

ANEXO X

ANALISIS DE FRECUENCIAS DE LAS CRECIDAS DEL RIO SALADO

INDICE GENERAL

1 INTRODUCCION.....	8
2 REVISIÓN DE ASPECTOS BASICOS SOBRE EL ANÁLISIS DE FRECUENCIA....	11
2.1 <i>Conceptos Básicos de Hidrología Estadística</i>	11
2.2 <i>Enfoques Metodológicos Alternativos</i>	13
2.3 <i>Revisión de Conceptos Asociados a Tests de Hipótesis y de Adecuación</i>	15
3 EMPLEO DE INFORMACIONES HISTORICAS EN EL ANÁLISIS DE FRECUENCIA DE CAUDALES MAXIMOS	33
3.1 <i>Introducción</i>	33
3.2 <i>El Uso de la Distribución GEV en Hidrología</i>	33
3.3 <i>La distribución GEV</i>	34
3.4 <i>Estimación de los Parámetros de la Distribución GEV</i>	36
3.5 <i>El Método de Máxima Verosimilitud (MLE) en Muestras Pequeñas</i>	36
3.6 <i>Estimadores de Máxima Verosimilitud Generalizados (GMLE)</i>	38
3.7 <i>Incorporación de Información Histórica</i>	40
3.8 <i>Análisis de Máxima Verosimilitud con Información Histórica Basado en Series Anuales</i>	41
3.9 <i>Resultados de Experimentos Monte Carlo</i>	43
3.10 <i>Otros Modelos Probabilísticos Considerados</i>	43
4 ANÁLISIS DE FRECUENCIA SOBRE SERIES DE DATOS HASTA EL AÑO 2002.	47
4.1 <i>Análisis de Frecuencia sobre Series Sistemáticas (S)</i>	47
4.2 <i>Análisis de Frecuencia sobre Series Reconstruidas y Sistemáticas (R+S)</i>	63
4.3 <i>Análisis de Frecuencia sobre Series Históricas y Sistemáticas (H+S)</i>	72
5 ANÁLISIS DE FRECUENCIA SOBRE SERIES DE DATOS HASTA EL AÑO 2005.	81
5.1 <i>Introducción</i>	81
5.2 <i>Análisis de Frecuencia sobre Series Sistemáticas (S)</i>	81
5.3 <i>Análisis de Frecuencia sobre Series Reconstruidas y Sistemáticas (R+S)</i>	93
5.4 <i>Análisis de Frecuencia sobre Series Históricas y Sistemáticas (H+S)</i>	101
5.5 <i>Análisis Comparativo de los Estudios de Frecuencia Efectuados</i>	108
6 MAGNITUD COMPARADA DE LA CRECIDA DEL AÑO 2003	111
7 ANÁLISIS DEL RIESGO DE FALLA DE LA OBRA DE DEFENSA OESTE	113
7.1 <i>Introducción</i>	113
7.2 <i>Análisis Estadístico Clásico del Riesgo de Falla</i>	114
7.3 <i>Probabilidad de Crecidas del Río Salado Condicionadas al Sistema Paraná</i>	115
7.4 <i>Análisis de la Vulnerabilidad Hídrica del Tramo II del Terraplén Oeste</i>	118
8 CONCLUSIONES.....	120

INDICE DE FIGURAS

Figura X. 1. Modelos de Descomposición de las Series Temporales (Fuente: Catalini, 2004).	29
Figura X. 2. Remoción de tendencias y saltos en una serie de tiempo anual. (Fuente Maidment, 1992).	32
Figura X. 3. Relación entre el parámetro de forma κ y el coeficiente de asimetría de la distribución GEV (fuente: Meylan y Musy, 1996).	35
Figura X. 4. Distribución verdadera ($\kappa = -0,20$) y “estimada” ($\hat{\kappa} = -2,00$) para una pequeña muestra de tamaño $n = 15$ (valores muestrales indicados por cuadrados). (Fuente: Martins, 2005).	37
Figura X. 5. Funciones de densidad de probabilidad (fdps) de la distribución GEV para $\kappa = -0,3, -0,1$ e $+0,3$ (donde $\xi = 10$ e $\alpha = 2,6$) (Fuente: Martins, 2005).	38
Figura X. 6. <i>Prior Geofísico</i> para el parámetro κ (de forma) de la distribución GEV. Distribución Beta con media $= -0,10$ y variancia $= (0,122)^2$ (Fuente: Martins, 2005).	39
Figura X. 7. Serie anual con H años de información histórica, con K = 4 picos de crecidas en el período histórico superiores al umbral (o nivel) de percepción y S años de registros sistemáticos (Fuente: Martins, 2005).	41
Figura X. 8. Gráfico de tendencia en las series de caudales máximos RP70 (Período 1954-2002).	49
Figura X. 9. Prueba de datos dudosos en RP70 (periodo: 1954-2002).	50
Figura X. 10. Prueba de datos dudosos en RP70 (periodo: 1954-1972).	51
Figura X. 11. Prueba de datos dudosos en RP70 (periodo: 1972-2002).	51
Figura X. 12. Prueba de datos dudosos en RP70 (periodo: 1954-2002, serie estandarizada mediante corrección de media y varianza).	52
Figura X. 13. Análisis gráfico de normalidad de la serie 1954-2002.	54
Figura X. 14. Análisis gráfico de normalidad de la serie 1954-2002.	55
Figura X. 15. Análisis gráfico de normalidad de la serie 1954-2002.	55
Figura X. 16. Posición de ploteo de las series analizadas de acuerdo a la formula de Gringorten.	57
Figura X. 17. Grafica de los ajustes realizados sobre la serie RP70 1954-2002 estandarizada.	59
Figura X. 18. Grafica de los ajustes sobre la serie estandarizada 1954-2002 (RP70), con Tr asociada.	59
Figura X. 19. Grafica de los ajustes realizados sobre la serie 1972-2002 (RP70).	62
Figura X. 20. Grafica de los ajustes realizados sobre la serie 1972-2002, con recurrencia asociada (RP70).	62
Figura X. 21. Gráfico de tendencia en las series de caudales máximos RP70 (Período 1928-2002).	64
Figura X. 22. Prueba de datos dudosos en RP70 (periodo: 1928-2002).	65
Figura X. 23. Prueba de datos dudosos en RP70 (periodo: 1928-2002 serie corregida por la media).	66
Figura X. 24. Posición de ploteo de las series analizadas de acuerdo a la formula de Gringorten.	69
Figura X. 25. Grafica de los ajustes realizados sobre la serie estandarizada 1928-2005 (RP70).	71

Figura X. 26 Grafica de los ajustes realizados con recurrencia asociada sobre la serie estandarizada 1928-2002.	71
Figura X. 27. Cuantiles de crecida del modelo GEV/MLE y sus límites de confianza del 95%, asociados a diferentes tiempos de retorno calculados a partir del uso de información sistemática (GEV-SIS) y del uso adicional de información histórica (GEV-BCD y GEV-CD).	75
Figura X. 28. Cuantiles de crecida del modelo GEV/GMLE y sus límites de confianza del 95%, asociados a diferentes tiempos de retorno calculados a partir del uso de información sistemática (GEV-SIS) y del uso adicional de información histórica (GEV-BCD y GEV-CD).	76
Figura X. 29. Cuantiles de crecida LN/MLE y sus límites de confianza de 95%, asociados a diferentes tiempos de retorno calculados a partir del uso de información sistemática (LN-SIS) y del uso adicional de información histórica (LN-BCD y LN-CD).	77
Figura X. 30. Cuantiles de crecida LPIII/Abordaje Bayesiano y sus límites de confianza de 95%, asociados a diferentes tiempos de retorno calculados a partir del uso de información sistemática (LPIII-SIS) y del uso adicional de información histórica (LPIII-BCD y LPIII-CD).	78
Figura X. 31 Comparación de los cuantiles de crecida GEV/GML y sus límites de confianza de 95%, con los cuantiles correspondientes y sus límites de confianza de 95% obtenidos con LPIII/Abordaje Bayesiana calculados a partir del uso de información sistemática e histórica (Caso CD).	78
Figura X. 32. Cuantiles GEV/GML-CD de caudales máximos y sus límites de confianza de 95% asociados a diferentes tiempos de retorno y caudales observados asociados a sus respectivas probabilidades de excedencia calculadas para la serie 1875-2002 (el valor de la crecida de 2003 ha sido incorporado solo a los efectos gráficos).	80
Figura X. 33. Gráfico de tendencia en las series de caudales máximos RP70 (Período 1954-2005).	83
Figura X. 34. Prueba de datos dudosos en RP70 (periodo: 1954-2005)	84
Figura X. 35. Prueba de datos dudosos en RP70 (periodo: 1972-2005)	84
Figura X. 36. Prueba de datos dudosos en RP70 (periodo: 1954-2005, serie estandarizada por corrección de media y varianzas).	85
Figura X. 37. Posición de ploteo de las series analizadas de acuerdo a la formula de Gringorten.	88
Figura X. 38. Grafica de los ajustes realizados sobre la serie estandarizada 1954-2005 (RP70).	90
Figura X. 39. Grafica de los ajustes realizados sobre la serie estandarizada 1954-2005 (RP70).	90
Figura X. 40. Grafica de los ajustes realizados sobre la serie 1972-2005 (RP70)	92
Figura X. 41.. Grafica de los ajustes realizados con recurrencia asociada sobre la serie 1972-2005 (RP70).	92
Figura X. 42. Gráfico de tendencia en las series de caudales máximos RP70 (Período 1928-2005).	94
Figura X. 43. Prueba de datos dudosos en RP70 (periodo: 1928-2005)	95

Figura X. 44. Prueba de datos dudosos en RP70 (periodo: 1928-2005 serie corregida por la media).....	95
Figura X. 45. Posición de ploteo de las series analizadas de acuerdo a la formula de Gringorten.....	98
Figura X. 46. Gráfica de los ajustes realizados sobre la serie estandarizada 1928-2005 (RP70).....	100
Figura X. 47. Gráfica de los ajustes realizados con Tr asociada sobre la serie estandarizada 1928-2005.....	100
Figura X. 48. Cuantiles de crecida del modelo GEV/MLE y sus límites de confianza del 95%, asociados a diferentes tiempos de retorno calculados a partir del uso de información sistemática (GEV-SIS) y del uso adicional de información histórica (GEV-BCD y GEV-CD). Informaciones hasta el año 2005.	103
Figura X. 49. Cuantiles de crecida del modelo GEV/GMLE y sus límites de confianza del 95%, asociados a diferentes tiempos de retorno calculados a partir del uso de información sistemática (GEV-SIS) y del uso adicional de información histórica (GEV-BCD y GEV-CD). Informaciones hasta el año 2005.	104
Figura X. 50. Cuantiles de crecida LN/MLE y sus límites de confianza de 95%, asociados a diferentes tiempos de retorno calculados a partir del uso de información sistemática (LN-SIS) y del uso adicional de información histórica (LN-BCD y LN-CD). Informaciones hasta el año 2005.	105
Figura X. 51. Cuantiles de crecida LPIII/Abordaje Bayesiano y sus límites de confianza de 95%, asociados a diferentes tiempos de retorno calculados a partir del uso de información sistemática (LPIII-SIS) y del uso adicional de información histórica (LPIII-BCD y LPIII-CD). Informaciones hasta el año 2005.....	106
Figura X. 52. Comparación de los cuantiles de crecida GEV/GML y sus límites de confianza de 95%, con los cuantiles correspondientes y sus límites de confianza de 95% obtenidos con LPIII/Abordaje Bayesiana calculados a partir del uso de información sistemática e histórica (Caso CD). Informaciones hasta el año 2005.	106
Figura X. 53. Cuantiles GEV/GML-CD de caudales máximos y sus límites de confianza de 95% asociados a diferentes tiempos de retorno y caudales observados asociados a sus respectivas probabilidades de excedencia calculadas para la serie 1875-2005.....	107
Figura X. 54. Curva envolvente de los caudales extremos de crecidas registrados en el mundo, en función del área de la cuenca de aporte.....	112
Figura X. 55. Magnitud comparada de la crecida del año 2003 con los máximos registrados a nivel mundial.....	112

INDICE TABLAS

Tabla X. 1. Valores de Kn de un lado con un nivel de significancia del 10% para la distribución normal.....	30
Tabla X. 2. Muestra generada de tamaño n = 15 de una GEV($\xi=0$, $\alpha=1$, $\kappa = -0.20$) (Fuente: Martins, 2005).....	36
Tabla X. 3. Resumen estadístico de las series de caudales máximos (RP70, periodo 1954-2002).....	49
Tabla X. 4. Prueba t, suponiendo varianzas iguales.....	53
Tabla X. 5. Prueba t', suponiendo varianzas desiguales.....	53
Tabla X. 6. Prueba F para varianzas de dos muestras.....	54
Tabla X. 7. Prueba de Independencia (Wald-Wolfowitz).....	56
Tabla X. 8. Prueba de Estacionalidad (Kendall).....	56
Tabla X. 9. Prueba de Homogeneidad (Wilcoxon).....	57
Tabla X. 10. Resultados de los ajustes realizados sobre la serie estandarizada 1954-2002.....	58
Tabla X. 11. Resultados de los ajustes realizados sobre la serie 1972-2002.....	61
Tabla X. 12. Estadística Descriptiva de la serie completa.....	64
Tabla X. 13. Prueba t, suponiendo varianzas iguales (periodo 1928-2002).....	67
Tabla X. 14. Prueba t', suponiendo varianzas desiguales (periodo 1928-2002).....	67
Tabla X. 15. Prueba F para varianzas de dos muestras.....	68
Tabla X. 16. Prueba de Independencia (Wald-Wolfowitz).....	68
Tabla X. 17. Prueba de Estacionalidad (Kendall).....	68
Tabla X. 18. Prueba de Homogeneidad (Wilcoxon).....	69
Tabla X. 19. Resultados de los ajustes realizados sobre la serie estandarizada 1928-2002.....	70
Tabla X. 20. Serie completa de caudales máximos anuales empleada en los estudios estadísticos.....	73
Tabla X. 21. Resultados obtenidos a partir del Modelo GEV/MLE (informaciones hasta el año 2002).....	74
Tabla X. 22. Resultados obtenidos a partir del Modelo GEV/GMLE (informaciones hasta el año 2002).....	75
Tabla X. 23. Resultados para el Modelo LN/MLE (informaciones hasta el año 2002).....	76
Tabla X. 24. Resultados para el Modelo LPIII/Abordaje bayesiano (informaciones hasta el año 2002).....	77
Tabla X. 25. Tiempos de retorno [años] correspondientes a caudales máximos del río Salado estimados mediante la distribución GEV-GMLE (informaciones hasta el año 2002).....	80
Tabla X. 26. Resumen estadístico de las series de caudales máximos (RP70, Periodo 1954-2005).....	82
Tabla X. 27. Prueba t, suponiendo varianzas iguales (periodo 1954-2005).....	86
Tabla X. 28. Prueba t', suponiendo varianzas desiguales (periodo 1954-2005).....	86
Tabla X. 29. Prueba F para varianzas de dos muestras (periodo 1954-2005).....	87
Tabla X. 30. Prueba de Independencia Wald-Wolfowitz (periodo 1954-2005).....	87
Tabla X. 31. Prueba de Estacionalidad Kendall (periodo 1954-2005).....	87
Tabla X. 32. Prueba de Homogeneidad Wilcoxon (periodo 1954-2005).....	88
Tabla X. 33. Resultados de los ajustes realizados sobre la serie estandarizada 1954-2005.....	89
Tabla X. 34. Resultados de los ajustes realizados sobre la serie 1972-2005.....	91

Tabla X. 35. Estadística Descriptiva de la serie completa.....	93
Tabla X. 36. Prueba t, suponiendo varianzas iguales (periodo 1928-2005).....	96
Tabla X. 37. Prueba t', suponiendo varianzas desiguales (periodo 1928-2005).....	96
Tabla X. 38. Prueba F para varianzas de dos muestras	97
Tabla X. 39. Prueba de Independencia (Wald-Wolfowitz)	97
Tabla X. 40. Prueba de Estacionalidad (Kendall)	97
Tabla X. 41. Prueba de Homogeneidad (Wilcoxon)	98
Tabla X. 42. Resultados de los ajustes realizados sobre la serie estandarizada 1928-2005	99
Tabla X. 43. Resultados obtenidos a partir del Modelo GEV/MLE (informaciones hasta el año 2005).	102
Tabla X. 44. Resultados para el Modelo GEV/GMLE (informaciones hasta el año 2005). ..	103
Tabla X. 45. Resultados para el Modelo LN/MLE (informaciones hasta el año 2005).	104
Tabla X. 46. Resultados para el Modelo LPIII/Abordaje bayesiano (info. hasta el año 2005). ..	105
Tabla X. 47. Tiempos de retorno [años] correspondientes a caudales máximos del río Salado estimados mediante la distribución GEV-GMLE (informaciones hasta el año 2005). ..	108
Tabla X. 48. Recurrencias estimadas para la crecida de 2003 ($Q= 3.954 \text{ m}^3/\text{s}$) según distintas series de análisis.	110
Tabla X. 49. Frecuencia de ocurrencia (en %) de distintas situaciones diarias conjuntas en los ríos Paraná-Salado (base: período 1928-2004).	117

1 INTRODUCCION

La previsión hidrológica de crecidas en una determinada sección de un río puede ser efectuada a *corto plazo* o a *largo plazo*.

La *previsión de corto plazo se denomina pronóstico hidrológico en tiempo real* y consiste en indicar tanto el momento en el que la misma ocurrirá como su magnitud más probable.

Por el contrario, la *previsión de largo plazo* consiste en indicar la probabilidad de que ocurra una determinada crecida (o una superior a ella), sin precisar el tiempo exacto en el cual la misma se producirá. En otras palabras, la previsión de largo plazo es de tipo estadística y consiste en estimar la probabilidad de que una determinada crecida se presente en un año cualquiera. La previsión de largo plazo se asocia, por tanto, al *grado de previsibilidad (estadística) de la crecida*.

La previsión estadística de crecidas se basa fundamentalmente en el análisis de frecuencia de datos de caudales y/o niveles en la sección del río bajo análisis.

En este Anexo se presentan los trabajos realizados con el objeto de definir, por un lado, la previsibilidad estadística de la crecida del río Salado de abril/mayo de 2003 (o, dicho en otras palabras, el *tiempo de recurrencia* o *tiempo de retorno* de la citada crecida) y, por otro lado, las probabilidades a considerar en el análisis de riesgo de falla de la obra de Defensa Oeste de la ciudad de Santa Fe.

Cabe indicar que los estudios de frecuencia se han basado en la consideración del año hidrológico (1° de septiembre-31 de agosto). Al hacerse referencia de las series empleadas, se indica el año inicial y final de cada período, es decir, la serie 1954-2002 corresponde al período comprendido entre el 1° de septiembre de 1954 y el 31 de agosto de 2002.

Este Anexo se divide en siete partes principales. Cabe mencionar que a lo largo del Anexo se presentan algunos ítems destinados a facilitar la comprensión del lector menos avezado en temas de hidrología frecuencial. Otros, en cambio, se orientan a la justificación técnica de los trabajos realizados, hecho que exige un buen conocimiento previo de los temas abordados.

La primera parte de este Anexo se inicia con la revisión de una serie de conceptos básicos de hidrología estadística. Posteriormente se identifican distintos enfoques metodológicos alternativos asociados al análisis de frecuencias de crecidas de un río. También se revisan algunos conceptos asociados a los test paramétricos y no paramétricos empleados en el estudio.

La segunda parte es consagrada al enfoque derivado de la introducción de marcas históricas y umbrales de percepción. El mismo es abordado con detalle dado que su aplicación ha sido determinante para estimar las probabilidades asociadas a la crecida de abril/mayo de 2003.

La tercera parte de este Anexo sintetiza los estudios de frecuencia realizados con base en las informaciones disponibles hasta el año 2002, es decir, todas aquellas informaciones disponibles previo a la ocurrencia de la crecida de abril/mayo de 2003. Se incluyen en este ítem:

- (i) el estudio de frecuencia basado en la serie sistemática de caudales correspondiente al período 1954-2005, registrada en la sección de medición de la Ruta Provincial 70 (RP70);
- (ii) los estudios de frecuencia efectuados considerando tanto la serie sistemática de caudales antes indicada como la serie reconstruida de caudales correspondiente al período 1928-1954. Esta última se basa en el análisis de los registros existentes en la sección de medición RP70 y en la sección del puente Empalme San Carlos (ESC). Una descripción detallada del proceso mediante el cual fueron estimados los caudales de la serie reconstruida se presenta en el Anexo IX;
- (iii) los estudios de frecuencia efectuados considerando el enfoque que contempla las marcas hidrológicas de crecidas históricas y la inclusión de niveles o umbrales de percepción (o de selección de crecidas). De acuerdo a lo citado en el Anexo IX, la serie así conformada corresponde al período 1875-2002.

Cabe mencionar que el objetivo de presentación de todos estos estudios ha sido la de conformar una base de análisis que permita justificar de manera clara la adopción del enfoque de marcas históricas para definir el grado de previsibilidad estadística de la crecida de abril/mayo de 2003.

La cuarta parte de este Anexo sintetiza los estudios de frecuencia realizados con base en todas las informaciones disponibles hasta la actualidad, es decir, incorporando los datos correspondientes a los años 2003, 2004 y 2005. De la comparación de resultados obtenidos con aquellos presentados en el apartado precedente se deduce el impacto que la crecida de abril/mayo de 2003 tuvo sobre la estimación del tiempo de recurrencia o retorno de las crecidas del río Salado.

La quinta parte de este Anexo se refiere al análisis comparado de la magnitud de la crecida del año 2003 con registros máximos similares observados a nivel mundial.

La sexta parte se refiere a la estimación de las probabilidades asociadas al riesgo hídrico de falla de la obra de Defensa Oeste de la ciudad.

La séptima y última parte se refiere a las conclusiones acerca de los principales aspectos tratados a lo largo de este Anexo.

A fin de definir los alcances y limitaciones de los estudios presentados en este Anexo, cabe advertir que:

- La estimación de las crecidas históricas fue realizada con base en todas las informaciones que fue posible recolectar durante el tiempo de desarrollo de este estudio. Ello no exime de la conveniencia de producir actualizaciones toda vez que nuevas investigaciones conduzcan a la obtención de informaciones de peso para los fines estadísticos.
- La adopción del enfoque de marcas históricas y umbrales de percepción redujo significativamente la relevancia de la determinación exacta de los caudales correspondientes al período 1875-1953. En efecto, siendo preciso disponer del orden de magnitud de los caudales ocurridos en aquel período, se estima que los estudios hidráulicos efectuados han sido compatibles con los fines perseguidos.

2 REVISIÓN DE ASPECTOS BASICOS SOBRE EL ANÁLISIS DE FRECUENCIA

2.1 Conceptos Básicos de Hidrología Estadística

Probabilidad (P) y Tiempo de Retorno (Tr):

En su concepción más elemental la probabilidad es la relación (o el cociente) entre el número de casos favorables y el número de casos posibles. Para ejemplificar, considérese un dado que posee seis caras (números 1 a 6). En una jugada cualquiera la probabilidad de salir el número 4 es $P = 1/6$ (1 chance en 6 posibles).

El tiempo de retorno es el número *promedio* de jugadas luego de las cuales el número deseado se repite. Se define como la inversa de la probabilidad:

$$Tr = 1/P \quad (X. 1)$$

donde Tr es el tiempo de retorno y P es la probabilidad de ocurrencia en una jugada cualquiera.

En el caso analizado, usando la ecuación (X.1) resulta $Tr = 1/(1/6) = 6$. Por lo tanto, *en promedio*, el número 4 se repite a cada seis jugadas. Se sabe que ese número no ocurre exactamente a cada seis jugadas. Sin embargo, si el juego se repite un gran número de veces y se realiza el promedio, ciertamente este será igual a seis. En síntesis, el número 4 puede aparecer dos veces seguidas y pasar muchas jugadas sin aparecer, pero *en promedio* se repetirá a cada seis jugadas.

Probabilidad (P) y Tiempo de Retorno (Tr) de una crecida de un río:

Utilizando los datos históricos de caudales o niveles en un determinado lugar de interés puede ser estimada la probabilidad de que un determinado caudal (o nivel) sea igualado o superado en un año cualquiera. Para ello es fundamental disponer de datos hidrológicos que permitan realizar esta estimación.

Supóngase, por simplicidad que un río posee un régimen hidrológico tal que provoca una crecida por año. Realizando una analogía, cada año representa para las crecidas del río lo mismo que cada jugada del dado para el número 4. Un tiempo de retorno (o de recurrencia) de 10 años significa que, en promedio, la crecida se puede repetir a cada 10 años o que en cada año la misma posee el 10 % de chance de ocurrir. De modo similar a lo ocurrido con el número 4, la crecida con 10 años de tiempo de recurrencia puede presentarse en dos años seguidos, como también pueden pasar muchos años sin ocurrir, pero *en promedio* se espera que se produzca una vez a cada diez años.

Riesgo Hídrico:

Es la probabilidad de que ocurra un determinado fenómeno hidrológico en un determinado período de tiempo. Para ejemplificar puede averiguarse cuál es la probabilidad de que la crecida de $Tr = 10$ años ocurra en los próximos 5 años. Es decir, aquí interesa conocer cuál es la probabilidad de ocurrencia en todo un período y no sólo para un año cualquiera. La ecuación a aplicar es la siguiente:

$$P_n = 1 - (1 - 1/Tr)^n \quad (X. 2)$$

donde n es el número de años en el cual se desea estimar la probabilidad; P_n es la probabilidad deseada. Para la pregunta arriba citada la respuesta resulta:

$$P_n = 1 - (1 - 1/10)^5 = 0,41 \text{ o } 41\% \quad (X. 3)$$

Es decir, existe un 41% de probabilidad de que la crecida se presente en los próximos 5 años. Cuando se desea estimar el riesgo hídrico de falla de una obra se adopta a n como la vida útil de la obra, siendo que P_n resulta el riesgo hídrico de falla de la misma.

Clasificación Estadística de las Crecidas:

Diversos autores clasifican a las crecidas de los cursos como *ordinarias* si su período de retorno no excede 10 años; *extraordinarias* cuando está entre dicho valor y 100 años y *excepcionales* para recurrencias mayores. Se reserva la denominación de *episódicas* para aquellas crecidas generalmente asociadas a tiempos de retorno significativamente altos, de las cuales generalmente existen rasgos en el paisaje que se advierten en los análisis geológicos y geomorfológicos.

Estudio de frecuencia de caudales máximos:

El estudio de frecuencia de caudales máximos es uno de los tópicos más estudiados de la Hidrología, dada la necesidad de estimar la probabilidad de ocurrencia de crecidas para el diseño de obras hidráulicas, protección de ciudades, etc.

El enfoque clásico del análisis de frecuencia se basa en el empleo de una serie de datos observados de manera sistemática en una sección o punto de interés de un río. Con base en dicha serie (denominada muestra de datos) se analizan la capacidad de distintos modelos probabilísticos (o distribuciones teóricas) para representar las características de la serie sistemática observada. De la misma manera que existen distintos modelos probabilísticos posibles, también existen distintos métodos para determinar el valor de los parámetros de cada distribución o modelo. Este hecho le otorga a cada modelo mayor o menor posibilidad de adecuación a los datos sistemáticos disponibles. Cabe consignar que no todos los métodos de ajuste de parámetros se adecuan a todos los modelos posibles.

El método de ajuste de parámetros más clásico se denomina “método de los momentos”. Este método ha sido frecuentemente utilizado en décadas pasadas debido a la facilidad de su implementación. Sin embargo, a medida que se ha venido disponiendo de computadoras más poderosas y de programas más accesibles, otros métodos se han impuesto por ser mucho más precisos que aquel. Entre estos últimos cabe citar al método de máxima verosimilitud (MLE por sus siglas en inglés, *Maximum Likelihood Estimation*) y el método de máxima verosimilitud generalizado (GMLE, *Generalized Maximum Likelihood Estimation*). Las virtudes de estos métodos de estimación de parámetros frente al tradicional método de los momentos son relatadas en numerosos libros y publicaciones científicas internacionales.

Los dos últimos métodos indicados (MLE y GMLE) han sido empleados mayormente en este estudio. El método de los momentos ha sido empleado al solo efecto de efectuar comparaciones con estudios preexistentes. Un tratamiento más pormenorizado de los métodos de ajuste de parámetros de las distribuciones probabilísticas se presenta en el apartado 3.

2.2 Enfoques Metodológicos Alternativos

Existen distintas alternativas para encarar el estudio de frecuencia de crecidas de un río. A continuación se presentan en forma sintética los tres enfoques básicos empleados en este estudio.

Enfoque Clásico:

Conforme indicado precedentemente, el enfoque más clásico se basa en el empleo de una serie de valores observados de manera sistemática en la sección o punto de interés del río. Este enfoque será referenciado en este Anexo como estudio de frecuencia sobre la serie sistemática de caudales del río Salado, representado por la letra (S).

Enfoque de Población Mezcla:

Cuando las series de caudales muestran cambios sustanciales en la relación Caudal-Frecuencia (Q-Tr) suele ser de interés el análisis de la posibilidad de los datos provenir, en realidad, de una población “mezcla”. Esta posibilidad surge al admitirse que la serie de datos disponibles reúne valores correspondiente a dos (o más) poblaciones estadísticas diferentes. Un caso típico de existencia de población mezcla surge cuando se analizan crecidas provocadas por huracanes y por tormentas simples; otro caso de población mezcla se verifica al analizar crecidas originadas por lluvias y por deshielo, etc.

Conforme se ilustra posteriormente, en el caso del río Salado la relación antes indicada (Q-Tr) muestra un cambio de pendiente. Este hecho permitió considerar inicialmente la posibilidad de existencia de una población mezcla en la serie de caudales máximos sistemáticos: crecidas con aportes de los Bajos Submeridionales y crecidas sin aporte de los Bajos Submeridionales. Sin embargo, un análisis pormenorizado del origen de diversas

crecidas permitió verificar la existencia de crecidas de uno y otro origen en distintas partes del gráfico, no siendo posible corroborar la hipótesis de existencia de distintas poblaciones.

Enfoque Basado en Informaciones Históricas (marcas históricas y umbrales de percepción):

Consiste en incorporar, además de la serie de valores sistemáticos, aquellas estimaciones de caudales máximos asociados a marcas históricas de crecidas de gran envergadura que ocurrieron en el pasado. También pueden incluirse otras informaciones tales como una serie de valores de caudales reconstruidos a partir del análisis de todas las informaciones disponibles sobre el comportamiento del curso en períodos anteriores al inicio de la toma sistemática de observaciones.

Pese a tratarse de un tópico abordado en diferentes publicaciones científicas y técnicas, el enfoque de marcas históricas ha recibido menor atención por parte de los técnicos. Entre los motivos que justifican este comportamiento deben citarse las dificultades que a menudo existen para recolectar y combinar todas las informaciones históricas, hecho que requiere un sustancial tiempo de revisión en bibliotecas, archivos, diarios de la época, documentos históricos, etc. Entre las ventajas que otorga la incorporación de marcas históricas se destaca la estimación más precisa de la recurrencia (o retorno) de las crecidas máximas de baja frecuencia (*extraordinarias, excepcionales u episódicas*) como así también la disminución del intervalo de incertidumbre alrededor de las estimaciones efectuadas.

En este trabajo se aplican dos metodologías basadas en la incorporación de marcas históricas, ambas basadas en los trabajos propuestos por Martins y Stedinger (2001) y Reis y Stedinger (2005). El trabajo se basa en el estudio de la serie 1875-2005 conformada luego de efectuarse los trabajos indicados en el Anexo IX. Las metodologías alternativas empleadas utilizan:

- a. el número de crecidas que superan un determinado umbral (o valor) de crecidas además de la serie sistemática disponible. Esta alternativa es referenciada en este informe como BCD (por sus siglas en inglés, *Binomial Censored Data*).
- b. la magnitud de las crecidas que superan un determinado umbral (o valor) de crecidas además de la serie sistemática disponible. Esta alternativa es referenciada en este informe como CD (por sus siglas en inglés, *Censored Data*).

Un tratamiento más pormenorizado de las metodologías que emplean informaciones históricas se presenta en el apartado 3.

2.3 Revisión de Conceptos Asociados a Tests de Hipótesis y de Adecuación

2.3.1 Procedimientos paramétricos

Los procedimientos de prueba paramétricos pueden definirse como aquellos que requieren que el nivel de medición obtenido con los datos recolectados esté en forma de una escala de intervalos o de una escala de cocientes, o que implican la prueba de hipótesis de valores de parámetros especificados (como puede ser la media) y requieren un conjunto limitante de suposiciones. Es preciso decidir qué tipo de procedimientos se aplica según si:

1. las mediciones obtenidas con los datos son solamente categóricas (es decir, están escaladas nominalmente) o en rangos (es decir, están escaladas ordinalmente);
2. las suposiciones subyacentes en el uso de los métodos paramétricos no se cumplen estrictamente;
3. la situación requiere el estudio de características tales como aleatorización, independencia, simetría o bondad de ajuste en lugar de la prueba de hipótesis con respecto a valores específicos de parámetros de población en particular.

2.3.2 Procedimientos sin distribución y no paramétricos

Cuando no se pueden aplicar los métodos paramétricos de la prueba de hipótesis, se pueden elegir métodos libres de distribución o no paramétricos de prueba de hipótesis.

Los *procedimientos de prueba sin distribución* pueden definirse ampliamente como aquellos cuya estadística de prueba no depende de la forma de la distribución de la población subyacente de la cual se tomó la muestra de datos o como aquellos para los cuales los datos no tienen fuerza suficiente (es decir, están escalados nominalmente u ordinalmente) para garantizar operaciones aritméticas significativas.

Los *procedimientos no paramétricos* pueden definirse como aquellos que no tienen que ver con los parámetros de una población.

A continuación se presentara un procedimiento libre de distribución, la prueba de rangos con signo de Wilconxon, para una mediana supuesta y un procedimiento no paramétrico, la prueba de corridas de una muestra de Wald-Wolfowitz, para la aleatoriedad.

Existen cinco principales ventajas del uso de procedimientos sin distribución o de procedimientos no paramétricos:

1. Pueden utilizarse en todo tipo de datos; categóricos (escalados nominalmente), en forma de rangos (escalados ordinalmente), así como en datos que han sido medidos con más precisión (escalados de intervalo o de cociente)

2. Por lo general son fáciles de aplicar y rápidos de calcular, cuando los tamaños de la muestra son pequeños.
3. Implican un menor número de suposiciones menos limitantes (que se pueden cumplir con más facilidad) que los métodos paramétricos. En consecuencia, tiene una mayor aplicabilidad y producen un conjunto de conclusiones más general y con una base más amplia.
4. Los métodos no paramétricos permiten la resolución de problemas que no implican la prueba de parámetros de la población.
5. Dependiendo del procedimiento particular elegido, los métodos sin distribución pueden ser igualmente (o casi) poderosos que el procedimiento paramétrico correspondiente cuando las suposiciones del último se cumplen, y cuando no se cumplen puede ser bastante más poderos.

A pesar de que los procedimientos sin distribución y los no paramétricos pueden ser empleados de manera ventajosa en una variedad de situaciones, poseen tres principales desventajas, a saber:

1. Resulta desventajoso utilizar métodos libres de distribución cuando se pueden cumplir todas las suposiciones del procedimiento paramétrico.
2. Conforme aumenta el tamaño de la muestra, la manipulación de datos requerida para efectuar los procedimientos sin distribución y los no paramétricos es, a menudo, laboriosa, a menos que se disponga de un paquete de software adecuado.
3. Con frecuencia se necesitan tablas especiales de valores críticos para las estadísticas de prueba obtenidas mediante el uso del procedimientos sin distribución o de procedimientos no paramétricos, y no es tan fácil disponer de tales tabla como en el caso de las requeridas para obtener los valores críticos para las estadísticas de prueba de los procedimientos paramétricos (Z , t y χ^2).

2.3.3 Importancia de las suposiciones en la selección de la prueba

La sensibilidad de los procedimientos paramétricos a la violación de las suposiciones ha sido tema de consideración en la literatura estadística. Se dice que algunos procedimientos de prueba paramétricos son *robustos* por que son relativamente insensibles a ligeras violaciones en las suposiciones.

Sin embargo, cuando las violaciones son grandes, tanto el nivel de significancia (α) como la potencia de la prueba ($1-\beta$) reales pueden diferir grandemente de lo que cabría de esperar en otras circunstancias. En tales casos, una prueba paramétrica seria no válida y debería seleccionarse, en su lugar, un procedimiento sin distribución.

Por otro lado, no es ventajoso utilizar un procedimiento libre de distribución cuando todas las suposiciones de la correspondiente prueba paramétrica pueden lograrse. A menos que se utilice un procedimiento paramétrico en estos casos, no se estaría aprovechando los datos en toda su capacidad.

2.3.4 Prueba t de varianza conjunta para diferencias entre dos medias

Introducción:

Se parte de la suposición que se dispone de dos poblaciones independientes, cada una con una media y una desviación estándar propia (μ_1 , μ_2 y σ_1 , σ_2)

Supóngase también que se toma una muestra aleatoria de tamaño n_1 de la primera población y una muestra de tamaño n_2 de la segunda población.

La estadística de la prueba utilizada para determinar la diferencia entre las medias de las poblaciones está basada en la diferencia entre las medias de las muestras. Debido al teorema del límite central, esta estadística de prueba seguirá la distribución normal estándar para tamaños de muestra lo suficientemente grandes. La estadística de prueba Z es:

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \quad (\text{X. 4})$$

Desarrollo de la prueba t de varianza conjunta:

En la mayoría de los casos no se conoce la desviación estándar real de ninguna de las dos poblaciones. La única información que, por lo general, se puede obtener son las medias de las muestras y las desviaciones estándar de las muestras (S_1 y S_2). Si se hacen suposiciones acerca de que las muestras fueron tomadas de manera aleatoria e independiente, de sus respectivas poblaciones, que están distribuidas normalmente, y, además, que las varianzas de la población son iguales (esto es σ_1^2 y σ_2^2), se puede utilizar una prueba t de varianza conjunta para determinar si existe alguna diferencia significativa entre la media de dos poblaciones.

Para probar la hipótesis nula de no diferencia en las medias de dos poblaciones independientes,

$$H_0 \rightarrow \mu_1 = \mu_2,$$

contra la alternativa de que las medias no son iguales

$$H_1 \rightarrow \mu_1 \neq \mu_2,$$

se puede calcular la siguiente estadística de la prueba t de varianza conjunta:

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{S_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \quad (\text{X. 5})$$

donde la varianza conjunta es definida por:

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{(n_1 - 1) + (n_2 - 1)} \quad (\text{X. 6})$$

La prueba t de varianza conjunta debe su nombre a que la estadística de prueba requiere la unión o combinación de dos varianzas de muestra, para obtener la varianza conjunta, la mejor estimación de la varianza común a las dos poblaciones, bajo la suposición de que las dos varianzas de población son iguales.

La estadística de prueba t de varianza conjunta sigue una distribución t con n_1+n_2-2 grados de libertad. Para un nivel de significancia, α dado, se puede rechazar la hipótesis nula si la estadística de prueba t calculada es mayor que el valor crítico de extremo superior, de la distribución y, o si la estadística de la prueba calculada cae por debajo del valor crítico de extremo inferior de la distribución, es decir que la regla de decisión es:

$$\text{Rechazar } H_0 \Rightarrow \text{si } t > t_{n_1+n_2-2} \text{ o si } t < -t_{n_1+n_2-2}$$

Consideraciones:

En la prueba de diferencias entre las medias, se ha supuesto que se está tomando las muestras de poblaciones distribuidas normalmente que tienen la misma varianza. Se debe examinar las consecuencias que tienen las violaciones a esta suposición sobre la prueba t de varianza conjunta. En situaciones en las que no se pueda (o no se desee) realizar la suposición de que las dos poblaciones con igual varianza realmente están distribuidas de manera normal, la prueba de varianza conjunta es robusta (es decir, no sensible) con respecto a violaciones moderadas de la suposición de normalidad, siempre y cuando el tamaño de las muestras sea grande. En tales situaciones, la prueba t de varianza conjunta puede emplearse sin que se vea seriamente afectada en su potencia.

Por otro lado, si el tamaño de las muestras es pequeño y no se pueda (o no se desee) realizar la suposición de que los datos de cada grupo fueron tomados de una población normalmente distribuida, se tiene dos alternativas. Ya sea que se lleve a cabo, en cada uno de los resultados, alguna transformación normalizante, y después de utilice la prueba t de varianza conjunta, o que se siga un procedimiento libre de distribución, como la prueba de suma de rangos de Wilcoxon, que no depende de la suposición de normalidad de las dos poblaciones.

2.3.5 Prueba t' de varianza separada para diferencia entre medias

En situaciones en la que no se pueda (o no se desee) realizar la suposición de que las dos poblaciones, normalmente distribuidas, de las cuales se tomaron las muestras tienen varianzas iguales, se dice que se tiene un problema de Behrens-Fisher, y se puede utilizar la prueba t' de varianza separada (independientes), desarrollada por Satterhwaite.

Desarrollo

Para probar la hipótesis nula de no diferencia en las medias de dos poblaciones independientes,

$$H_0 \rightarrow \mu_1 = \mu_2 \text{ o } \mu_1 - \mu_2 = 0$$

contra la alternativa de que las medias no son iguales

$$H_1 \rightarrow \mu_1 \neq \mu_2 \text{ o } \mu_1 - \mu_2 \neq 0$$

se puede calcular la siguiente estadística de prueba t' de varianzas independientes:

$$t' = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)}} \quad (\text{X. 7})$$

La estadística de prueba t' de varianzas independientes puede ser aproximada por una distribución t con los grados de libertad, ν , tomados como la parte entera del siguiente cálculo:

$$\nu = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{(n_1 - 1)} + \frac{\left(\frac{S_2^2}{n_2}\right)^2}{(n_2 - 1)}} \quad (\text{X. 8})$$

Para un nivel dado de significancia, α , se puede rechazar la hipótesis nula si la estadística de prueba t' calculada es mayor que el valor crítico de extremo superior, de la distribución t, o si la estadística de la prueba calculada cae por debajo del valor crítico de extremo inferior de la distribución t. Es decir, la regla de decisión es:

$$\text{Rechazar } H_0 \Rightarrow \text{si } t > t_\nu \text{ o si } t < -t_\nu$$

Consideraciones

La principal diferencia entre la prueba t y t' es que la primera supone que las poblaciones de las cuales se toma la muestra está normalmente distribuida y tienen la misma varianza, mientras que la última solamente supone que las poblaciones están distribuidas de manera normal.

Entonces si la hipótesis nula fuera verdadera y no hubiera una diferencia en la media de las dos poblaciones, el uso de la prueba t , nos podría producir un error de tipo I; mientras que el uso de la prueba t' tendría como resultado una decisión correcta. En el otro extremo, si la hipótesis nula fuera en realidad falsa y si hubiera diferencias en las medias de dos poblaciones, el uso de la prueba t tendría como resultado la decisión correcta de rechazar la hipótesis nula; mientras que el empleo de la prueba t' haría que se cometiera un error de tipo II, no existiendo capacidad de detectar una diferencia promedio verdadera en los dos grupos.

Resolver un dilema como éste es parte de un buen análisis de datos. ¿Qué prueba estadística, t o t' , es más verosímil? Para llegar al núcleo de la cuestión, se debe llevar a cabo un análisis exploratorio de datos y evaluar la plausibilidad de las suposiciones necesarias para utilizar las pruebas t y t' . Una forma de detectar la diferencia en las varianzas es por medio de la prueba F . Basándose en los resultados de este prueba, se puede analizar cuál de las pruebas t es más adecuada.

Por otro lado, si el análisis exploratorio de datos revela que la suposición de normalidad de las poblaciones muestrales es cuestionable, esto podría guiar a la conclusión de que ni la prueba t ni la t' son apropiadas. En tal situación se llevaría a cabo una transformación de datos o se emplearía un procedimiento libre de distribución que no tuviera necesidad de hacer estas suposiciones limitantes.

2.3.6 Prueba de estacionalidad de Kendall

Cuando es necesario determinar si una serie temporal presenta una tendencia, se suelen realizar sobre la misma distintos test de tipo paramétrico y no paramétricos. En este caso la prueba Kendall, del tipo no paramétrico, responde a los siguientes lineamientos, en donde i , representa el registro. Para llevarlo a cabo se calcula:

$$S = \sum_{i=1}^{N-1} s(i) \quad (X. 9)$$

en donde $s(i)$ es el número de los valores de $x_j > x_i$, siendo $i < j \leq n$, o sea para cada valor de x_i ordenado cronológicamente, se calcula cuantos valores posteriores de la misma variable lo superan. Para el último valor no se define $s(i)$ por que no existen valores posteriores; por esa razón la sumatoria de los $s(i)$ se realiza hasta los $n-1$ valores.

Así mismo se calcula:

$$T = \sum_{i=1}^{N-1} t(i) \quad (\text{X. 10})$$

En forma análoga a la anterior $t(i)$ es el número de valores $x_j < x_i$, siendo $i < j \leq n$, o sea para cada valor ordenado cronológicamente, se calculan cuantos valores posteriores son menores.

Una vez que se han calculado los valores de S y T, se define el índice $I = S - T$, que debe ser próximo a cero si la hipótesis nula (H_0) es verdadera. Se deben verificar las siguientes condiciones:

1. $I < I_{\text{crit}}(\alpha)$, o sea que el valor de I calculado sea menor que un valor de I crítico tabulado en función del tamaño de la muestra y según diferentes niveles de significación.
2. Para tamaños de muestra mayores a 10, el valor del estadístico se aproxima a una distribución normal. Por lo que se debe utilizar una forma estandarizada del estadístico en donde se transforma el índice I en el índice K, definido como:

$$K = \frac{I - 1}{\left(\frac{n \cdot (n - 1) \cdot (2n + 5)}{18} \right)^{0.5}} \quad I > 0$$

$$K = 0 \quad I = 0 \quad (\text{X. 11})$$

$$K = \frac{I + 1}{\left(\frac{n \cdot (n - 1) \cdot (2n + 5)}{18} \right)^{0.5}} \quad I < 0$$

La hipótesis nula es rechazada para un nivel de significancia α si $|K| > K_{(1-\alpha/2)}$ donde $K_{(1-\alpha/2)}$ es el valor de la distribución normal estándar.

2.3.7 Prueba de suma de rangos de Wilcoxon para diferencia entre dos medianas

Introducción

Si no se puede (o no se desea) realizar la suposición de que los datos de cada grupo son tomados de poblaciones distribuidas normalmente, entonces se tienen dos alternativas.

Se puede utilizar la prueba t o t', la que sea más apropiada después de realizar alguna transformación normalizante sobre los datos, o se puede seguir algún procedimiento libre de distribución. Como puede ser la prueba de suma de rangos de Wilcoxon, un procedimiento libre de distribución poderoso, muy sencillo y ampliamente utilizado, para probar las diferencias entre las medianas de dos poblaciones.

Esta prueba ha resultado ser casi tan poderosa como su contraparte paramétrica (pruebas t y t') en condiciones apropiadas para estas última, y es probable que sea más poderosa cuando las suposiciones limitantes de tales pruebas no se cumplen.

Procedimiento

Para efectuar la prueba de suma de rangos de Wilcoxon se debe sustituir las observaciones de las dos muestras de tamaños n_1 y n_2 , por sus rangos combinados. Los rangos son asignados de tal manera que el rango 1 se asigna a la más pequeña de las $n = n_1 + n_2$ observaciones combinadas, el rango 2 se le asigna a la siguiente más alta y así sucesivamente, de modo que el rango n queda asignado a la observación más grande. Si varias observaciones tienen el mismo valor, se asigna a cada una de estas el promedio de los rangos que, en otra circunstancia, se les habría asignado.

La estadística de la prueba de Wilcoxon T_1 , es simplemente la suma de los rangos asignados a las n_1 observaciones de la muestra más pequeña.

Para cualquier valor entero n , la suma de los n primeros enteros consecutivos pueden calcularse fácilmente como $n(n+1)/2$. La estadística de prueba T_1 , más la suma de los rangos asignados a los n_2 elementos de la segunda muestra, T_2 , por consiguiente, debe ser igual a este valor, es decir;

$$T_1 + T_2 = \frac{n(n+1)}{2} \quad (\text{X. 12})$$

De modo que esta ecuación puede ser empleada como verificación del procedimiento de asignación de rangos.

La prueba de la hipótesis nula puede ser de dos extremos o de un extremo, dependiendo de si se está probando si las dos medianas de población son exclusivamente diferentes o si una es mayor que la otra.

La estadística de prueba, T_1 , está distribuida de manera aproximadamente normal. Se puede emplear la siguiente fórmula de aproximación de muestra grande para probar la hipótesis nula:

$$|W| = \frac{T_1 - \mu_{T_1}}{\sigma_{T_1}} \quad (\text{X. 13})$$

El valor medio de la estadística de prueba T_1 , puede calcularse con:

$$\mu_{T_1} = \frac{n_1(n+1)}{2} \quad (\text{X. 14})$$

y la desviación estándar de la estadística de prueba T_1 , se puede calcular con:

$$\sigma_{T_i} = \sqrt{\frac{n_1 \cdot n_2 (n+1)}{12}} \quad (\text{X. 15})$$

Basándose en α , el nivel de significación seleccionado, la hipótesis nula puede ser rechazada si el valor Z calculado cae en la región de rechazo apropiada.

2.3.8 Prueba F para diferencias entre dos varianzas

Introducción

En las tres secciones anteriores, se examinaron algunos procedimientos para probar las diferencias en la tendencia central (es decir, diferencias en las medias o en las medianas) entre dos poblaciones independientes. En muchas situaciones, sin embargo, puede interesar también probar si dos poblaciones independientes tienen la misma variabilidad.

Se puede estar interesados en estudiar las varianzas de dos poblaciones como un “medio para llegar a un fin”, es decir, probar la suposición de varianzas iguales con el propósito de determinar cual de las pruebas t es la más apropiada, o se puede existir interés real en el estudio de las varianzas de dos poblaciones como un “fin en si mismo”

Desarrollo

Con el fin de probar la igualdad de varianzas de dos poblaciones independientes, se ha diseñado un procedimiento estadístico basado en el cociente de dos varianzas de muestra.

Si se supone que los datos obtenidos de cada población están distribuidos normalmente, entonces S_1^2 y S_2^2 sigue una distribución conocida como F. Los valores críticos de la distribución F dependen de dos conjuntos de grados de libertad de las muestras. La estadística de prueba F para probar la igualdad entre dos varianzas sería:

$$F = \frac{S_1^2}{S_2^2} \quad (\text{X. 16})$$

Al probar la igualdad de dos varianzas, se pueden emplear pruebas de dos extremos o de un extremo, dependiendo de si se está probando si las varianzas de las dos poblaciones son diferentes o si una de ellas es mayor o igual que la otra.

Para un nivel dado de significancia, α , para probar la hipótesis nula de igualdad de varianza

$$H_0 \rightarrow \sigma_1^2 = \sigma_2^2$$

contra la hipótesis alternativa de que las dos varianzas de poblaciones no son iguales

$$H1 \rightarrow \sigma_1^2 \neq \sigma_2^2$$

Se puede rechazar la hipótesis nula si la estadística de prueba calculada F , es mayor que el valor crítico de extremo superior, de la distribución F , o si la estadística de prueba calculada cae por debajo del valor crítico de extremo inferior, la regla de decisión es:

$$\text{Rechazar } H_0 \text{ si } F > F_{U(n_1-1), (n_2-1)} \text{ o si } F < F_{L(n_1-1), (n_2-1)}$$

Precaución:

Al probar la igualdad de dos varianzas de población, se debe tener en cuenta que la prueba supone que cada una de las distribuciones están distribuidas de manera normal. Esto es, si se cumple con la suposición de normalidad para cada población, la estadística de prueba F sigue una distribución F con n_1-1 y n_2-1 grados de libertad.

Desafortunadamente, esta prueba no es robusta con respecto a violaciones a esta suposición, en particular cuando los tamaños de las muestras no son iguales. Por tanto, si las poblaciones no están, por lo menos, distribuidas de manera aproximadamente normal, la precisión del procedimiento puede verse afectada.

2.3.9 Prueba de corridas de una muestra de Wald-Wolfwitz

Introducción

Por lo general se supone que los datos recolectados en un estudio constituyen una muestra aleatoria, de modo que cada observación o medida es tomada de la población de manera aleatoria e independiente. Tal suposición, sin embargo, puede ser probada mediante el empleo de un procedimiento no paramétrico conocido como prueba de corridas de una muestra de Wald-Wolfwitz para aleatoriedad. Este procedimiento no paramétrico no tiene que ver con la prueba de cualquier parámetro en particular y, por tanto, no tiene una contraparte paramétrica.

Para probar la aleatoriedad, la hipótesis nula sería:

H_0 : El proceso que genera el conjunto de datos numéricos es aleatorio.

y la hipótesis alternativa sería:

H_1 : El proceso que genera el conjunto de datos numéricos no es aleatorio.

La hipótesis nula de aleatoriedad puede probarse mediante la observación del orden o de la secuencia en que se obtienen los elementos de la muestra. Si a cada elemento se le asigna uno de dos términos, como E y F (por éxito y fracaso), dependiendo de si la medida cae por arriba o por debajo de un cierto valor, la aleatoriedad de la secuencia puede ser investigada.

Si ésta es generada de manera aleatoria, el valor (E o F) de un elemento será independiente tanto de su posición en la secuencia como del valor de los elementos que le preceden. Por otra parte, si el valor de un elemento de la secuencia es afectado por los valores de los demás elementos, o si la probabilidad de su ocurrencia depende de su posición en la secuencia, el proceso que la genera no es considerado aleatorio.

En los casos no aleatorios los elementos parecidos tenderán a agruparse (del mismo modo que cuando existe una tendencia presente en los datos) o se mezclarán de manera alternada, de modo que se presentaría algún efecto periódico sistemático.

Para estudiar si una secuencia observada es aleatoria o no, se considera como estadístico de prueba al número de corridas presente en los datos. Una *corrida* se define como una serie consecutiva de elementos similares que están limitados por elementos de un tipo diferente o por el inicio o fin de la secuencia.

Desarrollo:

Para probar la hipótesis nula de aleatoriedad, podemos dividir el tamaño completo de la muestra, n , en dos partes, n_1 , el número de éxitos, n_2 , el número de fracasos. La estadística de prueba, representada por la letra U , para una prueba de dos extremos, si U es mayor o menor de lo que cabría esperar en una serie aleatoria de datos, se rechazaría la hipótesis nula de aleatoriedad a favor de la hipótesis alternativa. Si para una combinación dada de n_1 y n_2 , U es mayor o igual al valor crítico superior, o menor o igual al valor crítico inferior, la hipótesis nula de aleatoriedad puede ser rechazada al nivel de significación α . Sin embargo, si U se encuentra entre estos límites, la hipótesis nula de aleatoriedad puede rechazarse.

Por otra parte, las pruebas de aleatoriedad no siempre son de dos extremos, si se está interesado en probar la aleatoriedad contra una alternativa específica de un efecto de tendencia (de que hay una tendencia de agrupamiento de los elementos parecidos), se necesita una prueba de un extremo. En este caso, se rechaza la hipótesis nula solamente si se presenta muy pocas corridas (u). En el otro extremo, si se está interesado en probar la aleatoriedad contra un efecto sistemático o periódico, se utiliza una prueba de un extremo que rechaza la hipótesis nula sólo si se presentan demasiadas corridas.

Independientemente si la prueba es de un extremo o de dos extremos, sin embargo, para una muestra de tamaño n , mayor a 40, el estadístico U está distribuido de manera aproximadamente normal. Por consiguiente la fórmula de aproximación se muestra a continuación:

$$U = \frac{u - \mu_u}{\sigma_u} \quad (\text{X. 17})$$

donde u es el número total de corridas observadas, el valor medio de u es dado por:

$$\mu_u = \frac{2n_1n_2}{n} + 1 \quad (\text{X. 18})$$

y la desviación estándar de U es dada por:

$$\sigma_u = \sqrt{\frac{2n_1n_2(2n_1n_2 - n)}{n^2(n-1)}} \quad (\text{X. 19})$$

2.3.10 Prueba de la bondad de ajuste χ^2

La bondad de ajuste de una distribución de probabilidad puede probarse comparando los valores teóricos de una función de distribución y los muestrales de las funciones de frecuencia relativa o de frecuencia acumulada.

Aspectos matemáticos:

El valor muestral de la frecuencia relativa del intervalo i es, dado por la expresión:

$$f_s = \frac{n_i}{n} \quad (\text{X. 20})$$

Mientras que el valor teórico es $p(x_i) = F(x_i) - F(x_{i-1})$. La prueba estadística χ^2 , χ_c^2 , está dada por:

$$\chi_c^2 = \sum_{i=1}^m \frac{n[f_s(x_i) - p(x_i)]^2}{p(x_i)} \quad (\text{X. 21})$$

donde m es el número de intervalos, debe notarse que $nf_s(x_i) = n_i$, el número de ocurrencias observadas en el intervalo i, y $np(x_i)$ es el correspondiente número esperados de ocurrencias en el intervalo i; luego el cálculo se limita a elevar al cuadrado la diferencia entre el número de ocurrencias observadas y esperadas, dividiendo por el número de ocurrencias esperadas en el intervalo y sumando el resultado para todos los intervalos.

Para describir la prueba χ^2 , debe definirse la distribución de probabilidad χ^2 . Una distribución χ^2 con ν grados de libertad es la distribución para la suma de los cuadrados de ν variables aleatorias normales estándar independientes z_i ; esta suma es la variable aleatoria:

$$\chi_\nu^2 = \sum_{i=1}^{\nu} z_i^2 \quad (\text{X. 22})$$

La hipótesis nula será de la forma $H_0 \rightarrow F = F_0$ donde F_0 es la función de distribución teórica adoptada, en el caso contrario se rechaza la hipótesis.

2.3.11 Análisis de frecuencia

Posición de Ploteo

Una posición de ploteo se refiere al valor de probabilidad asignada a cada uno de los datos a graficarse. Se han propuesto numerosos métodos para la determinación de las posiciones de graficación, la mayoría de los cuales son empíricos. Si n es el número total de valores que van a ser graficados y m es la posición de un valor en una lista ordenada por magnitud descendente, la probabilidad de excedencia del m -ésimo valor mayor a , x_m , es, para un n grande,

$$P(X \geq x_m) = \frac{m}{n} \quad (\text{X. 23})$$

Sin embargo, esta fórmula simple produce una probabilidad del ciento por ciento para $m = n$, que puede ser difícil graficar en una escala de probabilidad. Como un ajuste, la anterior ecuación puede modificarse a

$$P(X \geq x_m) = \frac{m-1}{n} \quad (\text{X. 24})$$

Aún cuando esta ecuación no produce una probabilidad del ciento por ciento, si produce una probabilidad de cero (para $m=1$), lo cual también puede ser difícil de graficar en un papel de probabilidad.

Las dos ecuaciones anteriores representan los límites dentro de los cuales deberían localizarse las posiciones de graficación apropiadas. La mayor parte de las fórmulas de posición de graficación están representadas en la siguiente forma:

$$P(X \geq x_m) = \frac{m-b}{n+1-2b} \quad (\text{X. 25})$$

donde b es un parámetro. Por ejemplo, $b=0.5$ para la fórmula de Hazen, $b=0.3$ para la fórmula de Chegodayev, y $b=0$ para Weibull. También, para algunos otros ejemplos $b=3/8$ para la fórmula de Blom, $1/3$ para Tukey y 0.44 para la de Gringorten.

Cunnane estudió los diferentes métodos disponibles para las posiciones de graficación utilizando criterios de varianza mínima y no sesgo. Un método de graficación no sesgado es aquel que, si se utiliza para la graficación de un número grande muestras de igual tamaño, resultará en que el promedio de los puntos graficados para cada valor m cae en la línea de distribución teórica. Un método de graficación con varianza mínima es aquel que minimiza la varianza de los puntos graficados alrededor de la línea teórica. Cunnane concluyó que la ecuación de graficación de Weibull es sesgada y grafica los valores máximos de una muestra con periodos de retorno demasiado pequeños. Para datos normalmente distribuidos la posición de graficación de Blom está bastante cercana a ser no sesgada, mientras que para

datos distribuidos de acuerdo con la distribución de Valor Extremo Tipo I, la fórmula de Gringorten es la mejor. Para la distribución log-Pearson Tipo III el valor óptimo de b depende del valor del coeficiente de asimetría, siendo mayor que $3/8$ cuando los datos tienen asimetría positiva y menor que $3/8$ cuando los datos tienen asimetría negativa. Las mismas posiciones de graficación pueden aplicarse a los logaritmos de los datos, cuando se utiliza la distribución lognormal, por ejemplo.

De todos modos, cabe resaltar que todas las posiciones son, en cierto modo, imprecisas, fundamentalmente en los extremos de las distribuciones, es decir, en correspondencia con tiempos de retorno muy altos.

2.3.12 Definición de las Series de Tiempo

En muchas áreas del conocimiento las observaciones de interés son obtenidas en instantes sucesivos del tiempo, por ejemplo, a cada hora, durante 24 horas, mensuales, trimestrales, semestrales o bien registradas por algún equipo en forma continua. Se denomina Serie de Tiempo a un conjunto de mediciones de cierto fenómeno o experimento registradas secuencialmente en el tiempo. Estas observaciones serán denotadas por $\{x(t_1), x(t_2), \dots, x(t_n)\} = \{x(t) : t \in T \subseteq R\}$ con $x(t_i)$ el valor de la variable x en el instante t_i .

Si $T = Z$ se dice que la serie de tiempo es discreta y si $T = R$ se dice que la serie de tiempo es continua. Cuando $t_{i+1} - t_i = k$ para todo $i = 1, \dots, n-1$, se dice que la serie es *equiespaciada*, en caso contrario será *no equiespaciada*.

Por lo general, en Hidrología se trabaja con series de tiempo discreta, equiespaciadas en cuyo caso se asume sin pérdida de generalidad que: $\{x(t_1), x(t_2), \dots, x(t_n)\} = \{x(1), x(2), \dots, x(n)\}$.

2.3.13 Modelos clásicos de series de tiempo

Un modelo clásico para una serie de tiempo, supone que una serie $x(1), \dots, x(n)$ puede ser expresada como suma o producto de cuatro componentes: *tendencia*, *estacionalidad*, *ciclicidad* y un término de *error aleatorio*.

Existen por lo menos tres modelos de series de temporales, que normalmente se aceptan como buenas aproximaciones de las relaciones existentes entre los componentes de los datos observados, ellos son:

1. *Aditivo*: $Y(t) = T(t) + E(t) + C(t) + A(t)$
2. *Multiplcativo*: $Y(t) = T(t) \cdot E(t) \cdot C(t) \cdot A(t)$
3. *Mixto*: $Y(t) = T(t) \cdot E(t) \cdot C(t) + A(t)$

en donde:

$Y(t)$: serie observada en instante t

$T(t)$: componente de tendencia

$E(t)$: componente estacional

$C(t)$: componente cíclico

$A(t)$: componente aleatoria (accidental)

Una suposición usual es que $A(t)$ sea una componente aleatoria o *ruido blanco* con media cero y varianza constante.

Un modelo aditivo (1), es adecuado, por ejemplo, cuando $E(t)$ no depende de otras componentes, como $T(t)$; si por el contrario la estacionalidad varía con la tendencia, el modelo más adecuado es un modelo multiplicativo (2). Es claro que el modelo 2 puede ser transformado en aditivo, tomando logaritmos. El problema que se presenta es modelar adecuadamente las componentes de la serie.

La Figura X. 1 muestra los posibles patrones que podrían seguir series representadas por los modelos (1), (2) y (3).

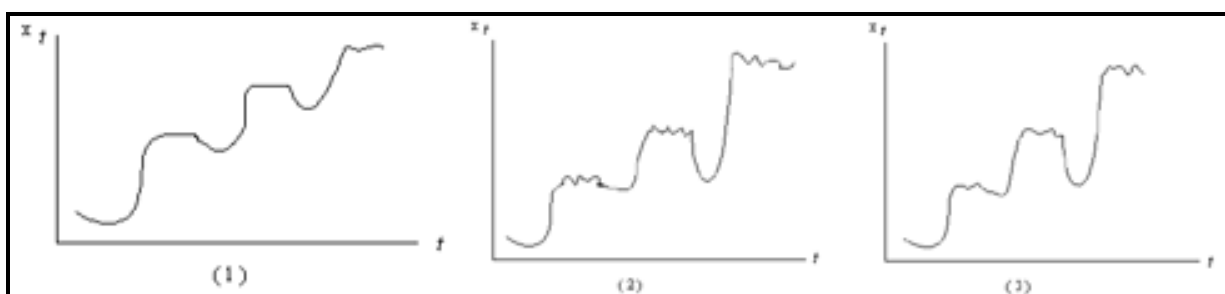


Figura X. 1. Modelos de Descomposición de las Series Temporales (Fuente: Catalini, 2004).

Conceptos Básicos de Series de Tiempo

La planificación racional exige prever los sucesos del futuro que probablemente vayan a ocurrir. La previsión, a su vez, se suele basar en lo que ha ocurrido en el pasado, se tiene entonces un nuevo tipo de inferencia estadística que se hace acerca del futuro de alguna variable o compuesto de variables basándose en sucesos pasados. La técnica más importante para hacer inferencias sobre el futuro con base en lo ocurrido en el pasado, es el *análisis de series temporales*.

Uno de los problemas que intentan resolver las series de tiempo es el de la predicción ó pronóstico. Esto es, que dada una serie $\{x(t1), \dots, x(tn)\}$ los objetivos de interés se asocian a la descripción del comportamiento de la serie, investigar el mecanismo generador de la serie temporal y buscar posibles patrones temporales que permitan sobrepasar la incertidumbre del futuro.

En adelante se estudiará como construir un modelo para explicar la estructura y prever la evolución de una variable que se observa a lo largo del tiempo. Para validar este modelo es necesario contrastar la hipótesis de *homocedasticidad*, esto es, la varianza de los residuos es constante y no varía en los diferentes niveles del factor. La falta de homocedasticidad se denomina *heterocedasticidad*.

Prueba de Datos Dudosos - Outliers

Chow et al. (1995) establecen que los datos dudosos (outliers) son puntos de la información que se alejan significativamente de la tendencia que posee la serie de información. La retención o eliminación de estos datos puede afectar significativamente la magnitud de los parámetros estadísticos calculados para la serie temporal, especialmente en muestras pequeñas.

Los procedimientos para tratar los datos dudosos requieren un criterio que involucra consideraciones matemáticas e hidrológicas. De acuerdo al Water Resources Council (1981), si la asimetría de la estación se encuentra comprendida entre -0,4 y +0,4, se considera necesario la aplicación de pruebas para detectar los datos dudosos antes de eliminar cualquier otro valor del conjunto de datos.

La siguiente ecuación de frecuencia puede utilizarse para detectar los datos dudosos altos:

$$y_H = \bar{y} + K_n \cdot s_y \quad (\text{X. 26})$$

donde y_H es el umbral de dato dudoso alto en unidades logarítmicas y K_n , es tal como se presenta en la Tabla X. 1 para un tamaño de muestra n . Los valores de K_n dados se usan en pruebas de un lado, para detectar datos dudosos altos en el nivel de significancia del 10% de la información normalmente distribuida. Una ecuación similar puede utilizarse para detectar los datos dudosos bajos:

$$y_L = \bar{y} - K_n \cdot s_y \quad (\text{X. 27})$$

donde y_L es el umbral de datos dudosos altos bajos en unidades logarítmicas s_y es la desviación estándar de la muestra.

Tabla X. 1. Valores de K_n de un lado con un nivel de significancia del 10% para la distribución normal

<i>Tamaño de la Muestra</i>	<i>K_n</i>	<i>Tamaño de la Muestra</i>	<i>K_n</i>	<i>Tamaño de la Muestra</i>	<i>K_n</i>	<i>Tamaño de la Muestra</i>	<i>K_n</i>
10	2.036	24	2.467	38	2.661	60	2.837
11	2.088	25	2.486	39	2.671	65	2.866
12	2.134	26	2.502	40	2.682	70	2.893
13	2.175	27	2.519	41	2.692	75	2.917
14	2.213	28	2.534	42	2.700	80	2.940
15	2.247	29	2.549	43	2.710	85	2.961
16	2.279	30	2.563	44	2.719	90	2.981
17	2.309	31	2.577	45	2.727	95	3.000
18	2.335	32	2.591	46	2.736	100	3.017
19	2.361	33	2.604	47	2.744	110	3.049
20	2.385	34	2.616	48	2.753	120	3.078
21	2.408	35	2.628	49	2.760	130	3.104
22	2.429	36	2.639	50	2.768	140	3.129
23	2.448	37	2.650	55	2.804		

2.3.14 Descomposición de la estructura de una serie temporal

Las series hidrológicas pueden exhibir, en distintos grados, tendencias, saltos, estacionalidad, autocorrelación y falta de normalidad. Estos atributos son conocidos en las series de tiempo hidrológicas como *componentes*. Por lo que, para llevar un análisis estadístico sobre ella, ésta puede ser *particionada* o *descompuesta* en sus componentes.

Tendencias y Saltos

En general, factores naturales y humanos pueden producir en forma gradual o instantánea la aparición de tendencias o saltos en la serie hidrológica. Por ejemplo, el incendio de un gran bosque en la cuenca de un río puede afectar inmediatamente a la escorrentía, generando un salto en la serie de caudales, mientras que si la remoción del bosque es gradual entonces puede resultar en un cambio a lo largo del tiempo en la tendencia de las medias sobre la serie de caudales.

Algunas de las principales fuentes que pueden originar la aparición de tendencias o saltos en las series de caudales, son el cambio del uso del suelo en la cuenca, la implantación de una presa o la construcción de alguna estructura en el curso que afecte el normal drenaje de la cuenca, al igual que el efecto producido por el calentamiento global y el cambio climático.

Remoción de tendencias

Una serie hidrológica puede presentar saltos en uno o más de sus parámetros estadísticos. La más común de estas es la tendencia en las medias y en la varianza.

La descomposición de la serie de tiempo es mostrada en forma esquemática en la Figura X. 2. Una tendencia lineal es mostrada en la Figura X. 2 (a). La tendencia \bar{y}_t puede ser removida mediante la diferencia $y_t - \bar{y}_t$ como se muestra en la Figura X. 2 (b). La diferencia en varianza de una serie puede ser expresada como, s_t^2 , la cual puede ser función del tiempo (en ese caso existe tendencia en la varianza) o puede ser constante, como se grafica en la Figura X. 2 (c). La tendencia en la varianza puede ser removida por medio de $(y_t - \bar{y}_t) / \sigma_t$ (el proceso de construcción de una nueva serie mediante la remoción de la media y el desvío estándar es conocido como *estandarización*). La serie residual, Figura X. 2 (d) puede tener aún otras propiedades como correlación de la estructura, la cual debe ser descompuesta y removida.

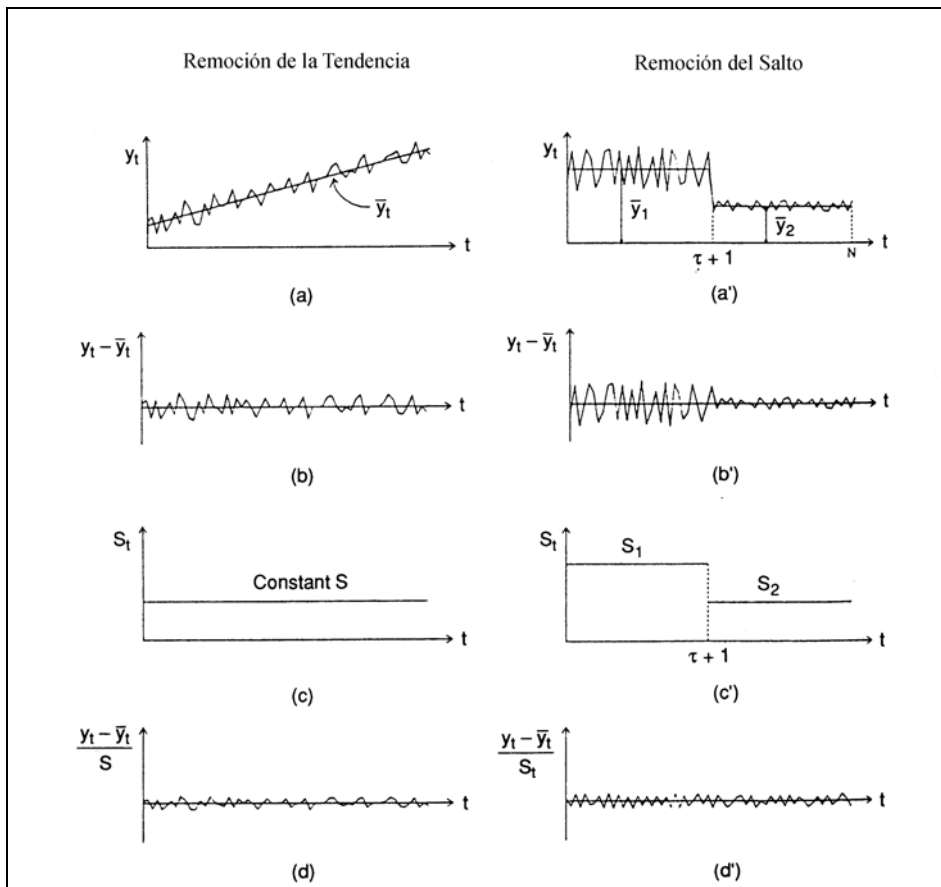


Figura X. 2. Remoción de tendencias y saltos en una serie de tiempo anual. (Fuente Maidment, 1992).

Remoción de Saltos

Las series hidrológicas, también pueden presentar en uno o más de sus parámetros estadísticos saltos positivos o negativos en la media y la varianza, las cuales son las más comúnmente analizadas. Las Figura X. 2 a' y d' son ejemplos de la remoción de estos saltos de una serie anual.

En la Figura X. 2 (a') el salto en la media y la varianza ocurre para el tiempo $\tau + 1$. Este salto en la media es removido por $y_t - \bar{y}_t$ como se muestra en la Figura X. 2 (b'), y el salto en la varianza es removido por medio de $(y_t - \bar{y}_t) / \sigma_t$. La grafica residual de la serie se presenta en la Figura X. 2 (d'). Ahora la serie cuenta con media cero y varianza uno, aún así la serie puede presentar otras propiedades como ser la autocorrelación.

3 EMPLEO DE INFORMACIONES HISTORICAS EN EL ANÁLISIS DE FRECUENCIA DE CAUDALES MAXIMOS

3.1 Introducción

Este apartado describe la metodología adoptada para estimar el tiempo de retorno de la crecida de abril/mayo de 2003 del río Salado, en las proximidades de la ciudad de Santa Fe. La misma hace uso de la información histórica levantada a través de registros y relatos de crecidas históricas, según fuera descripto en el Anexo IX. El material descripto a continuación se basa en un importante conjunto de publicaciones científicas internacionales y sigue de cerca los lineamientos teóricos y prácticos presentados por Martins y Stedinger (2001) y Reis y Stedinger (2005).

El abordaje adoptado se basa en el empleo de series anuales dado que estudios recientes mostraron que el uso de la información histórica en las mismas posee un desempeño próximo de aquel obtenido con el uso de series parciales.

El abordaje anual adoptado se basa en el uso de la distribución de valor extremo generalizada (GEV) de tres parámetros. Cabe indicar que, como caso particular de este modelo, se obtiene la distribución Gumbel de dos parámetros (cuando el parámetro de forma de la GEV es cero).

Conforme se desprende de la literatura científica actualizada y de los análisis realizados en este estudio, el valor de la información histórica puede ser muy destacado. Ello depende, en parte, de la magnitud del nivel de percepción de crecidas considerado y el período de registro de la información histórica.

Otros modelos también fueron ensayados, pero no adoptados. Ello fue como consecuencia de que la distribución GEV resulta ser muy flexible y recomendada como distribución de máximos. Entre los modelos utilizados cabe destacar a las distribuciones Lognormal y Log-Pearson III, siendo esta última adoptada en el marco de un abordaje del tipo *bayesiano*.

3.2 El uso de la distribución GEV en hidrología

Muchos problemas en recursos hídricos requieren la identificación de una relación de frecuencia. Estas relaciones son frecuentemente definidas por el ajuste de una distribución probabilística a un conjunto de datos, lo cual provee una representación compacta y suavizada de la distribución de frecuencia empírica.

Una vez que la distribución es ajustada, se pueden obtener estimadores de cuantiles correspondientes a distintas frecuencias. Aún así, cabe destacar que la estimación de cuantiles está sujeta a importantes fuentes de incertidumbre, incluyendo errores sistemáticos de las mediciones, la incertidumbre del modelo y la incertidumbre paramétrica.

Además de ello, también es importante destacar que la longitud de los registros de una estación restringe la precisión de los estimadores deducidos con base en la serie de la misma (o también denominados estimadores “locales”). Así, es importante emplear métodos eficientes para estimar cuantiles y parámetros, así como también emplear informaciones adicionales, como regional y/o histórica, cuando las mismas se encuentran disponibles (Jin y Stedinger, 1989; Martins y Stedinger, 2001). En este apartado se muestra cómo incorporar informaciones adicionales empleando datos de crecidas para describir el procedimiento natural de naturaleza aleatoria de generación de caudales.

El modelo empleado aquí para describir la distribución de crecidas es la distribución de probabilidades generalizada de extremos (GEV), introducida por Jenkinson (1955). Esta distribución ha sido empleada en varias aplicaciones en el campo de la hidrología. La misma fue recomendada para el análisis de frecuencia de crecidas a nivel local (*U.K. Flood Studies Report - Natural Environment Research Council, 1975*), para precipitaciones en los Estados Unidos (Willeke *et al.*, 1995), y para ondas del mar (Haan y de Ronde, 1998). Varios estudios han utilizado este modelo con base en informaciones locales (Madsen *et al.*, 1997a). En contextos regionales, la distribución GEV recibió atención especial desde el momento de su introducción en el procedimiento “*index-flood*” basado en momentos de pesos probabilísticos (PWM) de Wallis (1980), Greis e Wood (1981), y Hosking *et al.* (1985b).

Muchos estudios de frecuencia regionales han utilizado la distribución GEV (Hosking *et al.*, 1985b; Wallis y Wood, 1985; Lettenmaier *et al.*, 1987; Hosking y Wallis, 1988; Chowdhury *et al.*, 1991; Stedinger y Lu, 1995; Madsen *et al.*, 1997b; Hosking y Wallis, 1997; Muller y Fill, 1997; Fill y Stedinger, 1998). En la práctica la misma ha sido empleada para modelar una amplia variedad de extremos naturales, incluyendo crecidas, precipitación, velocidad del viento, altura de onda y otros máximos. El origen físico de estos máximos sugiere que sus distribuciones pueden ser de valores extremos (EV) englobados por la distribución GEV (EV tipos I, II, III).

Matemáticamente, la distribución GEV es muy atractiva, porque su inversa puede ser expresada analíticamente, y sus parámetros pueden ser fácilmente estimados por los métodos de los momentos y momentos-L (Hosking *et al.*, 1985a; Hosking, 1990).

3.3 La distribución GEV

La distribución de valores extremos generalizados (GEV) incorpora las distribuciones Tipo I de Gumbel ($\kappa = 0$), Tipo II de Frechet ($\kappa < 0$) y Weibull o Tipo III ($\kappa > 0$). La Figura X. 3 ilustra acerca de los tres tipos de distribuciones englobadas en la GEV.

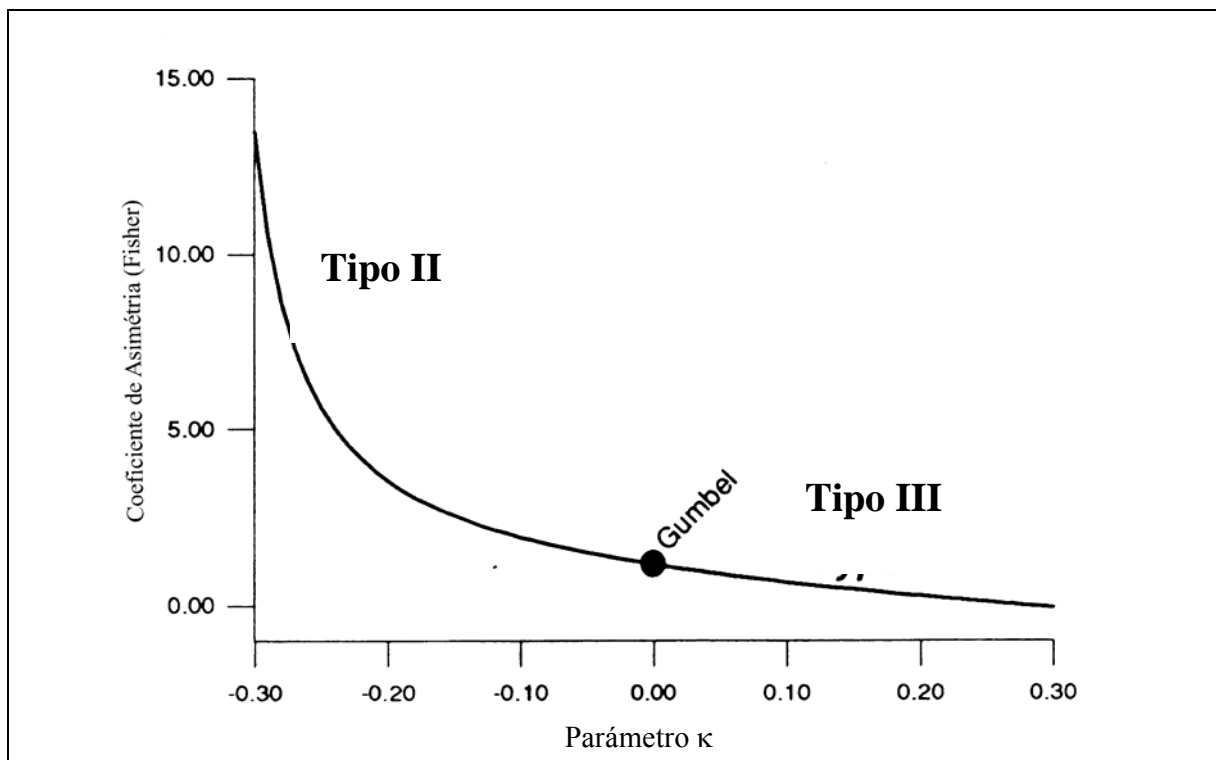


Figura X. 3. Relación entre el parámetro de forma κ y el coeficiente de asimetría de la distribución GEV (fuente: Meylan y Musy, 1996).

La distribución GEV posee función de probabilidad acumulada dada por la expresión:

$$F(x) = \exp\left(-\left(1 - \kappa \cdot \frac{(x-\xi)}{\alpha}\right)^{1/\kappa}\right) \quad \text{para } \kappa \neq 0 \quad (\text{X. 28})$$

$$= \exp\left(-\exp\left(-\frac{(x-\xi)}{\alpha}\right)\right) \quad \text{para } \kappa = 0 \quad (\text{X. 29})$$

donde $\xi + \alpha/\kappa \leq x < \infty$ para $\kappa < 0$, $-\infty < x < +\infty$ para $\kappa = 0$, e $-\infty < x \leq \xi + \alpha/\kappa$ para $\kappa > 0$. Aquí ξ , α y κ son los parámetros de posición, escala y forma, respectivamente. La expresión de los cuantiles de la distribución GEV, dados en términos de los parámetros y de la probabilidad acumulada p , resulta:

$$x_p = \xi + \frac{\alpha}{\kappa} \cdot \left[1 - [-\ln(p)]^\kappa\right] \quad \text{para } \kappa \neq 0 \quad (\text{X. 30})$$

$$= \xi - \alpha \ln(-\ln(p)) \quad \text{para } \kappa = 0 \quad (\text{X. 31})$$

3.4 Estimación de los Parámetros de la Distribución GEV

Hosking *et al.* (1985a) mostraron que los estimadores basados en el método de los momentos ponderados probabilísticos (PWM), o su equivalente, el método de los momentos-L (ML), para la distribución GEV son mejores que los estimadores de máxima verosimilitud (MLE) en términos de sesgo (*bias*) y variancia para tamaños muestrales variando de 15 a 100. Mas recientemente, Madsen *et al.* (1997a) mostraron que los estimadores de cuantiles basados en el método de los momentos (MOM) poseen menor raíz del error cuadrático medio (REQM) para $-0,25 < \kappa < 0,30$ en relación a aquellos basados en los métodos ML y MLE cuando se estima el evento de 100 años de tiempo de retorno para tamaños muestrales (*n*) entre 10 y 50 años. Estimadores de máxima verosimilitud (MLE, por sus siglas en inglés *Maximum Likelihood Estimation*) son preferibles solamente cuando $\kappa > 0,3$ y $n \geq 50$.

Martins y Stedinger (2000) explicaron el porqué de los estimadores MLE se comportan tan pobremente cuando se los compara a los otros métodos. El comportamiento citado de los estimadores MLE en muestras pequeñas puede ser explicada por el hecho que la incertidumbre en κ puede ser muy grande, y puede resultar en valores absurdos del parámetro de forma κ da la distribución GEV. Esto explica el comportamiento inestable de los estimadores de cuantiles basados en MLE en el caso de pequeñas muestras (Hosking *et al.*, 1985a). En contraposición a la estimación MLE, los estimadores MOM restringen $\hat{\kappa}$ a valores mayores que $-1/3$, en cuanto los estimadores ML restringen $\hat{\kappa}$ a valores mayores que -1 (Coles y Dixon, 1998). Ambos estimadores MOM y ML incorporan implícitamente alguna restricción en los valores de κ en el procedimiento de estimación, en cuanto el estimador MLE no realiza ninguna restricción.

3.5 El Método de Máxima Verosimilitud (MLE) en Muestras Pequeñas

El método de máxima verosimilitud está basado en la Teoría de Grandes Muestras. Hosking *et al.* (1985a) reconocieron que el desempeño de sus estimadores (GEV-MLE) en muestras pequeñas y moderadas no fue muy bien estudiado. Resultados de Hosking *et al.* (1985a) indican que las estimaciones de cuantiles obtenidos por Momentos-L (ML) son tendenciosas, aunque preferibles a los correspondientes estimadores de máxima verosimilitud (MLE). La razón es que estos últimos son más variables. El análisis detallado del comportamiento de los estimadores MLE en muestras pequeñas demuestra que, efectivamente, pueden ser generados valores absurdos del parámetro de forma κ de la distribución GEV. Para ilustrar el problema de la estimación de κ en muestras pequeñas, considérese la muestra en la Tabla X. 2, la cual fue generada en forma aleatoria durante el desarrollo de este estudio.

Tabla X. 2. Muestra generada de tamaño $n = 15$ de una GEV($\xi=0$, $\alpha=1$, $\kappa = -0.20$) (Fuente: Martins, 2005).

-0,4154	-0,3999	-0,3973	-0,3397	-0,3094
-0,0997	0,1546	0,4539	1,0572	1,4462
1,4480	2,3660	2,4663	2,8629	4,6820

Los estimadores de Máxima Verosimilitud (MV) para los parámetros de la distribución GEV son:

$$\hat{\xi} = -0,26 \quad \hat{\alpha} = 0,33 \quad \text{y} \quad \hat{\kappa} = -2,00$$

El verdadero valor para el cuantil 0,999 ($\xi=0$, $\alpha=1$, $\kappa=-0,20$) es 14,9, en cuanto el valor estimado ($\hat{\xi}=-0,26$, $\hat{\alpha}=0,33$ e $\hat{\kappa}=-2,00$) para el mismo cuantil es del orden de 1×10^5 . La Figura X. 4 compara la verdadera distribución y la estimada para esta pequeña muestra.

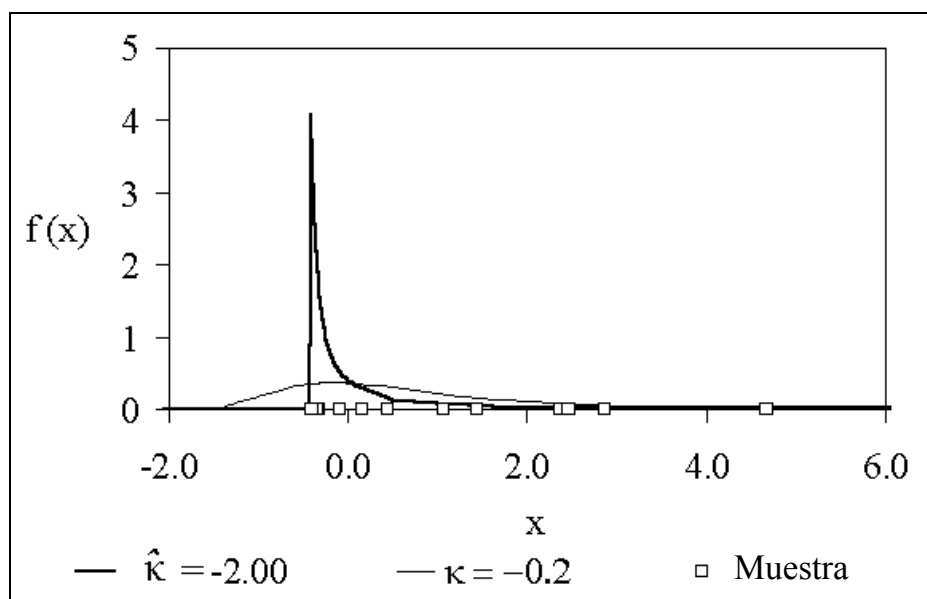


Figura X. 4. Distribución verdadera ($\kappa = -0,20$) y “estimada” ($\hat{\kappa} = -2,00$) para una pequeña muestra de tamaño $n = 15$ (valores muestrales indicados por cuadrados). (Fuente: Martins, 2005).

Martins y Stedinger (2000) analizaron la incertidumbre paramétrica de la distribución GEV causada por los errores muestrales en el parámetro de forma. Una estructura *bayesiana* fue empleada para incorporar la incertidumbre de κ en el análisis de MLE. El parámetro de forma κ fue restringido a una amplitud de variación estadística y físicamente razonable para series de crecidas anuales. En el trabajo citado esa amplitud para κ es descripta como una información *a priori* (*priori* informativo), lo cual asocia pesos a diferentes valores de κ dentro de la amplitud permitida. Los autores demuestran que tal información *a priori*, denominada *priori* Geofísico, combinada con datos muestrales locales, resulta en estimadores de cuantiles, sobre la amplitud de valores κ de interés, más precisos que aquellos derivados de los estimadores MOM e ML con muestras de tamaño pequeño y moderado. El nuevo estimador fue denominado estimador Generalizado de Máxima Verosimilitud (GMLE), y así como los estimadores de Máxima Verosimilitud clásicos, el mismo puede incorporar fácilmente la información histórica.

3.6 Estimadores de Máxima Verosimilitud Generalizados (GMLE)

Representación de la incertidumbre en κ : una distribución a priori para κ

Supóngase que el parámetro verdadero κ de la distribución GEV, con función de densidad de probabilidad $f(x|\xi, \alpha, \kappa)$, es una variable aleatoria variando en el intervalo $[\kappa_L, \kappa_U]$ y con función de densidad a priori (*priori*): $(\pi(\kappa))$.

La experiencia hidrológica indica que el intervalo $-0,30 \leq \kappa \leq 0$ es el más probable para κ . Muchos estudios muestran que las crecidas parecen poseer colas de distribución más importantes (o más pesadas; $\kappa < 0$) que aquellas estimadas por una distribución Gumbel ($\kappa = 0$) (Farquharson et al., 1992; Madsen y Rosbjerg, 1997). Adicionalmente, el examen de las funciones de densidad de probabilidad (fdps) GEV revela que formas razonables para las mismas son obtenidas posiblemente para κ entre $-0,3$ y $+0,3$.

La Figura X. 5 muestra las fdps de la distribución GEV para diferentes valores de κ . En dicha Figura, con $\kappa = -0,1$ y $-0,3$, ambas fdps son asimétricas positivamente con una larga cola superior, como se puede esperar para picos de crecida y fenómenos similares estrictamente positivos como, por ejemplo, la precipitación máxima anual. Adicionalmente, la distribución GEV solamente tiene variancia finita cuando $\kappa > -0,5$, y asimetría finita cuando $\kappa > -1/3$.

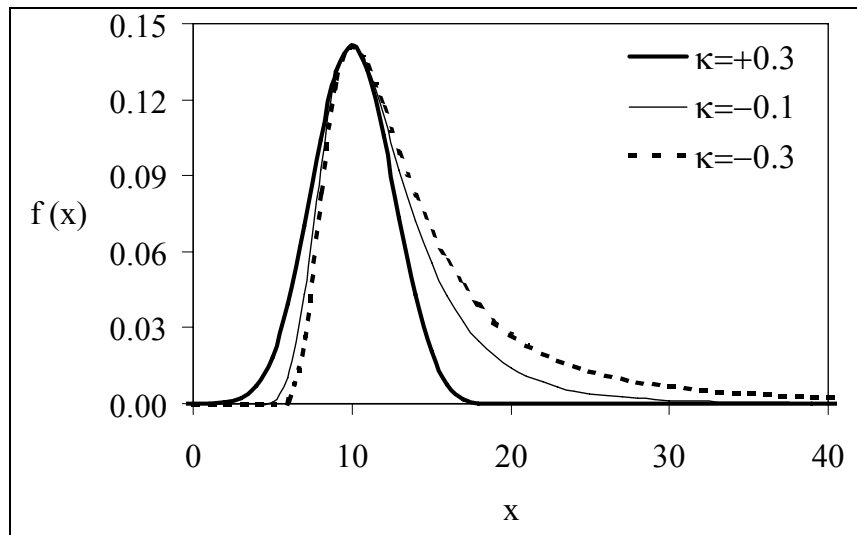


Figura X. 5. Funciones de densidad de probabilidad (fdps) de la distribución GEV para $\kappa = -0,3, -0,1$ e $+0,3$ (donde $\xi = 10$ e $\alpha = 2,6$) (Fuente: Martins, 2005).

Para $\kappa < 0$, la distribución de valor extremo Tipo II posee una cola de tipo Pareto, tal que: $F(x) = 1 - [1 + |\kappa/\alpha \cdot (\xi - x)]^{-1/|\kappa|}$ a medida que $x \rightarrow \infty$, y para $\kappa > 0$, la distribución de extremo Tipo III presenta un contacto de tipo polinomial tal que:

$F(x) = 1 - [(\kappa/\alpha) \cdot (\xi + \alpha/\kappa - x)]^{1/\kappa}$ a medida que $x \rightarrow \xi + \alpha/\kappa$. De un modo general, una

distribución *a priori* para el factor de forma κ , en el caso de crecidas máximas anuales parecería como la fdp de la Figura X. 6.

La información *a priori* (o *priori*) empleado aquí es una distribución Beta, $\pi(\kappa) = (0.5+\kappa)^{p-1}(0.5-\kappa)^{q-1}/B(p,q)$, entre $[-0,5; +0,5]$, con $p = 6$, y $q = 9$, donde:

$$B(p,q) = \Gamma(p)\Gamma(q)/\Gamma(p+q) \quad (X.3)$$

Esta distribución posee $E[\kappa] = -0,10$ y $\text{Var}[\kappa] = (0,122)^2$. La Figura X. 6 muestra el *priori* para κ , denominado aquí como *priori Geofísico*. Este *priori* es relativamente achatado para valores de κ entre $-0,30$ y $0,15$, lo cual representa el intervalo de interés. Si la información regional de un número de estaciones pueden ser utilizadas para desarrollar un *priori* más informativo para κ (Madsen and Rosbjerg, 1997b), serán obtenidas mejoras en los estimadores de cuantiles extremos (probabilidad de excedencia $\leq 1\%$). Lo mismo se espera del uso de la información histórica.

Una vez que el *priori* ($\pi(\kappa)$) é elegido, la densidad conjunta (función de verosimilitud generalizada) es calculada como $GL(\xi, \alpha, \kappa | x) = L(\xi, \alpha, \kappa | x) \pi(\kappa)$, la cual traduce la relación entre la función de verosimilitud generalizada y la función verosimilitud. Se podría ser más riguroso e incluir también un *priori* conjunto para ξ y α . Sin embargo, ello no fue realizado en este estudio. Un procedimiento *bayesiano* completo en análisis de frecuencia de crecidas fue presentado por Kuczera [1999]. Tal procedimiento es apropiado cuando se posee información realmente regional sobre los parámetros de posición y escala de la distribución, lo cual por lo general no se dispone en casos como los analizados en este trabajo.

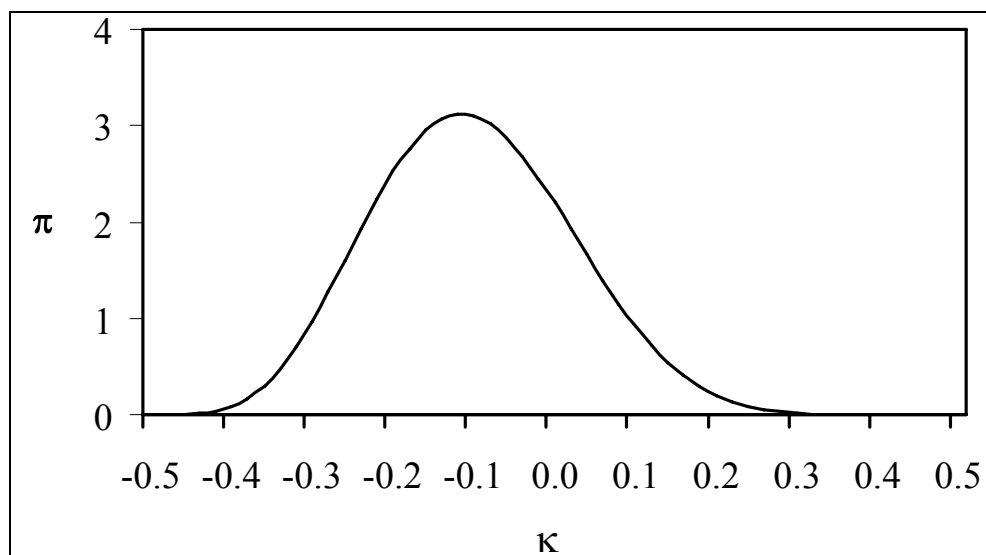


Figura X. 6. *Prior Geofísico* para el parámetro κ (de forma) de la distribución GEV. Distribución Beta con media = $-0,10$ y variancia = $(0,122)^2$ (Fuente: Martins, 2005).

Los estimadores de máxima verosimilitud generalizada (GMLE) de ξ , α y κ pueden ser identificados por la maximización de la función log-verosimilitud generalizada, lo cual corresponde a la moda de la distribución *a posteriori* bayesiana de los parámetros (Berger, 1985, p. 133). El procedimiento de Newton-Raphson es habitualmente empleado para calcular los estimadores de máxima verosimilitud.

Con los estimadores GMLE a medida que el tamaño muestral aumenta, la información contenida en la verosimilitud debe dominar la información provista por el *priori* (Robert, 1994; p. 138), tal que asintóticamente los dos poseen las mismas propiedades. En general los estimadores GMLE tendrán las mismas propiedades óptimas asintóticas si tanto la verosimilitud como el *priori* satisfacen unas pocas condiciones de regularidad, y pueden hasta realmente poseer estas propiedades en otros casos (Lehman & Casella, 1996). Con el *priori geofísico* el estimador GMLE irá a heredar las propiedades óptimas asintóticas del GEV-MLE para $-0,5 < \kappa < 0,5$.

3.7 Incorporación de Información Histórica

Métodos basados en máxima verosimilitud pueden emplear eficientemente información histórica, aumentando la información provista por los registros sistemáticos existentes. Cabe mencionar, sin embargo, que cuando se trabaja con la distribución GEV, el análisis de máxima verosimilitud clásico puede ser afectado por valores absurdos identificados para el parámetro de forma κ .

Tal como fuera mencionado, Martins y Stedinger (2000ab) resolvieron este problema con el uso de estimadores de máxima verosimilitud generalizados. Este es el abordaje adoptado en este estudio. El uso de la información histórica puede mejorar la precisión de tales parámetros, así como los estimadores de los cuantiles de interés.

La información histórica tiene su valor demostrado en varios estudios que hacen uso de series anuales (Leese, 1973; Condie y Lee, 1982; Condie, 1986; Stedinger y Cohn, 1986, 1987; Cohn y Stedinger, 1987; Jin y Stedinger, 1989; Francés, Salas y Boes, 1994; Martins y Stedinger, 2001). La ventaja de los métodos basados en máxima verosimilitud radica en que la información histórica puede ser fácilmente incorporada en la función de verosimilitud, aún si las magnitudes de las crecidas no pueden ser determinadas de manera precisa.

Si este último tipo de información pudiese ser determinada precisamente, entonces las magnitudes de los picos observados juntamente con el nivel (o umbral) considerado, el cual provee un límite para los valores no observados, son fácilmente incorporadas como datos truncados en la función de verosimilitud. Este caso es denominado *caso de datos truncados* (CD, del inglés *Censored Data*). Si, por el contrario, la única información disponible es el número de años en los cuales las crecidas excedieron un umbral de percepción (x_T), esta información puede también ser incorporada en la función de verosimilitud, según lo indicado por Stedinger y Cohn (1986). Este caso es denominado *datos truncados binomiales* (BCD, del inglés *Binomial Censored Data*).

El estudio de frecuencia que contempla la incorporación de datos históricos de crecidas del río Salado descrito en este estudio sigue la metodología propuesta por Martins y Stedinger (2001). El mismo emplea los estimadores de máxima verosimilitud generalizados en la estimativa de los cuantiles de la crecida y, por lo tanto, en la definición del tiempo de retorno de la crecida de abril/mayo de 2003.

3.8 Análisis de Máxima Verosimilitud con Información Histórica Basado en Series Anuales

El abordaje que será expuesto en este ítem se basa en el uso de series anuales con información histórica. La Figura X. 7 muestra H años de registro histórico con K = 4 crecidas superiores al umbral de percepción y un registro sistemático de S años.

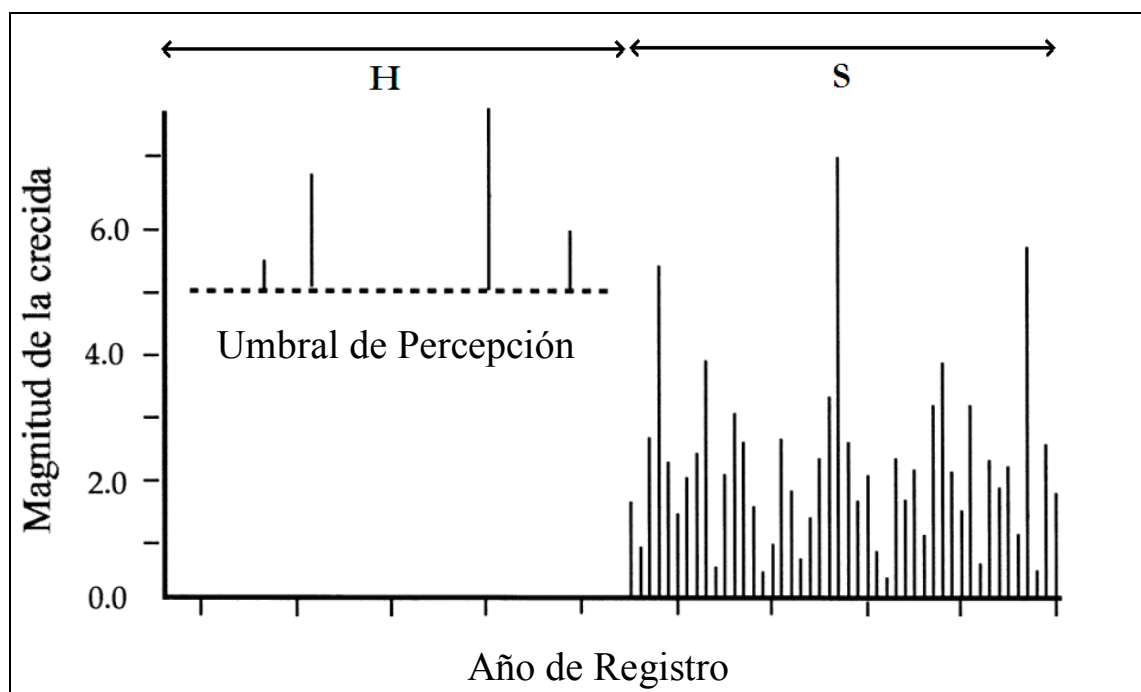


Figura X. 7. Serie anual con H años de información histórica, con K = 4 picos de crecidas en el período histórico superiores al umbral (o nivel) de percepción y S años de registros sistemáticos (Fuente: Martins, 2005).

La función log-verosimilitud:

$$L'(\theta^A | x) = \sum_{i=1}^S \ln(f(x_i | \theta^A)) \quad (X. 32)$$

para el conjunto de crecidas del registro sistemático $\{x_i\}$, iid, de una GEV con tamaño muestral s es:

$$L_{AMS}(\theta^A | x) = -s \ln(\alpha) + \sum_{i=1}^s \left[\left(\frac{1}{\kappa} - 1 \right) \ln(y_i) - (y_i)^{1/\kappa} \right] \quad (X. 33)$$

donde $y_i = 1 - \frac{\kappa}{\alpha}(x_i - \xi)$ y $\theta^A = (\xi, \alpha, \kappa)$ [Hosking et al., 1985].

Nuevamente, el uso de información histórica es incorporado: el caso BCD, donde solamente el número de crecidas (K) por encima de x_T en H años es conocida, y el caso CD donde las magnitudes de tales crecidas $\{x_j^*\}$ pueden ser determinadas de manera precisa. La expresión genérica para la log-verosimilitud del caso BCD resulta:

$$L_{AMS}^{BCD}(\theta^A | x) = \sum_{i=1}^s \ln(f(x_i | \theta^A)) + K \ln[1 - F(x_T | \theta^A, x \geq x_T)] + (H - K) \ln[F(x_T | \theta^A, x \geq x_T)] \quad (X. 34)$$

lo cual para la GEV resulta:

$$L_{AMS}^{BCD}(\theta^A | x) = L_{AMS}(\theta^A | x) + K \ln \left[1 - \exp \left(- (y_T)^{1/\kappa} \right) \right] - (H - K) (y_T)^{1/\kappa} \quad (X. 35)$$

donde $y_T = 1 - (\kappa/\alpha)(x_T - \xi)$.

Para el caso CD es:

$$L_{AMS}^{CD}(\theta^A | x) = \sum_{i=1}^s \ln(f(x_i | \theta^A)) + \sum_{j=1}^K \ln(f(x_j^* | \theta^A, x \geq x_T)) + (H - K) \ln[F(x_T | \theta^A, x \geq x_T)] \quad (X. 36)$$

lo cual para la GEV resulta:

$$L_{AMS}^{CD}(\theta^A | x) = L_{AMS}(\theta^A | x) - K \ln(\alpha) + \sum_{j=1}^K \left[\left(\frac{1}{\kappa} - 1 \right) \ln(y_j^*) - (y_j^*)^{1/\kappa} \right] - (H - K) (y_T)^{1/\kappa} \quad (X. 37)$$

El primer término de ambas ecuaciones de log-verosimilitud de los casos BCD y CD corresponde al registro sistemático. La función de densidad de probabilidad condicional (*pdf condicional*), $f(x_j^* | \theta^A, x \geq x_T)$ es dada por:

$$f(x_j^* | \theta^A, x \geq x_T) = \frac{f(x_j^* | \theta^A)}{1 - F(x_T | \theta^A)} \quad (X. 38)$$

3.9 Resultados de Experimentos Monte Carlo

Martins & Stedinger (2001) realizaron una serie de experimentos para comparar las ventajas derivadas del uso de información histórica con series anuales y parciales. De estos experimentos se puede concluir para las series anuales que:

- (i) cuando el parámetro de forma de la distribución GEV debe ser estimado (sin usar un valor fijo, como en el caso Gumbel, donde $\kappa = 0$), toda la información sobre la frecuencia de grandes crecidas se encuentra en los propios datos, y un año de información histórica (CD) posee casi el mismo valor que un año de registro sistemático;
- (ii) cuando la información histórica precisa no está disponible, mejoras sustanciales pueden ser aún obtenidas a partir del uso de datos truncados binomiales (BCD);
- (iii) cuando el umbral (o nivel) de percepción es próximo al cuantil de interés, los datos BCD valen casi lo mismo que los datos CD cuando se estiman los tres parámetros de la distribución GEV (y no dos, como en el caso particular de la distribución Gumbel).

3.10 Otros Modelos Probabilísticos Considerados

3.10.1 Lognormal (LN)

Para obtener las ecuaciones de log-verosimilitud para observaciones del conjunto de crecidas del registro sistemático, iid, que siguen un modelo Log Normal, basta sustituir en las ecuaciones, X.32, X.34, X.36 y X.38 la función de densidad de probabilidad de la distribución Log Normal, a saber:

$$f(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(\ln(x) - \mu)^2\right] \quad (\text{X. 39})$$

donde μ e σ^2 son la media y la variancia del logaritmo de los caudales máximos respectivamente.

3.10.2 Log-Pearson III (LPIII)

Esta sección presenta un análisis *bayesiano* para la distribución LPIII utilizando el algoritmo Metropolis-Hastings. En primer lugar se describe la distribución. Posteriormente, se presentan las distribuciones *a priori* de los parámetros, la función de verosimilitud y las distribuciones auxiliares utilizadas para proponer nuevos valores de los parámetros en el algoritmo Metropolis-Hastings.

El modelo:

Si las crecidas $\{Q_1, \dots, Q_N\}$ se distribuyen de acuerdo con la distribución LPIII, entonces $X = \log(Q)$ posee una distribución Pearson tipo III cuya función de densidad de probabilidad (*pdf*) es descripta de la siguiente forma

$$f_X(x) = \frac{|\beta|}{\Gamma(\alpha)} [\beta(x-\tau)]^{\alpha-1} e^{-\beta(x-\tau)} \quad (\text{X. 40})$$

donde α , β , y τ son los parámetros de forma, escala y posición respectivamente, y $\Gamma(\alpha)$ es la función Gamma. Si $\beta > 0$, la distribución posee una asimetría positiva y τ es el límite inferior de la distribución; se $\beta < 0$, la asimetría es negativa con el límite superior τ . Cuando el parámetro de forma tiende a infinito, el coeficiente de asimetría tiende a cero, y la LPIII se transforma en la distribución Log Normal. Como el interés está generalmente en las distribuciones con coeficiente de asimetría pequeño, algunos problemas numéricos pueden ocurrir cuando se ajusta la distribución utilizando el método MV. Una otra parametrización basada en la media, en el desvío estándar y la asimetría es normalmente utilizada para el cálculo de los cuantiles de la crecida:

$$x_p = \mu + \sigma K_p(\gamma) \quad (\text{X. 41})$$

donde $K_p(\gamma)$ es el factor de frecuencia con asimetría γ , y es igual al cuantil de la distribución PIII cuando la media es cero y el desvío estándar es igual a la unidad. El factor de frecuencia K_p puede ser aproximado a través de la transformación de Wilson-Hilferty (Kirby, 1972) para $|\gamma| < 2$ y $0.01 < p < 0.99$.

$$K_p(\gamma) = \frac{2}{\gamma} \left(1 + \frac{\gamma z_p}{6} - \frac{\gamma^2}{36} \right)^3 - \frac{2}{\gamma} \quad (\text{X. 42})$$

donde z_p es el cuantil de la distribución normal estándar.

Abordaje Bayesiana con simulación tipo MCMC (Monte Carlo Markov Chains):

El algoritmo Metropolis-Hastings es utilizado para generar una muestra empírica de los parámetros que represente la distribución *a posteriori* de μ , σ , γ , y de los cuantiles x_p . Este procedimiento requiere la definición de las distribuciones *a priori* y un auxilio de propuesta de los parámetros y de la función de verosimilitud.

Distribuciones “a priori”

En el ejemplo que será presentado en la próxima sección se empleó una distribución *a priori* cuasi no informativa para μ , σ , mientras que una más informativa para γ . Dado que μ puede ser negativo, la distribución *a priori* utilizada fue una distribución normal con media cero y variancia igual a 1.000. Zellner (1971) presenta una discusión de cómo utilizar una distribución *a priori* no informativa para σ , y sugiere el uso de una distribución proporcional al inverso del desvío estándar. En relación a γ , estudios anteriores y limitaciones físicas relacionadas a las crecidas indican que el coeficiente de asimetría debe variar como máximo entre $\pm 1,4$. Si los valores de asimetría poblacional fuesen uniformemente distribuidos entre $\pm 1,0$, la variancia sería igual a 0,33, ciertamente un número muy grande. La distribución *a*

priori adoptada para γ fue una distribución normal con media cero y variancia igual a 0,3 (Tasker y Stedinger, 1986; Reis *et al.*, 2003, Reis, 2005).

$$\xi(\mu) \sim N(0,1000)$$

$$\xi(\sigma) \propto \frac{1}{\sigma} \quad (\text{X. 43})$$

$$\xi(\gamma) \sim N(0,0.3)$$

Función de Verosimilitud:

La sección anterior describe la función verosimilitud que incorpora las informaciones históricas. Tanto la X.34 como la X.36 pueden ser utilizadas. La estimación de la función requiere el cálculo de la $f_X(x)$ de todas las observaciones y la función de distribución acumulada (cdf) del límite de percepción $F_X(x_o)$. Utilizando la transformación de Wilson-Hilferty para relacionar la variable X_g con la variable normal estándar z_n se tiene:

$$z_n = \frac{\gamma}{6} + \frac{6}{\gamma} \left[\frac{\gamma}{2} \left(\frac{X_g - \mu}{\sigma} \right) + 1 \right]^{1/3} - \frac{6}{\gamma} \quad (\text{X. 44})$$

Por lo tanto, la distribución de densidad de probabilidad de la función LPIII para X_g puede ser obtenida a través de la distribución de densidad de probabilidad de z_n

$$f_{X_g}(x_g) = \phi(z_n) \frac{dz_n}{dX_g} = \frac{\phi(z_n)}{\sigma \left[\frac{\gamma}{2} \left(\frac{x_g - \mu}{\sigma} \right) + 1 \right]^{2/3}} \quad (\text{X. 45})$$

Sustituyéndose las ecuaciones X.44 y X.45 en las ecuaciones X.34 o X.36, se obtiene la función de verosimilitud utilizada en el algoritmo Metropolis-Hastings. Esta aproximación es relativamente adecuada en la región próxima a $\gamma = 0$, de manera que el algoritmo es estable desde el punto de vista computacional.

Distribución propuesta

Los tres parámetros μ , σ , y γ son propuestos de manera independiente entre si, aunque los valores propuestos dependan de los valores de cada parámetro en la última iteración. La distribución propuesta para la media μ y una distribución normal con media igual a μ_{t-1} y variancia igual a σ_{t-1}^2 dividida por el número de observaciones N . En el caso de que las magnitudes de las crecidas históricas sean utilizadas, el número de observaciones es igual a $(s + k)$, en caso contrario, N es igual a s ,

$$\mu_t \sim N(\mu_{t-1}, \sigma_{t-1}^2 / N) \quad (\text{X. 46})$$

La distribución propuesta para σ es la distribución Gamma con media igual a σ_{t-1} y variancia modelada como una función de σ_{t-1} y γ_{t-1} , de acuerdo a lo descrito por Stedinger y Tasker (1986):

$$\begin{aligned} \sigma_t &\sim \text{gamma}(a, b) \\ a &= \frac{\sigma_{t-1}^2}{\text{Var}(\sigma_{t-1})} \\ b &= \frac{\text{Var}(\sigma_{t-1})}{\sigma_{t-1}} \end{aligned} \quad (\text{X. 47})$$

$$\text{Var}(\sigma_{t-1}) = \frac{\sigma_{t-1}^2 (1 + 0.75\gamma_{t-1}^2)}{2N} \quad (\text{X. 48})$$

La distribución propuesta para γ es una distribución normal con media igual a γ_{t-1} y variancia modelada como una función de γ_{t-1} e N , según descrito en el U.S. Bulletin 17B (USWRC, 1982):

$$\begin{aligned} \gamma_t &\sim N[\gamma_{t-1}, \text{Var}(\gamma)] \\ \text{Var}(\gamma) &= \left[1 + \frac{6}{N}\right]^2 10^{[a-b \log(N/10)]} \\ a &= \begin{cases} -0.33 + 0.08|\gamma_{t-1}| & \text{if } |\gamma_{t-1}| < 0.90 \\ -0.52 + 0.30|\gamma_{t-1}| & \text{if } |\gamma_{t-1}| > 0.90 \end{cases} \\ b &= \begin{cases} 0.94 - 0.26|\gamma_{t-1}| & \text{if } |\gamma_{t-1}| < 1.50 \\ 0.55 & \text{if } |\gamma_{t-1}| > 1.50 \end{cases} \end{aligned} \quad (\text{X. 49})$$

Como comentado anteriormente, τ puede ser tanto un límite superior o inferior, dependiendo del signo de β . Siempre que el valor de τ es mayor que la menor observación, en el caso de $\beta > 0$, o menor que la mayor observación cuando $\beta < 0$, los valores propuestos de los parámetros son rechazados ya que la función de verosimilitud sería nula en estos casos.

4 ANÁLISIS DE FRECUENCIA SOBRE SERIES DE DATOS HASTA EL AÑO 2002

4.1 Análisis de Frecuencia sobre Series Sistemáticas (S)

4.1.1 Análisis de Frecuencia con Base en la Serie 1954-2002

Para la realización de este análisis se empleó la serie de caudales y niveles diarios de la Ruta Provincial 70 (RP70) correspondiente al período 1954-2002. Esta serie presenta un vacío de información en el periodo 1987-1993, razón por la cual la misma incluye 41 valores.

Análisis preliminar de la serie

Previo al estudio de frecuencia se realizó un análisis de la serie a los efectos de verificar sus cualidades estadísticas. Este análisis previo incluyó la determinación de los parámetros descriptivos de la serie. El estudio incluyó también el análisis de los valores dudosos (“outliers”) según la metodología propuesta por el U.S. Water Resources Council (1981). La misma se basa en el análisis de la asimetría de la serie analizada. Del mismo modo, fueron realizados distintos tipos de tests estadísticos de tipo paramétrico y no paramétrico para analizar las condiciones de independencia, tendenciosidad y homogeneidad de la serie.

A continuación se indican las metodologías que fueron empleadas en estos análisis:

- Análisis de Independencia de la Serie (prueba Wald-Wolfowitz)
- Prueba t y t' para medias de dos muestras
- Prueba F para varianzas de dos muestras
- Análisis de Tendencia (prueba Kendall)
- Análisis de Homogeneidad (prueba Wilcoxon)

Los resultados de estos estudios son sintetizados a continuación.

Estadística descriptiva

En la Tabla X. 3 se presentan los principales parámetros estadísticos de la serie en estudio.

Detección de tendencias en la serie

Aunque el análisis gráfico carece del peso de las pruebas de tipo paramétrico y no paramétrico, el mismo representa la forma más práctica y rápida disponible para visualizar posibles tendencias de una serie. La tendencia representa el comportamiento predominante de la serie; esta puede ser definida como el cambio de la media a lo largo de un periodo. En la Figura X. 8 se presentan en forma gráfica los análisis realizados. En ella se observa la existencia de una tendencia creciente, con un salto muy marcado de las medias a partir de la década del 70. El salto en la media es del 65 %, ya que la misma varía de 276 m³/s (1954-1972) a 804.1 m³/s (1972-2002).

La presencia del salto en la media de los caudales máximos, evidenciada a partir del año 1972, indujo a analizar la totalidad de las posibles causas naturales y artificiales de dicho comportamiento.

Entre las primeras se analizaron las tendencias climáticas observadas, que dan cuenta de un período más húmedo en relación a los años 1954-1971 en diversas estaciones de la región, como también en caudales de ríos vecinos.

También fue considerada como causa natural del salto observado la probable existencia de un cambio climático que se hace más evidente a nivel mundial. La consideración de este último aspecto no impide remarcar que ya en años anteriores (más precisamente en el año 1914), se registraron en la cuenca inferior del río Salado precipitaciones diarias del orden de 600 mm (citadas en el Anexo IX), como también que en años recientes (1995) se registró el mínimo caudal diario en la estación RP70: 7 m³/s.

Entre las causas antrópicas analizadas se destacan: (i) el cambio en el uso del suelo, (ii) el paulatino impacto de las intervenciones humanas dentro de la red de drenaje de las cuencas santafesinas de aporte al río Salado y (iii) la probable influencia sobre las mediciones hidrométrica de la RP70 del remanso provocado por la presencia del terraplén y puente de la Autopista Santa Fe-Rosario (AU01). Este puente y sus accesos fueron ejecutados en 1971/1972, dejando libre una luz extremadamente exigua: apenas 155 m.

El análisis del impacto de las intervenciones humanas sobre la red de drenaje de la cuenca inferior del río Salado es tratado en el Anexo VII. Por su parte, en relación a la última causa apuntada, se procedió al análisis de los remansos provocados por el puente de la AU01 para distintos caudales máximos del río Salado y diferentes niveles del río Paraná. Los estudios permitieron concluir que la influencia del puente produce variaciones del orden del +5 % en los caudales observados en la estación RP70. Esta variación se encuentra dentro del orden de incertidumbre de los registros de campo, concluyéndose por lo tanto que la misma no representa una causa de peso en el cambio de la media de los caudales máximos.

En síntesis, el análisis realizado condujo al estudio estadístico de dos subseries correspondientes a los períodos 1954-1972 y 1972-2002 respectivamente. En la Tabla X. 3 se presentan las principales informaciones que describen a cada una de ellas.

Tabla X. 3. Resumen estadístico de las series de caudales máximos (RP70, periodo 1954-2002)

	1954-2002	1954-1972	1972-2002	Estandarizada
Media	572.4	276.4	804.1	804.1
Error típico	88.3	35.1	137.7	101.8
Mediana	406.6	251.8	662.9	662.9
Desviación estándar	565.3	148.8	660.5	652.1
Varianza de la muestra	319592.9	22140.9	436197.6	425292.7
Curtosis	6.5	0.2	3.1	1.5
Coefficiente de asimetría	2.5	0.9	1.9	1.4
Rango	2563.9	511.8	2517.8	2615.1
Mínimo	108.1	108.1	154.2	56.9
Máximo	2672.0	619.9	2672	2672.0
Suma	23469.0	4975.8	18493.2	32966.1
Cantidad de valores	41	18	23	41

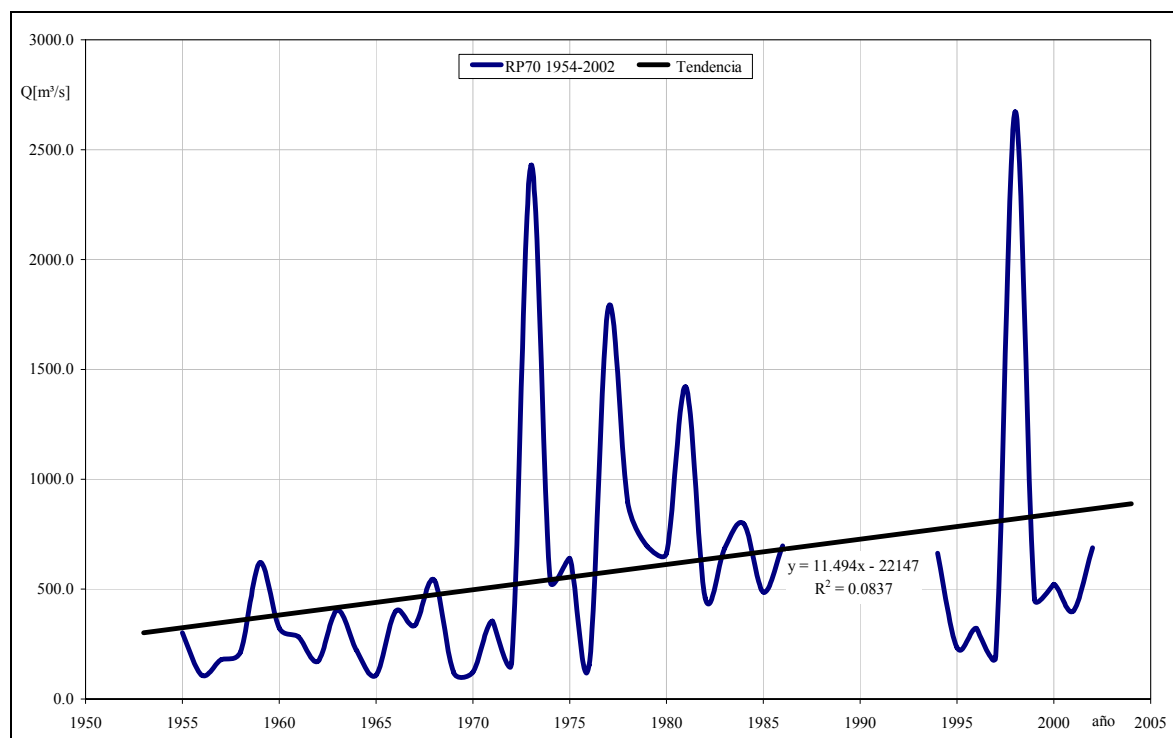


Figura X. 8. Gráfico de tendencia en las series de caudales máximos RP70 (Período 1954-2002).

Análisis de valores dudosos (outliers):

De acuerdo a lo previamente indicado, el análisis de valores dudosos se realizó tanto para la serie completa analizada, 1954-2002, como para las dos subseries en que la misma fue dividida, 1954-1972 y 1972-2002. Las Figura X. 9, Figura X. 10, Figura X. 11 y Figura X. 12 sintetizan los resultados obtenidos. En las mismas se observa que en todos los casos los puntos se encontraron dentro de la faja establecida por los límites superior (yh) e inferior (yL) respectivamente. Por lo tanto, se consideró la no existencia de valores dudosos o “outliers” evidentes en las series analizadas.

En la Figura X. 9 se presenta el análisis correspondiente a la serie 1954-2002 estandarizada por la media y la varianza. La necesidad de consideración de esta serie se realiza en los próximos apartados. La conclusión del gráfico de valores dudosos es similar a la comentada para las todas series precedentemente descriptas.

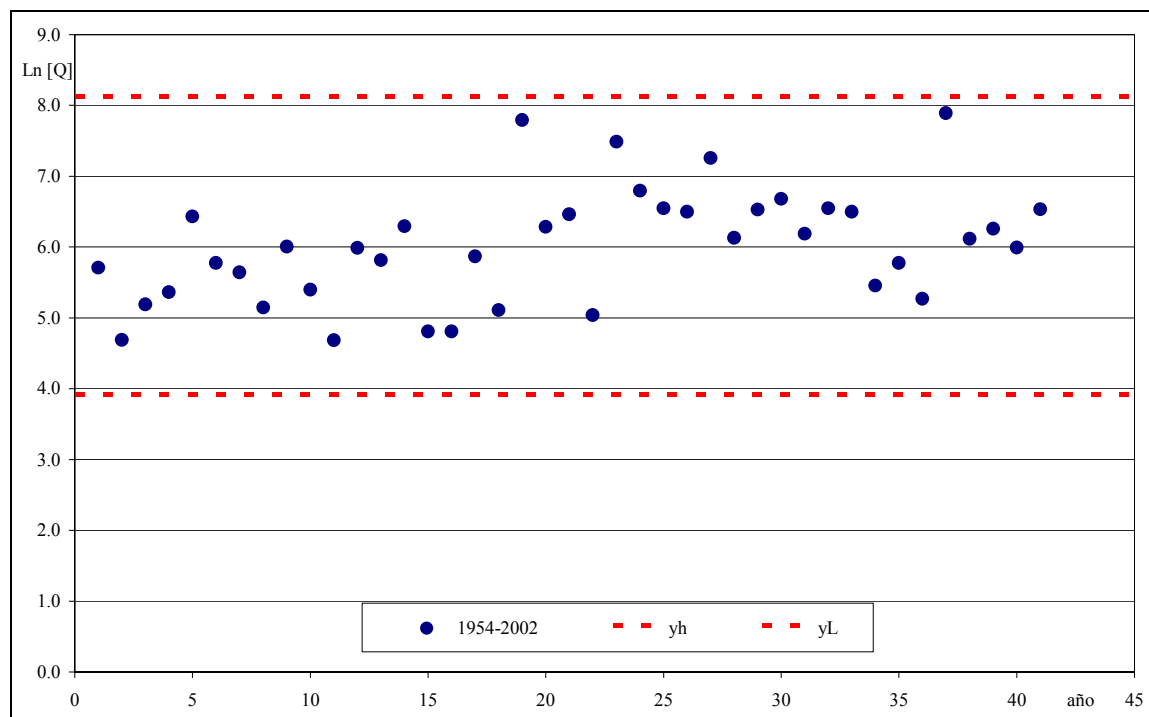


Figura X. 9. Prueba de datos dudosos en RP70 (periodo: 1954-2002).

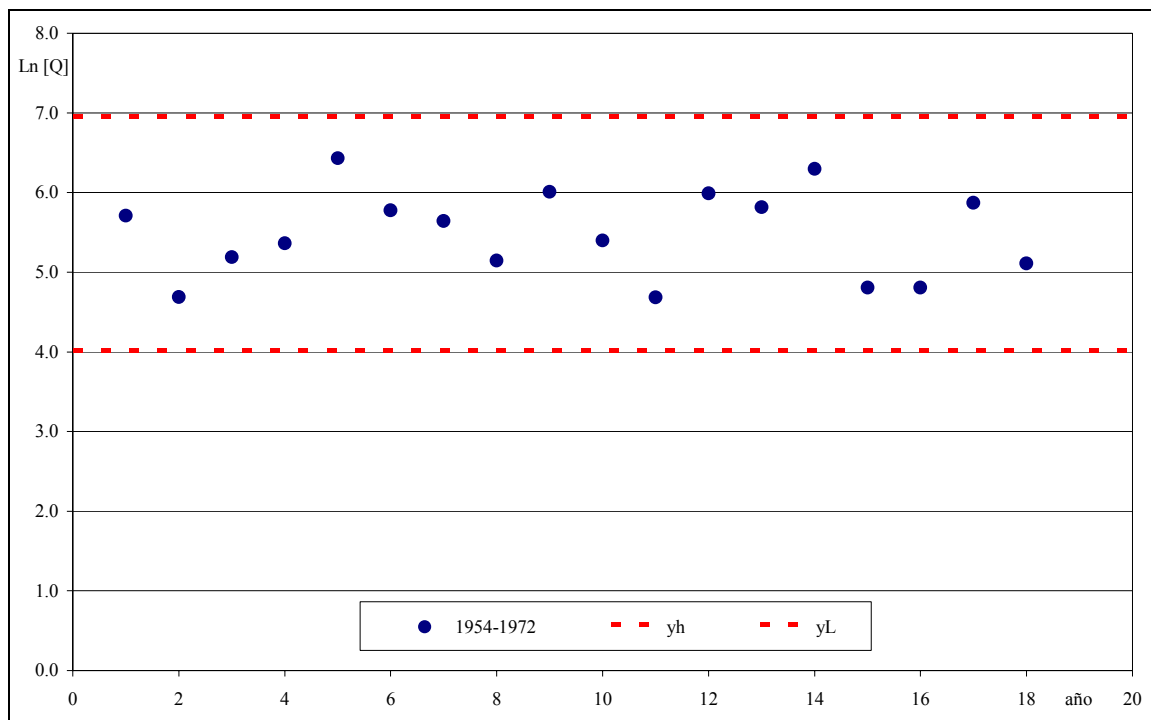


Figura X. 10. Prueba de datos dudosos en RP70 (periodo: 1954-1972).

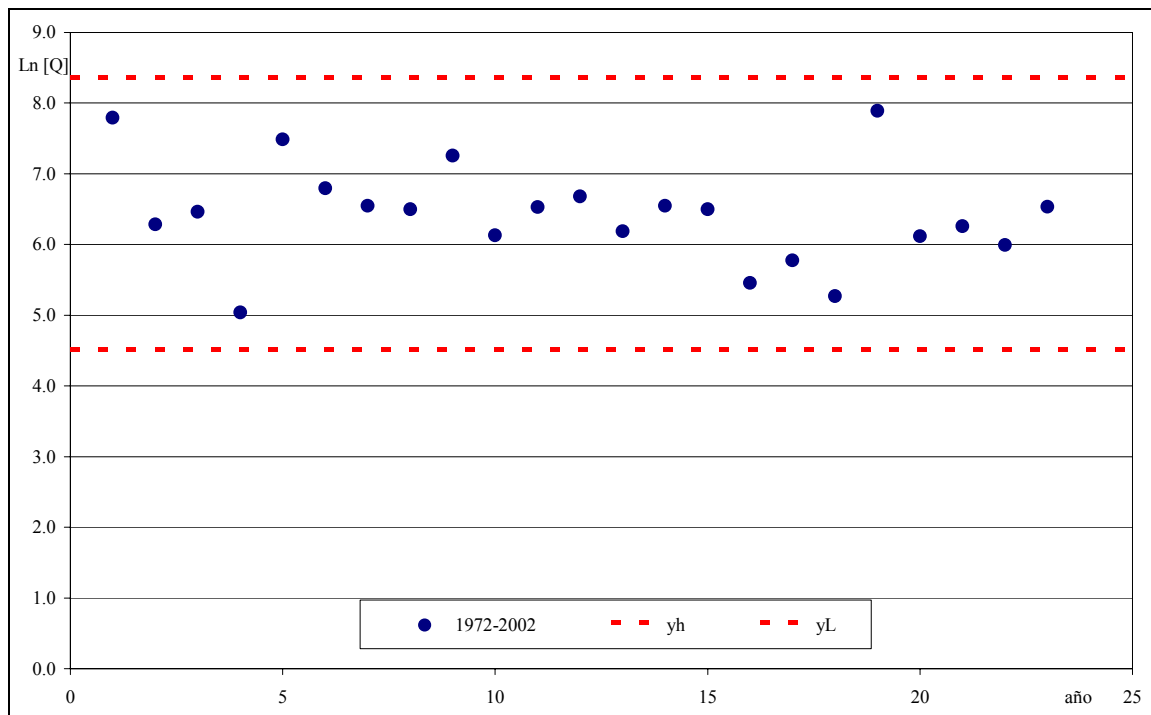


Figura X. 11. Prueba de datos dudosos en RP70 (periodo: 1972-2002).

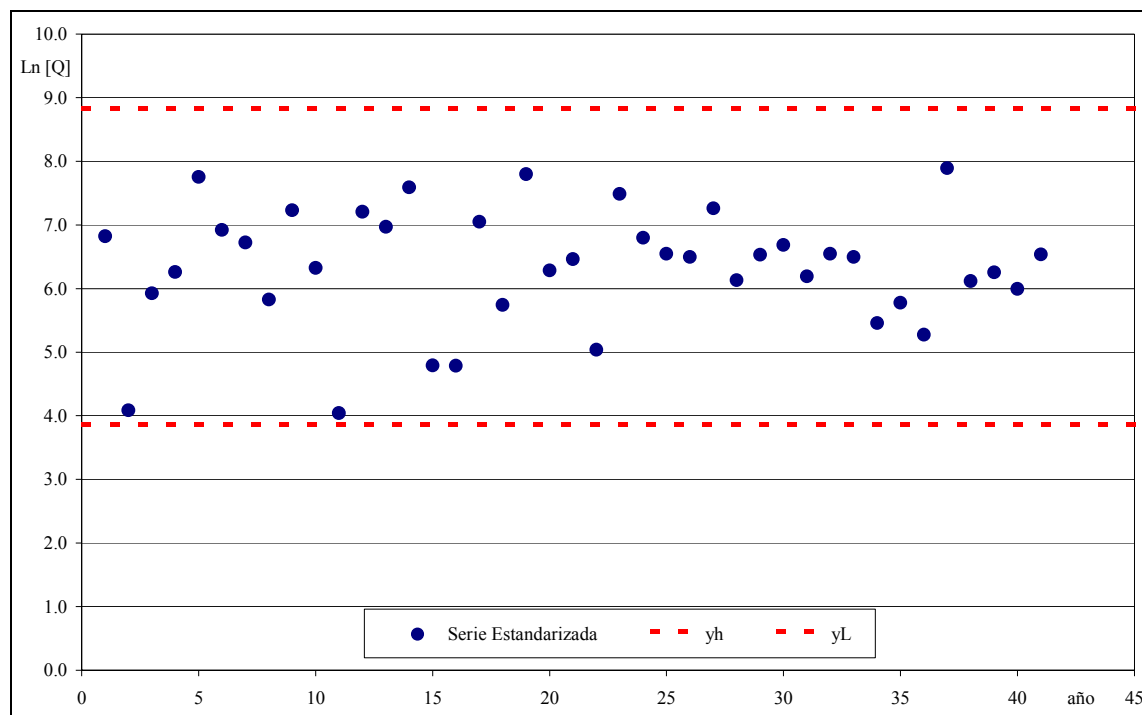


Figura X. 12. Prueba de datos dudosos en RP70 (periodo: 1954-2002, serie estandarizada mediante corrección de media y varianza).

Análisis de pruebas paramétricas

A los efectos de cuantificar las características de cada una de las dos subseries previamente indicadas, se procedió a la aplicación de la totalidad de tests paramétricos indicados en la primera parte de este Anexo. En las Tablas Tabla X. 4, Tabla X. 5 y Tabla X. 6, se presentan los resultados obtenidos.

Del análisis de los resultados se desprende que el salto que presenta la serie a partir de 1972 es significativo desde el punto de vista estadístico. El mismo se registra tanto en las medias como en las varianzas. Dado que estos tests se basan en la condición de normalidad de la muestra, la información que los mismos proveen debe ser considerada como no concluyente. Es preciso recordar que las pruebas t son relativamente robustas ante la ocurrencia de esta violación, pero no así la prueba F de Fischer para las varianzas. Ello justificó el análisis del ajuste de la serie a la distribución normal (condición de normalidad de la serie).

Según se ilustra en las Figura X. 13, Figura X. 14 y Figura X. 15, la serie analiza presenta un alejamiento de la distribución normal, fundamentalmente en los máximos (cola de la distribución). Ello justificó, a su vez, la realización de pruebas no paramétricas a los efectos de ratificar las indicaciones de los tests paramétricos.

Tabla X. 4. Prueba t, suponiendo varianzas iguales

	1954-1972	1972-2002
Media	276.4	804.1
Varianza	22140.9	436197.6
Observaciones	18	23
Varianza agrupada	255711.3	
Diferencia hipotética de las medias	0	
Grados de libertad	39	
Estadístico t	-3.316	
Significancia (α)	0.050	
P(T<=t) una cola	0.001	
Valor crítico de t (una cola)	1.685	
Significancia (α)	0.050	
P(T<=t) dos colas	0.002	
Valor crítico de t (dos colas)	2.023	
Hipótesis H0	Se Rechaza	

Tabla X. 5. Prueba t', suponiendo varianzas desiguales

	1954-1972	1972-2002
Media	276.4	804.1
Varianza	22140.9	436197.6
Observaciones	18	23
Diferencia hipotética de las medias	0	
Grados de libertad	25	
Estadístico t	-3.713	
Significancia (α)	0.050	
P(T<=t) una cola	0.001	
Valor crítico de t (una cola)	1.708	
Significancia (α)	0.050	
P(T<=t) dos colas	0.001	
Valor crítico de t (dos colas)	2.060	
Hipótesis H0	Se Rechaza	

Tabla X. 6. Prueba F para varianzas de dos muestras

	1954-1972	1972-2002
Media	276.4	804.1
Varianza	22140.9	436197.6
Observaciones	18	23
Grados de libertad	17	22
F	0.051	
Significancia (α)	0.050	
P(F<=f) una cola	0.000	
Valor crítico para F (una cola)	0.453	
Hipótesis H0	Se Rechaza	

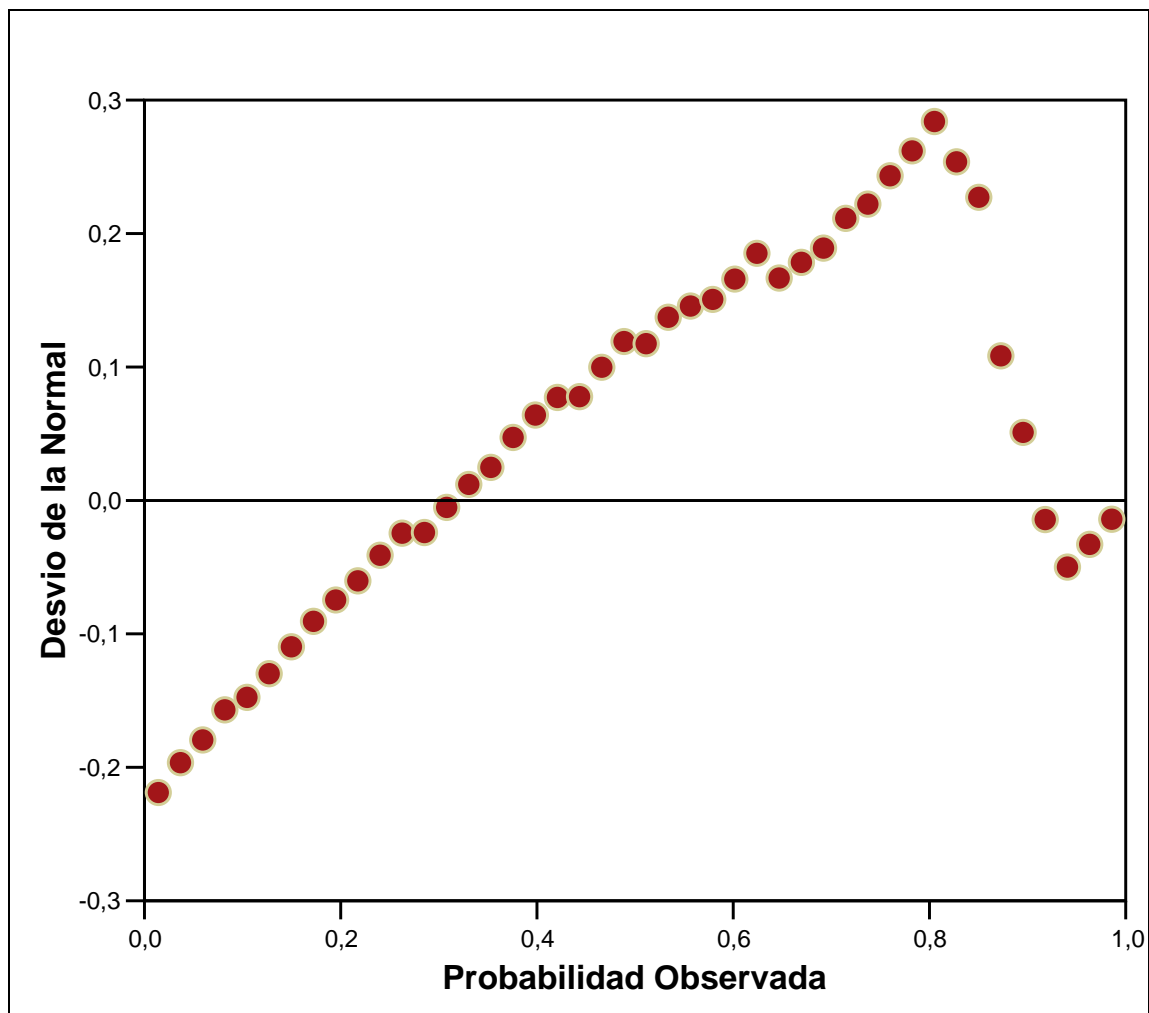


Figura X. 13. Análisis gráfico de normalidad de la serie 1954-2002.

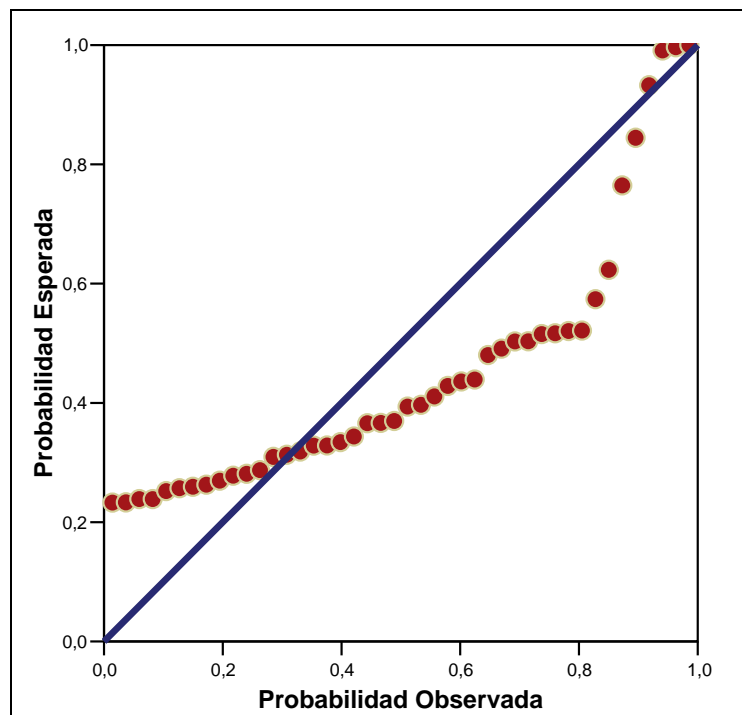


Figura X. 14. Análisis gráfico de normalidad de la serie 1954-2002.

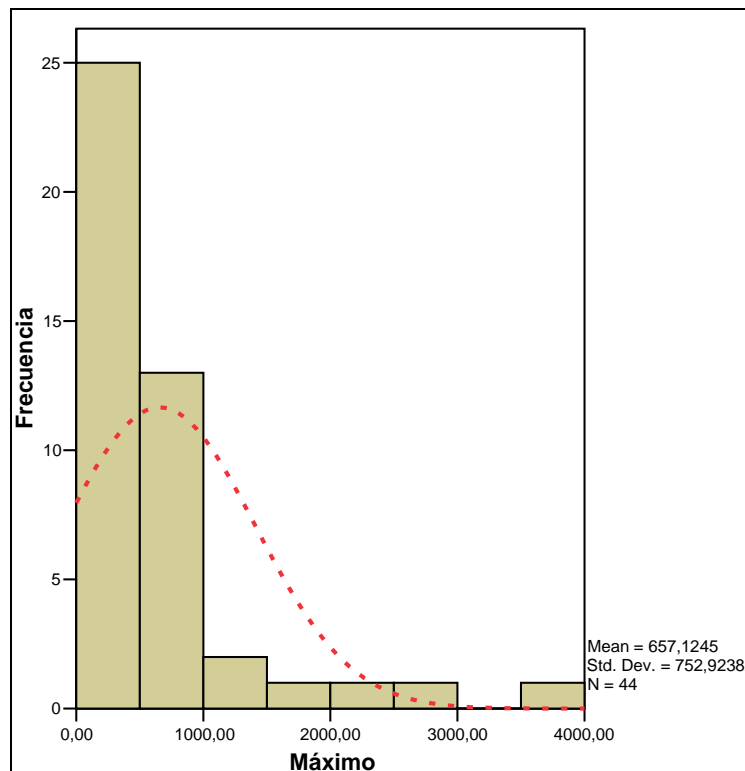


Figura X. 15. Análisis gráfico de normalidad de la serie 1954-2002.

Análisis de pruebas no paramétricas

De acuerdo a lo previamente indicado, fueron aplicados los tests de Wald-Wolfowitz (independencia), Kendall (tendencia) y Wilcoxon (Homogeneidad). En las Tabla X. 7, Tabla X. 8 y Tabla X. 9, se presentan los resultados alcanzados por la aplicación de estos tests para las series 1954-2002 y 1972-2002. Del análisis de todas las informaciones presentadas se dedujo la existencia de dos caminos alternativos para la realización del estudio de frecuencia con base en la serie sistemática. Ambos fueron aplicados en este trabajo:

- (iv) *estudio basado en la serie moderna 1972-2002*: se caracteriza por representar de manera directa condiciones más húmedas que aquellas observadas en décadas anteriores. Además, en relación a la serie 1954-2002, refleja una mayor influencia antrópica en la cuenca del río Salado como también síntomas de un probable cambio climático.
- (v) *estudio basado en la serie 1954-2002 estandarizada*: permite emplear la totalidad de las informaciones sistemáticas disponibles. En este estudio la estandarización se realizó sobre el periodo anterior a 1972, considerando la media y la varianza del periodo posterior. Conforme presentado previamente, esta serie estandarizada fue objeto de todos los análisis paramétricos y no paramétricos considerados en este estudio, como así también el test de datos dudosos (o “outliers”). Al estandarizarse la serie por la media y la varianza del período más reciente, se remarcan las condiciones dominantes de este último.

Tabla X. 7. Prueba de Independencia (Wald-Wolfowitz)

	1954-2002	1972-2002	Estandarizada
Valor del estadístico U	0.07	0.94	1.18
Valor de significancia (α)	0.05	0.05	0.05
Valor probabilidad (p)	0.943	0.347	0.238
Hipótesis H0	Se Acepta	Se Acepta	Se Acepta

Tabla X. 8. Prueba de Estacionalidad (Kendall)

	1954-2002	1972-2002	Estandarizada
Valor del estadístico K	2.77	1.48	0.57
Valor de significancia (α)	0.01	0.05	0.05
Valor probabilidad (p)	0.006	0.139	0.567
Hipótesis H0	Se Rechaza	Se Acepta	Se Acepta

Tabla X. 9. Prueba de Homogeneidad (Wilcoxon)

	1954-2002	1972-2002	Estandarizada
Valor del estadístico W	4.03	1.08	0.0919
Valor de significancia (α)	0.05	0.05	0.05
Valor probabilidad (p)	0.000	0.073	0.927
Hipótesis H0	Se Rechaza	Se Acepta	Se Acepta
División de la serie	1971-1972	1971-1972	1971-1972

Los resultados de los tests no paramétricos permitieron ratificar que el salto que presenta la serie 1954-2002 no solo afecta estadísticamente a la estacionalidad de la serie sino que también influye en la homogeneidad de la misma. Esta última característica fue asociada al hecho de corresponder el periodo 1954-1972 a un ciclo relativamente seco, mientras que la serie 1972-2002, tal como ya fuera apuntado, representa un ciclo de mayor humedad, con posibles reflejos de una mayor influencia antrópica en la cuenca del río Salado y también síntomas de un probable cambio climático.

Posición de ploteo (distribución empírica)

Al considerarse la fórmula de Gringorten, las series empleadas en el presente análisis adoptaron las posiciones de ploteo presentadas en la Figura X. 16. En la Figura se puede apreciar la existencia de un quiebre que origina dos tramos: uno se asocia a los máximos recurrentes (tramo inferior) y el restante a los máximos menos frecuentes (tramo superior).

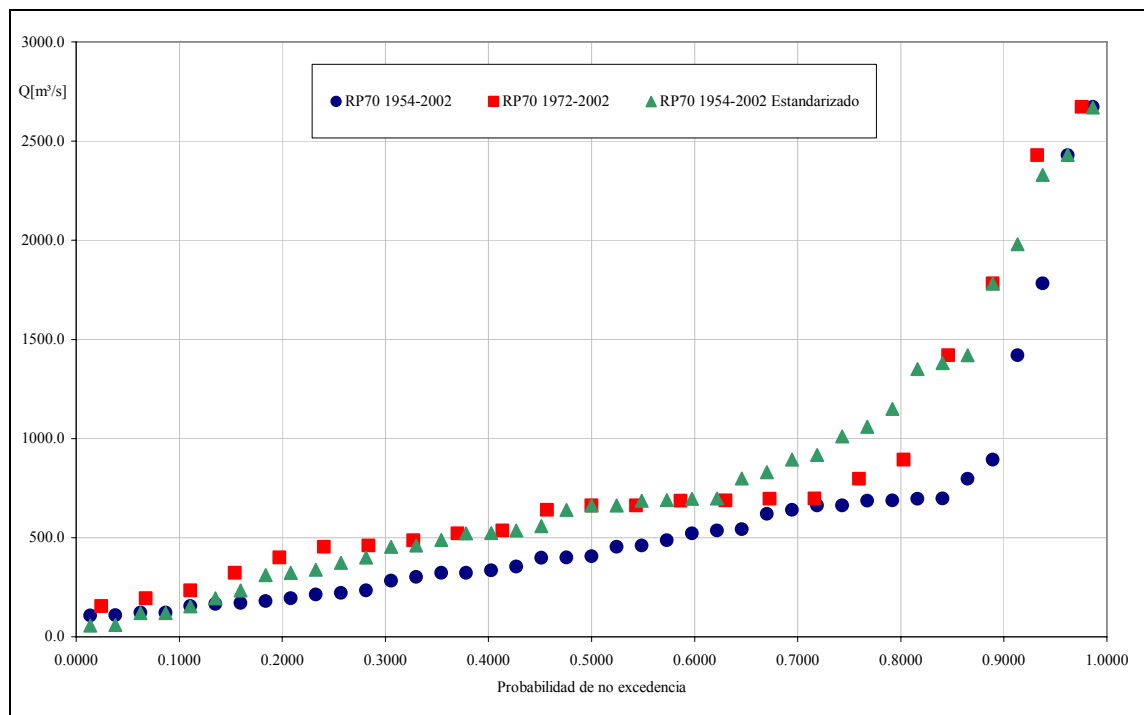


Figura X. 16. Posición de ploteo de las series analizadas de acuerdo a la fórmula de Gringorten.

4.1.2 Análisis de Frecuencia con base en la serie 1954-2002 estandarizada

A los efectos del análisis de frecuencia fueron ajustadas distintas funciones de frecuencia (o modelos probabilísticos). Los mismos incluyeron los modelos Exponencial, Log Pearson III, de Valores Extremos Generalizados (GEV) y LogNormal.

Todas las distribuciones fueron ajustadas empleando el método de máxima verosimilitud (MLE, por sus siglas en inglés *Maximum Likelihood Estimation*). Para estos análisis fue empleado el paquete estadístico HYFRAN[©], desarrollado por la Cátedra de Hidrología Estadística CRNSG-INRS-Eau/Hydro-Québec (Département d'Hydrologie) /Alcan.

En las Tabla X. 10, Figura X. 17 y Figura X. 18 se presentan las informaciones correspondientes al ajuste de la serie 1954-2002 estandarizada. Las Figuras indicadas se diferencian en la escala de frecuencia empleadas (abscisas), permitiendo contar con dos visiones del ajuste de cada distribución. Luego de la aplicación del test de adecuación χ^2 se obtuvo para todas las funciones ajustadas la aceptación de la adecuación con una significancia del 5 %. Cabe consignar que para determinadas distribuciones y algunos valores de recurrencia, T_r , el paquete HYFRAN[©] no ofrece los valores correspondientes al intervalo de confianza del 5%. Esta situación se refleja en las Tablas con la indicación "N/D".

Del análisis gráfico se desprende que las distribuciones GEV, Exponencial y Log Pearson III son aquellas que podrían seleccionarse para explicar el comportamiento de la totalidad de la serie analizada, con ventajas de la última sobre las restantes.

Tabla X. 10. Resultados de los ajustes realizados sobre la serie estandarizada 1954-2002

Función de Ajuste		GEV		Log-Pearson III		Lognormal		Exponencial	
Tr [años]	q	Q _T	Intervalo	Q _T	Intervalo	Q _T	Intervalo	Q _T	Intervalo
10000	0.9999	14600	N/D	5710	N/D	17600	3460 -31800	7090	4910 - 9270
2000	0.9995	9330	N/D	4880	N/D	11900	3280 -204000	5860	4060 - 7660
1000	0.9990	7650	N/D	4500	N/D	9860	3090 -16600	5330	3690 - 6960
200	0.9950	4720	N/D	3570	N/D	6130	2500 - 9770	4100	2840 - 5350
100	0.9900	3790	N/D	3150	N/D	4870	2200 - 7550	3570	2480 - 4650
50	0.9800	3000	1370 - 4630	2720	N/D	3790	1880 - 5690	3030	2110 - 3960
20	0.9500	2150	1280 - 3030	2140	1400 - 2880	2600	1460 - 3730	2330	1630 - 3040
10	0.9000	1620	1100 - 2150	1680	1210 - 2160	1860	1140 - 2570	1800	1260 - 2340
5	0.8000	1160	856 - 1470	1230	872 - 1590	1240	829 - 1640	1270	893 - 1650
3	0.6667	857	646 - 1070	893	612 - 1170	846	596 - 1100	880	622 - 1140
2	0.5000	619	461 - 777	618	414 - 823	569	408 - 729	569	406 - 732

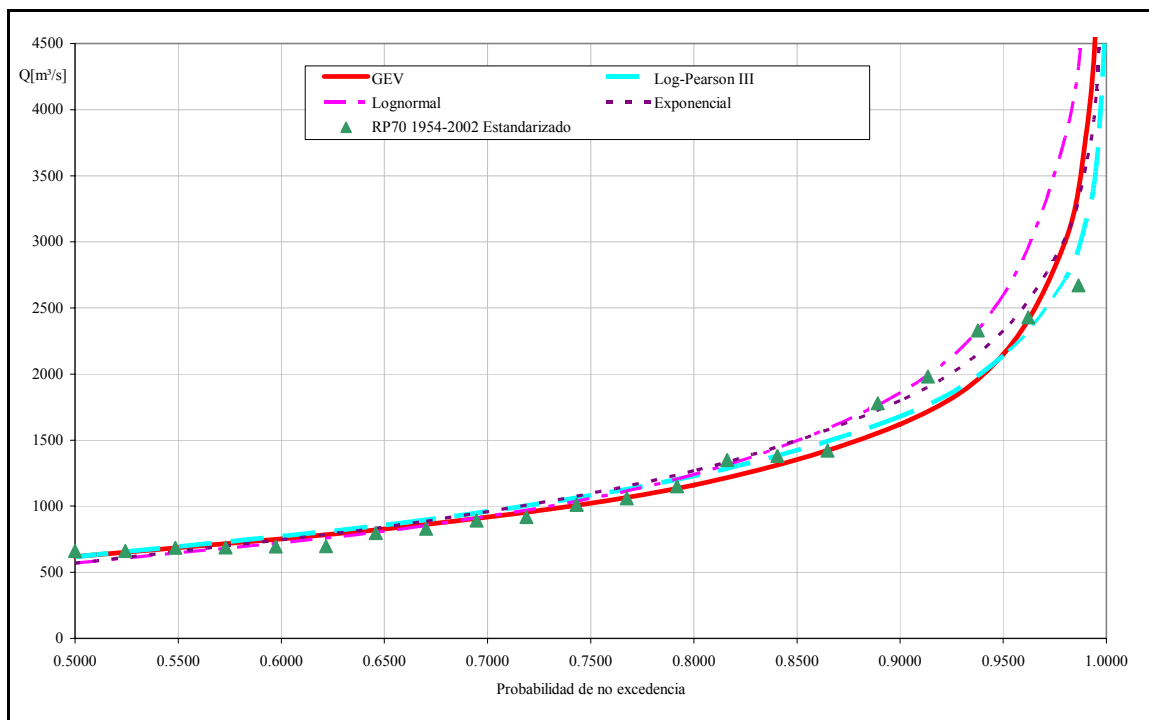


Figura X. 17. Gráfica de los ajustes realizados sobre la serie RP70 1954-2002 estandarizada.

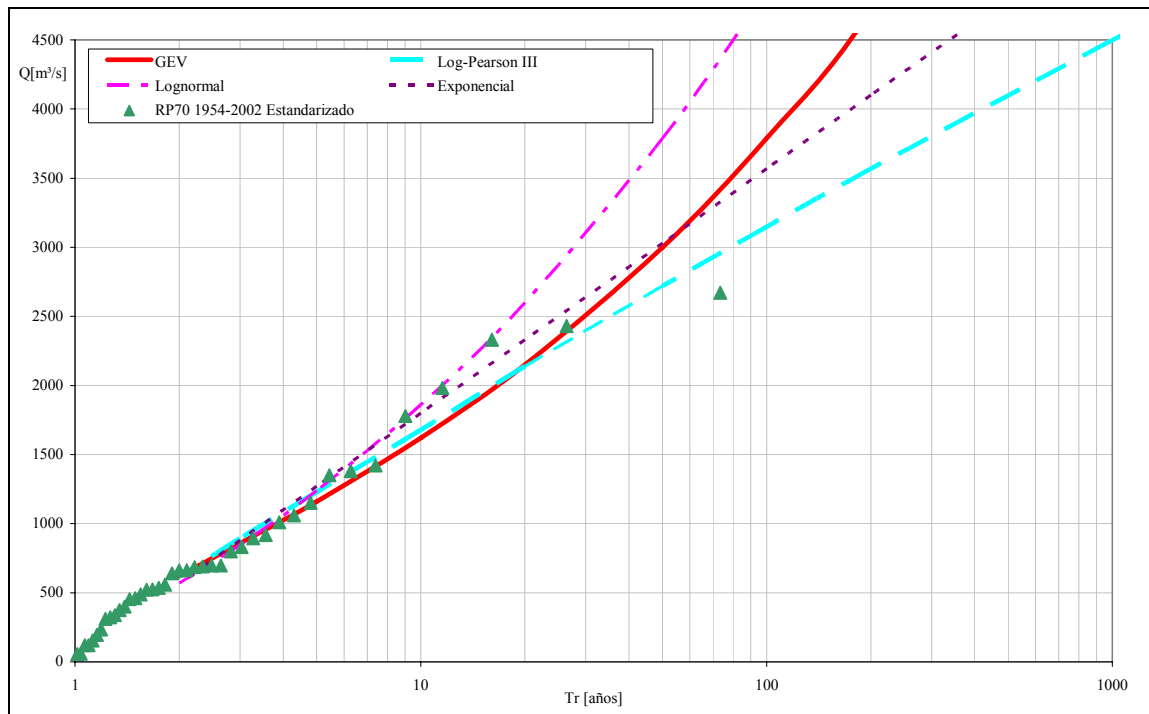


Figura X. 18. Gráfica de los ajustes sobre la serie estandarizada 1954-2002 (RP70), con Tr asociada.

La aplicación del método MLE a la serie 1954-2002 estandarizada condujo al siguiente ajuste de parámetros (distribución GEV):

$$\hat{\xi} = 474,16 \quad \hat{\alpha} = 377,198 \quad \text{y} \quad \hat{\kappa} = -0,256$$

Tal como se analiza en el apartado 3 de este Anexo, se observa la obtención de un valor del parámetro $\hat{\kappa}$ que resulta razonable en un análisis de crecidas.

4.1.3 Análisis de Frecuencia con Base en la Serie Moderna 1972-2002

Los estudios de frecuencia realizados sobre esta serie fueron similares a aquellos indicados previamente. En este caso fueron ajustadas las distribuciones GEV, Pearson III, Log Pearson III, Log Normal y Exponencial.

El método empleado para el ajuste de parámetros de las distribuciones fue el de máxima verosimilitud (MLE), a excepción de la distribución Pearson III, la cual fue ajustada por el método de los momentos. El empleo de este último método se debió al interés de comparar los resultados con trabajos preexistentes basados en esta serie. En el caso de la distribución Exponencial el paquete HYFRAN[©] solo permitió el ajuste de la misma por medio del método MLE.

Cabe señalar que el empleo del método MLE fue considerado solamente con fines comparativos, dados los inconvenientes ya citados en el apartado anterior, referidos a las limitaciones que presenta este método al aplicárselo en series cortas (en este caso la serie posee apenas 23 datos debido a la existencia de faltantes de información).

En la Tabla X. 11 se presentan los resultados correspondientes a la serie moderna 1972-2002. Del mismo modo que fuera indicado precedentemente, cabe recordar que para determinadas distribuciones y algunos valores de recurrencia, T_r , el paquete HYFRAN[©] no ofrece los valores correspondientes al intervalo de confianza del 5%. Esta situación se refleja en las Tablas con la indicación "N/D". Las Figuras indicadas se diferencian en la escala de frecuencia empleadas (abscisas), permitiendo contar con dos visiones del ajuste de cada distribución.

Del análisis gráfico se desprende que la totalidad de las distribuciones ensayadas podrían explicar el comportamiento de la serie moderna analizada.

Sin embargo, el empleo del test de adecuación χ^2 indicó el rechazo de la hipótesis de adecuación H_0 para las distribuciones Pearson III y Exponencial para un nivel de significancia del 5 %. Por el contrario, dicho test aceptó la hipótesis de adecuación de las restantes distribuciones consideradas. Considerando un nivel de significancia del 1 % el test aceptó la totalidad de las adecuaciones ensayadas.

Cabe señalar que estos dos modelos fueron adoptados por Paoli et al. (2003) para la representación estadística de esta serie 1972-2002. Los autores concluyeron acerca de la

conveniencia de la adopción de estas distribuciones debido al hecho de ambas proveer valores físicamente razonables para altos tiempos de retorno. Si bien dicha consideración es acertada, en este estudio las distribuciones Pearson III y Exponencial fueron desconsideradas como representativas de la serie 1972-2002 debido a los siguientes motivos:

- (i) el test de adecuación χ^2 permitió deducir que no es posible concluir que la muestra provenga de las distribuciones Pearson III y Exponencial.
- (ii) los caudales marcadamente altos deducidos de la consideración de las restantes distribuciones se corresponden, en realidad, con probabilidades de ocurrencia extremadamente bajas;
- (iii) la planicie de inundación natural del río Salado posee una capacidad máxima para escurrir caudales del orden de 11.000 m³/s (aunque probablemente la ocurrencia de caudales de esta magnitud se asocie a períodos geológicos más húmedos, en los cuales podrían haber existido, inclusive, cambios sustanciales en la conformación de la cuenca de aporte del río).

Un aspecto adicional que pesó negativamente a la hora de considerar esta serie moderna en el análisis de frecuencia fue lo reducido de la muestra (n = 23 años) para tratar de establecer los cuantiles de valores extremos máximos.

Tabla X. 11. Resultados de los ajustes realizados sobre la serie 1972-2002

Función de Ajuste		GEV		Pearson III		Log-Pearson III		Lognormal		Exponencial	
Tr [años]	q	Q _T	Intervalo	Q _T	Intervalo	Q _T	Intervalo	Q _T	Intervalo	Q _T	Intervalo
10000	0.9999	19900	N/D	6070	N/D	11100	N/D	8860	N/D	6380	3780 - 8990
2000	0.9995	11400	N/D	5050	N/D	7670	N/D	6530	N/D	5290	3140 - 7440
1000	0.9990	8920	N/D	4610	N/D	6480	N/D	5660	1620 - 9690	4820	2870 - 6770
200	0.9950	4990	N/D	3590	N/D	4240	N/D	3920	1500 - 6340	3720	2230 - 5220
100	0.9900	3860	N/D	3150	N/D	3470	N/D	3280	1410 - 5150	3250	1960 - 4550
50	0.9800	2960	N/D	2700	N/D	2790	N/D	2700	1290 - 4110	2780	1680 - 3880
20	0.9500	2050	840 - 3250	2120	1110 - 3130	2030	928 - 3140	2020	1100 - 2930	2160	1320 - 3000
10	0.9000	1510	830 - 2200	1670	1010 - 2330	1540	880 - 2200	1560	938 - 2180	1690	1040 - 2330
5	0.8000	1080	706 - 1460	1220	711 - 1730	1110	724 - 1500	1140	749 - 1530	1220	770 - 1670
3	0.6667	809	565 - 1050	895	433 - 1360	831	568 - 1100	848	589 - 1110	871	566 - 1180
2	0.5000	607	432 - 781	610	230 - 991	611	422 - 799	624	442 - 806	596	402 - 789

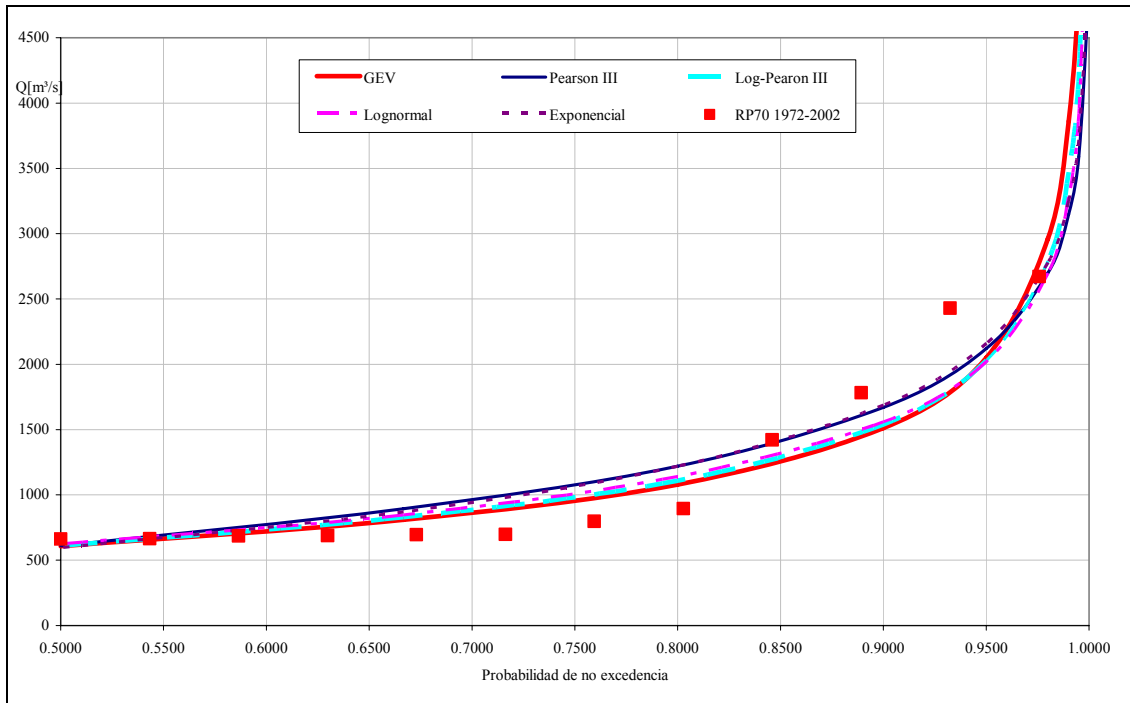


Figura X. 19. Grafica de los ajustes realizados sobre la serie 1972-2002 (RP70)

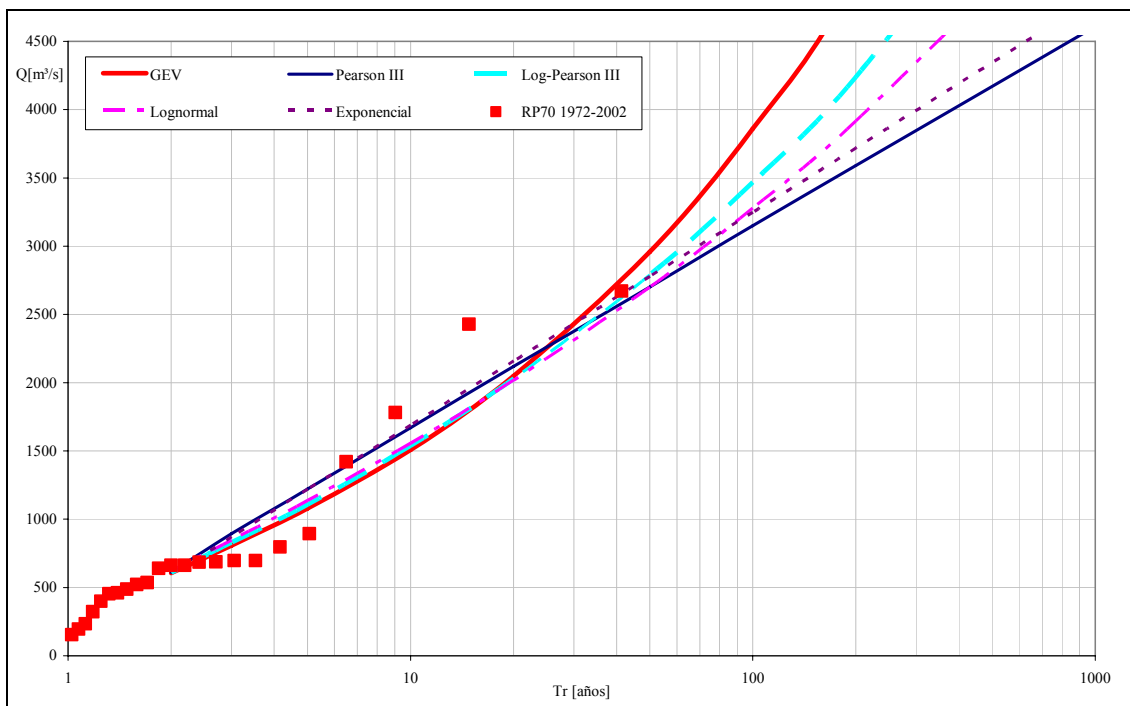


Figura X. 20. Grafica de los ajustes realizados sobre la serie 1972-2002, con recurrencia asociada (RP70).

La aplicación del método MLE a la serie corta 1972-2002 condujo al siguiente ajuste de parámetros de la distribución GEV:

$$\hat{\xi} = 488,15 \quad \hat{\alpha} = 304,07 \quad \text{y} \quad \hat{\kappa} = -0,339$$

Tal como se analiza en el apartado 3 de este Anexo, se observa un valor inferior a -0,3 para $\hat{\kappa}$, lo cual resulta sospechoso y probablemente tendencioso, generado por las limitaciones del método MLE al aplicárselo a una serie muy corta.

4.2 Análisis de Frecuencia sobre Series Reconstruidas y Sistemáticas (R+S)

4.2.1 Análisis de Frecuencia con Base en la Serie 1928-2002

De acuerdo a lo indicado en el Anexo IX, las informaciones disponibles de la Dirección Nacional de Construcciones Portuarias y Vías Navegables (DNCPyVN) correspondientes a la estación Empalme San Carlos (ESC) fueron empleadas para estimar la serie de caudales del período 1928-1953 en la Ruta Provincial 70 (RP70). Este trabajo se efectuó mediante aplicación del modelo de flujo gradualmente variado descrito en el Anexo IX. El objetivo de la reconstrucción de dicha serie fue la ampliación del espectro de informaciones disponibles sobre la estación RP70, con vistas a su incorporación en el análisis de marcas históricas y umbrales de percepción descritos en el apartado siguiente. Esa metodología estadística exige disponer del orden de magnitud de los caudales máximos históricos anteriores (o posteriores) a la serie sistemática, de modo de poder concluir acerca del número y/o magnitud de crecidas que superaron un determinado umbral de caudales máximos.

Los estudios efectuados, basados en un riguroso análisis de todas las informaciones disponibles, permitieron estimar los caudales máximos del período 1928-1953 con un grado de incertidumbre compatible con el requerido en el análisis estadístico. Por lo tanto, ello dio pie a la conformación de la serie 1928-2005, la cual posee 77 datos en lugar de los 41 datos de la serie sistemática originalmente analizada. Cabe consignar que esta serie posee también algunos vacíos de información en las décadas del 40 y 50, en años donde no fue posible definir con el nivel de precisión deseado la serie de caudales diarios.

Los estudios que a continuación se presentan, basados en una combinación de las series reconstruida y sistemática (R+S) fueron un paso intermedio hacia el análisis de frecuencia combinando de series sistemáticas y de marcas históricas (H+S)

Estadística descriptiva

Al igual que en los casos anteriores, mediante la observación de la serie histórica de valores se puede observar el salto en la media entre los periodos 1928-1972 y 1972-2005. En este caso la diferencia en la media es menor ya que alcanza el 55 % entre ambos periodos, en lugar del 65 % observado en la serie sistemática. La explicación radica en el hecho del período 1954-1972 corresponder a un ciclo más seco en relación a los restantes períodos analizados.

Tabla X. 12. Estadística Descriptiva de la serie completa

	1928-2002	Estandarizada
Media	512.5	612.6
Error típico	58.6	61.4
Mediana	401.0	538.8
Moda	467.4	832.5
Desviación estándar	461.1	483.2
Varianza de la muestra	212578.0	233507.7
Curtosis	11.5	6.9
Coefficiente de asimetría	3.0	2.2
Rango	2600.1	2600.1
Mínimo	71.9	71.9
Máximo	2672	2672
Suma	31777.63456	37983.79414
Cantidad de datos	62	62

Detección de tendencias en la serie

La Figura X. 21 ilustra acerca de los análisis realizados. En ella se observa la tendencia ya citada, de crecimiento de los caudales a partir de la década del 70. En este caso la pendiente de la recta que indica la tendencia es menor en relación al caso basado simplemente en la serie sistemática 1954-2002, hecho derivado de las razones ya apuntadas (ciclo seco 1954-1972).

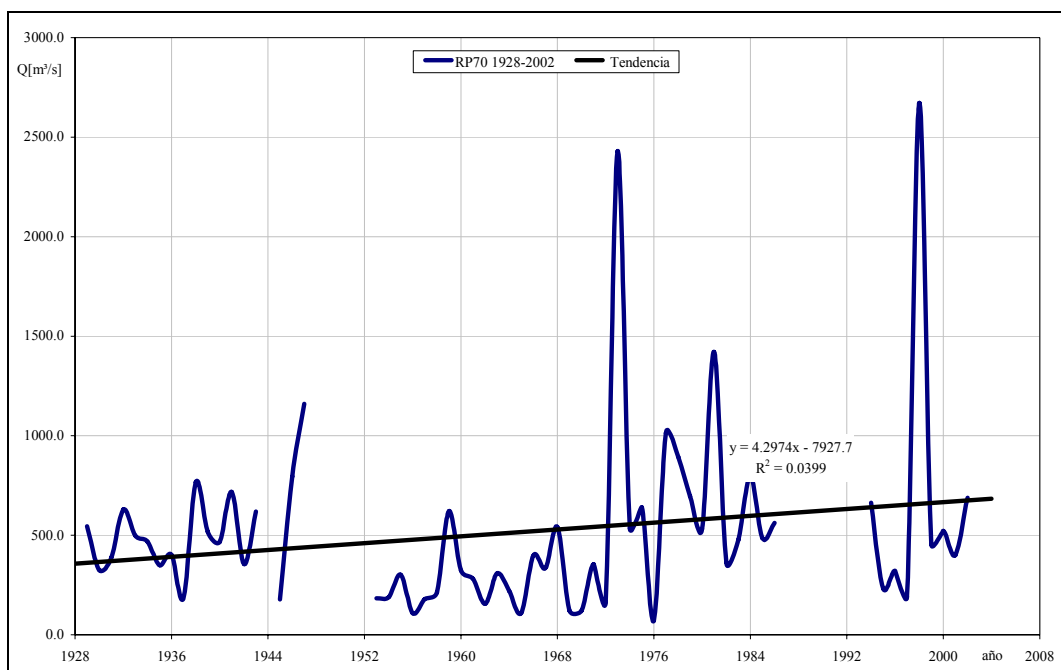


Figura X. 21. Gráfico de tendencia en las series de caudales máximos RP70 (Período 1928-2002)

Análisis de valores dudosos (outliers)

Procediendo de forma similar a la realizada sobre la serie sistemática, se verificó que todos los puntos de la serie 1928-2002 (serie R+S) cayeron dentro de la banda de confianza dentro de la cual puede suponerse que los datos no son dudosos (o “outliers”). Los resultados se presentan en la Figura X. 22

Por su parte, en la Figura X. 23 se presenta el análisis de datos dudosos (o “outliers” evidentes) efectuado sobre la serie 1928-2002 estandarizada. La descripción de esta serie se efectúa en los apartados posteriores.

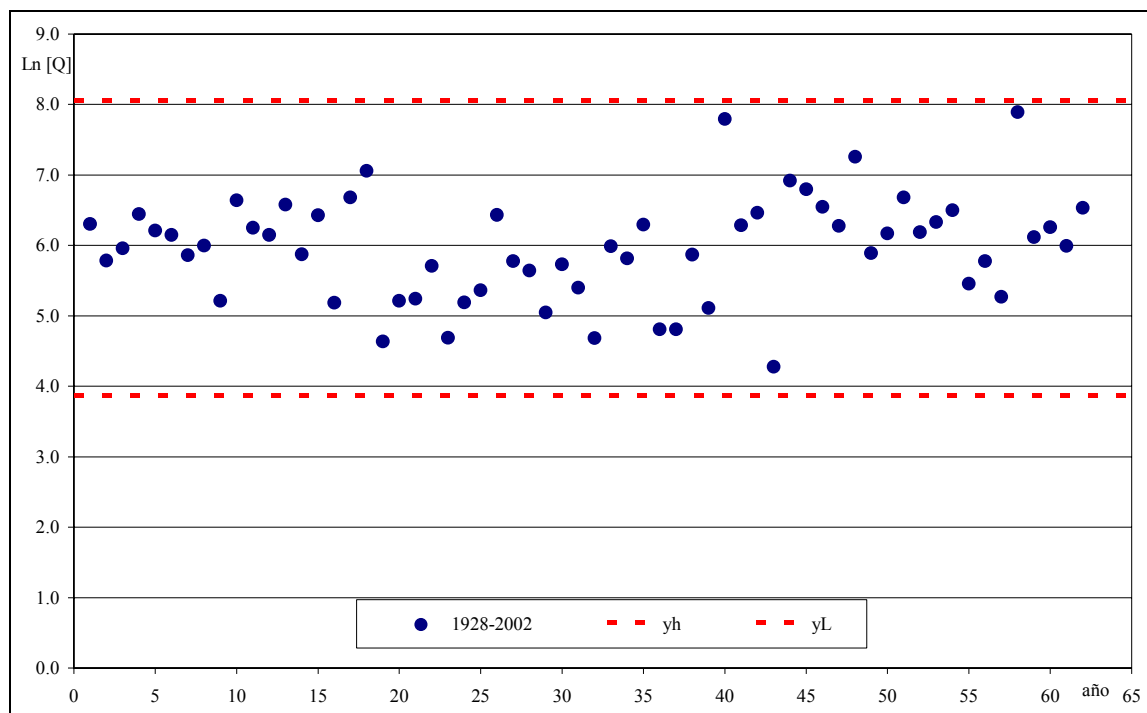


Figura X. 22. Prueba de datos dudosos en RP70 (periodo: 1928-2002)

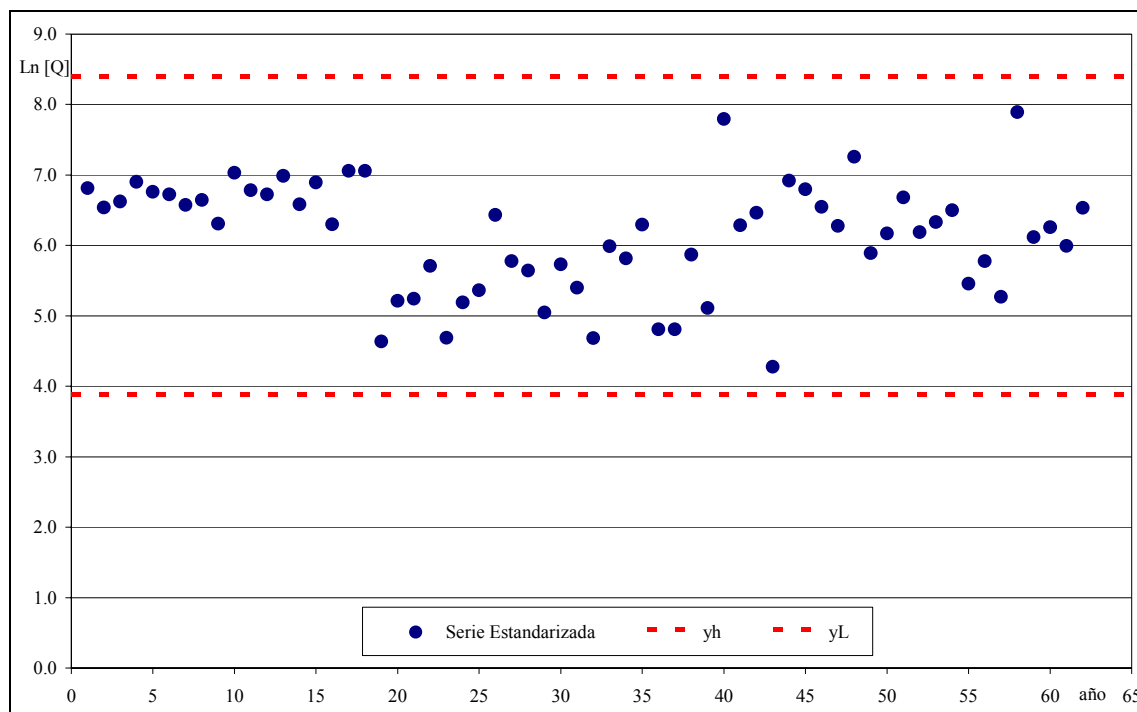


Figura X. 23. Prueba de datos dudosos en RP70 (periodo: 1928-2002 serie corregida por la media).

Análisis de pruebas paramétricas

Del mismo que fuera tratada la serie sistemática, se procedió a la aplicación de la totalidad de tests paramétricos indicados en la primera parte de este Anexo. En las Tabla X. 13, Tabla X. 14 y Tabla X. 15 se presentan los resultados obtenidos.

Del análisis de los resultados se desprende que la serie R+S disponible presenta un salto que es significativo desde el punto de vista estadístico a partir del año 1972. En este caso el mismo se registra en las medias. Sin embargo, para ratificar los resultados de los tests paramétricos fueron aplicados también en este caso los tests no paramétricos. Los resultados obtenidos se describen a continuación.

Tabla X. 13. Prueba t, suponiendo varianzas iguales
 (periodo 1928-2002)

	1928-1972	1972-2002
Media	377.1	742.2
Varianza	54170.5	408208.7
Observaciones	39	23
Varianza agrupada	183984.5	
Diferencia hipotética de las medias	0	
Grados de libertad	60	
Estadístico t	-3.237	
Significancia (α)	0.050	
P(T<=t) una cola	0.001	
Valor crítico de t (una cola)	1.671	
Significancia (α)	0.050	
P(T<=t) dos colas	0.002	
Valor crítico de t (dos colas)	2.000	
Hipótesis H0	Se rechaza	

Tabla X. 14. Prueba t', suponiendo varianzas desiguales
 (periodo 1928-2002)

	1928-1972	1972-2002
Media	377.1	742.2
Varianza	54170.5	408208.7
Observaciones	39	23
Diferencia hipotética de las medias	0	
Grados de libertad	25	
Estadístico t	-2.639	
Significancia (α)	0.050	
P(T<=t) una cola	0.007	
Valor crítico de t (una cola)	1.708	
Significancia (α)	0.050	
P(T<=t) dos colas	0.014	
Valor crítico de t (dos colas)	2.060	
Hipótesis H0	Se rechaza	

Tabla X. 15. Prueba F para varianzas de dos muestras

	1928-1972	1972-2002
Media	377.1	742.2
Varianza	54170.5	408208.7
Observaciones	39	23
Grados de libertad	38	22
F	0.133	
Significancia (α)	0.050	
P(F<=f) una cola	0.000	
Valor crítico para F (una cola)	0.547	
Hipótesis H0	Se rechaza	

Análisis de pruebas no paramétricas

En las Tabla X. 16, Tabla X. 17 y Tabla X. 18, se presentan los resultados alcanzados por la aplicación de los tests no paramétricos descriptos en la primera parte de este Anexo sobre la serie 1928-2005. Del análisis de estos resultados se concluyó acerca de la necesidad de estandarizar la serie en el periodo 1928-1972, pero en este caso considerando solamente una estandarización en la media (remoción del salto entre las medias de los periodos 1928-1972 y 1972-2002) debido a que el test de estacionalidad de Kendall no detectó un salto de las varianzas que sea significativo desde el punto de vista estadístico.

Tabla X. 16. Prueba de Independencia (Wald-Wolfowitz)

	1928-2002	Estandarizada
Valor del estadístico U	0.22	0.62
Valor de significancia (α)	0.05	0.05
Valor probabilidad (p)	0.828	0.538
Hipótesis H0	Se Acepta	Se Acepta

Tabla X. 17. Prueba de Estacionalidad (Kendall)

	1928-2002	Estandarizada
Valor del estadístico K	1.06	1.40
Valor de significancia (α)	0.05	0.05
Valor probabilidad (p)	0.291	0.161
Hipótesis H0	Se Rechaza	Se Acepta

Tabla X. 18..Prueba de Homogeneidad (Wilcoxon)

	1928-2002	Estandarizada
Valor del estadístico W	3.18	0.98
Valor de significancia (α)	0.01	0.05
Valor probabilidad (p)	0.002	0.329
Hipótesis H0	Se Rechaza	Se Acepta
División de la serie	1971-1972	1971-1972

Posiciones de ploteo (distribución empírica)

Empleando la formula de Gringorten se obtuvo la Figura X. 24. La misma permite advertir el mismo quiebre que se detectara con anterioridad, debido a que la incorporación de las nuevas estimaciones de caudal incorporó mayor información en los tramos bajos y medios, es decir, se incorporaron más eventos del tipo frecuente.

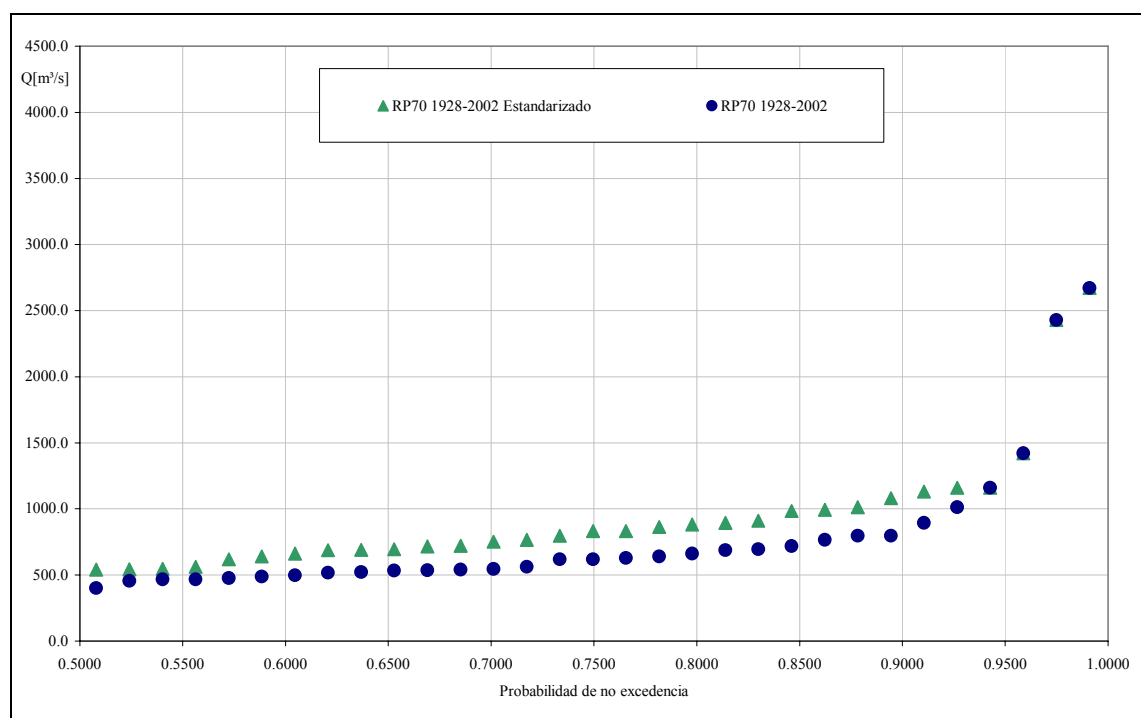


Figura X. 24. Posición de ploteo de las series analizadas de acuerdo a la formula de Gringorten

4.2.2 Análisis de Frecuencia con Base en la Serie 1928-2002 Estandarizada

En este caso fueron aplicadas las distribuciones GEV, Pearson III, Log Pearson III, Lognormal y Exponencial, empleando en la mayoría de los casos el método de ajuste MLE. La única distribución que fue ajustada por el método de los momentos fue la Pearson III. De acuerdo a lo precedentemente indicado, la adopción de este método se debió exclusivamente a los fines de comparación de resultados con estudios preexistentes.

Del mismo que fuera indicado precedentemente, cabe recordar que para determinadas distribuciones y algunos valores de recurrencia T_r , el paquete HYFRAN[®] no ofrece los valores correspondientes al intervalo de confianza del 5%. Esta situación se refleja en las Tablas con la indicación “N/D”. Las Figuras indicadas se diferencian en la escala de frecuencia empleadas (abscisas), permitiendo contar con dos visiones del ajuste de cada distribución.

En la Tabla X. 19 y en la Figura X. 25 y Figura X. 26 se presentan los resultados obtenidos sobre la serie estandarizada 1928-2002.

Del empleo del test de adecuación χ^2 para las distribuciones ajustadas se obtuvo la aceptación de las mismas con una significancia del 5 %, con excepción de las distribuciones Exponencial y Pearson III, para la cual se rechazó la hipótesis H_0 de adecuación del ajuste.

Cabe señalar que empleando un nivel de significancia del 1 % fue posible obtener la aceptación de todas las hipótesis de adecuación en relación a las distribuciones ensayadas. Por lo tanto, se determinó que no era posible concluir si la serie 1928-2002 estandarizada provenía de una distribución del tipo Exponencial y/o Pearson III.

En términos gráficos se observa un comportamiento similar de las distribuciones ensayadas para bajos tiempos de retorno, aunque se destaca nuevamente la distribución GEV como aquella con mayores cualidades para representar a la distribución empírica. Cabe destacar, sin embargo, la presencia del valor correspondiente al año 1973, el cual se aparta de las tendencias analíticas seguidas por las distribuciones teóricas.

Tabla X. 19. Resultados de los ajustes realizados sobre la serie estandarizada 1928-2002

Función de Ajuste		GEV		Pearson III		Log-Pearson III		Lognormal		Exponencial	
T_r [años]	q	Q_T	Intervalo	Q_T	Intervalo	Q_T	Intervalo	Q_T	Intervalo	Q_T	Intervalo
10000	0.9999	9030	N/D	4760	N/D	5310	N/D	8830	3890 - 13800	5130	3860 - 6390
2000	0.9995	6010	N/D	3930	N/D	4220	N/D	6280	3120 - 9450	4240	3190 - 5290
1000	0.9990	5010	N/D	3570	N/D	3780	N/D	5360	2800 - 7920	3860	2910 - 4810
200	0.9950	3220	N/D	2750	1290 - 4200	2810	N/D	3570	2090 - 5040	2980	2250 - 3700
100	0.9900	2620	1310 - 3940	2400	1270 - 3520	2410	1360 - 3470	2930	1810 - 4040	2590	1960 - 3230
50	0.9800	2120	1240 - 2990	2040	1230 - 2860	2040	1300 - 2770	2360	1530 - 3180	2210	1680 - 2750
20	0.9500	1550	1070 - 2030	1580	1100 - 2060	1560	1130 - 1990	1700	1190 - 2220	1710	1300 - 2120
10	0.9000	1190	901 - 1490	1230	925 - 1540	1220	935 - 1510	1280	935 - 1620	1330	1010 - 1640
5	0.8000	875	699 - 1050	893	639 - 1150	895	705 - 1090	901	693 - 1110	948	727 - 1170
3	0.6667	658	535 - 781	656	426 - 886	665	527 - 804	650	516 - 784	667	517 - 817
2	0.5000	487	394 - 580	455	279 - 631	478	377 - 579	462	371 - 553	444	349 - 539

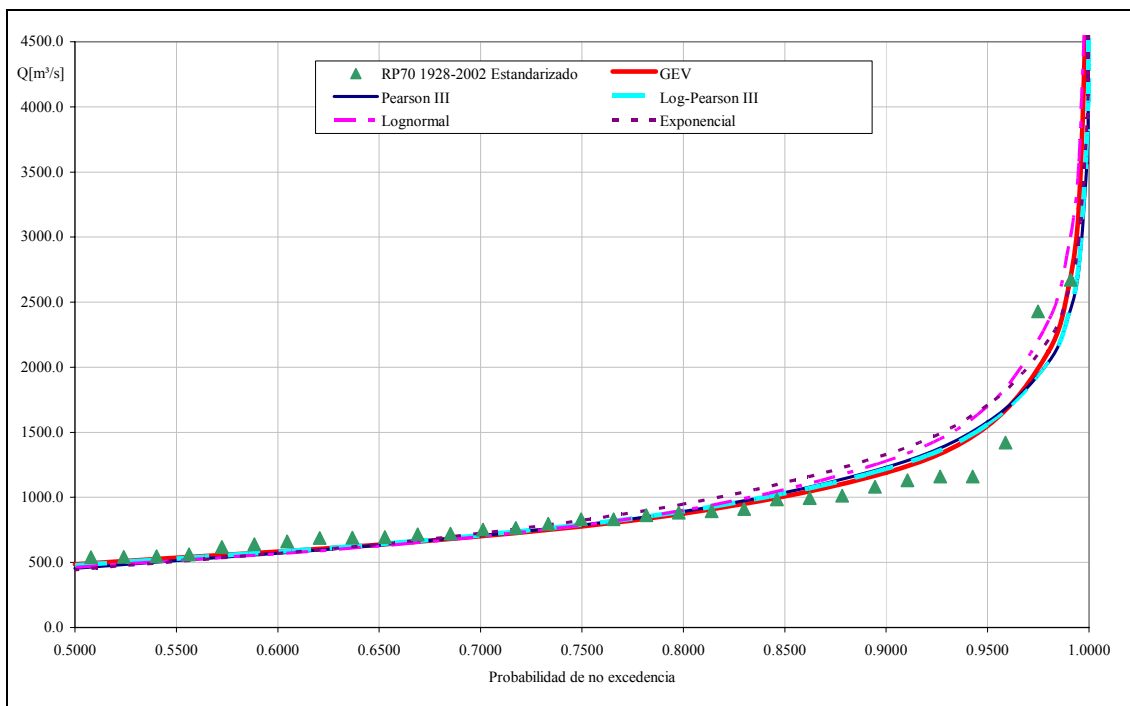


Figura X. 25. Grafica de los ajustes realizados sobre la serie estandarizada 1928-2005 (RP70)

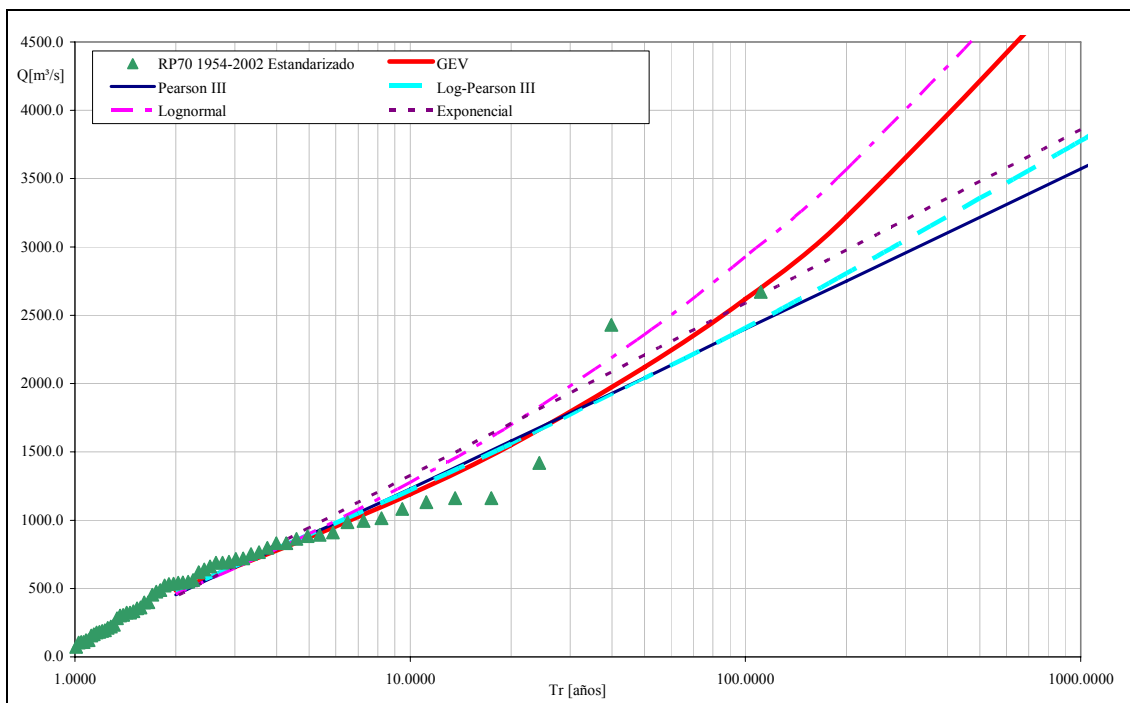


Figura X. 26 Grafica de los ajustes realizados con recurrencia asociada sobre la serie estandarizada 1928-2002.

La aplicación del método MLE a la serie 1928-2002 estandarizada condujo al siguiente ajuste de parámetros de la distribución GEV:

$$\hat{\xi} = 380,93 \quad \hat{\alpha} = 276,42 \quad \text{y} \quad \hat{\kappa} = -0,227$$

De acuerdo a lo indicado en el apartado 3 de este Anexo, se observa la obtención de un valor razonable para el parámetro $\hat{\kappa}$.

4.3 Análisis de Frecuencia sobre Series Históricas y Sistemáticas (H+S)

4.3.1 Análisis de frecuencia con base en la serie 1875-2002

Para la realización de este estudio fueron empleadas todas las informaciones históricas que fue posible recopilar, relacionar y deducir durante el desarrollo de este estudio. De acuerdo a lo indicado en el Anexo IX, fue posible reconstruir las principales marcas históricas del período previo a la existencia de la serie sistemática desde el año 1875. Ello permitió tener una visión general del comportamiento del río Salado en relación con sus principales marcas históricas dentro del período 1875-2005. En este apartado se presentan los estudios basados en la serie 1875-2002, es decir, basados en todas las informaciones disponibles con anterioridad a la crecida de abril/mayo de 2003. Posteriormente, en el apartado correspondiente, se analizan los resultados de los estudios estadísticos basados en la totalidad de las informaciones disponibles hasta la actualidad (agosto 2005).

Conforme citado en el Anexo IX, las dos principales marcas históricas consideradas correspondieron al período 1875-1927. Más precisamente, las mismas correspondieron a los años 1886 (caudal máximo de la crecida estimado en el orden de 1.750 m³/s) y 1914 (caudal máximo estimado del orden de 2.750 m³/s).

La metodología empleada para el análisis de las marcas históricas fue la propuesta, respectivamente, por Martins y Stedinger (2001) para el tratamiento estadístico de informaciones históricas con series anuales y parciales y por Reis y Stedinger (2005) para el tratamiento *bayesiano* de la información. La aplicación siguió el tratamiento indicado en el apartado 3 de este Anexo.

Dado el interés por estimar los cuantiles correspondientes a altos tiempos de retorno (valores extremos de caudales), el umbral de percepción fue fijado en 1.700 m³/s. La adopción de este umbral permitió considerar las dos marcas históricas estimadas durante el desarrollo del estudio, como así también la serie sistemática de caudales 1972-2002.

4.3.2 Datos empleados en el análisis

Los datos considerados en este análisis fueron las dos marcas históricas antes indicadas (período 1875-1927), los caudales máximos reconstruidos (período 1928-1953) y los caudales observados de manera sistemática (período 1954-2002). Los mismos se indican en la Tabla X. 20.

Tabla X. 20. Serie completa de caudales máximos anuales empleada en los estudios estadísticos.

Año Hidrológico	Q [m³/s]	Tipo de dato	Año Hidrológico	Q [m³/s]	Tipo de dato	Año Hidrológico	Q [m³/s]	Tipo de dato	Año Hidrológico	Q [m³/s]	Tipo de dato
1886/1887	1750.0	H	1946/47	1160.8	R	1966-1967	334.9	S	1986-1987		
1913/1914	2759.0	H	1947/48			1967-1968	541.9	S	1987-1988		
1928/29	546.2	R	1948/49	103.0	R	1968-1969	122.4	S	1988-1989		
1929/30	325.3	R	1949/50			1969-1970	122.3	S	1989-1990		
1930/31	387.1	R	1950/51			1970-1971	354.1	S	1990-1991		
1931/32	629.7	R	1951/52			1971-1972	165.5	S	1991-1992		
1932/33	498.1	R	1952/53	183.3	R	1972-1973	2429.0	S	1992-1993		
1933/34	467.4	R	1953/54	189.15	R	1973-1974	535.6	S	1993-1994	662.9	S
1934/35	350.7	R	1954-1955	301.9	S	1974-1975	640.3	S	1994-1995	234.4	S
1935/36	402.0	R	1955-1956	108.7	S	1975-1976	71.9	S*	1995-1996	322.6	S
1936/37	183.3	R	1956-1957	179.5	S	1976-1977	1013.9	S*	1996-1997	194.7	S
1937/38	766.7	R	1957-1958	213.3	S	1977-1978	893.4	S	1997-1998	2672.0	S
1938/39	518.1	R	1958-1959	619.9	S	1978-1979	696.0	S	1998-1999	454.3	S
1939/40	467.4	R	1959-1960	322.8	S	1979-1980	532.7	S*	1999-2000	521.5	S
1940/41	717.8	R	1960-1961	282.5	S	1980-1981	1420.0	S	2000-2001	400.0	S
1941/42	355.8	R	1961-1962	155.7	S*	1981-1982	361.8	S*	2001-2002	688.6	S
1942/43	619.6	R	1962-1963	308.2	S*	1982-1983	477.8	S*	2002-2003	3954.0	S
1943/44			1963-1964	221.0	S	1983-1984	797.6	S	2003-2004	290.5	S
1944/45	178.6	R	1964-1965	108.1	S	1984-1985	487.5	S	2004-2005	1050.0	S
1945/46	795.9	R	1965-1966	398.9	S	1985-1986	561.7	S*			

Observaciones: H = Deducido de marca histórica; R = Caudal Estimado S = Dato Sistemático S* = Dato Sistemático corregido por remanso
 Los caudales corregidos por efectos de remanso no corresponden necesariamente al mismo día de observación del valor de caudal originalmente considerado como extremo anual. Ello es una consecuencia directa del análisis de los efectos de remanso por el sistema Paraná, resultando que no necesariamente los caudales máximos se correspondieron con los niveles máximos.

Datos Originales (sin corrección por efectos de remanso del sistema Paraná)							
1961-1962	171.5	1962-1963	406.6	1975-1976	154.2	1976-1977	1782
1979-1980	663.4	1981-1982	460.0	1982-1983	685.9	1985-1986	697.3

4.3.3 Resultados obtenidos empleando el enfoque de informaciones históricas

Cuantiles de Crecida Estimados:

Las Tabla X. 21, Tabla X. 22, Tabla X. 23 y Tabla X. 24 presentan los parámetros estimados, los cuantiles de crecidas y los intervalos de confianza para estos cuantiles, asociados a varios tiempos de retorno para los modelos GEV/ML, GEV/GML, LN/ML y LPIII/Abordaje bayesiano, respectivamente. Las referidas Tablas muestran los parámetros y los cuantiles identificados a partir del uso de registros sistemáticos solamente, como así también con el uso adicional de información histórica (casos BCD y CD).

Comparando la Tabla X. 21 (GEV/ML) y la Tabla X. 22 (GEV/GML) se puede observar que la metodología de máxima verosimilitud (GML) adoptada corrigió las inconsistencias del método de la máxima verosimilitud clásico (ML), a saber:

- valor sospechoso del factor de forma (κ muy negativo, lo cual implica en este caso, $\kappa < -0,36$, que el momento de orden 3, asimetría, no existe);
- límites de confianza negativos para tiempos de retorno ≥ 500 años.

La Figura X. 28 muestra que las estimativas puntuales del modelo GEV/GMLE para datos sistemáticos son superiores a aquellas que emplean la información histórica, existiendo una ganancia considerable en términos de precisión de estas estimaciones como consecuencia del uso de la información histórica. Esto puede ser fácilmente verificado si se observa la reducción de los límites de confianza de 95% asociados a las mismas. Las diferencias entre las estimaciones puntuales para los casos GEV-BCD y GEV-CD son despreciables, siendo que las ganancias en la precisión (amplitud del intervalo de confianza del 95%) derivadas del uso de la información CD en lugar de la BCD, resultan pequeñas dado que el umbral elegido fue relativamente alto ($1.700 \text{ m}^3/\text{s}$).

En el caso de la distribución LPIII/Abordaje bayesiano (Figura X. 30) las diferencias entre las estimaciones puntuales con y sin el uso de la información histórica son más significativas, ocurriendo problemas en las estimaciones de cuantiles de crecidas con información sistemática para elevados tiempos de retorno. Esto último se evidencia cuando se analizan los límites de confianza de 95 % para las estimaciones realizadas. Las estimaciones puntuales LPIII-BCD y LPIII-CD son bastante próximas, pero en este caso las ganancias en la precisión debido al uso de datos tipo CD no son despreciables para elevados tiempos de retorno.

La Figura X. 31 compara los resultados de los dos modelos que emplean la información histórica (Caso CD), GEV/GML-CD y LPIII/Abordaje bayesiano-CD. Esta Figura muestra que las estimaciones puntuales para los dos modelos son bastante próximas. Sin embargo, la precisión de la estimativa GEV/GML es mayor, en especial para elevados Tr.

Tabla X. 21. Resultados obtenidos a partir del Modelo GEV/MLE (informaciones hasta el año 2002).

Tr [años]	Parám.	Sistemático			BCD			CD		
	ξ	453.9			440.6			443.9		
	α	298.6			267.7			269.8		
	κ	-0.36			-0.23			-0.22		
	Probab.	2.5%	moda	97.5 %	2.5 %	Moda	97.5%	2.5 %	moda	97.5%
5	0.8	873	1049	1225	709	920	1131	715	922	1129
10	0.9	1012	1493	1973	949	1229	1510	954	1226	1499
20	0.95	1039	2048	3057	1164	1580	1996	1177	1568	1960
50	0.98	793	3020	5248	1370	2130	2889	1412	2098	2785
100	0.99	261	3997	7734	1441	2625	3810	1522	2572	3622
500	0.998	-3164	7471	18106	1179	4126	7072	1440	3983	6527
1000	0.999	-6261	9714	25689	795	4962	9129	1191	4759	8327

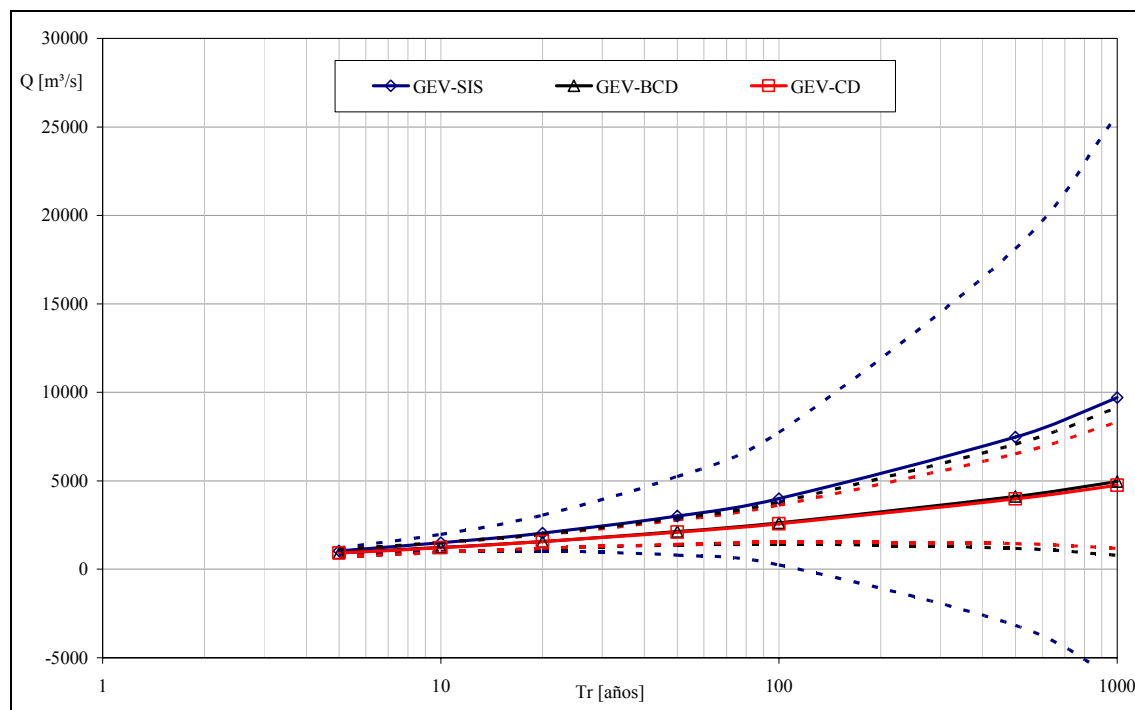


Figura X. 27. Cuantiles de crecida del modelo GEV/MLE y sus límites de confianza del 95%, asociados a diferentes tiempos de retorno calculados a partir del uso de información sistemática (GEV-SIS) y del uso adicional de información histórica (GEV-BCD y GEV-CD).

Tabla X. 22. Resultados obtenidos a partir del Modelo GEV/GMLE (informaciones hasta el año 2002).

Tr [años]	Parám.	Sistemático			BCD			CD		
	ξ	478.9			459.5			455.8		
	α	310.3			282.0			280.3		
	κ	-0.20			-0.17			-0.17		
	Probab.	2.5%	moda	97.5 %	2.5 %	Moda	97.5%	2.5 %	moda	97.5%
5	0.8	728	1022	1317	843	942	1041	731	936	1141
10	0.9	938	1363	1787	1048	1235	1422	961	1227	1494
20	0.95	1141	1741	2342	1250	1554	1858	1192	1545	1897
50	0.98	1380	2321	3261	1501	2030	2559	1485	2019	2552
100	0.99	1525	2832	4140	1673	2441	3210	1684	2428	3172
500	0.998	1643	4331	7018	1957	3604	5251	2013	3585	5158
1000	0.999	1549	5141	8732	2004	4212	6421	2072	4191	6311

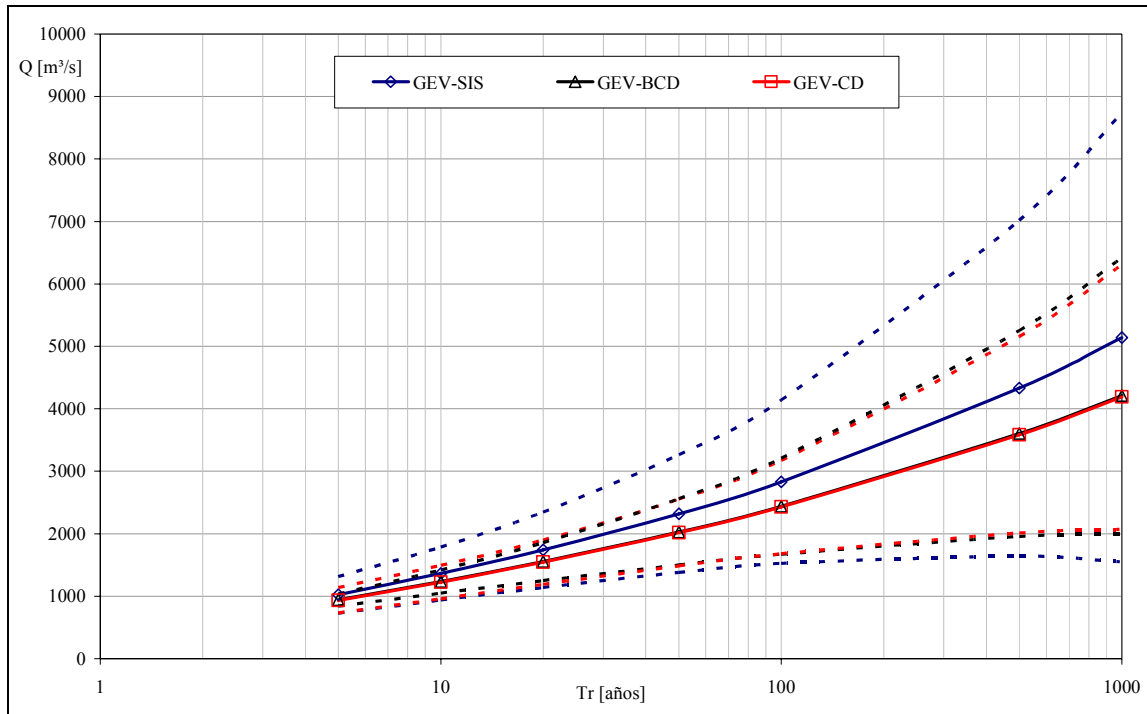


Figura X. 28. Cuantiles de crecida del modelo GEV/GMLE y sus límites de confianza del 95%, asociados a diferentes tiempos de retorno calculados a partir del uso de información sistemática (GEV-SIS) y del uso adicional de información histórica (GEV-BCD y GEV-CD).

Tabla X. 23. Resultados para el Modelo LN/MLE (informaciones hasta el año 2002).

Tr [años]	Parám.	Sistemático			BCD			CD		
	μ	6.381			6.288			6.288		
	σ	0.733			0.662			0.661		
	Probab.	2.5%	moda	97.5 %	2.5 %	Moda	97.5%	2.5 %	moda	97.5%
50	0.98	1808	2661	3916	1590	2098	2768	1654	1808	2661
100	0.99	2131	3249	4954	1852	2512	3406	1932	2131	3249
200	0.995	2475	3901	6151	2127	2964	4130	2224	2475	3901
333	0.997	2742	4425	7143	2338	3322	4719	2447	2742	4425
1000	0.999	3360	5688	9630	2816	4166	6164	2954	3360	5688

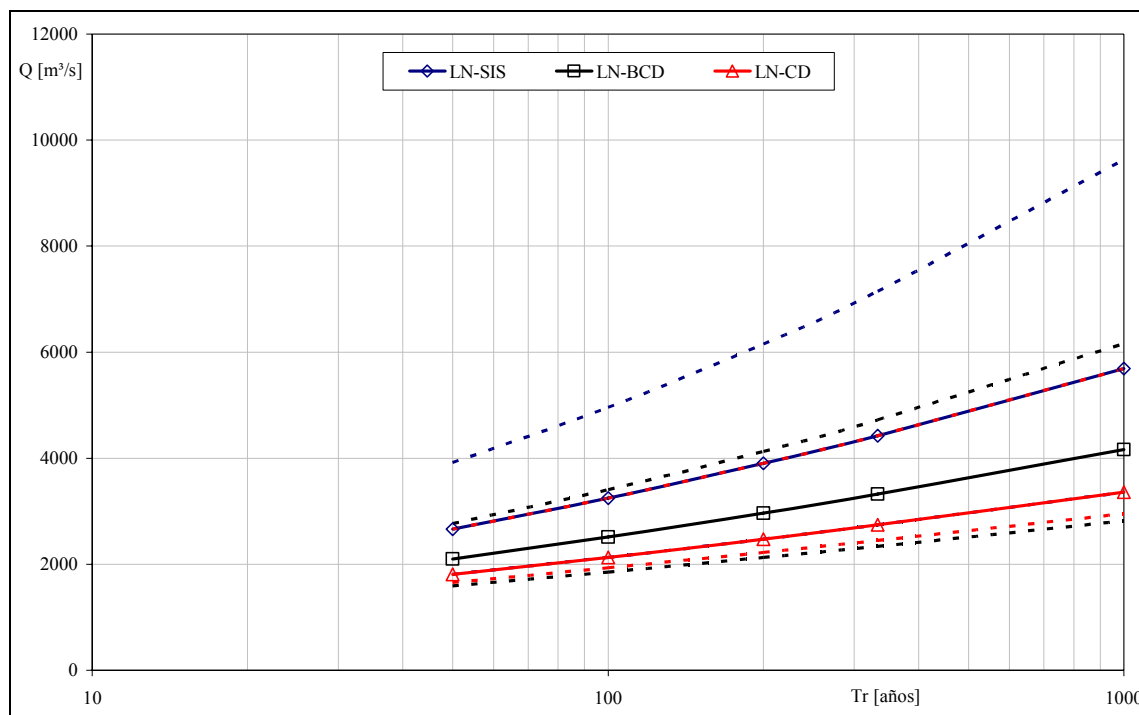


Figura X. 29. Cuantiles de crecida LN/MLE y sus límites de confianza de 95%, asociados a diferentes tiempos de retorno calculados a partir del uso de información sistemática (LN-SIS) y del uso adicional de información histórica (LN-BCD yLN-CD).

Tabla X. 24. Resultados para el Modelo LPIII/Abordaje bayesiano (informaciones hasta el año 2002).

Tr [años]	Parámetros	Sistemático			BCD			CD		
	μ	2.773			2.7			2.718		
	σ	0.347			0.3			0.3		
	γ	0.09			0.02			0.01		
	Probabilidad	2.50%	media	97.50%	2.50%	media	97.50%	2.50%	media	97.50%
5	0.8	807	1135	1763	731	936	1175	729	935	1172
10	0.9	1110	1608	2856	1002	1272	1626	994	1265	1599
20	0.95	1420	2141	4434	1283	1630	2204	1272	1618	2134
50	0.98	1835	2962	7654	1652	2149	3248	1640	2129	3077
100	0.99	2154	3687	11528	1930	2588	4332	1916	2554	4050
500	0.998	2834	5762	28717	2520	3774	8142	2536	3693	7413
1000	0.999	3117	6865	41104	2766	4370	10642	2779	4257	9493

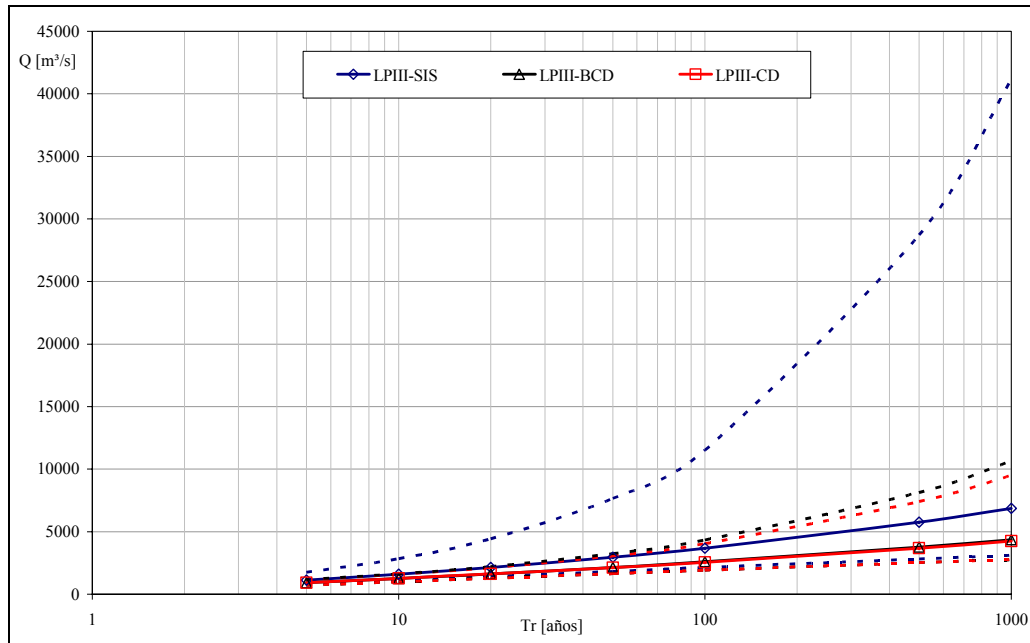


Figura X. 30. Cuantiles de crecida LPIII/Abordaje Bayesiano y sus límites de confianza de 95%, asociados a diferentes tiempos de retorno calculados a partir del uso de información sistemática (LPIII-SIS) y del uso adicional de información histórica (LPIII-BCD y LPIII-CD).

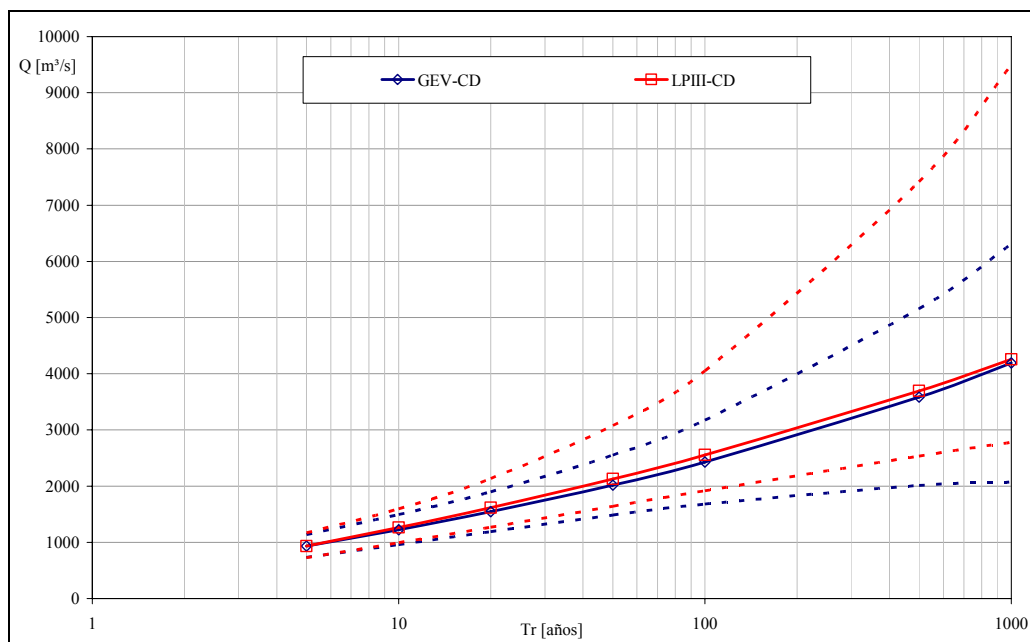


Figura X. 31 Comparación de los cuantiles de crecida GEV/GML y sus límites de confianza de 95%, con los cuantiles correspondientes y sus límites de confianza de 95% obtenidos con LPIII/Abordaje Bayesiana calculados a partir del uso de información sistemática e histórica (Caso CD).

Posiciones de Ploteo con la Presencia de Información Histórica

Supóngase que entre n muestras, donde $n = s+h$ (el período total del registro), un cierto umbral (o nivel) de percepción es excedido r veces, ya sea durante el período de información histórica ($h=86$) o sistemática ($s=24$). Así, un estimador intuitivo para la probabilidad de excedencia del umbral de percepción ($x_T = 1.700 \text{ m}^3/\text{s}$) es $q_e = r/n = 0,045$. Si los r valores que exceden este umbral son colocados en orden decreciente e indexada de la forma $i = 1, \dots, r$, las posiciones de ploteo aproximadas para las probabilidades de excedencia (q_i) dentro del intervalo $(0, q_e)$ resultan (Martins, 2005):

$$q_i = q_e \left(\frac{i-a}{r+1-2.a} \right) = \frac{r}{n} \left(\frac{i-a}{r+1-2.a} \right) \quad (\text{X. 50})$$

pudiéndose utilizar la estimación de Gringorten ($a=0,44$) para la GEV. Elecciones razonables del parámetro “ a ” se traducen en diferencias despreciables en las posiciones de ploteo resultantes. Es preciso considerar, además, que todas las posiciones de ploteo son imprecisas, en especial en la región de interés, o sea, en la zona de los extremos. Stedinger et al. (1993) y Hirsch y Stedinger (1987) realizaron un análisis de las posiciones de ploteo ante la presencia de informaciones históricas (o truncadas).

Aún resta definir las probabilidades de excedencia q_j para los caudales máximos del período sistemático que están abajo del umbral, es decir, aquellas asociadas a probabilidades de excedencia en el intervalo $(q_e, 1)$. Así, ocurriendo e ($e=3$) crecidas durante el período sistemático por encima del umbral de percepción, incluidas en r , restan entonces $s-e$ ($s-e=21$ en este caso) crecidas abajo del umbral considerado. Para estas crecidas colocadas en orden decreciente e indexadas por $j = 1, \dots, s-e$, las posiciones de ploteo aproximadas son dadas por la expresión (Martins, 2005)

$$q_j = q_e + (1-q_e) \left(\frac{j-a}{s-e+1-2.a} \right) = \frac{r}{n} + \frac{n-r}{n} \left(\frac{j-a}{s-e+1-2.a} \right) \quad (\text{X. 51})$$

La Figura X. 28 muestra los cuantiles de crecida GEV/GMLE-CD y sus límites de confianza de 95 % asociados a diferentes tiempos de retorno y los caudales observados asociados a sus respectivas probabilidades de excedencia calculadas conforme las ecuaciones X.50 y X.51, empleando las posiciones de ploteo. En la misma figura se localiza la crecida del año 2003 al solo efecto de observar la capacidad del modelo para estimar su valor. Esta crecida no fue considerada en los análisis a los efectos de emular las condiciones previas a su ocurrencia.

Tiempo de retorno de la crecida de abril/mayo de 2003

De acuerdo a los análisis presentados en el apartado 3 como también a los resultados ilustrados en las Figuras precedentes, se adoptó la combinación GEV-GMLE a los fines de estimar el tiempo de retorno de la crecida de abril/mayo de 2003, como así también sus intervalos de confianza.

En la Tabla X. 25 se presentan los resultados obtenidos al aplicarse los enfoques BCD y CD.

Tabla X. 25. Tiempos de retorno [años] correspondientes a caudales máximos del río Salado estimados mediante la distribución GEV-GMLE (informaciones hasta el año 2002).

Caudal considerado		Enfoque BCD	Enfoque CD
3.954	m ³ /s	787	810
3.010		240	246
2.500		112	115

Aunque el tiempo de retorno para la crecida de 2003 de 3.954 m³/s estimado a partir del enfoque CD sea de 810 años, vale mencionar el hecho de que el intervalo de confianza de 95 % del cuantil relativo al caudal de aproximadamente 180 años ya engloba al caudal de 3.954 m³/s (distribución GEV-GMLE).

Por otra parte, cabe también mencionar que en el caso de la distribución LPIII ajustada según un enfoque bayesiano, el caudal de 3.954 m³/s estaría próximo del límite superior del intervalo de credibilidad de 95% correspondiente al caudal de 100 años de tiempo de retorno.

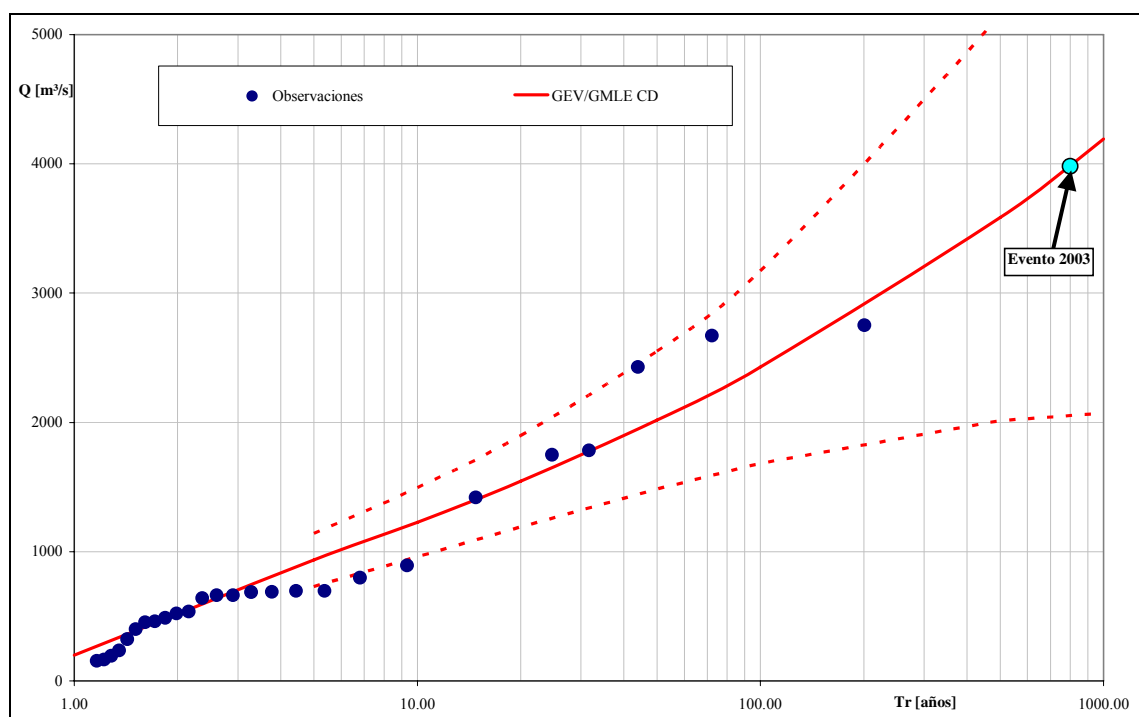


Figura X. 32. Cuantiles GEV/GML-CD de caudales máximos y sus límites de confianza de 95% asociados a diferentes tiempos de retorno y caudales observados asociados a sus respectivas probabilidades de excedencia calculadas para la serie 1875-2002 (el valor de la crecida de 2003 ha sido incorporado solo a los efectos gráficos).

5 ANÁLISIS DE FRECUENCIA SOBRE SERIES DE DATOS HASTA EL AÑO 2005

5.1 Introducción

El hecho de realizar un tratamiento estadístico detallado de las nuevas series surgidas al considerar los últimos tres años (solo 3 datos adicionales en relación a las precedentemente analizadas), se justificó por el hecho de corresponder el año 2003 a la crecida máxima *maximorum* que se registró hasta el presente en la sección RP70 dentro del período 1875-2005. Por otra parte, estos análisis permiten considerar la totalidad de los datos existentes hasta la actualidad (agosto de 2005).

El interés de la realización de estos estudios radica también en el hecho de analizar el impacto que la crecida del año 2003 posee sobre la previsibilidad estadística de los caudales extremos máximos del río Salado en la sección RP70.

A continuación se presenta un análisis similar al efectuado en el apartado precedente, con todas las series extendidas hasta el año 2005.

5.2 Análisis de Frecuencia sobre Series Sistemáticas (S)

5.2.1 Análisis de Frecuencia con Base en la Serie 1954-2005

Para la realización de este análisis se empleó la serie de caudales y niveles diarios de la Ruta Provincial 70 (RP70) correspondiente al período 1954-2005. Esta serie presenta un vacío de información en el período 1987-1993, razón por la cual la misma incluye 44 valores.

Análisis preliminar de la serie

De manera similar a lo oportunamente indicado para los estudios de las series hasta el año 2002, previo al estudio de frecuencia se realizó un análisis de la serie a los efectos de verificar sus cualidades estadísticas. Este análisis previo incluyó la determinación de los parámetros descriptivos de la serie. El estudio incluyó también el análisis de los valores dudosos (“*outliers*”) según la metodología propuesta por el U.S. Water Resources Council (1981). La misma se basa en el análisis de la asimetría de la serie analizada. Del mismo modo, fueron realizados distintos tipos de tests estadísticos de tipo paramétrico y no paramétrico para analizar las condiciones de normalidad, aleatoriedad, homogeneidad y tendencia de la serie.

Las metodologías que fueron empleadas en estos análisis se corresponden con aquellas utilizadas previamente para las series hasta el año 2002, es decir:

- Análisis de Independencia de la Serie (prueba Wald-Wolfowitz)
- Prueba t y t' para medias de dos muestras
- Prueba F para varianzas de dos muestras

- Análisis de Tendencia (prueba Kendall)
- Análisis de Homogeneidad (prueba Wilcoxon)

Los resultados de estos estudios son sintetizados a continuación.

Estadística descriptiva

En la Tabla X. 26 se presentan los principales parámetros estadísticos de la serie en estudio.

Tabla X. 26..Resumen estadístico de las series de caudales máximos (RP70, Periodo 1954-2005).

	<i>1954-2005</i>	<i>1954-1972</i>	<i>1972-2005</i>	<i>Estandarizada</i>
Media	653.7	276.4	914.9	914.9
Error típico	113.2	35.1	173.3	131.7
Mediana	430.5	251.8	663.2	663.2
Desviación estándar	750.7	148.8	883.7	873.4
Varianza de la muestra	563618.1	22140.9	780929.4	762768.2
Curtosis	9.2	0.2	5.1	2.8
Coficiente de asimetría	2.9	0.9	2.2	1.6
Rango	3845.9	511.8	3799.8	4038.8
Mínimo	108.1	108.1	154.2	-84.8
Máximo	3954.0	619.9	3954	3954.0
Suma	28763.5	4975.8	23787.68	40256.1
Cantidad de datos	44	18	26	44

De los parámetros estadísticos descriptivos presentados en la Tabla X. 26, se puede apreciar que la diferencia entre las medias del periodo 1954-1972 y 1972-2005 es del 70 %, es decir, la incorporación de los tres últimos años (y especialmente la introducción de la crecida de 2003) significó un aumento del 5 % en relación al análisis correspondiente hasta el año 2002.

Detección de tendencias en la serie

Conforme citado previamente en este Anexo, aunque el análisis gráfico carece del peso de las pruebas de tipo paramétrico y no paramétrico, el mismo representa la forma más práctica y rápida disponible para visualizar posibles tendencias de una serie. En la Figura X. 33 se presentan en forma gráfica los análisis realizados.

La tendencia observada y el cambio en las medias condujeron al análisis de dos subseries correspondientes a los períodos 1954-1972 y 1972-2005 respectivamente. En la Tabla X. 26 se presentan las principales informaciones que describen a cada una de ellas. Del mismo modo, se presentan las informaciones correspondientes a la serie 1954-2005 estandarizada tanto por la media como por la varianza.

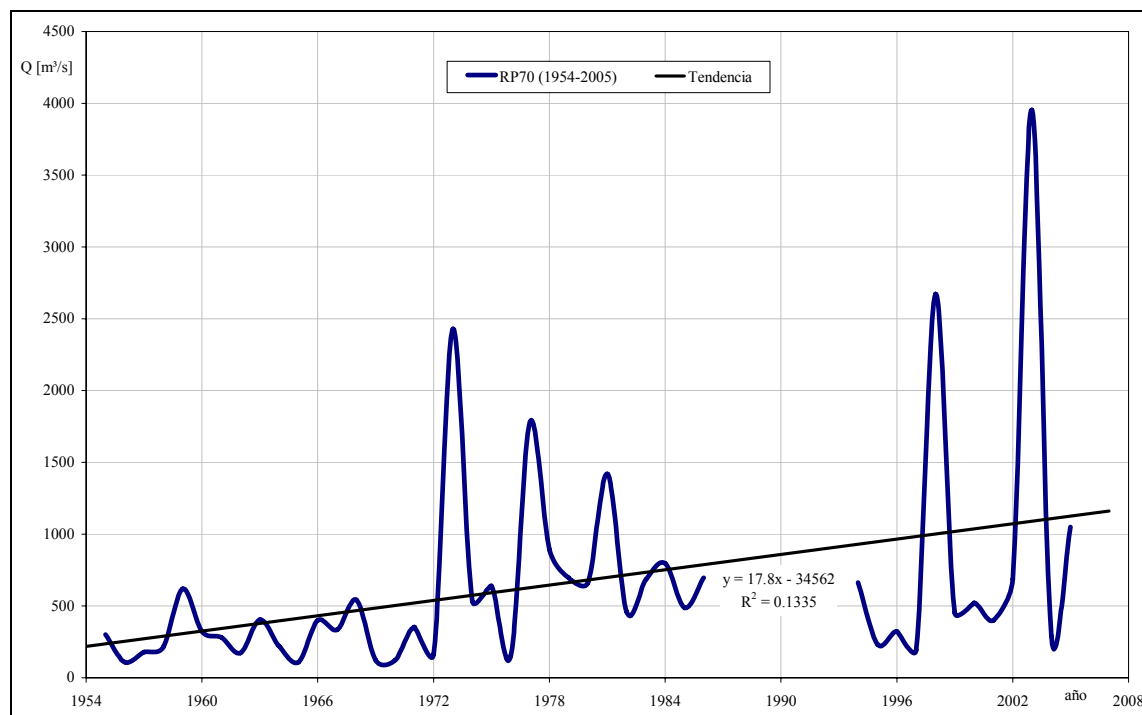


Figura X. 33. Gráfico de tendencia en las series de caudales máximos RP70 (Período 1954-2005).

Análisis de valores dudosos (outliers)

Se procedió de forma similar a la realizada sobre las series previas, de acuerdo a la metodología de Chow et al. (1995) para establecer los valores de la serie que podrían ser clasificados como datos dudosos (o “outliers” evidentes).

Las Figura X. 34, Figura X. 35 y Figura X. 36 sintetizan los resultados obtenidos correspondientes a las series 1954-2005, 1972-2005 y 1954-2005 estandarizada por la media y la varianza. En las mismas se observa que en todos los casos los puntos se encontraron dentro de la faja establecida por los límites superior (yh) e inferior (yL) respectivamente. Una excepción lo constituye la serie 1954-2005 estandarizada en donde el test indica que el mínimo caudal máximo de la serie podría ser considerado como un “outlier”. Sin embargo, este aspecto carece de interés práctico para el presente estudio, dado el objetivo de estudio sobre el otro extremo de la serie.

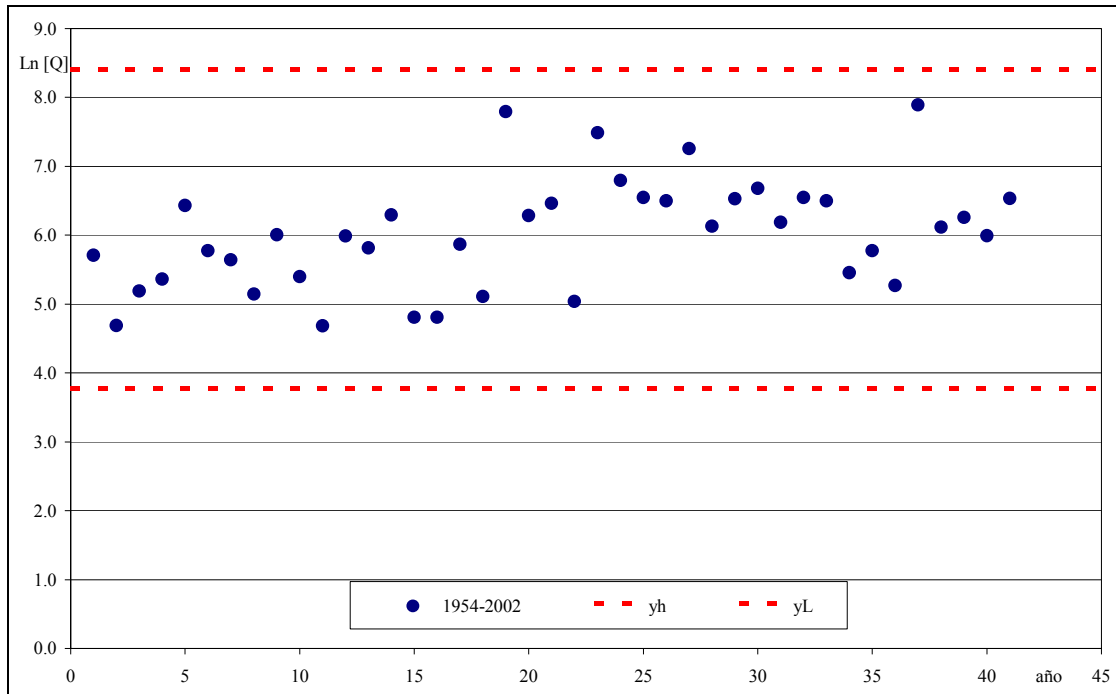


Figura X. 34. Prueba de datos dudosos en RP70 (periodo: 1954-2005)

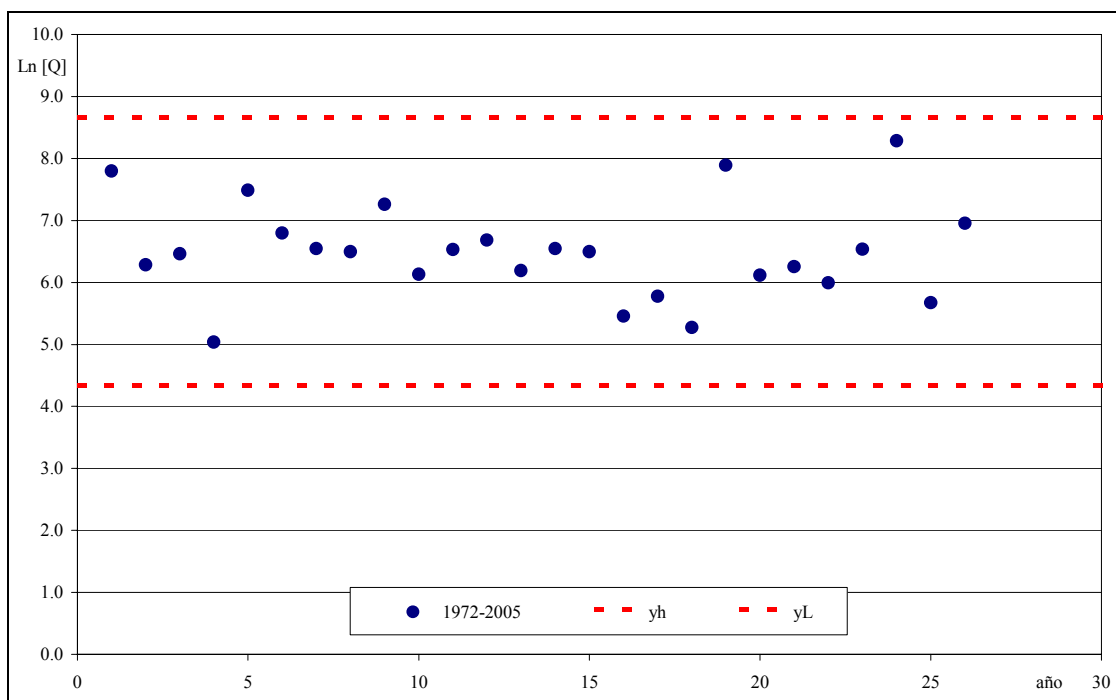


Figura X. 35. Prueba de datos dudosos en RP70 (periodo: 1972-2005)

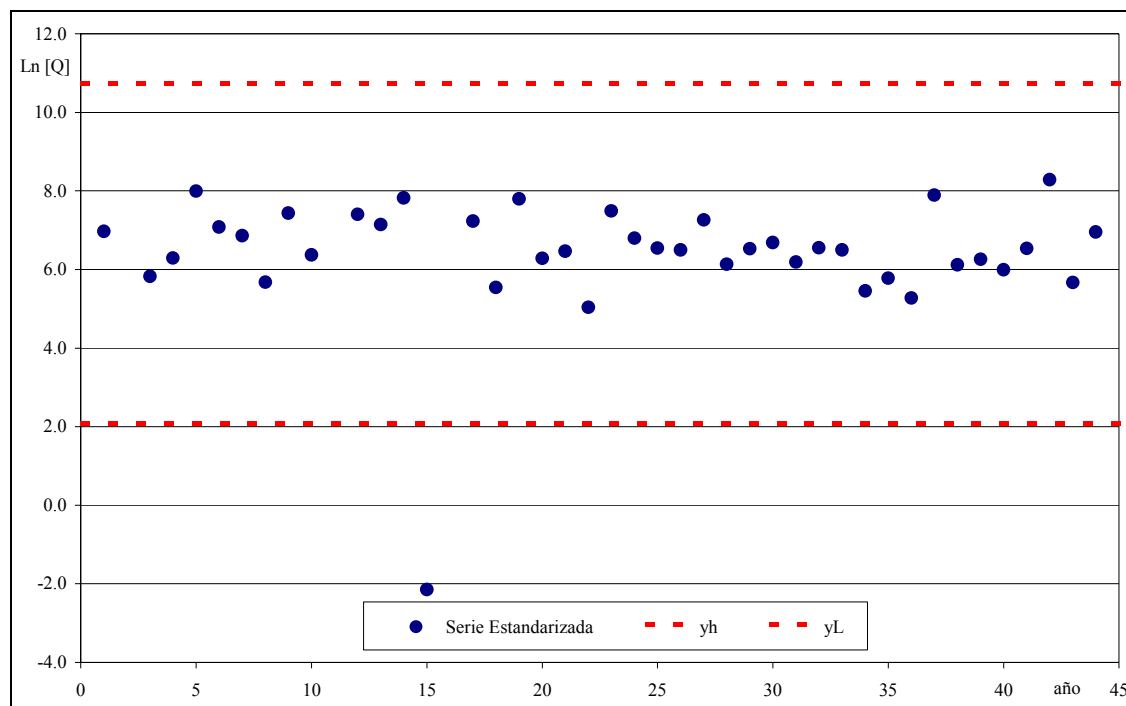


Figura X. 36. Prueba de datos dudosos en RP70 (periodo: 1954-2005, serie estandarizada por corrección de media y varianzas).

Análisis de pruebas paramétricas

A los efectos de cuantificar las características de cada una de las series previamente indicadas, se procedió a la aplicación de la totalidad de tests paramétricos indicados en la primera parte de este Anexo. En las Tabla X. 27, Tabla X. 28 y Tabla X. 29 se presentan los resultados obtenidos.

De estos análisis de los mismos se pueden extraer las mismas conclusiones que con el empleo de la serie anterior al evento del 2003, es decir, que en la serie se registra un salto que no solo afecta las medias de la serie si no que también influye sobre las varianzas. Por su parte, la incorporación de los últimos tres años de la serie en el periodo 1972-2005, origina un aumento de la media de aproximadamente un 12 % en relación al periodo 1972-2002.

Tabla X. 27. Prueba t, suponiendo varianzas iguales
 (periodo 1954-2005).

	1954-1972	1972-2005
Media	276.4	914.9
Varianza	22140.9	780929.4
Observaciones	18	26
Varianza agrupada	473800.7	
Diferencia hipotética de las medias	0	
Grados de libertad	42	
Estadístico t	-3.025	
Significancia (α)	0.050	
P(T<=t) una cola	0.002	
Valor crítico de t (una cola)	1.682	
Significancia (α)	0.050	
P(T<=t) dos colas	0.004	
Valor crítico de t (dos colas)	2.018	
Hipótesis H0	Se Rechaza	

Tabla X. 28. Prueba t', suponiendo varianzas desiguales
 (periodo 1954-2005)

	1954-1972	1972-2005
Media	276.4	914.9
Varianza	22140.9	780929.4
Observaciones	18	26
Diferencia hipotética de las medias	0	
Grados de libertad	27	
Estadístico t	-3.611	
Significancia (α)	0.050	
P(T<=t) una cola	0.001	
Valor crítico de t (una cola)	1.703	
Significancia (α)	0.050	
P(T<=t) dos colas	0.001	
Valor crítico de t (dos colas)	2.052	
Hipótesis H0	Se Rechaza	

Tabla X. 29. Prueba F para varianzas de dos muestras
(periodo 1954-2005)

	1954-1972	1972-2005
Media	276.4	914.9
Varianza	22140.9	780929.4
Observaciones	18	26
Grados de libertad	17	25
F	0.028	
Significancia (α)	0.050	
P(F<=f) una cola	0.000	
Valor crítico para F (una cola)	0.458	
Hipótesis H0	Se Rechaza	

Pruebas no paramétricas

De acuerdo a lo previamente indicado, entre los tests no paramétricos posibles, fueron aplicados los tests de Wald-Wolfowitz (independencia), Kendall (tendencia) y Wilcoxon (Homogeneidad).

En las Tabla X. 30, Tabla X. 31 y Tabla X. 32, se presentan los resultados alcanzados por la aplicación de estos tests para las series 1954-2005, 1972-2005 y 1954-2005 estandarizada. Del análisis de resultados se dedujo precisamente la necesidad de estandarizar la serie en el periodo anterior a 1972, llevando las medias y varianzas a la del periodo posterior (a los efectos de aprovechar todas las informaciones disponibles de esa serie).

Tabla X. 30. Prueba de Independencia Wald-Wolfowitz
(periodo 1954-2005)

	1954-2005	1972-2005	Estandarizada
Valor del estadístico U	0.31	0.96	1.20
Valor de significancia (α)	0.05	0.05	0.05
Valor probabilidad (p)	0.761	0.336	0.230
Hipótesis H0	Se Acepta	Se Acepta	Se Acepta

Tabla X. 31. Prueba de Estacionalidad Kendall (periodo 1954-2005)

	1954-2005	1972-2005	Estandarizada
Valor del estadístico K	3.08	0.79	0.31
Valor de significancia (α)	0.01	0.05	0.05
Valor probabilidad (p)	0.002	0.427	0.754
Hipótesis H0	Se Rechaza	Se Acepta	Se Acepta

Tabla X. 32. Prueba de Homogeneidad Wilcoxon
(periodo 1954-2005)

	1954-2005	1972-2005	Estandarizada
Valor del estadístico W	4.12	1.21	0.08
Valor de significancia (α)	0.01	0.05	0.05
Valor probabilidad (p)	0.000	0.227	0.933
Hipótesis H0	Se Rechaza	Se Acepta	Se Acepta
División de la serie	1971-1972	1971-1972	1971-1972

Posición de ploteo (distribución empírica):

La Figura X. 37 ilustra acerca de la posición de ploteo de las series empíricas consideradas en este subítem (1954-2005; 1972-2005 y 1954-2005 estandarizada) al emplear la fórmula de Gringorten.

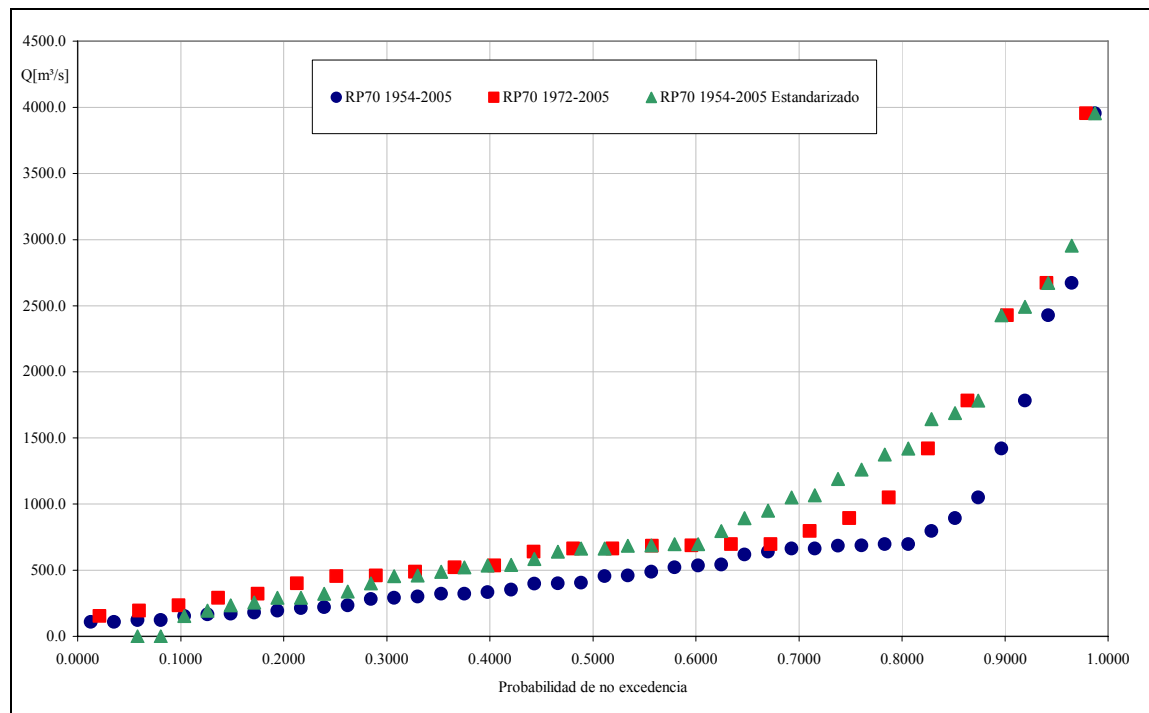


Figura X. 37. Posición de ploteo de las series analizadas de acuerdo a la fórmula de Gringorten.

5.2.2 Análisis de Frecuencia con base en la serie 1954-2005 estandarizada

Cabe indicar que para la serie estandarizada fue posible ajustar solamente las distribuciones GEV y Pearson III. Ello resulta de las características de los datos una vez que los mismos fueron tratados mediante la estandarización de la serie y la varianza.

Las distribuciones fueron ajustadas empleando el método de máxima verosimilitud (MLE). Los resultados obtenidos sobre la serie estandarizada estadísticamente (Tabla X. 33) se

presentan a continuación. También en este caso merece indicarse que para determinadas distribuciones y algunos valores de recurrencia T_r , el paquete HYFRAN[®] no ofrece los valores correspondientes al intervalo de confianza del 5%. Esta situación se refleja en las Tablas con la indicación “N/D”.

Ambas distribuciones superaron el test de adecuación de χ^2 con significancia del 5 %.

Del análisis gráfico se desprende un comportamiento semejante de ambas distribuciones para retratar los valores extremos máximos de la serie 1954-2005 estandarizada.

Tabla X. 33. Resultados de los ajustes realizados sobre la serie estandarizada 1954-2005

Función de Ajuste		GEV		Pearson III	
T_r	q	Q_T	Intervalo	Q_T	Intervalo
10000	0.9999	18700	N/D	7370	N/D
2000	0.9995	11900	N/D	6170	N/D
1000	0.9990	9750	N/D	5640	2040 - 9250
200	0.9950	5970	N/D	4420	2040 - 6800
100	0.9900	4760	N/D	3890	2000 - 5780
50	0.9800	3750	1710 - 5780	3350	1920 - 4780
20	0.9500	2650	1560 - 3740	2630	1730 - 3530
10	0.9000	1960	1310 - 2620	2070	1480 - 2670
5	0.8000	1370	990 - 1760	1500	1060 - 1940
3	0.6667	978	715 - 1240	1070	675 - 1470
2	0.5000	671	475 - 867	689	342 - 1030

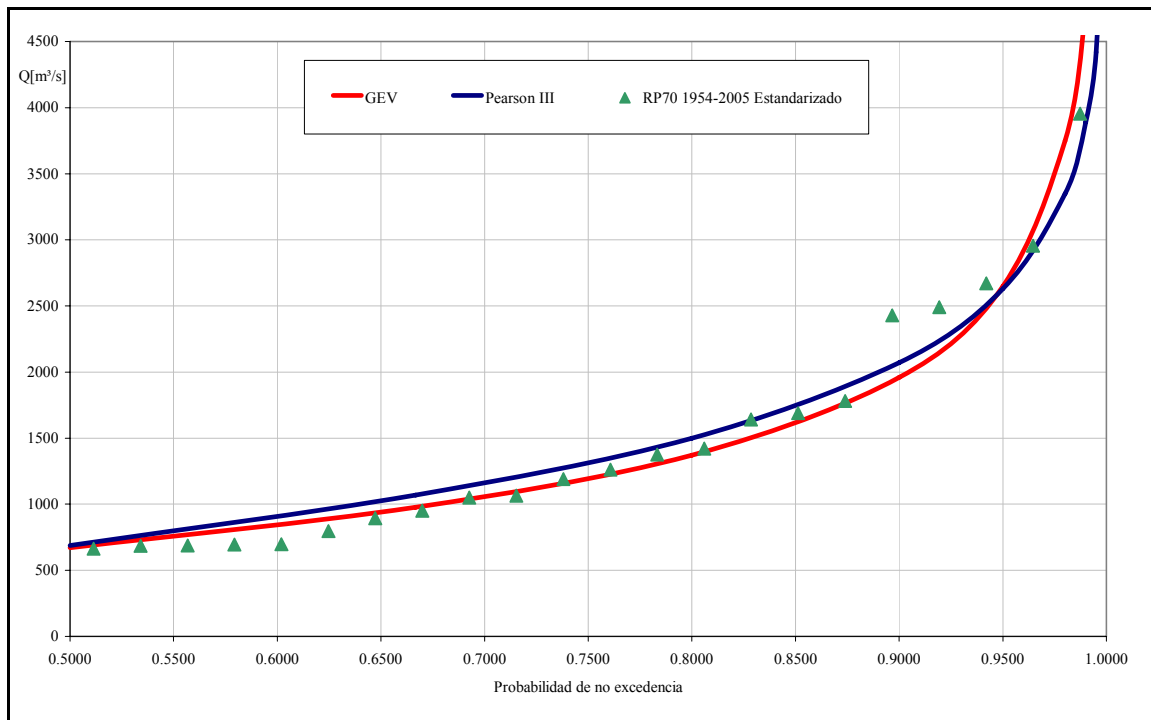


Figura X. 38. Grafica de los ajustes realizados sobre la serie estandarizada 1954-2005 (RP70).

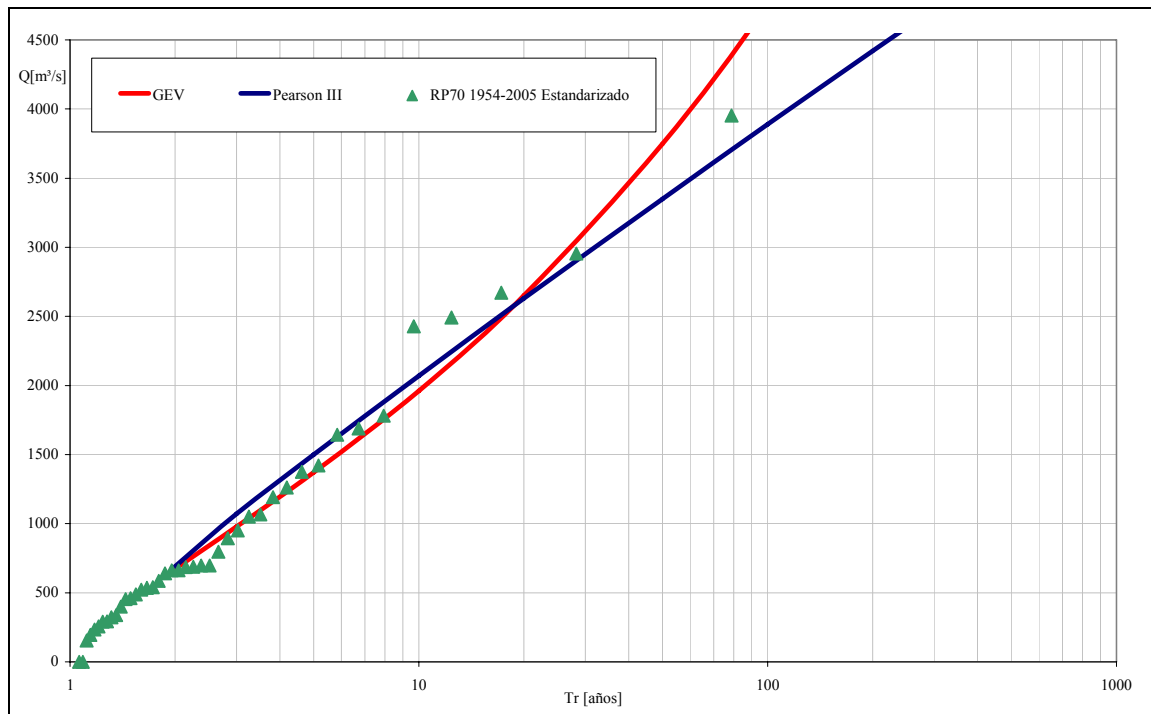


Figura X. 39. Grafica de los ajustes realizados sobre la serie estandarizada 1954-2005 (RP70).

La aplicación del método MLE a la serie 1954-2005 estandarizada condujo al siguiente ajuste de parámetros de la distribución GEV:

$$\hat{\xi} = 484,18 \quad \hat{\alpha} = 486,045 \quad \text{y} \quad \hat{\kappa} = -0,257$$

De acuerdo a lo indicado en el apartado 3 de este Anexo, se observa la obtención de un valor razonable para el parámetro $\hat{\kappa}$.

5.2.3 Análisis de Frecuencia con Base en la Serie Moderna 1972-2005

Los ajustes de totalidad de las distribuciones ensayadas correspondientes a la serie 1972-2005 se presentan en la Tabla X. 34, Figura X. 40 y Figura X. 41,. En todos los casos se empleó como método de ajuste el MLE, salvo para el caso de la distribución Pearson III la cual fue ajustada por el método de los momentos a los fines de comparación previamente apuntados. De igual modo que aquello expresado en el apartado precedente, la aplicación del método MLE a esta serie corta (n = 26 años) tuvo por objetivo analizar el valor del parámetro de forma, κ , de la distribución GEV, con el fin de evidenciar potenciales problemas de estimación.

Cabe consignar que el empleo del test de adecuación χ^2 para todas las funciones ajustadas a la serie 1972-2005 permitió observar el rechazo de la hipótesis de adecuación H_0 para la distribución Pearson III para un grado de significancia del 5 %.

Contrariamente, con un grado de significancia del 1 % todas las distribuciones fueron no rechazadas. Por lo tanto, no fue posible concluir que la muestra proviene de una distribución Pearson III.

Tabla X. 34. Resultados de los ajustes realizados sobre la serie 1972-2005

Función de Ajuste		GEV		Pearson III		Log-Pearson III		Lognormal		Exponencial	
Tr	q	Q _T	Intervalo	Q _T	Intervalo	Q _T	Intervalo	Q _T	Intervalo	Q _T	Intervalo
10000	0.9999	52000	N/D	8620	N/D	20800	N/D	12200	N/D	7410	4560 - 10300
2000	0.9995	24400	N/D	7070	N/D	13000	N/D	8750	N/D	6140	3790 - 8480
1000	0.9990	17600	N/D	6400	N/D	10500	N/D	7480	1970 - 13000	5590	3460 - 7720
200	0.9950	8150	N/D	4860	N/D	6140	N/D	5000	1820 - 8180	4320	2680 - 5950
100	0.9900	5830	N/D	4200	N/D	4790	N/D	4110	1690 - 6530	3770	2350 - 5180
50	0.9800	4140	N/D	3550	N/D	3670	N/D	3320	1530 - 5110	3220	2020 - 4420
20	0.9500	2610	N/D	2690	1320 - 4060	2510	N/D	2410	1280 - 3540	2490	1570 - 3410
10	0.9000	1800	868 - 2730	2050	1160 - 2930	1810	937 - 2680	1810	1070 - 2560	1950	1240 - 2650
5	0.8000	1210	739 - 1670	1420	679 - 2150	1240	775 - 1700	1280	831 - 1740	1400	906 - 1890
3	0.6667	860	582 - 1140	984	317 - 1650	892	593 - 1190	930	637 - 1220	993	659 - 1330
2	0.5000	623	438 - 808	621	N/D	634	428 - 840	664	464 - 864	672	460 - 884

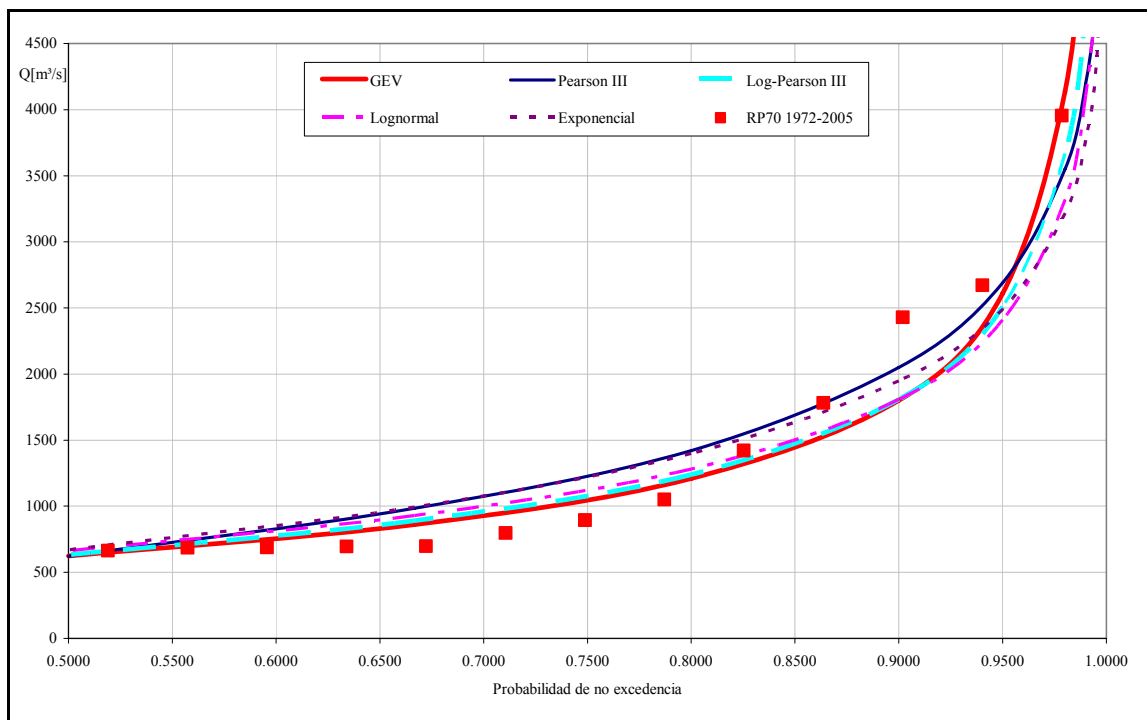


Figura X. 40. Grafica de los ajustes realizados sobre la serie 1972-2005 (RP70)

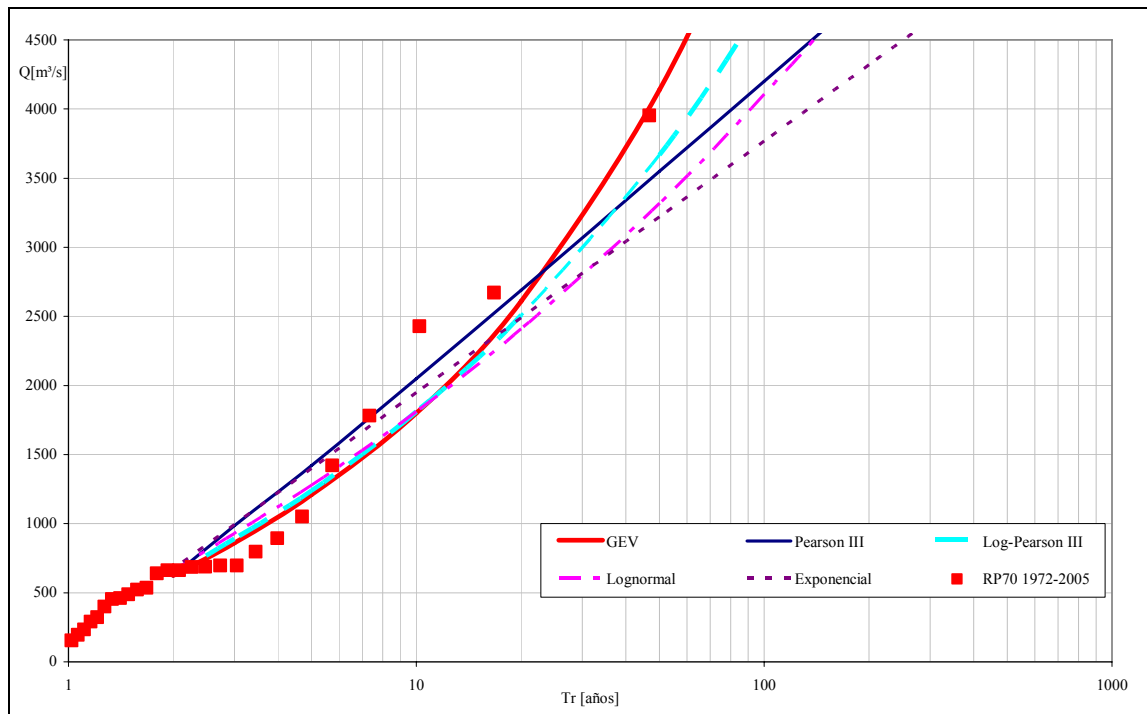


Figura X. 41.. Grafica de los ajustes realizados con recurrencia asociada sobre la serie 1972-2005 (RP70).

La aplicación del método MLE a la serie corta 1972-2005 condujo al siguiente ajuste de parámetros de la distribución GEV:

$$\hat{\xi} = 491,68 \quad \hat{\alpha} = 328,137 \quad \text{y} \quad \hat{\kappa} = -0,467$$

De acuerdo a lo indicado en el apartado 3 de este Anexo, se observa la obtención de un valor inadecuado para el parámetro $\hat{\kappa}$, demostrando las falencias de ajuste que posee el método al aplicarse a series cortas, con $n < 50$ años.

Pese a ello, el análisis gráfico permite observar un buen comportamiento de la distribución GEV. Este ajuste, sin embargo, no posee un adecuado sustento teórico.

5.3 Análisis de Frecuencia sobre Series Reconstruidas y Sistemáticas (R+S)

5.3.1 Análisis de Frecuencia con Base en la Serie 1928-2005

En este caso los análisis estadísticos fueron efectuados sobre la serie de datos reconstruidos y sistemáticos, considerando las informaciones hasta el año 2005.

Estadística descriptiva

Al igual que en los casos anteriores, mediante la observación de la serie histórica de valores se desprende el cambio en la media entre los periodos 1928-1972 y 1972-2005. En este caso la diferencia en las medias es menor, siendo del 55 % entre ambos periodos de comparación (en los análisis previos, considerando solamente la información sistemática, esta diferencia era del 70 % para la serie 1928-2002).

Tabla X. 35. Estadística Descriptiva de la serie completa

	1928-2005	Estandarizada
Media	570.3	696.7
Error típico	77.4	79.7
Mediana	402.0	541.9
Moda	467.4	950.5
Desviación estándar	624.1	642.6
Varianza de la muestra	389561.3	412898.8
Curtosis	15.4	10.9
Coefficiente de asimetría	3.6	2.8
Rango	3882.1	3882.1
Mínimo	71.9	71.9
Máximo	3954	3954
Suma	37072.11456	45284.23065
Cantidad de valores	65	65

Detección de tendencias en la serie

En la Figura X. 42 se presenta en forma gráfica el análisis realizado. En la misma se observa la tendencia creciente a partir de la década del 70. Como ya fuera citado, la pendiente de la recta es menor en relación a la obtenida con base en la serie sistemática equivalente.

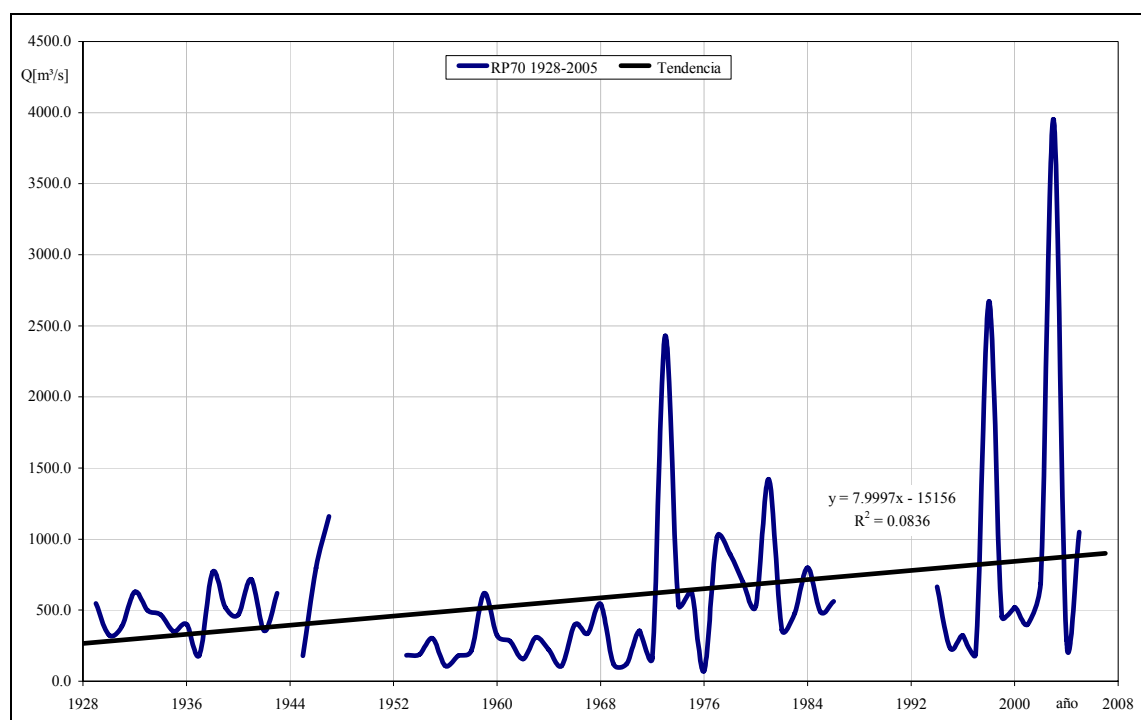


Figura X. 42. Gráfico de tendencia en las series de caudales máximos RP70 (Período 1928-2005)

Análisis de valores dudosos (outliers):

Se procedió de forma similar a la realizada sobre las series precedentemente analizadas. Se verificó que ningún valor fue identificado como un posible “outlier” evidente, tanto para el caso de la serie 1928-2005 como para su correspondiente serie estandarizada (la necesidad de estandarización de la serie surgió del análisis de los test paramétricos y no paramétricos posteriormente presentados).

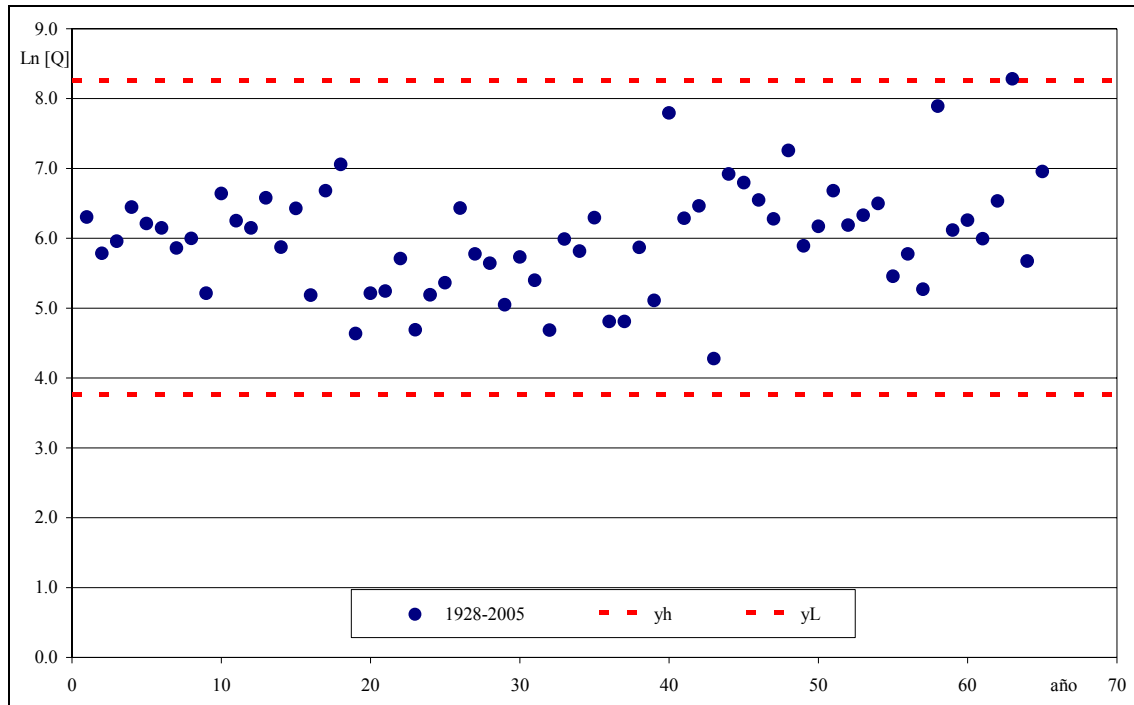


Figura X. 43. Prueba de datos dudosos en RP70 (periodo: 1928-2005)

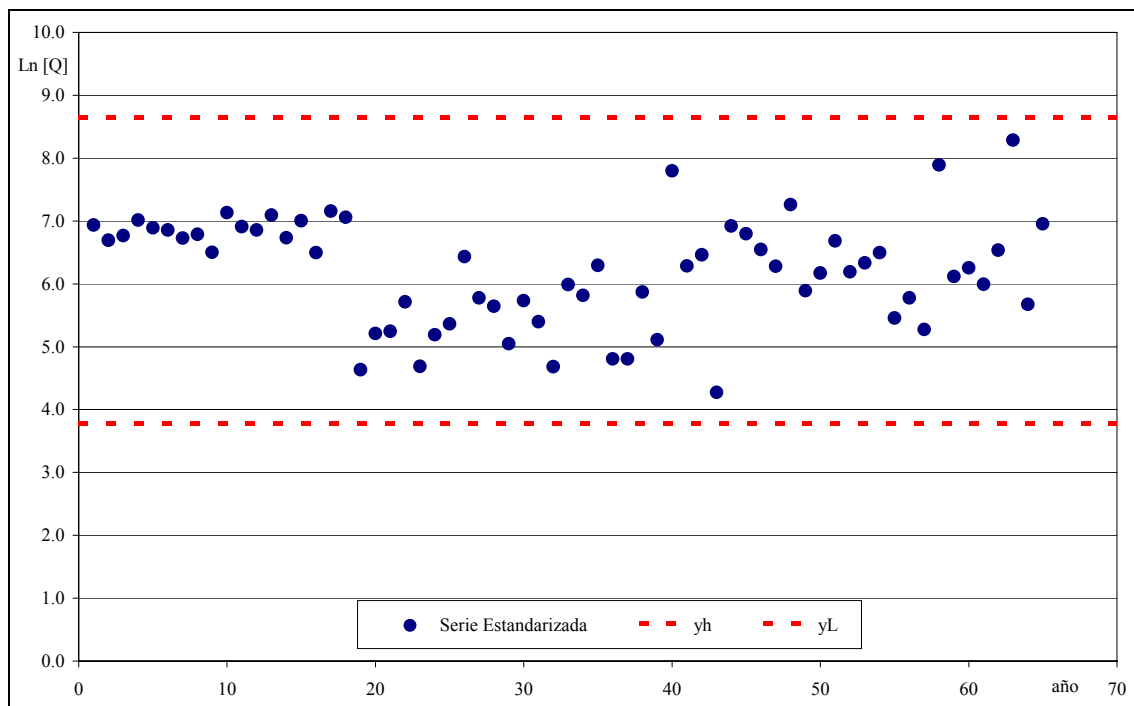


Figura X. 44. Prueba de datos dudosos en RP70 (periodo: 1928-2005 serie corregida por la media)

Análisis de pruebas paramétricas

A continuación se presentan en las Tabla X. 36, Tabla X. 37 y Tabla X. 38 los resultados de los tests paramétricos aplicados a las series 1928-2005 y 1972-2005.

Tabla X. 36. Prueba t, suponiendo varianzas iguales
(periodo 1928-2005)

	1928-1972	1972-2005
Media	377.1	860.2
Varianza	54170.5	769325.8
Observaciones	39	26
Varianza agrupada	337962.3	
Diferencia hipotética de las medias	0	
Grados de libertad	63	
Estadístico t	-3.282	
Significancia (α)	0.050	
P(T<=t) una cola	0.001	
Valor crítico de t (una cola)	1.669	
Significancia (α)	0.050	
P(T<=t) dos colas	0.002	
Valor crítico de t (dos colas)	1.998	
Hipótesis H0	Se rechaza	

Tabla X. 37. Prueba t', suponiendo varianzas desiguales
(periodo 1928-2005)

	1928-1972	1972-2005
Media	377.1	860.2
Varianza	54170.5	769325.8
Observaciones	39	26
Diferencia hipotética de las medias	0	
Grados de libertad	27	
Estadístico t	-2.744	
Significancia (α)	0.050	
P(T<=t) una cola	0.005	
Valor crítico de t (una cola)	1.703	
Significancia (α)	0.050	
P(T<=t) dos colas	0.010	
Valor crítico de t (dos colas)	2.051	
Hipótesis H0	Se rechaza	

Tabla X. 38. Prueba F para varianzas de dos muestras

	1928-1972	1972-2005
Media	377.1	860.2
Varianza	54170.5	769325.8
Observaciones	39	26
Grados de libertad	38	25
F	0.070	
Significancia (α)	0.050	
P(F \leq f) una cola	0.000	
Valor crítico para F (una cola)	0.556	
Hipótesis H0	Se rechaza	

Análisis de pruebas no paramétricas:

En las Tabla X. 39, Tabla X. 40 y Tabla X. 41, se presentan los resultados alcanzados por la aplicación de los test no paramétricos de Wald-Wolfowitz (independencia), Kendall (estacionalidad) y Wilcoxon (homogeneidad) para las serie 1928-2005 y 1928-2005 estandarizada. La necesidad de estandarización de la primera de ellas surge del análisis de los resultados de los test paramétricos y no paramétricos. Como particularidad cabe señalar que del resultado del test de Kendall se desprende que es necesario estandarizar la serie en el periodo anterior a 1972, pero solo por las medias, debido a que el test no detecta un cambio sustancial sobre las varianzas.

Tabla X. 39. Prueba de Independencia
(Wald-Wolfowitz)

	1928-2005	Estandarizada
Valor del estadístico U	0.23	0.37
Valor de significancia (α)	0.05	0.05
Valor probabilidad (p)	0.815	0.714
Hipótesis H0	Se Acepta	Se Acepta

Tabla X. 40. Prueba de Estacionalidad (Kendall)

	1928-2005	Estandarizada
Valor del estadístico K	1.49	1.13
Valor de significancia (α)	0.05	0.05
Valor probabilidad (p)	0.136	0.257
Hipótesis H0	Se Acepta	Se Acepta

Tabla X. 41. Prueba de Homogeneidad (Wilcoxon)

	1928-2005	Estandarizada
Valor del estadístico W	3.37	1.01
Valor de significancia (α)	0.01	0.05
Valor probabilidad (p)	0.001	0.312
Hipótesis H0	Se Rechaza	Se Acepta
División de la serie	1971-1972	1971-1972

Posición de ploteo (distribución empírica)

En la Figura X. 45 se presenta la distribución empírica de las series 1928-2005 y 1928-2005 estandarizada (los cambios corresponden solamente a las magnitudes de los caudales). Se advierte el quiebre entre los caudales máximos frecuentes y los caudales máximos de baja frecuencia.

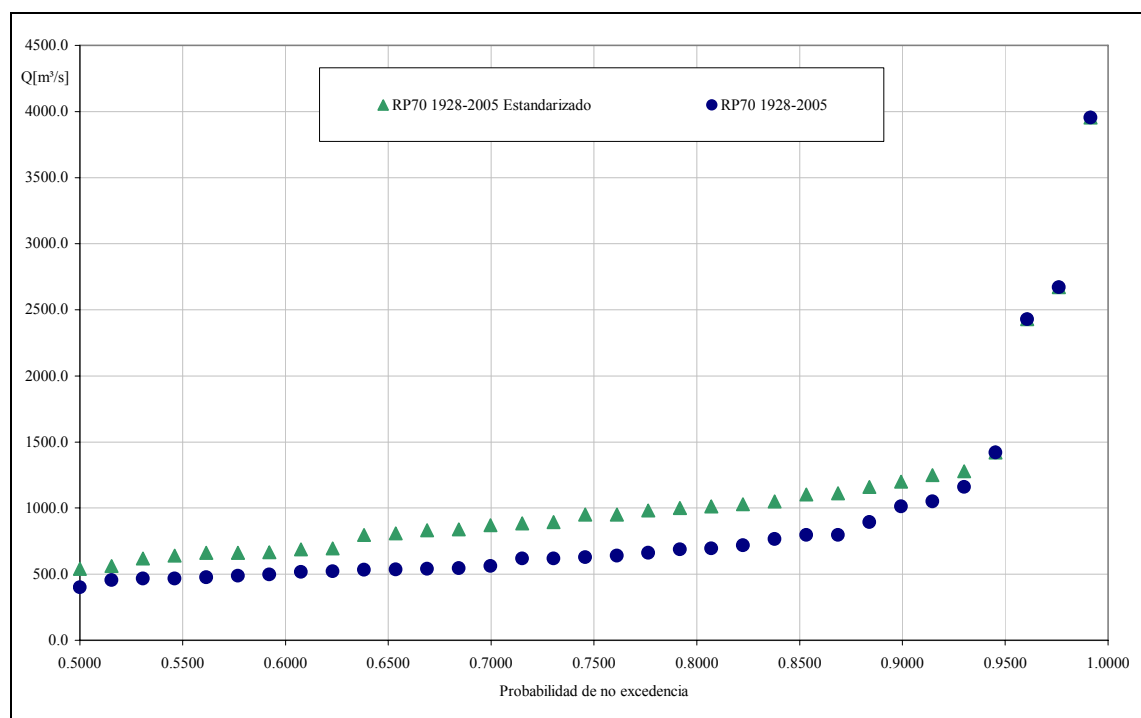


Figura X. 45. Posición de ploteo de las series analizadas de acuerdo a la fórmula de Gringorten

5.3.2 Análisis de Frecuencia con Base en la Serie 1928-2005 Estandarizada

Los distintos ajustes realizados son presentados a continuación (Tabla X. 42, Figura X. 46 y Figura X. 47). Salvo la distribución Pearson III que fuera ajustada por el método de los momentos, las restantes distribuciones fueron ajustadas por el método MLE. La selección del método de los momentos para la función Pearson III se debió a los fines comparativos con estudios precedentes previamente indicados.

El empleo del test de adecuación χ^2 rechazó la hipótesis de adecuación H0 para la distribución Pearson III para un nivel de significancia del 5 %. El resto de las distribuciones superaron este test. Para el nivel de significación del 1 % la hipótesis H0 no fue rechazada para ninguna de las distribuciones. De lo anterior se concluyó que no era posible concluir que la muestra provenía de una distribución del tipo Pearson III.

Del análisis gráfico sobre papel probabilístico se confirma la buena adecuación de la distribución GEV a los valores extremos máximos, hecho ya denotado previamente con base en series que dispusieron de menores informaciones en relación a la analizada.

Tabla X. 42. Resultados de los ajustes realizados sobre la serie estandarizada 1928-2005

Función de Ajuste		GEV		Pearson III		Log-Pearson III		Lognormal		Exponencial	
T	q	Q _T	Intervalo	Q _T	Intervalo	Q _T	Intervalo	Q _T	Intervalo	Q _T	Intervalo
10000	0.9999	21300	N/D	7060	N/D	8920	N/D	11800	4880 - 18700	5910	4480 - 7340
2000	0.9995	11900	N/D	5710	N/D	6570	N/D	8180	3870 - 12500	4890	3710 - 6060
1000	0.9990	9210	N/D	5130	N/D	5680	N/D	6900	3450 - 10400	4450	3370 - 5520
200	0.9950	5030	N/D	3800	N/D	3880	N/D	4450	2530 - 6380	3420	2600 - 4250
100	0.9900	3840	1490 - 6190	3250	N/D	3210	1630 - 4800	3600	2160 - 5050	2980	2270 - 3700
50	0.9800	2900	1440 - 4370	2700	1430 - 3960	2610	1550 - 3670	2860	1810 - 3900	2540	1940 - 3150
20	0.9500	1960	1230 - 2700	1990	1300 - 2680	1900	1310 - 2500	2020	1370 - 2660	1960	1500 - 2430
10	0.9000	1420	1010 - 1840	1480	1010 - 1940	1430	1060 - 1810	1480	1070 - 1900	1520	1170 - 1880
5	0.8000	984	758 - 1210	994	552 - 1440	1010	775 - 1240	1020	772 - 1260	1080	835 - 1330
3	0.6667	710	564 - 855	688	312 - 1060	726	567 - 886	717	562 - 873	759	590 - 929
2	0.5000	508	405 - 611	451	230 - 673	506	395 - 617	498	395 - 600	502	395 - 609

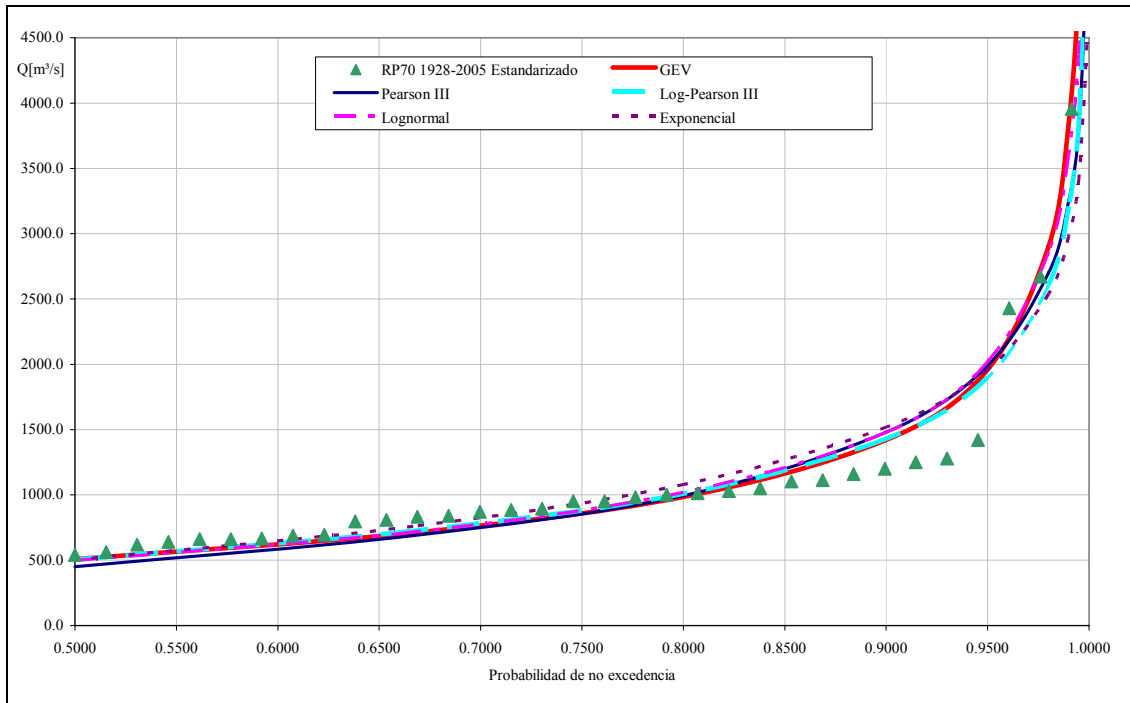


Figura X. 46. Gráfica de los ajustes realizados sobre la serie estandarizada 1928-2005 (RP70).

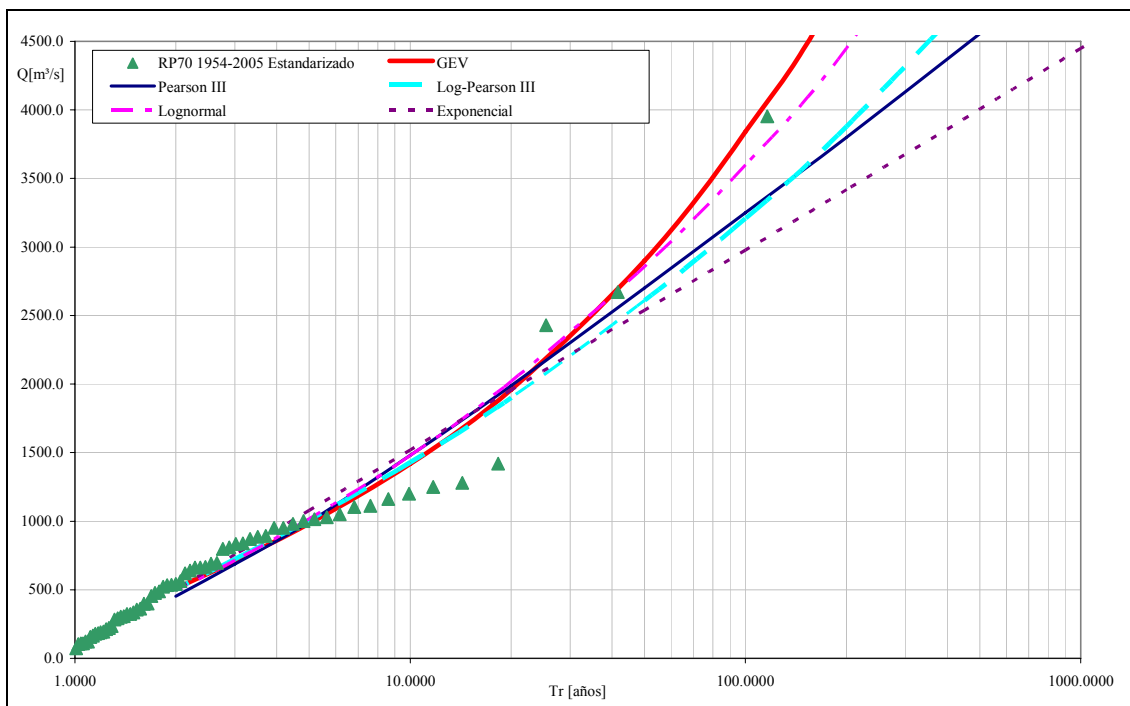


Figura X. 47. Gráfica de los ajustes realizados con Tr asociada sobre la serie estandarizada 1928-2005.

La aplicación del método MLE a la serie 1928-2005 estandarizada condujo al siguiente ajuste de parámetros de la distribución GEV:

$$\hat{\xi} = 390,23 \quad \hat{\alpha} = 300,75 \quad \text{y} \quad \hat{\kappa} = -0,35$$

De acuerdo a lo indicado en el apartado 3 de este Anexo, se observa la obtención de un valor sospechoso para el parámetro $\hat{\kappa}$, hecho que denotaría falencias en el ajuste.

5.4 Análisis de Frecuencia sobre Series Históricas y Sistemáticas (H+S)

5.4.1 Análisis de Frecuencia con Base en la Serie 1875-2005

En este apartado se presentan los estudios basados en la serie 1875-2005, es decir, basados en todas las informaciones disponibles hasta la actualidad (agosto 2005).

La metodología empleada para el análisis de las marcas históricas fue la propuesta, respectivamente, por Martins y Stedinger (2001) para el tratamiento estadístico de informaciones históricas con series anuales y parciales y por Reis y Stedinger (2005) para el tratamiento bayesiano de la información.

Cuantiles de Crecida Estimados:

Las Tabla X. 43, Tabla X. 44, Tabla X. 45 y Tabla X. 46 presentan los parámetros estimados, los cuantiles de crecidas y los intervalos de confianza para estos cuantiles, asociados a varios tiempos de retorno para los modelos GEV/ML, GEV/GML, LN/ML y LPIII/Abordaje *bayesiano*, respectivamente. Las referidas Tablas muestran los parámetros y los cuantiles identificados a partir del uso de registros sistemáticos solamente, como así también con el uso adicional de información histórica (casos BCD y CD).

Tal como ocurriera previamente, comparando la Tabla X. 43 (GEV/MLE) y la Tabla X. 44 (GEV/GMLE) se puede observar que la metodología de máxima verosimilitud (GMLE) adoptada corrigió las inconsistencias del método de la máxima verosimilitud clásico (MLE), a saber:

- valor sospechoso del factor de forma (κ muy negativo, lo cual implica en este caso, $\kappa < -0,46$, que el momento de orden 3, asimetría, no existe);
- límites de confianza negativos para tiempos de retorno ≥ 500 años.

La Tabla X. 44 y la Figura X. 49 muestra que las estimativas puntuales del modelo GEV/GMLE para datos sistemáticos son superiores a aquellas que emplean la información histórica, existiendo una ganancia considerable en términos de precisión de estas últimas estimaciones como consecuencia del uso de la información histórica. Esto puede ser fácilmente verificado si se observa la reducción de los límites de confianza de 95% asociados a las mismas. Las diferencias entre las estimaciones puntuales para los casos

GEV-BCD y GEV-CD son despreciables, siendo que las ganancias en la precisión (amplitud del intervalo de confianza del 95%) derivadas del uso de la información CD en lugar de la BCD, resultan pequeñas dado que el umbral elegido fue relativamente alto ($1.700 \text{ m}^3/\text{s}$).

En el caso de la distribución LPIII/Abordaje bayesiano (Figura X. 51) las diferencias entre las estimaciones puntuales con y sin el uso de la información histórica son más significativas, ocurriendo nuevamente problemas en las estimaciones de cuantiles de crecidas con información sistemática para elevados tiempos de retorno. Esto último se evidencia cuando se analizan los límites de confianza de 95 % para las estimaciones realizadas. Las estimaciones puntuales LPIII-BCD y LPIII-CD son bastante próximas, pero al igual que en caso del estudio realizado con informaciones hasta el año 2002, las ganancias en la precisión debido al uso de datos tipo CD no son despreciables para elevados tiempos de retorno.

La Figura X. 52 compara los resultados de los dos modelos que emplean la información histórica (Caso CD), GEV/GMLE-CD y LPIII/Abordaje Bayesiano-CD. Esta Figura muestra que las estimaciones puntuales para los dos modelos son bastante próximas. Sin embargo, la precisión de la estimativa GEV/GMLE es mayor, en especial para elevados tiempos de retorno.

La Figura X. 53 muestra los cuantiles de crecida GEV/GMLE-CD y sus límites de confianza de 95 % asociados a diferentes tiempos de retorno y los caudales observados asociados a sus respectivas probabilidades de excedencia calculadas conforme las ecuaciones X.50 y X.51, empleando las posiciones de ploteo.

Tabla X. 43. Resultados obtenidos a partir del Modelo GEV/MLE (informaciones hasta el año 2005).

Tr [años]	Parám.	Sistemático			BCD			CD		
	ξ	479.3			445.5			450.1		
	α	336.5			277.1			279.5		
	κ	-0.46			-0.29			-0.27		
	Probab.	2.5%	moda	97.5 %	2.5 %	Moda	97.5%	2.5 %	moda	97.5%
5	0.8	992	1207	1422	741	965	1189	749	967	1185
10	0.9	1191	1810	2429	1009	1322	1636	1015	1316	1616
20	0.95	1247	2621	3995	1256	1746	2236	1268	1724	2180
50	0.98	898	4164	7429	1503	2442	3381	1535	2384	3233
100	0.99	43	5842	11640	1593	3100	4608	1659	3000	4341
500	0.998	-6330	12575	31481	1232	5238	9244	1490	4958	8426
1000	0.999	-12745	17414	47572	670	6510	12349	1097	6100	11103

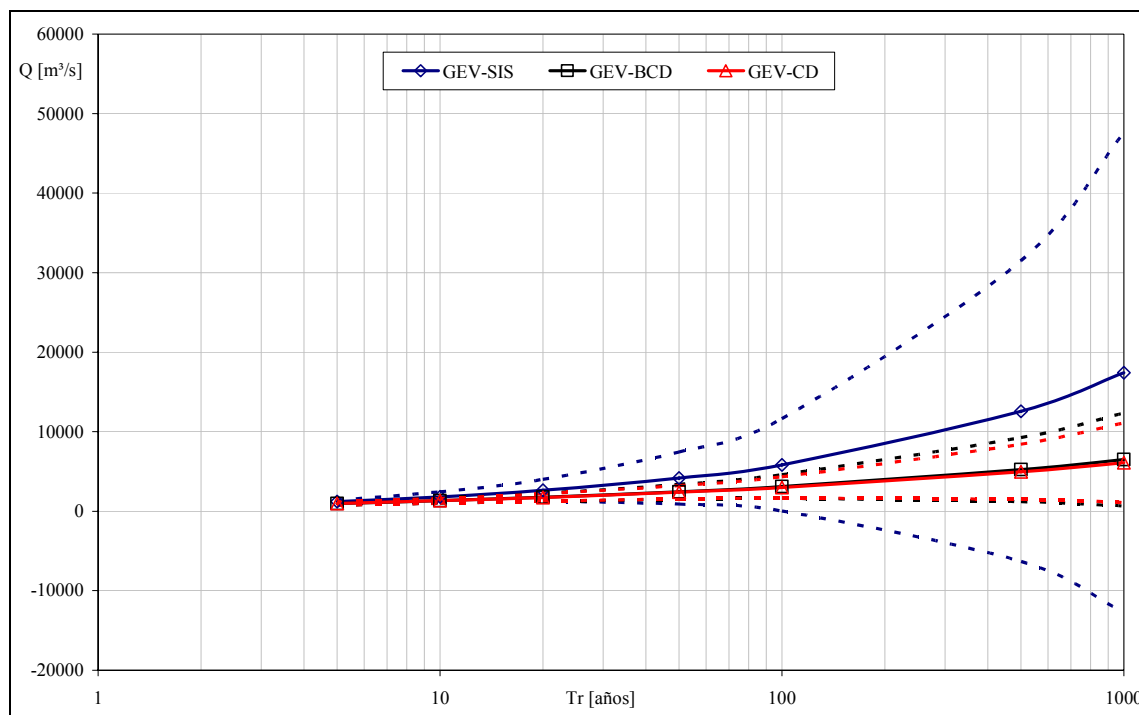


Figura X. 48. Cuantiles de crecida del modelo GEV/MLE y sus límites de confianza del 95%, asociados a diferentes tiempos de retorno calculados a partir del uso de información sistemática (GEV-SIS) y del uso adicional de información histórica (GEV-BCD y GEV-CD). Informaciones hasta el año 2005.

Tabla X. 44. Resultados para el Modelo GEV/GMLE (informaciones hasta el año 2005).

Tr [años]	Parám.	Sistemático			BCD			CD		
	ξ	514.4			467.1			468.3		
	α	352.2			294.5			295.0		
	κ	-0.24			-0.21			-0.20		
	Probab.	2.5%	moda	97.5 %	2.5 %	Moda	97.5%	2.5 %	moda	97.5%
5	0.8	808	1149	1490	776	986	1196	777	986	1195
10	0.9	1066	1562	2057	1032	1314	1595	1031	1311	1592
20	0.95	1330	2033	2736	1295	1680	2065	1291	1674	2057
50	0.98	1672	2776	3879	1639	2244	2850	1631	2230	2829
100	0.99	1908	3450	4993	1882	2746	3609	1873	2722	3571
500	0.998	2256	5511	8766	2324	4227	6130	2318	4168	6019
1000	0.999	2250	6670	11089	2424	5034	7644	2425	4953	7481

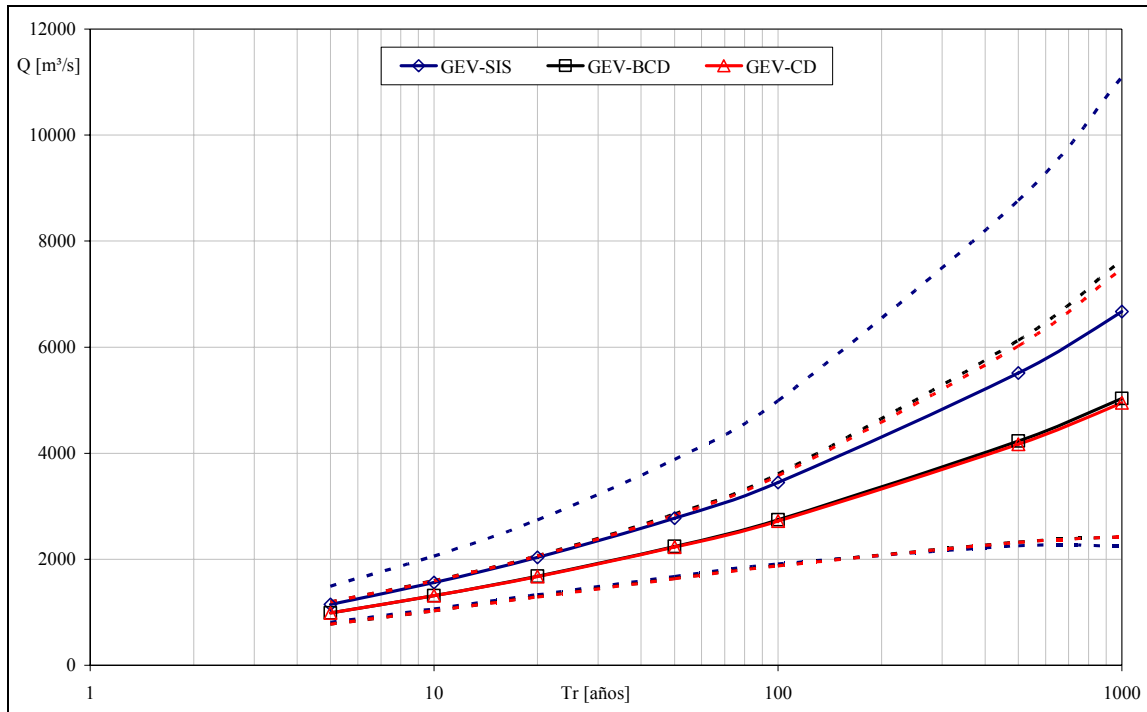


Figura X. 49. Cuantiles de crecida del modelo GEV/GMLE y sus límites de confianza del 95%, asociados a diferentes tiempos de retorno calculados a partir del uso de información sistemática (GEV-SIS) y del uso adicional de información histórica (GEV-BCD y GEV-CD). Informaciones hasta el año 2005.

Tabla X. 45. Resultados para el Modelo LN/MLE (informaciones hasta el año 2005).

Tr [años]	Parám.	Sistemático			BCD			CD		
	μ	6.476			6.322			6.323		
	σ	0.799			0.700			0.696		
	Probab.	2.5%	moda	97.5 %	2.5 %	Moda	97.5%	2.5 %	moda	97.5%
50	0.98	2199	3351	5105	1762	2341	3109	1828	2325	2958
100	0.99	2631	4166	6598	2068	2831	3876	2152	2810	3669
200	0.995	3096	5085	8353	2391	3372	4757	2494	3344	4482
333	0.997	3462	5834	9832	2639	3803	5481	2759	3769	5149
1000	0.999	4321	7670	13616	3207	4832	7281	3364	4782	6797

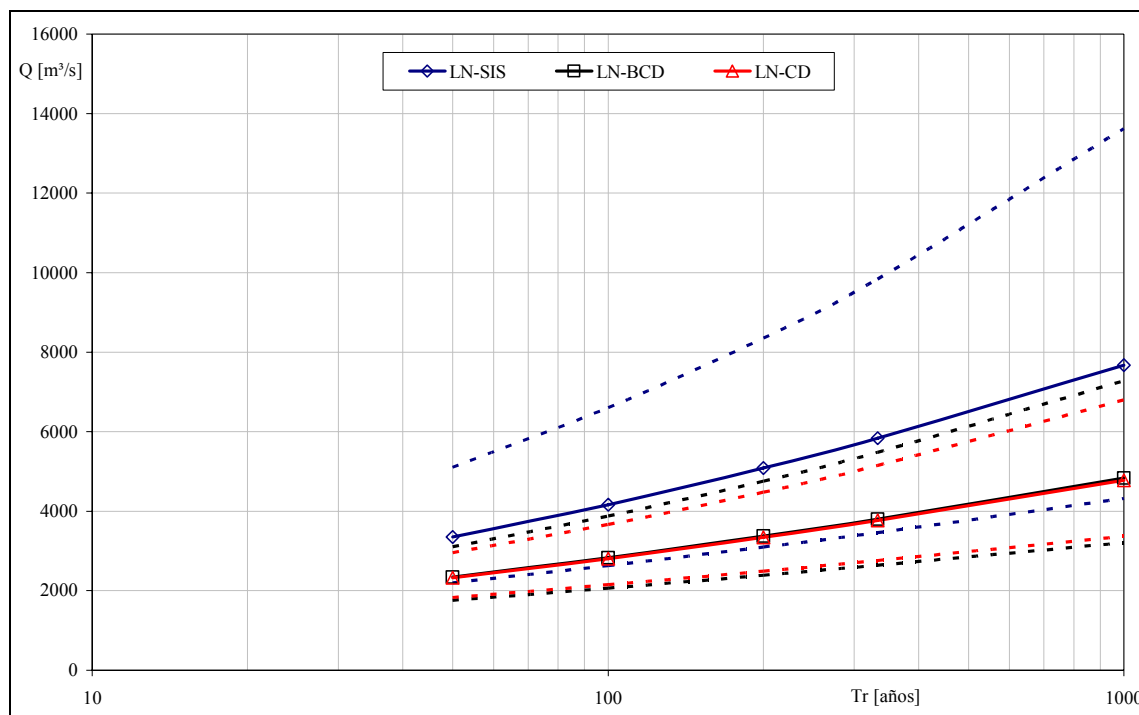


Figura X. 50. Cuantiles de crecida LN/MLE y sus límites de confianza de 95%, asociados a diferentes tiempos de retorno calculados a partir del uso de información sistemática (LN-SIS) y del uso adicional de información histórica (LN-BCD y LN-CD). Informaciones hasta el año 2005.

Tabla X. 46. Resultados para el Modelo LPIII/Abordaje bayesiano (info. hasta el año 2005).

Tr [años]	Parám.	Sistemático			BCD			CD		
	μ	2.81			2.73			2.73		
	σ	0.37			0.32			0.31		
	γ	0.16			0.13			0.11		
	Probab.	2.5%	moda	97.5 %	2.5 %	Moda	97.5%	2.5 %	moda	97.5%
5	0.8	910	1303	2093	774	987	1235	777	984	1228
10	0.9	1285	1916	3498	1080	1371	1760	1083	1361	1732
20	0.95	1690	2635	5701	1408	1802	2455	1402	1788	2394
50	0.98	2247	3802	10416	1855	2457	3747	1842	2423	3600
100	0.99	2705	4878	16104	2203	3030	5118	2189	2971	4848
500	0.998	3780	8173	42835	3033	4656	10253	3005	4519	9338
1000	0.999	4230	10037	64373	3393	5519	13617	3365	5330	12259

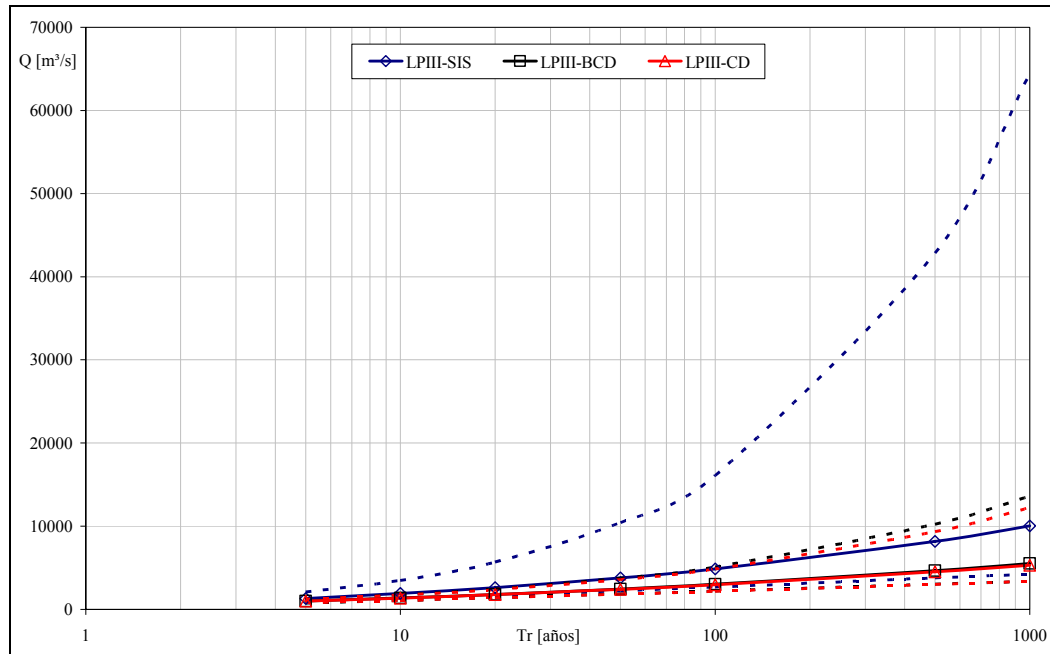


Figura X. 51. Cuantiles de crecida LPIII/Abordaje Bayesiano y sus límites de confianza de 95%, asociados a diferentes tiempos de retorno calculados a partir del uso de información sistemática (LPIII-SIS) y del uso adicional de información histórica (LPIII-BCD y LPIII-CD). Informaciones hasta el año 2005.

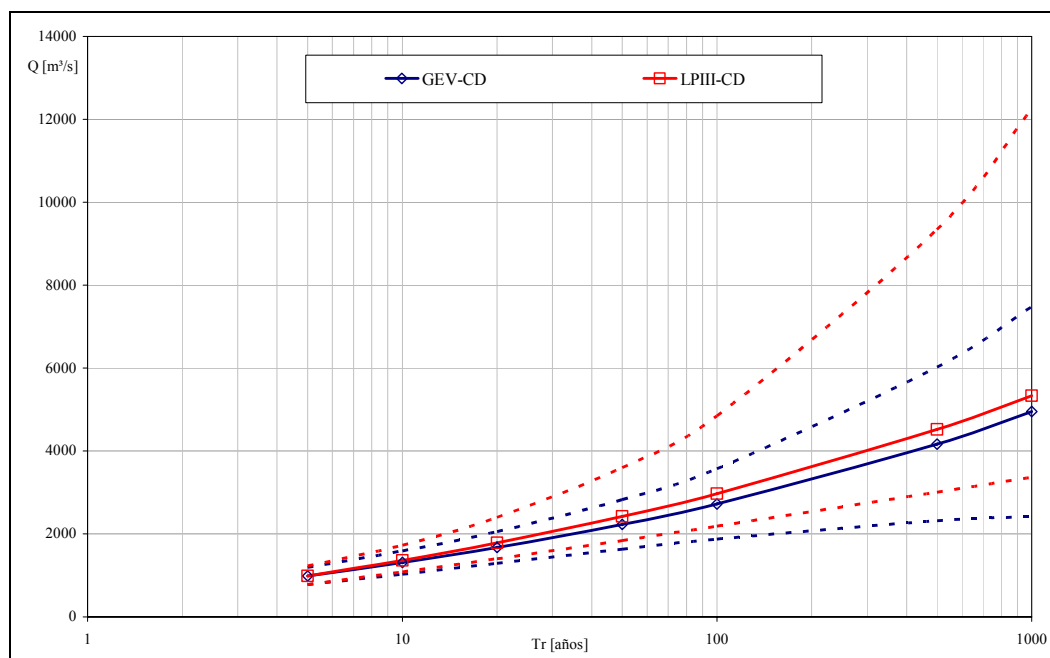


Figura X. 52. Comparación de los cuantiles de crecida GEV/GML y sus límites de confianza de 95%, con los cuantiles correspondientes y sus límites de confianza de 95% obtenidos con LPIII/Abordaje Bayesiana calculados a partir del uso de información sistemática e histórica (Caso CD). Informaciones hasta el año 2005.

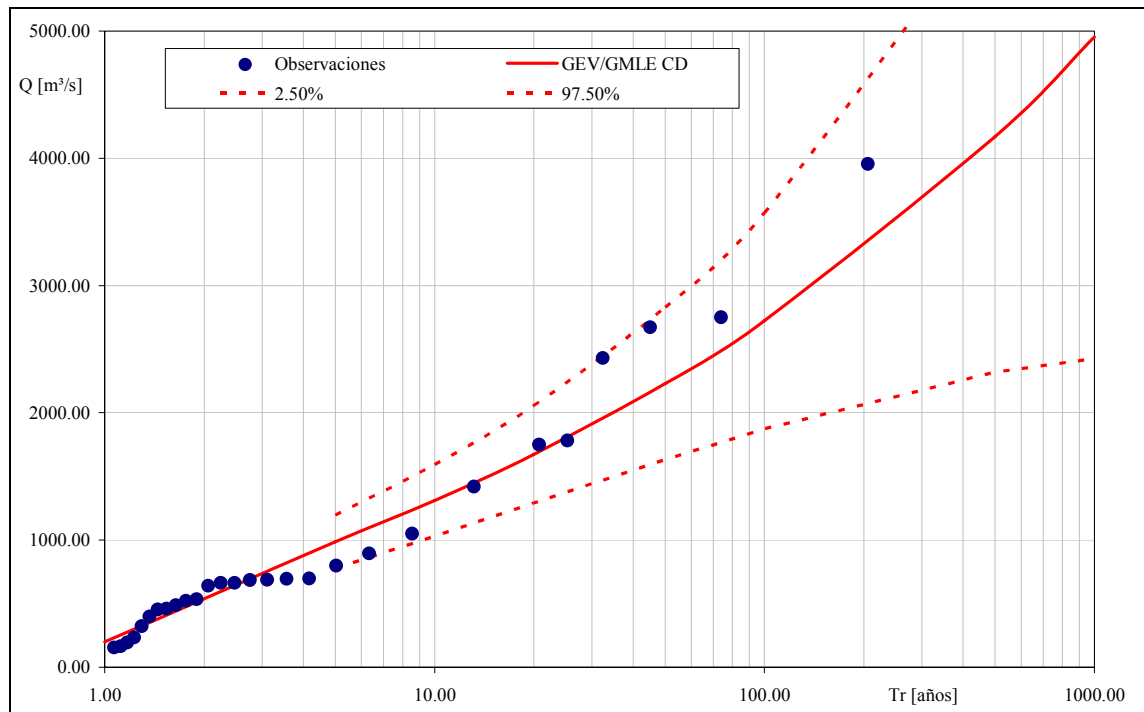


Figura X. 53. Cuantiles GEV/GML-CD de caudales máximos y sus límites de confianza de 95% asociados a diferentes tiempos de retorno y caudales observados asociados a sus respectivas probabilidades de excedencia calculadas para la serie 1875-2005.

Tiempo de retorno de la crecida de abril/mayo de 2003 (con informaciones hasta 2005)

De la misma forma en que se definió para la serie H+S disponible hasta el año 2002, en este caso se adoptó la combinación GEV-GMLE a los fines de estimar el tiempo de retorno de la crecida de abril/mayo de 2003, como así también sus intervalos de confianza.

En la Tabla X. 47 se presentan los resultados obtenidos al aplicarse los enfoques BCD y CD a la totalidad de las informaciones disponibles hasta el momento de concluirse este estudio (agosto 2005).

La figura permite apreciar que la totalidad de los puntos considerados se encuentran dentro del intervalo de confianza del 95%. El caudal de 3.954 m³/s posee un tiempo de retorno estimado por la distribución GEV-GMLE-CD igual a 431 años. Sin embargo, cabe resaltar que dicho valor ya aparece como extremo superior del intervalo del 95% para un caudal con un tiempo de retorno de aproximadamente 125 años.

Tabla X. 47. Tiempos de retorno [años] correspondientes a caudales máximos del río Salado estimados mediante la distribución GEV-GMLE (informaciones hasta el año 2005).

Caudal considerado		Enfoque BCD	Enfoque CD
3.954	m ³ /s	393	431
3.010		138	150
2.500		72	76

5.5 Análisis Comparativo de los Estudios de Frecuencia Efectuados

En la Tabla X. 48 se presentan las estimaciones correspondientes al tiempo de retorno (o recurrencia) del caudal máximo correspondiente al año 2003. Estas estimaciones han sido ordenadas de acuerdo a dos alternativas básicas: (i) con base en las informaciones disponibles hasta el año 2002 y (ii) con base en las informaciones disponibles hasta la actualidad (agosto 2005).

Se observa que, dada su magnitud, la incorporación de la crecida del año 2003 reduce en forma sistemática las estimaciones de tiempos de retorno correspondientes a los caudales extremos máximos.

El análisis de frecuencia basado únicamente en las series cortas modernas (a partir de 1972) presenta algunos inconvenientes tales como:

- a) la series poseen muy corta longitud (23 y 26 datos para las series hasta los años 2002 y 2005 respectivamente), hecho que impide estimar recurrencias del orden de 100 años o superiores con un nivel de incertidumbre acotado y compatible con los requerimiento de obras hidráulicas de cierta relevancia;
- b) al poseer menos de 50 datos poseen tendencia a producir ajustes inadecuados de la distribución GEV, hecho que se materializa en valores extremadamente bajos del parámetro de forma. Ello se traduce en estimaciones de recurrencias sensiblemente bajas;
- c) por tratarse de períodos muy cortos las estimaciones basadas en ellas pueden estar sujetas a tendencias propias de las variaciones observadas en los procesos hidrológicos naturales. En este caso particular, podrían ser tendenciosas por el hecho de basarse en un período húmedo. La falta de representatividad de una serie para estimar altos tiempos de retorno es un aspecto crucial del análisis de crecidas. En las zonas circundantes a la cuenca del Plata se han observado casos de falta de representatividad. Uno de ellos es el relatado por Tucci y Krebs (1986) para el río Itajaí (SC, Brasil). En la estación Blumenau se observó en el año 1852 un nivel máximo del río igual a 16,30 m. En 1888 se registró un máximo de 17,10 m. En el año 1911 la crecida alcanzó un nivel de 16.90 m. A partir de allí y por un período de 71 años el nivel no registró valores superiores a 13,50 m. Sin embargo, en 1983

el nivel volvió a superar dicha marca, al alcanzar 15,34 m. En 1984 alcanzó 15.50 m. En el caso del río Salado se produjo un período de 59 años relativamente secos (entre 1914 y 1973), con crecidas que rara vez superaron los 1.000 m³/s. En 1914 y en 1973 se registraron caudales máximos que superaron 2.400 m³/s. La secuencia de años húmedos a partir de 1973 es bien conocida. Se desprende que un análisis de frecuencia considerando un corto período de mínima (o de máxima) puede conducir a conclusiones inadecuadas por falta de representatividad de la serie.

Otro aspecto a destacar es que, en términos generales, la variación de la recurrencia de los caudales extremos máximos registrada en todas las alternativas analizadas (debido a la incorporación de la crecida del año 2003), es inferior a aquella indicada en estudios de frecuencia preexistentes.

La incorporación de marcas históricas permite extender la serie de análisis de modo de validar la inferencia acerca del comportamiento estadístico de altos valores de caudal. En otras palabras, con una serie de 130 años de extensión es más razonable realizar inferencias estadísticas hasta del orden de los 400 a 500 años.

Se ha observado que la incorporación de marcas históricas conduce a reducir las variaciones porcentuales del tiempo de retorno de los caudales extremos máximos. A su vez, conduce a reducir el intervalo de incertidumbre asociado a la estimación de diferentes cuantiles.

De los análisis efectuados se desprende que la distribución GEV se perfiló como aquella con mejor capacidad para representar las diferentes distribuciones empíricas consideradas al incorporarse la totalidad de las informaciones disponibles. Esta distribución es indicada en la literatura científica actualizada como una de las más aconsejadas para representar caudales extremos máximos.

Considerando la capacidad del método GMLE para el ajuste de parámetros de la distribución GEV empleando marcas históricas y umbrales de percepción, se concluye que:

- a) el empleo de todas las informaciones disponibles con anterioridad a la crecida máxima del año 2003 podría haber conducido a una estimación del tiempo de retorno de la misma del orden de 810 años. La probabilidad de ocurrencia anual correspondiente es del 0,12. Dicho de otro modo, un caudal de 3.954 m³/s corresponde al valor esperado de la crecida de 810 años. De todas maneras, es importante considerar que dicho caudal corresponde también al extremo superior del intervalo de confianza del 95 % para un caudal con recurrencia del orden de 180 años.
- b) con base en todas las informaciones que se disponen hasta la actualidad se estima que el tiempo de retorno o recurrencia de dicha crecida es del orden de 430 años. Ello implica que la misma posee una probabilidad de ocurrencia en un año cualquiera del 0,23 %.

La incorporación de nuevas informaciones que posean un grado de confianza aceptable podrá conducir a ajustes futuros de las probabilidades de ocurrencia de caudales máximos del río Salado y, por ende, de sus tiempos de retorno (o recurrencia).

Hasta tanto ello ocurra y a los efectos de considerar la previsibilidad estadística de crecidas máximas en el río Salado se sugiere la consideración de las estimaciones previamente indicadas.

Tabla X. 48. Recurrencias estimadas para la crecida de 2003 ($Q= 3.954 \text{ m}^3/\text{s}$) según distintas series de análisis.

Serie de análisis	Distribución de mejor ajuste	Método de ajuste	Tiempo de retorno [años]	Observaciones
Estimaciones basadas en las informaciones disponibles hasta el año 2002				
1954-2002 (S)	Log Pearson III	SAM	400	Estandarización (M y V)
1972-2002 (S)	GEV	MLE	109 (*)	(*): Ajuste teórico inadecuado
1928-2002 (R+S)	GEV	MLE	425	Estandarización (M)
1875-2002 (H+S)	GEV	GMLE	810	Análisis CD
Estimaciones basadas en las informaciones disponibles hasta el año 2005				
1954-2005 (S)	Pearson III	MLE	106	Estandarización (M y V)
1972-2005 (S)	GEV	MLE	46 (*)	(*): Ajuste teórico inadecuado
1928-2005 (R+S)	GEV	MLE	110	Estandarización (M)
1875-2005 (H+S)	GEV	GMLE	430	Análisis CD
Observaciones:				
S = serie sistemática ; R+S = serie reconstituida y sistemática ; H + S = serie de marcas históricas y sistemática				
Estand. (M y V) = estandarización de la serie por media y varianza				
Estand. (M) = estandarización de la serie por media				
Análisis CD = análisis basado en “ <i>censored data</i> ”				
SAM = método de estimación de parámetros mediante la metodología “ <i>Sundry Average Method</i> ” (o método de las tres medias, Bobée y Ashkar, 1991).				
MLE = método de estimación de parámetros por máxima verosimilitud				
GMLE = método de estimación de parámetros por máxima verosimilitud generalizada				

6 MAGNITUD COMPARADA DE LA CRECIDA DEL AÑO 2003

A los efectos de contar con un elemento adicional de comparación para cuantificar la magnitud de la crecida del río Salado en abril/mayo de 2003, fueron analizados los registros máximos mundiales y los estudios que sobre ellos describe la literatura. En particular, fueron analizados los trabajos de Francou y Rodier (1997) y Berga (2002).

En el primero de ellos fueron clasificados los registros de crecidas máximas a nivel mundial, estableciéndose la siguiente relación entre el área de la cuenca y los caudales máximos de crecida:

$$\frac{Q}{Q_0} = \left[\frac{A}{A_0} \right]^{(1 - \frac{K}{10})} \quad (\text{X. 52})$$

siendo Q el caudal máximo registrado, A el área de la cuenca, Q₀ y A₀ valores de referencia para el caudal y el área, y k el coeficiente de Francou-Rodier estimado según la siguiente expresión:

$$k = 10 * \left[1 - \frac{\log Q - 6}{\log A - 8} \right] \quad (\text{X. 53})$$

Los autores determinaron que el valor de k que corresponde a la curva envolvente de los caudales máximos corresponde a k = 6,4 (Figura X. 54).

Para la crecida de abril/mayo de 2003 corresponden los siguientes valores:

Q – caudal máximo (m³/s)= 3954

A – area de la cuenca (km²)= 41528

K - coeficiente de Francou-Rodier = 2.89

Q₀ = 10⁶ m³/s

A₀ = 10⁸ km²

Q específico= 0.095 m³/s/km²

Según se desprende del análisis, el valor de k correspondiente a la crecida del año 2003 es sustancialmente inferior al valor extremo del coeficiente correspondiente a la envolvente. Por lo tanto, se concluye que la crecida del río Salado del año 2003, si bien ha sido la máxima histórica registrada en la cuenca, no constituye una marca de gran significación a nivel mundial. En la Figura X. 55 se compara la posición del punto que representa a las condiciones del río Salado en relación a la curva envolvente a nivel mundial.

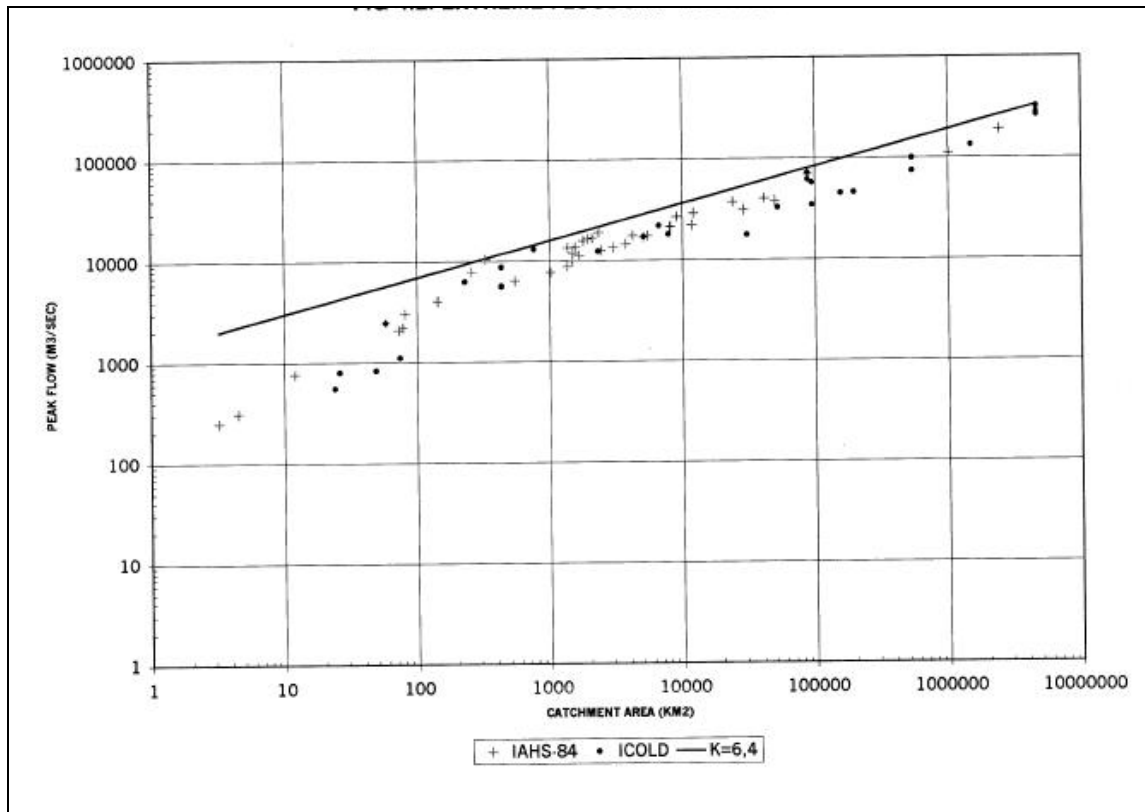


Figura X. 54. Curva envolvente de los caudales extremos de crecidas registrados en el mundo, en función del área de la cuenca de aporte.

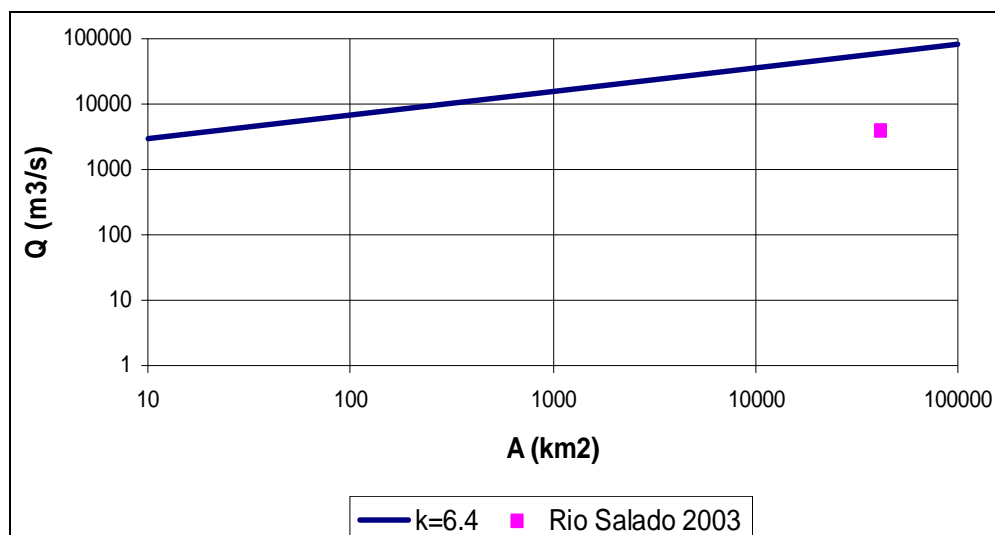


Figura X. 55. Magnitud comparada de la crecida del año 2003 con los máximos registrados a nivel mundial.

7 ANÁLISIS DEL RIESGO DE FALLA DE LA OBRA DE DEFENSA OESTE

7.1 Introducción

En toda decisión que se adopta en un proyecto de ingeniería, existe implícitamente una evaluación de riesgos.

En efecto, en una situación teórica y abstracta, si los recursos humanos y materiales, así como el tiempo, surgieran de una fuente inagotable, no existiría este problema. Pero en cambio, al ser siempre limitados los recursos y el tiempo, se debe adoptar una solución de compromiso entre costo y plazo versus seguridad.

Por simplicidad se puede considerar que mientras más costosa sea la obra y más tiempo se necesite para construirla, más segura será la misma. A la inversa, menor costo y menor plazo darán menor seguridad. A través de la evaluación de riesgos, se puede encontrar el punto óptimo entre ambos extremos.

Un ejemplo mundano es el de una persona que cruza una calle transitada y sin semáforos.

Podrá hacerlo sin mirar, corriendo el máximo riesgo pero sin demorar ni un segundo, ó podrá hacerlo en el momento que no haya ningún auto hasta donde su vista alcanza a observar, corriendo el mínimo riesgo, lo cual podrá requerir un considerable tiempo de espera.

Normalmente, se toma una decisión intermedia y se cruza observando y midiendo la distancia a los vehículos que se aproximan de uno y otro lado para evitar ser arrollado. El resultado es entonces una solución de compromiso.

En las obras de ingeniería, y muy especialmente en las que involucran aspectos hidráulicos, es fundamental la evaluación de riesgos.

En las obras hidráulicas, el comportamiento de los ríos no es totalmente definido ni manejable por el hombre en forma directa, por lo tanto existe un riesgo y éste consiste en que arribe un caudal superior al previsto en el punto en que se ejecuta la obra.

Es allí cuando el ingeniero debe evaluar los riesgos y tomar decisiones fundamentales, pues en general las obras hidráulicas implican elevados costos y plazos, siendo impracticable una solución de “riesgo cero”. Las soluciones de riesgo cero prácticamente no existen dentro y fuera de la ingeniería. De allí se desprende la importancia del análisis de riesgo de falla, en particular, en el caso de una obra hidráulica.

Este riesgo se analiza, en primer lugar, determinando cuál es la probabilidad de ocurrencia de un caudal que superara el valor de proyecto (o de diseño). Esta probabilidad, conforme fuera analizado en los apartados precedentes, surge de consideraciones hidrológicas y estadísticas.

A mayor probabilidad de ocurrencia de un caudal crítico, mayor será el riesgo de falla de la obra.

En segundo lugar, cuenta también en el análisis la evaluación del tiempo durante el cual la obra estará expuesta a ese riesgo. Es decir, cuenta la vida útil de la obra. A mayor exposición, mayor riesgo.

La adopción del caudal de proyecto (y por ende de su tiempo de retorno, o de su probabilidad de ocurrencia anual), surge de evaluar la magnitud de los daños que se producirían si ese caudal (o uno superior al mismo) se presenta durante la vida útil de la obra.

Corresponde abrir aquí un paréntesis y tomar nuevamente un ejemplo cualquiera: la falla de un medidor de combustible en un automóvil, puede no ser grave, en cambio en un avión puede ser fatal. Sin embargo la falla es básicamente la misma. Por eso los sistemas que evitan la falla son tan diferentes en uno y otro caso.

En el caso de una defensa costera de protección de una ciudad la magnitud de los daños generalmente es elevada por la existencia de zonas habitadas próximas. Por lo tanto, la relación entre probabilidad de ocurrencia de una crecida superior a la de proyecto, la vida útil de la obra y el riesgo de falla de la misma deben ser adecuadamente relacionados a la hora de proyectarse la obra.

7.2 Análisis Estadístico Clásico del Riesgo de Falla

En los apartados precedentes se concluyó que la crecida de abril/mayo de 2003 poseía un tiempo de retorno (o recurrencia) del orden de 810 años antes de su ocurrencia. Ello implica que la misma poseía una probabilidad de ocurrencia en un año cualquiera del 0,12 %.

La asignación de una vida útil de 100 años es habitual en obras hidráulicas de defensa ligadas a proyectos viales.

A partir de la ecuación (X.2) presentada en el inicio de este Anexo, se deduce que el riesgo hídrico de falla asumido al proyectarse una obra hidráulica con 100 años de vida útil y capaz de soportar hasta una crecida de 810 años de recurrencia, es del 12 %.

En otras palabras, si una obra de defensa es proyectada para resistir una crecida que posee una probabilidad anual igual al 0,12 %, dentro del período de 100 años existe el 12 % de probabilidad de que la obra falle por ocurrencia de una crecida superior a aquella considerada en su diseño.

De considerarse una vida útil de la obra igual a 500 años, el riesgo de falla sería del 67 %. Este resultado es considerado “clásico” dentro de los análisis de hidrología estadística, toda vez que en este caso el tiempo de retorno y la de vida útil de la obra son prácticamente similares.

De considerarse el riesgo asociado a un problema específico de una obra, tal como por ejemplo, una abertura en el sistema de defensas de una ciudad, lo apropiado es considerar el tiempo de exposición del problema y ya no la vida útil general de la obra.

El análisis precedente, si bien permite introducir el tema de probabilidad y riesgo de falla, no resulta estrictamente aplicable al caso de la defensa oeste de la ciudad de Santa Fe.

En efecto, en los análisis precedentes se han supuesto, implícitamente, dos hipótesis:

- c) la probabilidad de ocurrencia de una crecida del río Salado se relaciona con el riesgo hídrico de falla según la ecuación (X.2);
- d) la obra falla únicamente al ocurrir una crecida igual o superior a aquella de diseño.

Ambos hipótesis no son de cumplimiento estricto en la realidad, razón por la cual los análisis anteriores deben ser profundizados. Un análisis más detallado de estos aspectos se presenta en los ítems siguientes.

7.3 Probabilidad de Crecidas del Río Salado Condicionadas al Sistema Paraná

7.3.1 Análisis de la condicionalidad de los niveles

Desde el punto de vista hidrológico las cuencas de aportes de los ríos Salado y Paraná operan de manera independiente. En otras palabras, la ocurrencia de crecidas en una de ellas es independiente de la ocurrencia de crecidas en la otra.

Sin embargo, en la región de la desembocadura del río Salado en el sistema Paraná, los niveles de ambos ríos se encuentran relacionados entre sí. Por su dimensión y permanencia se destaca fundamentalmente la influencia del sistema Paraná sobre el río Salado.

De acuerdo al material presentado en el Anexo IX, la influencia de los niveles del sistema Paraná sobre el cauce del río Salado puede notarse hasta en secciones localizadas aguas arriba de la estación RP70. La extensión y magnitud de esta influencia depende tanto del nivel del río Paraná como de la magnitud de los caudales del río Salado.

En definitiva, la ocurrencia de una crecida del río Salado asociada a un cierto tiempo de retorno (o recurrencia) no implica necesariamente la ocurrencia de un determinado nivel del agua en el tramo próximo a la ciudad de Santa Fe. El nivel del agua en el río Salado en el tramo próximo a la ciudad de Santa Fe depende, además, de la altura que presente en esos momentos el sistema del río Paraná en las inmediaciones de la desembocadura del río Salado.

En tal caso no es posible analizar el riesgo de falla de una obra hidráulica mediante la consideración de probabilidades del río Salado independientes de los niveles del Paraná.

La cuestión obliga a abordar el problema mediante la teoría de la probabilidad condicional.

Admítase que el suceso A representa la ocurrencia de una crecida del río Salado. Supóngase, a su vez, que el suceso B representa la ocurrencia de un cierto nivel del sistema Paraná.

La probabilidad (anual) de ocurrencia de un caudal superior a un cierto valor se indica entonces como P(A). Del mismo modo, la probabilidad (anual) de ocurrencia de un nivel superior a un cierto valor se indica como P(B).

El problema analizado requiere considerar la probabilidad P(A/B). La misma se interpreta como la probabilidad de ocurrencia de una crecida en el río Salado, condicionada a la ocurrencia, en concomitancia en el tiempo, de un cierto nivel en el sistema Paraná (o de uno superior; suceso B).

La teoría estadística indica que la P(A/B) puede estimarse como:

$$P(A/B) = \frac{P(A \wedge B)}{P(B)} \quad (\text{X. 54})$$

donde $P(A \wedge B)$ representa la probabilidad anual de la ocurrencia de A y B.

Para llevar a cabo el análisis fueron consideradas las series diarias de caudales del río Salado y de niveles de la estación Colastiné, ambas para el período 1928-2005. Este período fue seleccionado debido a la disponibilidad conjunta de ambos tipos de informaciones. La selección de la estación Colastiné como representativa del sistema Paraná (disponible al momento de realizarse el estudio) se justificó debido tanto a su proximidad con el área de estudio como por su alta correlación con los niveles de la serie Setúbal (el estado del sistema Paraná en las proximidades a la desembocadura del río Salado actualmente se registra en el Dique II del Puerto Santa Fe). Las simplificaciones introducidas fueron consideradas compatibles con el objetivo del análisis estadístico.

7.3.2 Clasificación de situaciones típicas: niveles y caudales altos y bajos (a nivel diario)

Con base en las informaciones antes citadas y a los efectos de caracterizar el comportamiento conjunto de ambos ríos, fueron analizadas las situaciones derivadas de la ocurrencia diaria conjunta de niveles altos (o bajos) en el sistema Paraná con crecidas altas (o bajas) del Salado.

El trabajo realizado se basó en dos suposiciones básicas simplificadoras:

- a) fueron considerados “altos” los niveles del sistema Paraná (Colastiné) superiores a 13,57 m;
- b) fueron consideradas “altas” las crecidas del río Salado superiores a 1.700 m³/s.

La justificación de la adopción de estos umbrales se basó en las siguientes consideraciones:

- El valor de nivel igual a 13,57 m corresponde a aquel que se observó durante la ocurrencia del caudal pico de la crecida del año 2003 del río Salado. Cabe consignar que este valor del umbral representa, en términos prácticos, el promedio de la serie de niveles (el mismo es, en realidad, igual a 13,55 msnm). Del estudio de frecuencia de niveles del sistema Paraná se dedujo que este último nivel se asocia a una recurrencia de 2,5 años;
- El umbral de crecidas del río Salado igual a 1.700 m³/s fue considerado en todos los análisis estadísticos que contemplaron la incorporación de información histórica. Dicho caudal posee un tiempo de retorno estimado del orden de 28 años (serie S+H hasta el año 2002);

En la Tabla X. 49 se presentan los resultados obtenidos. La misma permite observar que la frecuencia relativa (o probabilidad empírica) de que ocurran niveles altos en el Paraná es del 8,50 %. Del mismo modo, la probabilidad de ocurrencia de crecidas altas en el Salado es del 0,24 %. Por su parte, la probabilidad empírica de obtener un día en el se verifiquen conjuntamente niveles altos en el Paraná y caudales altos en el Salado es del 0,11 %. Se desprende de las informaciones generadas que la situación más probable es la ocurrencia conjunta de niveles bajos en el sistema Paraná y de caudales bajos en el río Salado: la misma se presenta el 91,32 % de las veces.

La Tabla permite deducir también que las probabilidades empíricas de ocurrencia de crecidas altas del río Salado con niveles del río Paraná respectivamente altos o bajos son prácticamente similares (0,11 y 0,13 respectivamente). Las probabilidades obtenidas permiten realizar estimaciones de probabilidades condicionales.

Por ejemplo, la probabilidad de que se verifique un día con crecidas altas del río Salado, condicionada a la ocurrencia conjunta de un nivel alto en el río Paraná resulta igual a 1,29 %. De considerarse, sin embargo, que las probabilidades empíricas indicadas fueron obtenidas a nivel diario y no representan las probabilidades máximas anuales requeridas en el estudio del riesgo hídrico.

Tabla X. 49. Frecuencia de ocurrencia (en %) de distintas situaciones diarias conjuntas en los ríos Paraná-Salado (base: período 1928-2004).

Situaciones analizadas		Niveles del Sistema Paraná		Frec. totales referidos a crecidas
		“Altos”	“Bajos”	
Crecidas del río Salado	“Altas”	0,11	0,13	0,24
	“Bajas”	8,43	91,32	99,76
Frec. totales referidos a niveles		8,50	91,50	100

Observaciones:
 Nivel del sistema Paraná considerado en la clasificación: 13,57 msnm
 Caudal máximo medio diario del río Salado considerado en la clasificación: 1.700 m³/s

Estimación de las probabilidades condicionales a nivel anual

Para la estimación de las probabilidades condicionales a nivel anual se consideró la situación de mayor interés: la asociada a la ocurrencia de un nivel del sistema Paraná igual o superior a 13,57 msnm.

Cabe consignar que dicho valor del umbral representa, en términos prácticos, el promedio de la serie de niveles (13,55 msnm). Como ya fuera citado, del estudio de frecuencia de niveles del sistema Paraná se dedujo que este último nivel se asocia a una recurrencia de 2,5 años. Al tratarse de un nivel relativamente bajo (habitualmente igualado o superado por el sistema del río Paraná), el mismo implica, en términos prácticos, una condicionalidad muy reducida. Desde el punto de vista físico ello equivale a indicar que los niveles del río Salado en las inmediaciones de la ciudad de Santa Fe se encuentran, en tales circunstancias, con baja influencia del sistema Paraná. Desde el punto de vista estadístico equivale a indicar que la $P(A/B)$ no es muy diferente de aquella oportunamente estimada, $P(A)$.

Efectivamente, luego del estudio de frecuencia de máximos anuales correspondiente mediante la distribución GEV se concluyó en que, en términos prácticos, dicha situación no discrepa significativamente de aquella originalmente estudiada en relación a la estación de la RP70 para caudales del río Salado comprendido entre 2.000 m³/s y 4000 m³/s. En otras palabras, para el caso particular de un umbral relativamente bajo del sistema Paraná, lo cual implica considerar para el suceso B: *la ocurrencia de niveles en el río Paraná iguales o superiores a 13,57 msnm*, se concluye que, para crecidas entre 2.000 m³/s y 4.000 m³/s:

$$P(A/B) = P(A/13,57m) \approx P(A) \quad (X. 55)$$

siendo $P(A)$ la probabilidad de excedencia anual de los caudales del río Salado en RP70.

Cabe señalar que una explicación adicional de la aproximación arriba indicada está dada por el hecho de $P(A)$ haber sido estimada, implícitamente, para un nivel condicional del sistema Paraná del orden de 11.73 m (el mínimo valor del sistema Paraná en los años para los cuales se dispuso de registros durante este estudio).

7.4 Análisis de la Vulnerabilidad Hídrica del Tramo II del Terraplén Oeste

La suposición de falla de una obra de defensa habitualmente deriva de admitirse la superación del nivel de coronamiento del terraplén y/o por efectos de tubificación, volcamiento, etc. En todos los casos la variable fundamental en juego es la altura máxima del agua frente a la obra.

En el caso del Tramo II la falla de la obra de defensa no se asoció a la superación de su nivel de coronamiento sino al hecho de haber ingresado el agua a la ciudad. En efecto, si se considera que el objetivo fundamental de la obra de defensa era evitar el ingreso del agua a la ciudad, se concluye que el fracaso de la misma se produjo al ingresar el agua al recinto urbano. Por lo tanto, el riesgo de falla de la obra de defensa se asocia en forma directa a la

probabilidad del ingreso del agua por el extremo norte del tramo II, localizado sobre calle Gorostiaga.

Al considerarse dicho riesgo, lo apropiado es considerar el tiempo de exposición del problema (el no cierre del sistema de defensas de la ciudad) y ya no la vida útil general de la obra.

El análisis del riesgo del ingreso del agua por calle Gorostiaga para una situación semejante a la ocurrida durante la crecida del año 2003 se presenta en el Anexo XI.

8 CONCLUSIONES

Diversos estudios de frecuencia, basados en diferentes series de registros, fueron encarados con el fin de definir la previsibilidad estadística de la crecida del río Salado de abril/mayo de 2003 (o, dicho en otras palabras, el *tiempo de recurrencia* o *tiempo de retorno* de la citada crecida), como así también las probabilidades a considerar en el análisis de riesgo de falla de la obra de Defensa Oeste de la ciudad de Santa Fe.

Los datos considerados fueron:

- a) la serie sistemática de caudales correspondiente al período 1954-2002, registrada en la sección de medición de la Ruta Provincial 70 (RP70);
- b) la serie correspondiente al período 1928-1954, basada en la combinación de la serie sistemática antedicha y en la serie de caudales reconstituidos a partir del análisis de los registros existentes en la sección de medición RP70 y en la sección del puente Empalme San Carlos (ESC);
- c) la serie correspondiente al período 1875-2002, la cual combinó las marcas hidrológicas de crecidas históricas que superaron un umbral (o nivel) de percepción y las series antedichas.

Una extensión de todos estos estudios se realizó al incorporar las informaciones disponibles hasta el año 2005. El análisis comparativo permitió analizar el impacto que la crecida del año 2003 tuvo sobre los niveles de previsibilidad estadística de grandes crecidas.

Entre los aspectos de consideración asociados a los análisis realizados, se destacan:

- a) la incorporación de la crecida del año 2003 redujo en forma sistemática las estimaciones de tiempos de retorno correspondientes a los caudales extremos máximos deducidos con las distintas técnicas consideradas en este estudio. Sin embargo, las reducciones observadas fueron menores, en términos relativos, a aquellas citadas por estudios preexistentes;
- b) se observó que el análisis de frecuencia basado únicamente en las series cortas modernas (a partir de 1972) presenta diversos inconvenientes. Entre ellos cabe señalar la representatividad de los registros y la limitación para extender los análisis a recurrencias del orden de 100 años o superiores. La falta de representatividad de una serie para estimar altos tiempos de retorno es un aspecto crucial del análisis de crecidas. La literatura científica relata casos de falta de representatividad en zonas circundantes a la cuenca del Plata.

El abordaje adoptado para el empleo de las informaciones históricas se basó en el empleo de la distribución de valor extremo generalizada (GEV) de tres parámetros sobre las series anuales de valores máximos. La metodología aplicada se encuentra respaldada por un

importante conjunto de publicaciones científicas a nivel internacional. Para el ajuste de la distribución GEV fue empleado el método GMLE recomendado en la literatura científica. El mismo permitió estimar la recurrencia de la crecida objeto del análisis con mayor precisión que los restantes métodos ensayados. Entre estos últimos, se destacó la aplicación de la distribución Log Pearson III con base en un enfoque de tipo *bayesiano*.

Los estudios realizados permiten afirmar que:

- c) conforme citado en el Anexo IX, la crecida del año 2003 ($3.954 \text{ m}^3/\text{s}$) fue la mayor dentro del período 1875-2005. Aún así, se verificó que existían elementos con anterioridad al año 2003 que permitían la ocurrencia estadística de la misma.
- d) con base en la capacidad del método GMLE para el ajuste de parámetros de la distribución GEV empleando marcas históricas y umbrales de percepción, se concluye que el empleo de todas las informaciones disponibles con anterioridad a la crecida máxima del año 2003 podría haber conducido a estimar que un caudal de $3.954 \text{ m}^3/\text{s}$ (o uno superior al mismo) correspondía al valor más probable para la crecida que poseía un tiempo de retorno o recurrencia del orden de 810 años. Ello implica que el mismo poseía una probabilidad de ocurrencia en un año cualquiera del 0,12 %. Sin embargo, también es importante destacar que, alternativamente, dicho caudal podría haber sido previsto estadísticamente como aquel correspondiente al extremo superior del intervalo de confianza del 95 % para un caudal con recurrencia del orden de 180 años.
- e) con base en todas las informaciones que se disponen hasta la actualidad se estima que un caudal de $3.954 \text{ m}^3/\text{s}$ (o uno superior al mismo) corresponde al valor más probable para la crecida que posee un tiempo de retorno o recurrencia del orden de 430 años. Ello implica que el mismo posee una probabilidad de ocurrencia en un año cualquiera del 0,23 %. No obstante ello, es importante remarcar también que dicho caudal corresponde, a su vez, al límite superior del intervalo de confianza del 95 % para la crecida que posee una recurrencia del orden de 130 años.

Cabe resaltar que la incorporación de nuevas informaciones que posean un grado de confianza aceptable podrá conducir a ajustes futuros de las probabilidades de ocurrencia de caudales máximos del río Salado y, por ende, de sus tiempos de retorno (o recurrencia).

El análisis de las probabilidades de ocurrencia de ciertos niveles del río Salado en las inmediaciones de la ciudad de Santa Fe, condicionados a la ocurrencia de distintos niveles del sistema del río Paraná, fue realizado con base en el análisis estadístico de las series condicionales de caudales máximos anuales del río Salado y en el empleo de un modelo hidráulico.

El objetivo del estudio se centró en el nivel del sistema Paraná verificado en oportunidad de la ocurrencia del pico de la crecida de abril/mayo de 2003, es decir, 13,57 msnm. El análisis realizado permitió verificar que, por tratarse éste de un nivel relativamente bajo, la

condicionalidad impuesta por el sistema del río Paraná no fue significativa. Ello se tradujo en términos estadísticos en diferencias menores entre las probabilidades condicionales analizadas y la probabilidad de ocurrencia anual de los caudales máximos del río Salado en la sección RP70. Por lo tanto, se concluyó que para crecidas del río Salado comprendidas entre 2.000 m³/s y 4.000 m³/s son de aplicabilidad, a los fines del análisis del riesgo hídrico, las probabilidades generales antes mencionadas.

En síntesis, se concluye que en virtud de todos los antecedentes disponibles la crecida de abril/mayo de 2003 correspondió a un evento extremo, estadísticamente esperable, y de características excepcionales.