software estadístico *R*³.

Una vez hemos cargado el conjunto de datos desde *Rcmdr*, procedemos a la lectura de los mismos, mediante la inserción del siguiente código en la consola de *R*:

```
> attach(Datos)
> Datos
```

El siguiente paso consiste en estudiar la representación gráfica de la variable respuesta con respecto a las variables explicativas consideradas. Para ello será necesario la inserción del siguiente código en la consola de *R*:

```
> pairs(~Qp+Wavg+Vaha+hb+Bavg+Ver, main="Matriz de diagramas de dispersión",
      pch=16, col="red")
```

lo cual produce la siguiente matriz de gráficos de dispersión representada en la Figura 3.

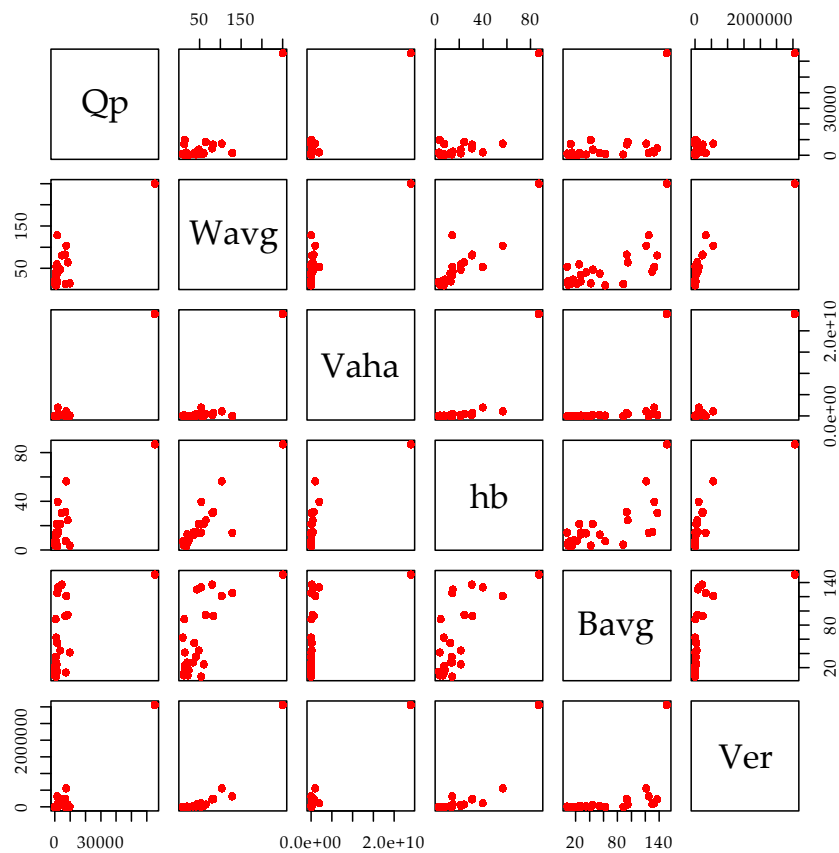


Figura 3. Matriz de diagramas de dispersión.

³ <https://www.r-project.org/>

Analizando la matriz de diagramas de dispersión no vemos a simple vista demasiados comportamientos lineales de unas variables con respecto a otras, excepto únicamente Wavg con hb. Sin embargo de este análisis si se se puede ver que existe un caso aparentemente “atípico” o lo que se denomina comúnmente *outlier*. Este caso pertenece a la rotura de la presa de Teton en Idaho, que tras un mes de su construcción colapsó debido a un proceso de tubificación. En este caso se trataba de una presa de grandes dimensiones (por ejemplo su altura llegaba casi a 100 metros), por lo que los datos de información sobre su colapso son demasiado grandes. Hemos procedido a eliminarlo del conjunto de datos y a efectuar nuevamente la matriz de diagramas de dispersión que se representa en la Figura 4.

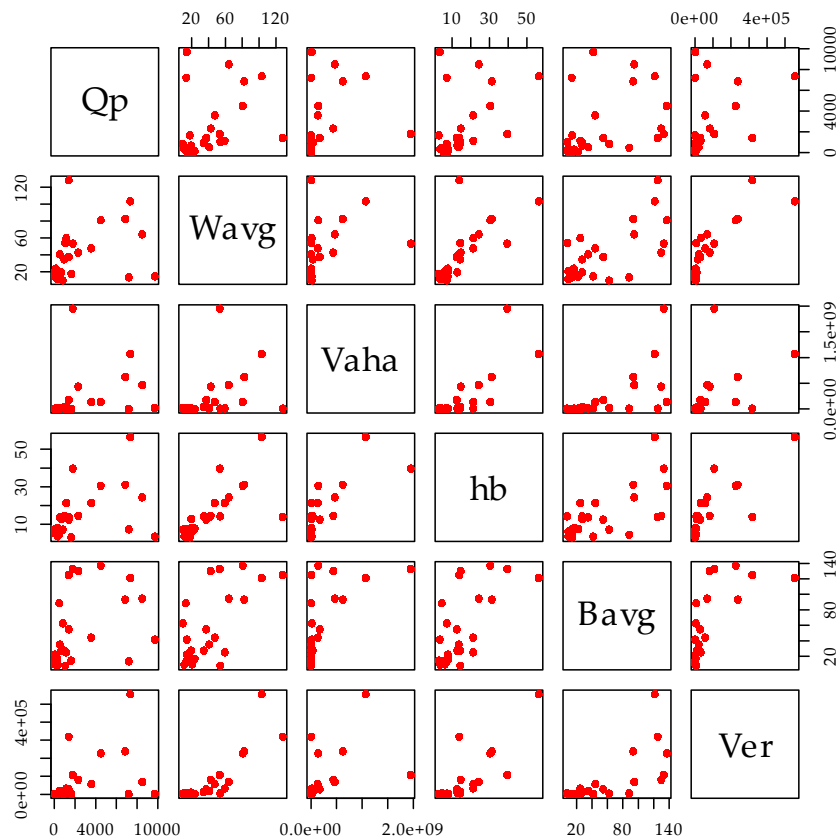


Figura 4. Matriz de diagramas de dispersión (con reducción de datos).

Podemos medir la intensidad de las relaciones lineales entre las distintas variables calculando sus *coeficientes de correlación* introduciendo el siguiente comando en la consola de R

```
> cor(Datos)
```

lo cual nos da la siguiente matriz:

	Qp	Wavg	Vaha	hb	Bavg	Ver
Qp	1.0000000	0.3307113	0.2925699	0.4313767	0.3650933	0.4233931
Wavg	0.3307113	1.0000000	0.3962958	0.7300872	0.6804174	0.8488001
Vaha	0.2925699	0.3962958	1.0000000	0.7752293	0.6267962	0.4951056
hb	0.4313767	0.7300872	0.7752293	1.0000000	0.6617470	0.8042378
Bavg	0.3650933	0.6804174	0.6267962	0.6617470	1.0000000	0.6971704
Ver	0.4233931	0.8488001	0.4951056	0.8042378	0.6971704	1.0000000

Los coeficientes de correlación toman valores entre -1 y 1 ; correspondiendo -1 a una relación lineal *negativa exacta* y 1 a una relación lineal *positiva exacta*. Un valor de 0 indica que *no existe relación lineal* ninguna entre las dos variables involucradas. Los elementos en la diagonal de la matriz son todos unos y representan el hecho de que la relación entre una variable y la misma variable es lineal positiva exacta (Pewsey, 2012). Además de los elementos de la diagonal principal que son iguales a 1 , pero que no tienen demasiado interés desde el punto de vista analítico, podemos observar que los coeficientes mayores en valor absoluto están entre las variables $Wavg \sim Ver(0,8488001)$ y $hb \sim Ver(0,8042378)$.

El siguiente paso del proceso de análisis consiste en considerar el modelo de regresión lineal múltiple, con todas las variables explicativas consideradas, esto es:

$$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_5 x_{5i} + \epsilon_i, \quad i = 1, \dots, n$$

donde

1. i identifica el *caso* y n el *número de casos*. En el caso del Conjunto de Datos 1 tenemos 25 observaciones distintas, es decir $n = 25$.
2. Y_i representa el valor de la variable respuesta para el caso i . En el caso de *Datos*, Y_i representa el caudal pico de descarga (medido en m^3).
3. β_0 es un *coeficiente de regresión* denominado *ordenada en el origen*, y representa el valor promedio de la variable respuesta cuando el valor de la variable explicativa es 0 .
4. x_{ji} es el valor observado de la *variable explicativa* j para el caso i . A diferencia del valor de la variable respuesta, se supone que el valor de la variable explicativa está medida con *exactitud* (es decir, *sin ningún tipo de perturbación*).
5. β_j ($j > 0$) son los *coeficientes de regresión* y representan el aumento en el *valor promedio* de la *variable respuesta* j correspondiente a un aumento de *una unidad* en el valor de la *variable explicativa*.
6. ϵ_i es una *perturbación* que proviene de una distribución normal con media cero y varianza σ^2 . Denotamos este hecho como $\epsilon_i \sim N(0, \sigma^2)$. Estas perturbaciones (o *errores*) describen la *variabilidad* de las observaciones alrededor del hiperplano $\beta_0 + \sum_{j=1}^8 \beta_j x_{ji}$ y son debidas a la influencia de otras variables *no especificadas* en la variable respuesta.

Vamos a proceder a realizar la estimación de los parámetros anteriormente mencionados que definen el *modelo de regresión lineal múltiple*. Para dicho ajuste consideramos la variable respuesta (caudal pico de descarga) como combinación lineal del resto de las variables explicativas, y a continuación generamos un resumen de dicho ajuste, para ello introducimos en la consola de R:

```
> lmDatos = lm(Qp~Wavg+Vaha+hb+Bavg+Ver)
> summary(lmDatos)
```

obteniendo los siguientes resultados:

```
Call:
lm(formula = Qp ~ Wavg + Vaha + hb + Bavg + Ver)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-1924.6 -1438.1 -1178.4  -241.8  8067.1
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1.104e+03	1.550e+03	0.713	0.485
Wavg	-2.451e+01	4.053e+01	-0.605	0.552
Vaha	-1.293e-06	2.708e-06	-0.477	0.639
hb	1.024e+02	1.245e+02	0.823	0.421
Bavg	1.299e+01	2.223e+01	0.584	0.566
Ver	5.518e-03	1.051e-02	0.525	0.606

Residual standard error: 2960 on 19 degrees of freedom

Multiple R-squared: 0.2246, Adjusted R-squared: 0.02055

F-statistic: 1.101 on 5 and 19 DF, p-value: 0.3924

Del análisis de estos resultados, podemos considerar las estimaciones de los parámetros de regresión, $\hat{\beta}_0 = 1,104 \times 10^3$, $\hat{\beta}_1 = -24,51, \dots$, y $\hat{\beta}_5 = 5,518 \times 10^{-3}$. Se considera un error estándar residual de $\sigma = 2960$, que explica el 22,46 % de la variabilidad del caudal pico de descarga estimado. Sin embargo, no todos los p -valores asociados a los coeficientes de regresión son menores de 0,05 (es decir tienen una *nivel de significación* mayor que el 5%. ¿Qué interpretación podemos hacer entonces? Por ejemplo si consideramos el p -valor más alto obtenido (0,639), corresponde a la variable Vaha. ¿Cuál es la hipótesis nula que estamos contrastando con este p -valor? Dicha hipótesis nula es que, *con todos los otros coeficientes de regresión incluidos en el modelo*, $\beta_1 = 0$. Entonces, estamos realizando un contraste sobre el efecto parcial de la variable Vaha. Con un p -valor tan grande, no podemos rechazar la hipótesis nula. Entonces, con todos los otros coeficientes de regresión incluidos en el modelo, parece perfectamente posible que β_1 sea 0, y por lo tanto esta variable no tenga un valor significativo en el ajuste realizado.

Del mismo modo no podemos considerar cierta la hipótesis de que $\beta_0 = 0$. Si en cualquiera de los contrastes de este tipo, el p -valor es *mayor* que 0,05 tenemos evidencia de que el modelo es *demasiado complejo* (es decir *contiene demasiados términos*). En consecuencia, debemos eliminar algunos de los términos del modelo. ¿Pero cuáles? No debemos quitar todas las variables no significativas “de golpe”, sino ir quitándolas, una a una, empezando con la variable *menos significativa*, es decir la de p -valor *más grande*. Entonces, en nuestro modelo debemos eliminar primero precisamente la variable Vaha (Pewsey, 2012).

Quitaremos la variable Vaha, reajustaremos el modelo reducido (sin Vaha pero con todas las otras variables) y veremos si quedan variables no significativas. Si las hay eliminaremos la variable menos significativa y seguiremos el proceso hasta que lleguemos a un modelo en el que todas las variables sean significativas (es decir que sus p -valores sean todos inferiores a 0,05). Si no pudiéramos quitar ninguna variable más, pero el p -valor asociado con la ordenada en el origen (Intercept) fuese mayor que 0,05, entonces finalmente deberíamos eliminar el término β_0 del modelo (Pewsey, 2012).

Vamos entonces a ajustar el siguiente modelo reducido, donde hemos eliminado la variable explicativa Vaha:

$$Y_i = \beta_0 + \beta_2 x_{2i} + \dots + \beta_5 x_{5i} + \epsilon_i, \quad i = 1, \dots, n$$

y volvemos a realizar un análisis de los resultados introduciendo en la consola de R:

```
> lmDatos2 = lm(Qp~Wavg+hb+Bavg+Ver)
> summary(lmDatos2)
```

obteniendo los siguientes resultados:

```
Call:
lm(formula = Qp ~ Wavg + hb + Bavg + Ver)
```

```

Residuals:
      Min       1Q   Median       3Q      Max
-2740.8 -1437.4 -1036.7   126.7  7976.4

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.439e+03  1.356e+03   1.062   0.301
Wavg         -1.739e+01  3.694e+01  -0.471   0.643
hb           5.763e+01  8.021e+01   0.719   0.481
Bavg         7.738e+00  1.894e+01   0.409   0.687
Ver          6.701e-03  1.001e-02   0.669   0.511

```

```

Residual standard error: 2902 on 20 degrees of freedom
Multiple R-squared: 0.2153,    Adjusted R-squared: 0.05837
F-statistic: 1.372 on 4 and 20 DF,  p-value: 0.2791

```

Podemos ver que el modelo aún es demasiado complejo y que por lo tanto puedo ser reducido aún más; en este caso eliminaremos la variable Bavg (la de p -valor mayor), y volvemos a repetir el proceso:

```

> lmDatos3 = lm(Qp~Wavg+hb+Ver)
> summary(lmDatos3)

```

```

Call:
lm(formula = Qp ~ Wavg + hb + Ver)

```

```

Residuals:
      Min       1Q   Median       3Q      Max
-2363.9 -1534.9 -1176.9   115.3  8090.0

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.577e+03  1.287e+03   1.226   0.234
Wavg         -1.427e+01  3.542e+01  -0.403   0.691
hb           6.451e+01  7.685e+01   0.839   0.411
Ver          7.409e-03  9.666e-03   0.766   0.452

```

```

Residual standard error: 2844 on 21 degrees of freedom
Multiple R-squared: 0.2088,    Adjusted R-squared: 0.09572
F-statistic: 1.847 on 3 and 21 DF,  p-value: 0.1696

```

Volvemos a realizar el proceso eliminando la variable Wavg (la de mayor p -valor):

```

> lmDatos4 = lm(Qp~hb+Ver)
> summary(lmDatos4)

```

obteniendo los siguientes resultados:

```

Call:
lm(formula = Qp ~ hb + Ver)

```

```

Residuals:

```

```

      Min      1Q  Median      3Q      Max
-2338.6 -1544.5 -1079.1  214.2  8239.6
    
```

Coefficients:

```

              Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.245e+03  9.682e+02  1.286  0.212
hb          5.984e+01  7.451e+01  0.803  0.430
Ver         4.900e-03  7.251e-03  0.676  0.506
    
```

```

Residual standard error: 2789 on 22 degrees of freedom
Multiple R-squared: 0.2026, Adjusted R-squared: 0.1302
F-statistic: 2.796 on 2 and 22 DF, p-value: 0.08283
    
```

Quitamos finalmente la última variable no significativa Ver, ya que su p -valor es mayor de 0,05.

```

> lmDatos5 = lm(Qp~hb)
> summary(lmDatos5)
    
```

obteniendo los siguientes resultados:

Call:

```
lm(formula = Qp ~ hb)
```

Residuals:

```

      Min      1Q  Median      3Q      Max
-3145.2 -1524.0 -956.6  460.9  8382.9
    
```

Coefficients:

```

              Estimate Std. Error t value Pr(>|t|)
(Intercept)  971.94      869.66  1.118  0.2753
hb           100.33      43.75  2.293  0.0313 *
---
    
```

```

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
    
```

```

Residual standard error: 2756 on 23 degrees of freedom
Multiple R-squared: 0.1861, Adjusted R-squared: 0.1507
F-statistic: 5.259 on 1 and 23 DF, p-value: 0.03131
    
```

En el último resumen reportado, podemos observar que el p -valor asociado a β_0 no es muy significativo (27,53%, por lo que el propio valor de β_0 pudiera ser considerado nulo ($\hat{\beta}_0 = 971,94$). Del anterior análisis debemos deducir que nuestro modelo no puede seguir siendo reducido. Las estimaciones puntuales de los parámetros de regresión son por lo tanto:

$$\hat{\beta}_0 = 971,94$$

$$\hat{\beta}_3 = 100,33$$

Además $\hat{\sigma} = 2756$, y la recta ajustada $Q_p = 971,94 + 100,33 \cdot hb$, explica el 18,61% de la variabilidad de la variable respuesta (caudal pico de descarga). Finalmente con un p -valor de 0,03131, se rechaza de manera rotunda la hipótesis nula de que $\beta_3 = 0$.

Para ver los intervalos de confianza del 95% para los coeficientes de regresión, basta introducir en la consola de R el siguiente código:

```
> confint(lmDatos5, level=0.95)

                2.5 %    97.5 %
(Intercept) -827.08973 2770.9709
hb           9.82219  190.8464
```

Para investigar las suposiciones sobre los errores necesitamos los valores ajustados para la variable respuesta y los residuos estudentizados, introduciendo en R:

```
> yajust= fitted(lmDatos5)
> sresDatos5=studres(lmDatos5)
> sresDatos5

      1      2      3      4      5      6
1.05883816 -0.72873178 -0.34749907 -0.32245522 0.16782250 -0.53510347
      7      8      9     10     11     12
-0.29284175 0.35451938 -0.51036917 2.02586973 -0.57565015 -0.04231475
     13     14     15     16     17     18
-0.54231438 -0.49416380 -0.67218320 -0.61318779 -0.36971380 -1.28408232
     19     20     21     22     23     24
0.13572116 4.11777684 -0.56228180 -0.60178984 -0.34518042 2.22028742
     25
0.17490015
```

Para investigar si existe estructura en los residuos estudentizados, producimos un diagrama de dispersión de ellos frente a los valores ajustados de la variable respuesta:

```
> plot(yajust, sresDatos5, main="Residuos estudentizados frente a los valores
ajustados", xlab="Valor ajustado", ylab="Residuo estudentizado", pch=16,
col="red")
```

Podemos observar que el residuo estudentizado correspondiente a la observación nº20 (Mossy Lake Dam, Georgia) es en valor absoluto mayor que 3, por lo que debiera ser considerado un *outliner*, y por lo tanto eliminado del análisis pues desvirtúa el mismo.

Con el modelo de regresión reducido obtenido anteriormente hay que realizar el estudio de los residuos una vez se ha eliminado de la serie del conjunto de datos el *outliner* anteriormente especificado. Eliminamos pues del conjunto de datos la información correspondiente a Mossy Lake Dam, Georgia, que es el que tiene un residuo estudentizado mayor que 3 (en valor absoluto), y volvemos a analizar los residuos:

```
> Datos6=Datos[-20,]
> attach(Datos6)
> lmDatos6=lm(Qp~hb)
> yajust=fitted(lmDatos6)
> sresDatos6=studres(lmDatos6)
> sresDatos6
```

lo que nos da el siguiente resultado:

```
      1      2      3      4      5      6
1.36773113 -0.85580750 -0.26271927 -0.13978626 0.31402789 -0.50888693
```

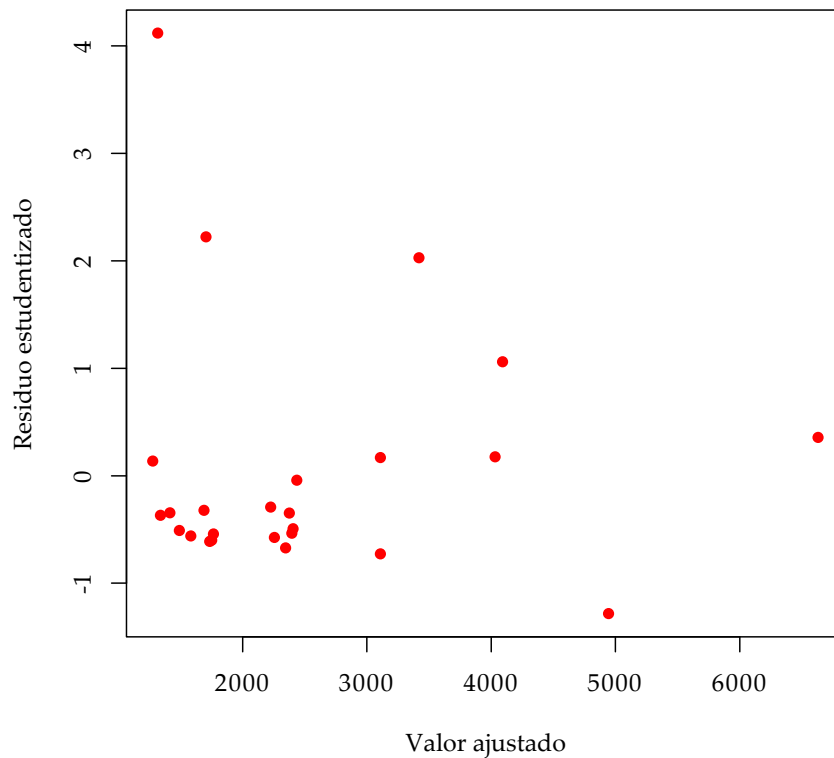



Figura 5. Residuos estudentizados frente a los valores ajustados.

7	8	9	10	11	12
-0.17233920	-0.01730428	-0.35572069	2.93854071	-0.54320402	0.12596967
13	14	15	16	17	18
-0.43473034	-0.45698433	-0.68080043	-0.52241042	-0.15248544	-1.91136139
19	20	21	22	23	24
0.51737101	-0.43561477	-0.50969873	-0.13142348	3.72938658	0.20322659

aún existe un *outlier*, el caso 23, tal y como muestra la Figura 6:

```
> plot(yajust, sresDatos6, main="Residuos estudentizados frente a los
valores ajustados", xlab="Valor ajustado", ylab="Residuo estudentizado",
pch=16, col="red")
```

Eliminamos el *outlier* observado en la Figura 6, y volvemos a repetir la operación.

```
> Datos7=Datos6[-23,]
> attach(Datos7)
> lmDatos7=lm(Qp~hb)
> yajust=fitted(lmDatos7)
> sresDatos7=studres(lmDatos7)
> sresDatos7
```

1	2	3	4	5	6
1.81165677	-0.96714134	-0.15429599	0.06146163	0.51142952	-0.46515140
7	8	9	10	11	12
-0.02763931	-0.27650259	-0.19119801	4.58317971	-0.49593507	0.33049711

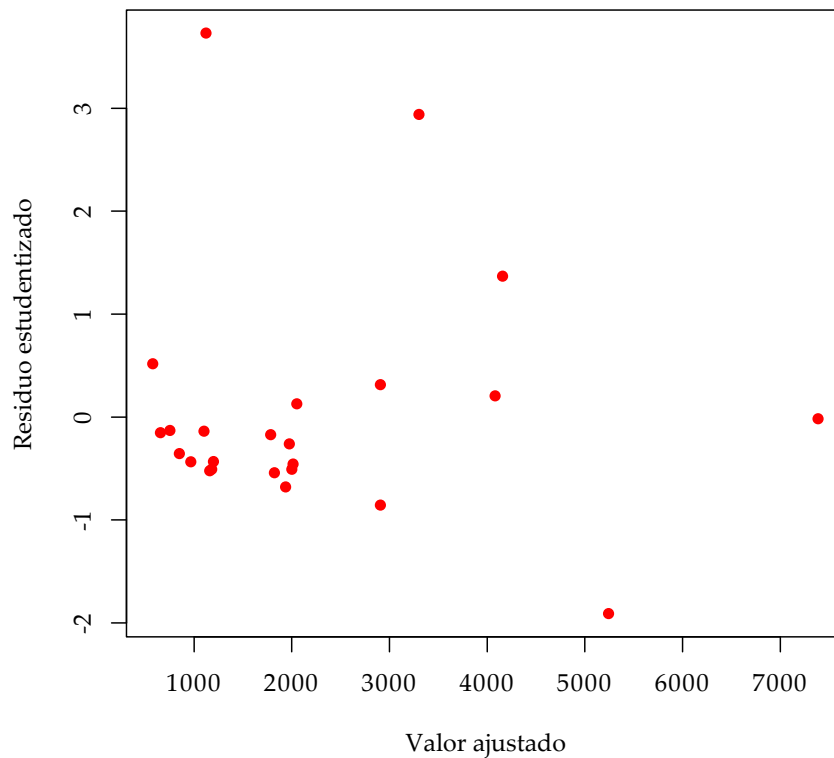


Figura 6. Residuos estudentizados frente a los valores ajustados (con reducción de datos).

13	14	15	16	17	18
-0.31583646	-0.40081380	-0.67713915	-0.42294320	0.07875159	-2.63084675
19	20	21	22	23	
0.94625692	-0.30000468	-0.40840219	0.09792709	0.29133360	

donde ahora podemos ver que existe un *outlier* en el caso 10, tal y como se muestra en la Figura 7.

```
> plot(yajust, sresDatos7, main="Residuos estudentizados frente a los
valores ajustados", xlab="Valor ajustado", ylab="Residuo estudentizado",
pch=16, col="red")
```

Nuevamente eliminamos el *outlier*, y repetimos el proceso.

```
> Datos8=Datos7[-10,]
> attach(Datos8)
> lmDatos8=lm(Qp~hb)
> yajust=fitted(lmDatos8)
> sresDatos8=studres(lmDatos8)
> sresDatos8
```

1	2	3	4	5	6
3.48480708	-1.09630691	-0.03318599	0.19900368	0.99343779	-0.46461458
7	8	9	10	11	12
0.12841396	0.45831495	-0.17385728	-0.52212342	0.65603757	-0.32003356
13	14	15	16	17	18

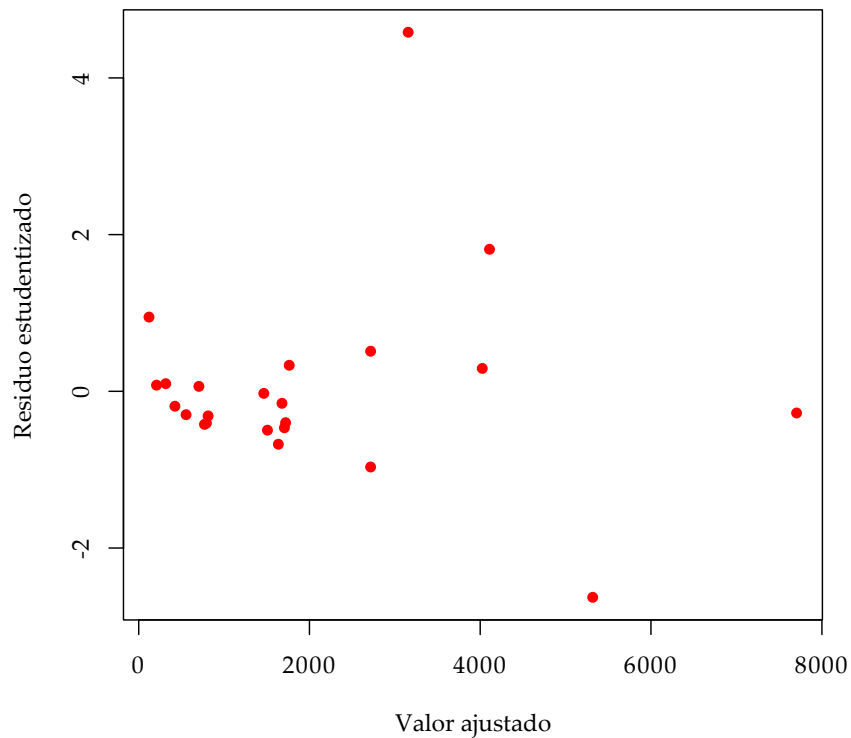


Figura 7. Residuos estudentizados frente a los valores ajustados (con reducción de datos).

```
-0.37373247 -0.76761563 -0.47296159 0.18752090 -3.43355087 1.43538707
          19          20          21          22
-0.31642297 -0.45104662 0.22214969 0.78243158
```

mediante la inserción del código

```
> plot(yajust, sresDatos8, main="Residuos estudentizados frente a los
      valores ajustados", xlab="Valor ajustado", ylab="Residuo estudentizado",
      pch=16, col="red")
```

En la Figura 8 se pueden identificar nuevamente dos nuevos *outliers* que eliminaremos y repetiremos el proceso.

```
> Datos9=Datos8[-1,]
> attach(Datos9)

> Datos10=Datos9[-16,]
> attach(Datos10)
> lmDatos10=lm(Qp~hb)
> yajust=fitted(lmDatos10)
> sresDatos10=studres(lmDatos10)
> sresDatos10
```

```
          1          2          3          4          5          6
-2.00976205 -0.04183506 0.41019203 1.63485458 -0.75373081 0.23705330
          7          8          9         10         11         12
```

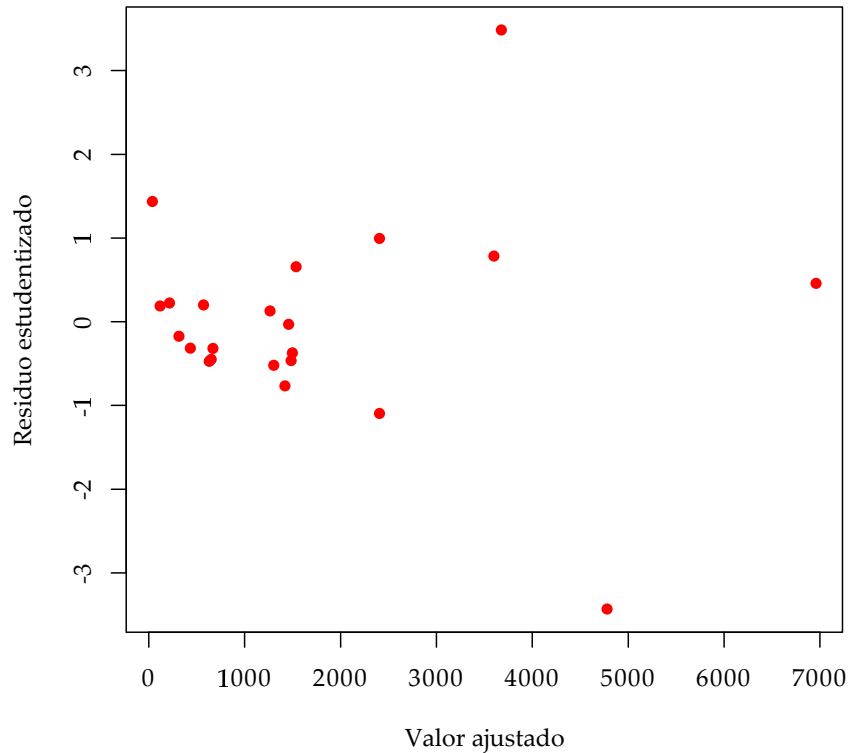


Figura 8. Residuos estudentizados frente a los valores ajustados (con reducción de datos).

```

0.18174984 -0.17500809 -0.83431830 1.09821665 -0.44355119 -0.60310084
      13         14         15         16         17         18
-1.27028565 -0.69249592 0.43102403 2.85426473 -0.41746511 -0.65779019
      19         20
0.47930885 1.15280771

```

ahora sí que no se observa ningún *outlier*. Si los representamos en un diagrama,

```

> plot(yajust, sresDatos10, main="Residuos estudentizados frente a los
valores ajustados", xlab="Valor ajustado", ylab="Residuo estudentizado",
pch=16, col="red")

```

Vemos por lo tanto que no existe ninguna estructura respecto a la distribución de los datos, por lo que se puede suponer cierta nuestra suposición en cuanto a *independencia* y homocedasticidad con respecto a los errores.

Reducido en gran medida nuestro conjunto de datos, veamos el listado resultante de los mismos:

```

> Datos10
  Qp  Wavg      Vaha  hb  Bavg  Ver
2  1130  59.60  11100000 21.30  25.00  31700
3  1420 128.00   6780000 14.00 125.00 319000
4   810   9.63  17000000  7.16  62.50  4310
5  3570  47.40 133000000 21.30  44.20  55700
6   929  34.30  33000000 14.20  27.40  13800
7  1420  37.30 173000000 12.50  54.60  28400

```

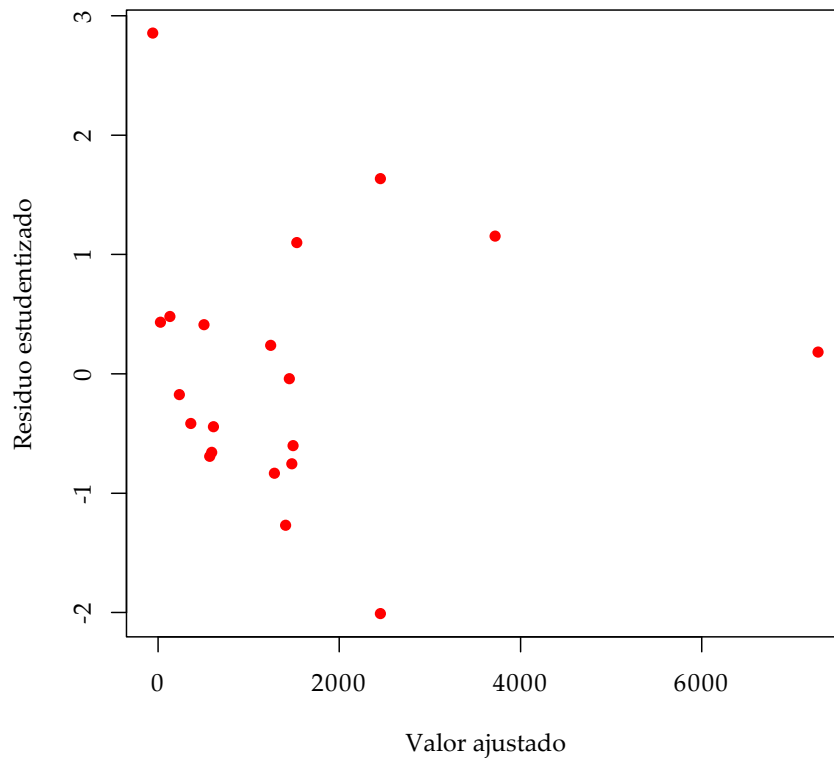


Figura 9. Residuos estudentizados frente a los valores ajustados (con reducción de datos definitivo).

8	7360	103.20	1070000000	56.40	121.00	555000
9	110	18.00	610000	5.18	13.50	1260
11	680	19.40	8780000	12.80	27.30	9940
12	2320	42.70	432000000	14.60	130.00	81000
13	290	19.80	4560000	7.92	16.80	2630
14	1050	53.90	3790000	14.30	7.62	5870
15	510	40.50	7830000	13.70	35.10	19500
16	71	14.20	5350000	7.62	22.20	2400
17	340	11.30	900000	3.66	9.14	378
19	1645	17.50	239000	3.05	14.20	758
21	60	20.60	545000	6.10	9.30	1170
22	116	23.50	816000	7.77	16.50	3010
23	480	13.10	5950000	4.42	88.40	5120
25	4500	80.80	135000000	30.50	137.00	227000

En la Figura 10 se puede ver representado un diagrama de dispersión de los datos de hb vs. Qp, es decir la altura de la brecha de rotura de presa (medido en m) en el eje de abcisas, frente al Caudal pico de descarga (medido en m³/s) en el eje de ordenadas, donde puede verse la fuerte linealidad de dichos datos del modelo reducido resultante. También se ha añadido la recta ajustada para los datos aparecidos en el modelo:

```
> plot(Qp, hb, main="Diagrama de dispersión hb vs. Qp", xlab="Altura de
brecha (m)", ylab="Caudal pico de descarga (m3/s)", pch=16, col="red")
> lmDatos=lm(Qp ~ hb)
> yajust=fitted(lmDatos)
> lines(hb, yajust, lwd=2, lty=2, col="blue")
```

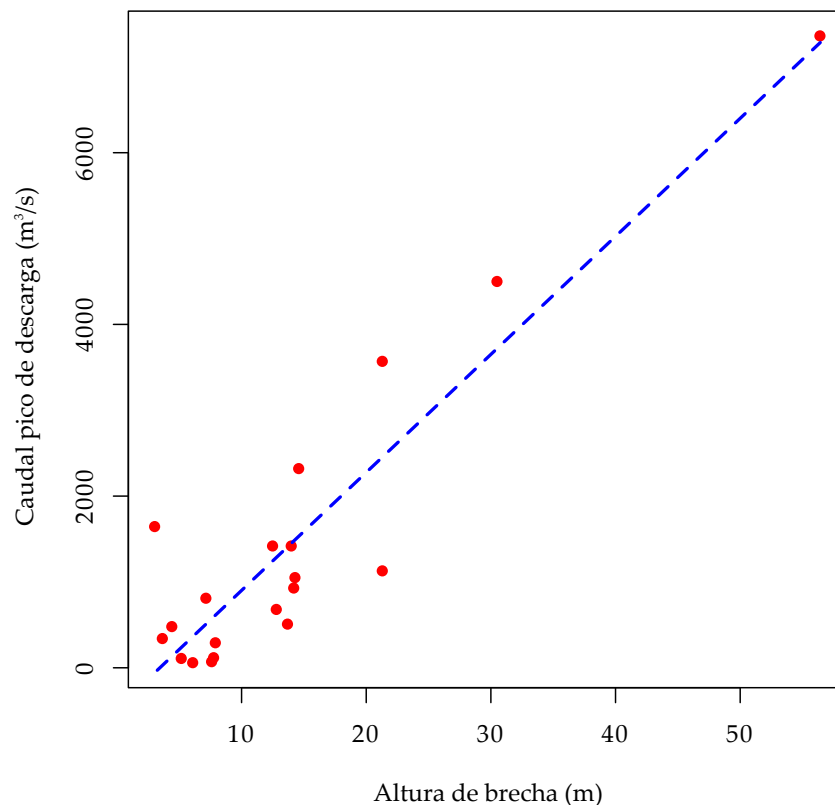


Figura 10. Diagrama de dispersión hb vs. Qp .

Una vez hemos reducido en gran medida nuestro conjunto de datos, veamos en que puede influir este hecho a la estimación realizada de los parámetros del modelo:

```
> lmDatos = lm(Qp~hb)
> summary(lmDatos)

Call:
lm(formula = Qp ~ hb)

Residuals:
    Min       1Q   Median       3Q      Max
-1325.20 -484.09  -79.36   320.24 1700.29

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -474.85     255.11  -1.861   0.0791 .
hb             137.56     13.96   9.855 1.12e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 739 on 18 degrees of freedom
Multiple R-squared:  0.8436,    Adjusted R-squared:  0.835
F-statistic: 97.12 on 1 and 18 DF,  p-value: 1.118e-08
```

En el último resumen reportado, podemos observar que el p -valor asociado a β_0 es muy

significativo (7,91 %), por lo que el propio valor de β_0 pudiera ser considerado nulo ($\hat{\beta}_0 = -474,85$). Del anterior análisis debemos deducir que nuestro modelo no puede seguir siendo reducido. Las estimaciones puntuales de los parámetros de regresión son por lo tanto:

$$\hat{\beta}_0 = -474,85$$

$$\hat{\beta}_3 = 137,56$$

Además $\hat{\sigma} = 739$, y la recta ajustada $Q_p = -474,85 + 137,56 \cdot hb$, explica el 84,36 % de la variabilidad de la variable respuesta (caudal pico de descarga). Finalmente con un p -valor de $1,118 \times 10^{-8}$, se rechaza de manera rotunda la hipótesis nula de que $\beta_3 = 0$.

Para ver los intervalos de confianza del 95 % para los coeficientes de regresión, basta introducir en la consola de R el siguiente código:

```
> confint(lmDatos5, level=0.95)

                2.5 %   97.5 %
(Intercept) -1010.8098  61.1132
hb           108.2352 166.8866
```

Como se puede ver, aunque pequeña, existe posibilidad de que $\beta_0 = 0$, ya que este valor está contenido en el intervalo de confianza anterior.

Nos queda por último realizar la suposición en cuanto a la *normalidad* de los datos. Para ello producimos un histograma de los residuos estudentizados con una estimación de densidad núcleo y la densidad de la distribución normal estándar superpuestas, mediante la introducción del siguiente código en la consola de R:

```
> yajust=fitted(lmDatos)
> sresDatos=studres(lmDatos)
> sresDatos
> hist(sresDatos, probability=TRUE, main="Histograma de Residuos
  Estudentizados", xlab="Residuo Estudentizado", ylab="Densidad", col="grey")
> lines(density(sresDatos, bw="SJ"), col="blue", lwd=2)
> x = seq(from=-3.5, to=3.5, by=0.05)
> lines(x, dnorm(x, 0, 1), col="red", lwd=2, lty=2)
```

A continuación representamos un gráfico Q-Q (Cuantil-Cuantil) normal con una línea de referencia como se representa en la Figura 12.

```
> qqnorm(sresDatos, main="Gráfico Q-Q normal", xlab="Cuantil teórico",
  ylab="Cuantil muestral", pch=16, col="red")
> qqline(sresDatos, lwd=2, lty=2, col="blue")
```

Finalmente, aplicamos el contraste de normalidad de Shapiro-Wilk, apropiada para conjuntos de datos en los que $n < 30$.

```
> shapiro.test(sresDatos)

Shapiro-Wilk normality test

data:  sresDatos
W = 0.9575, p-value = 0.4956
```

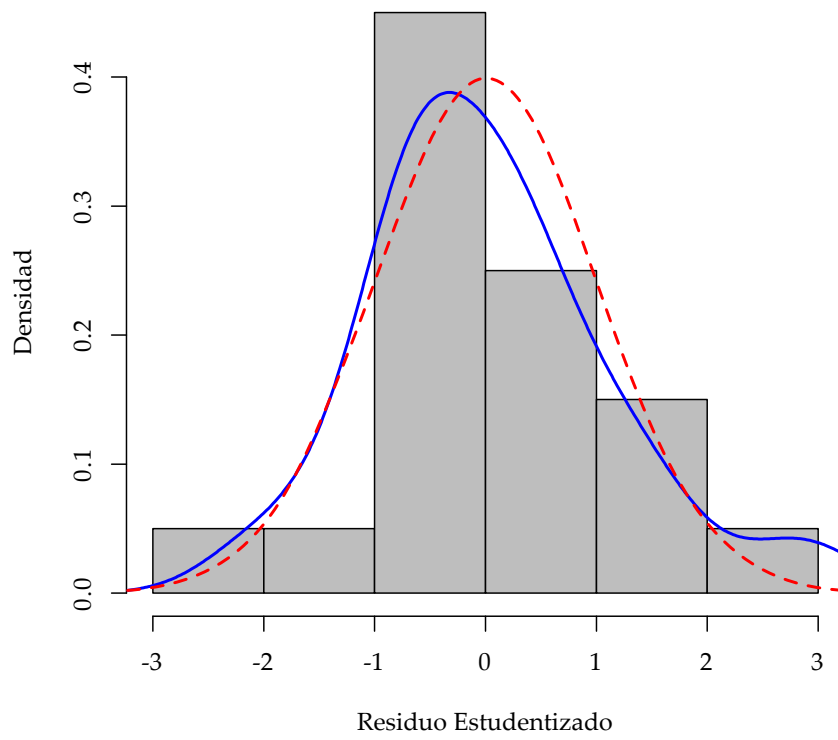


Figura 11. Histograma de Residuos Estudentizados.

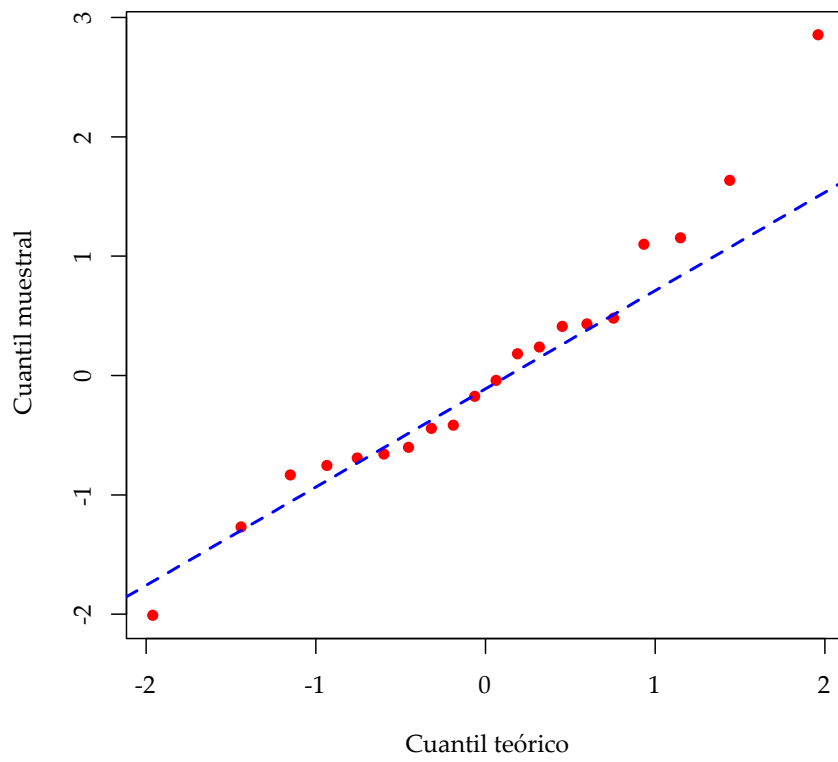


Figura 12. Gráfico Q-Q normal.

El histograma de los residuos se asimila a la forma de la campana de Gauss, los puntos del gráfico Q-Q normal parece que no se desvían demasiado de la diagonal principal. Además, en virtud de que el p -valor estimado en el Test de Shapiro-Wilk es mayor que 0,05 (en nuestro caso 0,4956), existe evidencia para considerar que la hipótesis nula no se cumple, por lo tanto nuestro conjunto de datos sigue una distribución normal.

8. Conclusiones

1. Con el conjunto de datos original analizado para desarrollar un modelo consistente, fue necesario llevar a cabo una “limpieza” de los mismos con el fin de eliminar los *outliers* que pudieran desvirtuar nuestro modelo resultante. Así conseguimos ajustar nuestro conjunto de datos con el modelo representado por la expresión

$$Q_p = -474,85 + 137,56 \cdot hb$$

que cumple todas las suposiciones de independencia, homocedasticidad y normalidad, ajustando más del 80 % de la variabilidad de la variable respuesta, en este caso el caudal pico de descarga.

2. Observando la Tabla 1, es conveniente utilizar modelos de regresión no lineal, con el fin de que el modelo supuesto para el primer conjunto de datos pueda ajustarse mejor. Es evidente que autores como MacDonald y Landgridge-Monopolis (1984) o Costa (1985) utilizaron este tipo de modelos para el análisis estadístico de los datos.
3. Quizás podríamos haber realizado un análisis más fino, considerando una serie de datos mayor ya que al haber tenido que utilizar un conjunto de datos completos para un mínimo de 6 variables, nuestro campo de actuación se ha restringido en mayor medida. Quizás sería interesante, realizar un análisis más pormenorizado con un número menor de variables explicativas de partida.

Referencias

- [1] BUREAU OF RECLAMATION: «Guidelines for defining inundated areas downstream from Bureau of Reclamation dams». Reclamation Planning Instruction N° 82-11, U.S. Department of the Interior, Bureau of Reclamation, Denver, 25, 1982.
- [2] BUREAU OF RECLAMATION: «Downstream hazard classification guidelines». ACER Tech. Memorandum N° 11. U.S. Department of the Interior, Bureau of Reclamation, Denver, 57, 1988.
- [3] COSTA, J. E.: «Floods from dam failures». U.S. Geological Survey, Open file Rep., 1985, 85, p. 560, Denver, 54.
- [4] EVANS, S. G.: «The maximum discharge of outburst floods caused by the breaching of man-made and natural dams». Canadian Geotech. Journal, 1986, 24 (4), pp. 385–387.
- [5] EZEKIEL, M.: *Methods of Correlation Analysis*. J. Wiley & Sons, inc, New York, 1930.
- [6] FREAD, D. L.: *DAMBRK: The NWS dam-break flood forecasting model*. Hydrologic Research Laboratory, National Weather Service, NOAA, 1984.
- [7] FREAD, D. L.: *BREACH, an erosion model for earthen dam failures*. Hydrologic Research Laboratory, National Weather Service, NOAA, 1988.

- [8] FROEHLICH, D. C.: «Embankment-dam breach parameters revisited». Water Resources Engineering, Proc. ASCE Conf. on Water Resources Engineering, New York, pp. 887–891, 1995a.
- [9] FROEHLICH, D. C.: «Peak outflow from breached embankment dam». Water Resources Plan. Manage. Div., Am. Soc. Civ. Eng., 121 (1), pp. 90–97, 1995b.
- [10] HAGEN, V. K.: «Re-evaluation of design floods and dam safety». Proc. 14th Congress of Int. Commission on Large Dams, International Commission on Large Dams, Paris, 1982.
- [11] KIRKPATRICK, G. W.: «Evaluation guidelines for spillway adequacy». The evaluation of dam safety, Engineering Foundation Conf., ASCE, New York, pp. 395–414, 1977.
- [12] MACDONALD, T. C. and LANDRIDGE-MONOPOLIS, J.: «Breaching characteristics of dam failure». Journal of Hydraulic Engineering, 1984, 110(5), pp. 567–586.
- [13] MACNEIL, D. R.: *Interactive Data Analysis*. J. Wiley & Sons, inc, New York, 1977.
- [14] MINISTERIO DE MEDIO AMBIENTE DE ESPAÑA: *Guía Técnica para la elaboración de Planes de Emergencia de Grandes Presas*. Secretaría de Estado de Aguas y Costas. Dirección General de Obras Hidráulicas y Calidad de las Aguas. Subdirección General de Gestión del Dominio Público Hidráulico, 2001.
- [15] MONTGOMERY, D. C.: *Diseño y Análisis de Experimentos*. Limusa Wiley, 2ª ed., 2003.
- [16] PETRASCHECK, A.W. and SYDLER, P.A.: «Routing of Dam Break Floods». International Water Power and Dam Construction, 1984, 36, pp. 29–32.
- [17] PEWSEY, A.: *Apuntes de Técnicas Estadísticas Avanzadas de Investigación en Geotecnologías*. Universidad de Extremadura, 2012.
- [18] SINGH, K. P. and SNORRASON, A.: *Sensitivity of outflow peaks and flood stages to the selection of dam breach parameters and simulation models*, 1982.
- [19] SOIL CONSERVATION SERVICE: «Simplified dam-breach routing procedure». Tech. Release, 1981, 66 (Rev. 1), p. 39.
- [20] THORNTON, C. I.; PIERCE, M. W. and ABT, S. R.: *Predicting Peak Outflow from Breached Embankment Dams*. National Dam Safety Review Board Steering Comitee on Dam Breach Equations. Colorado State University, 2010.
- [21] VON THUN, J. L. and GILLETTE, D. R.: «Guidance on breach parameters». Internal Memorandum, U.S. Dept. of the Interior, Bureau of Reclamation, Denver, 17, 1990.
- [22] WAHL, T. L.: «Predicting Embankment Dam Breach Parameters - A Needs Assessment». XXVIIth IAHR Congress San Francisco, California. U. S. Bureau of Reclamation, Denver, Colorado, USA, 1997.
- [23] WAHL, T. L.: *Prediction of embankment dam breach parameters: a literature review and needs assessment*. U.S. Department of the Interior, Bureau of Reclamation, Dam Safety Office, 1998.
- [24] WAHL, T. L.: «Uncertainty of Predictions of Embankment Dam Breach Parameters». Journal of Hydrologic Engineering, 2004, pp. 389–397.
- [25] WALDER, J. S. and O'CONNOR, J. E.: «Methods for predicting peak discharge of floods caused by failure of natural and constructed earth dams». Water Resour. Res., 1997, 33(10), p. 12.
- [26] WURBS, R. A.: «Dam-Breach Flood Wave Models». Journal of Hydraulic Engineering, Vol. 1987, 113, No. 29, pp. 29–46.

Sobre el autor:

Nombre: José Manuel Sánchez Muñoz

Correo electrónico: jmanuel.sanchez@gmx.es

Institución: Ingeniero de Caminos, Canales y Puertos. Profesor de Enseñanza Secundaria. Grupo de Innovación Educativa "Pensamiento Matemático", Universidad Politécnica de Madrid, España.

