

INTI

Instituto Nacional
de Tecnología Industrial



Centro de Investigación de los Reglamentos Nacionales
de Seguridad para las Obras Civiles del Sistema INTI

ESTIMACION DE VELOCIDADES DE RECURRENCIA DE VIENTOS BASADA EN VELOCIDADES MAXIMAS MENSUALES

DR. ALDO J. VIOLLAZ
LIC. STELLA M. SALVATIERRA
ING. ALICIA M. ARAGNO

Julio 1995

BIBLIOTECA DE APOYO

CIA SOC

ORGANISMOS PROMOTORES

Secretaría de Obras Públicas y Comunicaciones
Instituto Nacional de Tecnología Industrial
Secretaría de Estado de Desarrollo Urbano y Vivienda
Municipalidad de la Ciudad de Buenos Aires
Comisión Nacional de Energía Atómica

MIEMBRO ADHERENTE

Consejo Interprovincial de Ministros de Obras Públicas

AUTORES

ALDO J. VIOLLAZ

- Ph.D en Estadística, Universidad de California.
Ingeniero Mecánico, Universidad Nacional de Tucumán.
- Profesor Titular de Probabilidad y Estadística,
Facultad de Ciencias Exactas y Tecnología, Universidad
Nacional de Tucumán.
- Investigador Independiente del Consejo Nacional de
Investigaciones Científicas y Técnicas.

STELLA MARIS SALVATIERRA

- Magister en Estadística Aplicada,
Licenciada en Matemática, Universidad Nacional de Tucumán.
- Profesora Adjunta de Probabilidad y Estadística,
Facultad de Ciencias Exactas y Tecnología, Universidad
Nacional de Tucumán.

ALICIA ARAGNO

- Ingeniera Civil, a cargo del Area Acciones del Centro de
Reglamentos Nacionales de Seguridad para las Obras Civiles.

PROLOGO

La publicación de la **Estimación de Velocidades de Recurrencia de Vientos basada en Velocidades Máximas Mensuales**, realizada por el Dr. Aldo J. Viollaz, la Lic. Stella M. Salvatierra y la Ing. Alicia Aragno, tiene por objeto ampliar y dar un nuevo enfoque al tratamiento de las velocidades máximas de viento en la República Argentina.

En este trabajo se desarrolla la metodología estadística, que permite utilizar como observaciones básicas las velocidades máximas mensuales, suministradas por el Servicio Meteorológico Nacional, para el modelado de series de vientos con una componente estacional determinística.

El **CIRSOC** agradece a los autores la autorización para publicar este trabajo, que será un nuevo y valioso aporte para la comprensión del **Reglamento CIRSOC 102 "Acción del Viento sobre las Construcciones"**.

**Dirección Técnica
CIRSOC**

ESTIMACION DE VELOCIDADES DE RECURRENCIA DE VIENTOS **BASADA EN VELOCIDADES MAXIMAS MENSUALES**

Aldo Viollaz¹, Stella Salvatierra² y Alicia Aragno³

RESUMEN

En muchas aplicaciones de la Ingeniería la variable de interés es por lo general el máximo de la variable, considerada durante la vida útil de la estructura que está siendo diseñada.

Los registros de velocidades del viento de las estaciones meteorológicas pueden contener datos de velocidades máximas diarias, mensuales y hasta pueden proporcionar registros continuos. Estas series de observaciones poseen por lo general una componente estacional importante.

La metodología estadística clásica para la estimación de los parámetros de la distribución se basa en observaciones independientes e idénticamente distribuidas (i.i.d.). Para poder aplicar el modelo i.i.d., suelen tomarse máximos anuales como datos básicos eliminándose así la componente estacional pero perdiendo cierta cantidad de información.

En este trabajo realizamos la aplicación de la metodología desarrollada por Viollaz y Salvatierra para el modelado de series de vientos con una componente estacional determinística, utilizando como observaciones básicas a las velocidades máximas mensuales.

¹Universidad Nacional de Tucumán y Consejo Nacional de Investigaciones Científicas y Técnicas.

²Universidad Nacional de Tucumán.

³Centro de Investigación de Reglamentos de Seguridad para las Obras Civiles.

1.- INTRODUCCION

Las series de observaciones de variables meteorológicas tales como temperatura, velocidad del viento, cantidad de lluvia caída y de otras variables dependientes de variables meteorológicas, poseen por lo general una componente estacional importante.

En muchas aplicaciones de la Ingeniería la variable de interés es por lo general el máximo de la variable considerada durante la vida útil de la estructura que está siendo diseñada.

En este trabajo nosotros consideramos el problema de estimar las velocidades máximas del viento que tienen períodos de recurrencia de varias decenas de años.

Los registros de velocidades del viento de las estaciones meteorológicas pueden contener datos de velocidades máximas diarias, mensuales y hasta pueden proporcionar registros continuos.

La aplicación de modelos de Estadística de Extremos: la distribución de Gumbel y la de Frechet, requieren por lo general disponer de observaciones independientes e idénticamente distribuídas con el fin de poder estimar los parámetros de la distribución adoptada, para después calcular las velocidades de recurrencia.

Una forma de obtener observaciones independientes e idénticamente distribuídas consiste en tomar como variables básicas al máximo anual de la velocidad del viento. Así se elimina la componente estacional.

Es más o menos obvio que esta anualización de los datos produce cierta pérdida de información. En Salvatierra y Viollaz (1993) y Salvatierra (1994) se estudia la cuantía de esta pérdida, se demuestra que ésta es importante y se propone una metodología para estimar las velocidades de recurrencia del viento en base a velocidades máximas

mensuales.

En este trabajo se presenta esa metodología estadística y se la aplica para la estimación de los parámetros de la distribución del máximo anual del viento en quince estaciones de la República Argentina. En base a estas estimaciones se estiman las velocidades de recurrencia del viento para períodos de 20, 100 y 200 años.

En el trabajo se estudia la bondad de ajuste del modelo, encontrándose que el supuesto de distribución Gumbel es muy razonable.

2.- EL MODELO

Consideremos la matriz de observaciones:

$$\begin{array}{ccccccc} X_{1,1}, & X_{1,2}, & \dots, & X_{1,12} \\ X_{2,1}, & X_{2,2}, & \dots, & X_{2,12} \\ \dots & \dots & \dots & \dots \\ X_{n,1}, & X_{n,2}, & \dots, & X_{n,12} \end{array}$$

donde la fila i -ésima contiene las velocidades máximas mensuales del viento correspondiente al año i y n es el número de años observados.

Es obvio que la velocidad del viento contiene una componente estacional importante lo cual es un aspecto que incorporaremos al modelo.

La velocidad del viento en un lugar determinado se puede considerar como una variable aleatoria que depende del tiempo.

La primera hipótesis básica que hacemos es que la distribución de probabilidad de esta variable es función del día del año pero es independiente del año particular considerado. Obviamente esta hipótesis es sólo una primera aproximación pues es bien conocido que la actividad solar, la

cual tiene influencia sobre el clima, no es uniforme y presenta un cuasi período de aproximadamente once años. Esta hipótesis básica implica que las velocidades del viento correspondientes a un mismo mes j :

$$X_{ij}, \quad i=1, \dots, n$$

son variables aleatorias idénticamente distribuidas.

La segunda hipótesis que hacemos es que la distribución común de las variables X_{ij} , $i=1, \dots, n$, es Gumbel de parámetros ξ_j y θ , esto es, la correspondiente función de distribución está dada por:

$$G(x; \xi_j, \theta) = \exp\{-\exp[-(x-\xi_j)/\theta]\}$$

La tercera hipótesis básica es que las variables X_{ij} , $i=1, \dots, n$, $j=1, \dots, 12$, son independientes.

Para cada $i=1, \dots, n$ definimos el máximo durante el año i , por:

$$Y_i = \max\{X_{i1}, \dots, X_{i12}\}$$

Por la primera hipótesis básica los máximos anuales Y_i resultarán variables aleatorias independientes e idénticamente distribuidas. Llamando F_Y a la función de distribución común, se tiene que:

$$\begin{aligned} F_Y(y) &= \exp\left(-\sum_{j=1}^{12} \exp[-(y-\xi_j)/\theta]\right) \\ &= \exp\left(-\exp\left[-(y-\theta \ln \sum_{j=1}^{12} e^{\xi_j/\theta})/\theta\right]\right) \end{aligned}$$

lo que implica que cada Y_i tiene distribución de Gumbel de parámetros:

$$\xi_Y = \theta \ln \sum_{j=1}^{12} e^{\xi_j/\theta}$$

$$\theta_Y = \theta$$

La distribución F_Y del máximo anual es fundamental pues las velocidades de recurrencia se obtienen simplemente calculando los correspondientes percentiles superiores de la distribución.

Los parámetros ξ_Y y θ_Y de F_Y se pueden estimar en base a los máximos anuales usando la metodología clásica. Salvatierra y Viollaz (1993) y Salvatierra (1994) desarrollaron los métodos de estimación por el método de los momentos y por máxima verosimilitud en base a los máximos mensuales, mostrando que éstos son notablemente más eficientes que los correspondientes basados en la anualización de las observaciones.

En este trabajo sólo consideraremos el método de los momentos basado en máximos mensuales.

3.- ESTIMACION DE LOS PARAMETROS

En esta sección presentamos los estimadores por el método de los momentos de los parámetros ξ_Y y θ_Y del máximo anual.

De acuerdo a las hipótesis formuladas las observaciones máximas mensuales correspondientes al mes j ($j=1, \dots, 12$):

$$X_{1j}, X_{2j}, \dots, X_{nj}$$

son variables aleatorias independientes e idénticamente distribuidas con distribución Gumbel.

El estimador que tomaremos para $\theta = \theta_Y$ es:

$$\hat{\theta} = \frac{\sum_{j=1}^{12} \bar{X}_{pj}}{n}$$

donde

$$S_p^2 = \frac{1}{12} \sum_{j=1}^{12} S_j^2 = \frac{1}{12n} \sum_{i=1}^n \sum_{j=1}^{12} (X_{ij} - \bar{X}_{.j})^2, \quad \bar{X}_{.j} = \frac{1}{n} \sum_{i=1}^n X_{ij}$$

Análogamente ξ_j se estima por:

$$\xi_j = \bar{X}_{.j} - \gamma \theta$$

donde $\gamma = 0,57722\dots$ es la constante de Euler.

Finalmente los parámetros de la distribución del máximo anual se estiman por:

$$\begin{aligned} \theta_Y &= \theta = \frac{\sqrt{6}}{\pi} S_p \\ \xi_Y &= \theta \ln \sum_{j=1}^{12} \exp(\xi_j/\theta) \end{aligned} \quad (3.1)$$

Usando las expresiones (3.1) se han estimado los parámetros θ_Y y ξ_Y de la distribución del máximo anual correspondiente a quince estaciones de la República Argentina. Las varianzas y covarianza asintótica de los estimadores de los parámetros están dadas por (ver Salvatierra, 1994):

$$\begin{aligned} \text{Var}(\xi_Y) &\approx 1,100 \left(\ln \sum_{j=1}^k e^{\xi_j/\theta} - \frac{\sum_{j=1}^k \frac{\xi_j}{\theta} e^{\xi_j/\theta}}{\sum_{j=1}^k e^{\xi_j/\theta}} + 0,086 \right)^2 \frac{\theta^2}{nk} \\ &+ 1,645 \frac{\theta^2}{n} \frac{\sum_{j=1}^k e^{2\xi_j/\theta}}{\left(\sum_{j=1}^k e^{\xi_j/\theta} \right)^2} - 0,484 \frac{\theta^2}{nk} \end{aligned} \quad (3.2)$$

$$Var(\hat{\theta}) = \frac{6}{\pi^2} Var(S_p) \approx 1,100 \frac{\theta^2}{nk} \quad (3.2)$$

$$Cov(\hat{\xi}_Y, \hat{\theta}_Y) \approx 0,095 \frac{\theta^2}{nk} + 1,100 \left(\ln \sum_{j=1}^k e^{\xi_j/\theta} - \frac{\sum_{j=1}^k \frac{\xi_j}{\theta} e^{\xi_j/\theta}}{\sum_{j=1}^k e^{\xi_j/\theta}} \right) \frac{\theta^2}{nk}$$

En la Tabla 1 se muestran estas estimaciones y sus varianzas y covarianzas estimadas.

4.- CALCULO DE LAS VELOCIDADES DE RECURRENCIA

La velocidad de recurrencia correspondiente a un período de T años se ha obtenido por la expresión:

$$\hat{v}_T = \hat{\xi}_Y - \hat{\theta}_Y \ln \left(-\ln \left(1 - \frac{1}{T+1} \right) \right)$$

Usando las expresiones (3.2) se calcula la varianza asintótica de \hat{v}_T . Con estos valores se calcularon los valores de \hat{v}_T y sus errores estándar, correspondientes a períodos de T = 20, 100 y 200 años. Los resultados se muestran en la Tabla 2.

Estación	ξ_Y	θ_Y	$\hat{v}ar(\xi_Y)$	$\hat{v}ar(\theta_Y)$	$\hat{c}ov(\xi_Y, \theta_Y)$
Neuquén	63.87	7.029	1.800	0.238	0.600
Azul	52.869	6.904	1.691	0.230	0.551
Salta	41.017	5.528	1.080	0.147	0.339
San Miguel	44.446	5.045	0.902	0.123	0.296
Aeroparque	55.107	6.960	1.729	0.234	0.572
Villa Reynold	64.064	10.955	4.256	0.579	1.386
Junín	50.203	6.825	1.662	0.225	0.536
Ezeiza	51.612	6.775	1.614	0.221	0.528
Resistencia	52.796	6.447	1.464	0.201	0.479
San Juan	56.619	6.039	1.288	0.176	0.419
Córdoba	50.612	5.263	0.950	0.134	0.301
Trelew	73.974	8.990	2.941	0.390	0.978
Villa Dolores	51.686	6.572	1.519	0.208	0.489
9 de Julio	48.670	6.353	1.454	0.195	0.481
Lomitas	44.966	4.542	0.735	0.100	0.242

Tabla 1: Estimaciones de los parámetros de la distribución del máximo anual.

Estación	\hat{V}_{20}	\hat{V}_{100}	\hat{V}_{200}
Neuquén	85.103 (2.757)	96.279 (3.521)	101.133 (3.855)
Azul	73.721 (2.668)	84.697 (3.415)	89.466 (3.741)
Salta	57.713 (2.115)	66.502 (2.709)	70.320 (2.969)
San Miguel	59.683 (1.952)	67.704 (2.498)	71.189 (2.737)
Aeroparque	76.128 (2.704)	87.194 (3.459)	92.001 (3.789)
Villa Reynold	97.151 (4.232)	114.568 (5.417)	122.134 (5.934)
Junín	70.816 (2.636)	81.667 (3.373)	86.381 (3.696)
Ezeiza	72.074 (2.612)	82.846 (3.344)	87.525 (3.665)
Resistencia	72.267 (2.488)	82.518 (3.185)	86.970 (3.490)
San Juan	74.858 (2.329)	84.460 (2.982)	88.631 (3.267)
Córdoba	66.507 (1.997)	74.875 (2.562)	78.510 (2.810)
Trelew	101.126 (3.522)	115.419 (4.500)	121.628 (4.926)
Villa Dolores	71.535 (2.524)	81.984 (3.233)	86.523 (3.543)
9 de Julio	67.857 (2.477)	77.958 (3.166)	82.346 (3.467)
Lomitas	58.684 (1.762)	65.905 (2.255)	69.042 (2.470)

Tabla 2: Estimaciones de los velocidades de recurrencia correspondientes a un períodos de 20, 100 y 200 años con sus respectivos errores estándar (en nudos/hora).

5.- VALIDACION DEL MODELO

Se ha investigado empíricamente la validez de las hipótesis sobre las cuales se construyó el modelo, en base al estudio de los "residuos" $X_{ij} - \xi_j$.

Se representaron gráficamente los residuos ordenados según su magnitud en función de sus valores teóricos obtenidos de la distribución de Gumbel. Si el modelo adoptado es correcto, los puntos deberían localizarse, en promedio, sobre una recta y a medida que el modelo postulado es menos adecuado para modelar los datos, mayor sería el apartamiento de la gráfica de esta recta.

Cuando la gráfica muestra puntos de quiebre (cambios en la pendiente) en las proximidades de los extremos, esto debe tomarse como indicador de la existencia de observaciones con errores groseros y no como indicador de la falta de ajuste del modelo.

En las Figuras 1-15 hay algunas gráficas que corresponderían a conjuntos de datos conteniendo datos atípicos. Por ejemplo, en la Figura 1 el último punto parece sospechoso. En las Figuras 4 y 5 los dos últimos puntos son también dudosos. La Figura 6 presenta un punto de quiebre bien marcado pero como deja a la derecha 12 puntos es difícil admitir que todas son observaciones atípicas. En este último caso es necesario realizar un estudio minucioso de los datos de esa estación para ver si se puede encontrar una justificación para la forma particular de esta gráfica. Los datos de la Figura 14 parecen contener 2 observaciones atípicas.

Si se hace abstracción de las observaciones atípicas, se encuentra que el modelo de Gumbel proporciona un muy buen ajuste para 13 de las 15 estaciones consideradas. Vemos que el ajuste no es satisfactorio para las estaciones de Azul y Villa Reynold. Por otra parte, salvo estas 2 estaciones, el modelo Gumbel tiende a ser ligeramente conservador (valor observado \leq valor teórico) lo cual nos parece que es bueno desde el punto de vista de la Ingeniería pues nos coloca del lado de la seguridad.

También se ha probado ajustar a los datos la distribución de Frechet mediante el conocido expediente de ajustar el modelo Gumbel al logaritmo de las observaciones. El análisis de los correspondientes gráficos de las observaciones ordenadas en función de sus valores esperados

mostró un menor grado de ajuste que el modelo Gumbel.

Existen técnicas estadísticas que se aplican cuando hay observaciones atípicas, tales como: eliminación de las observaciones atípicas, reemplazo de estas observaciones por valores estimados y métodos robustos. En este trabajo se empleó la siguiente metodología: primero se reemplazaron las observaciones atípicas por valores estimados obtenidos ajustando por mínimos cuadrados una recta a los residuos en función de sus valores teóricos. A continuación se recalcularon los estimadores de los parámetros por el método de los momentos en base a los datos así modificados. Las diferencias entre las nuevas estimaciones y las primitivas fueron en todos los casos pequeñas respecto del desvío estándar de estas estimaciones. Por este motivo solo damos las estimaciones basadas en los datos sin modificar.

Validación del Modelo Gumbel

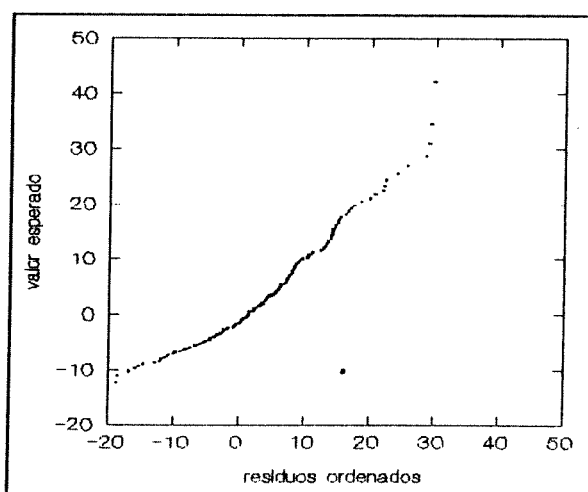


Figura 1: Neuquén

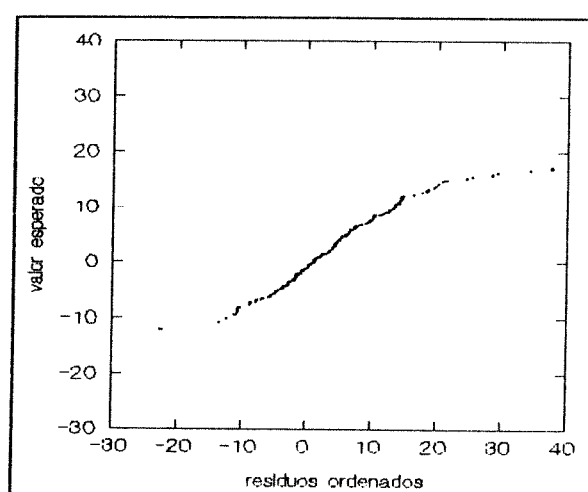


Figura 2: Azul

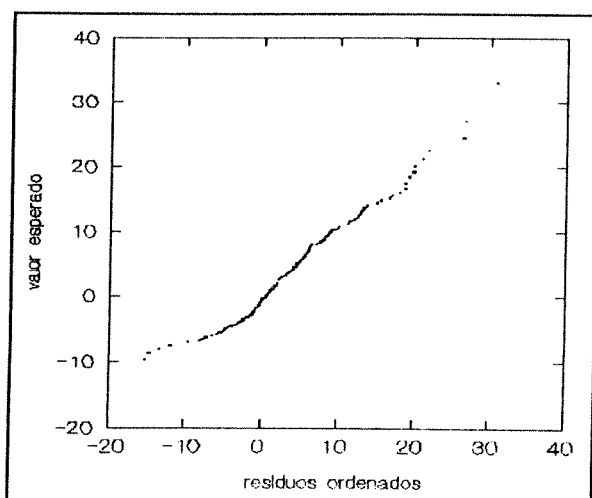


Figura 3: Salta

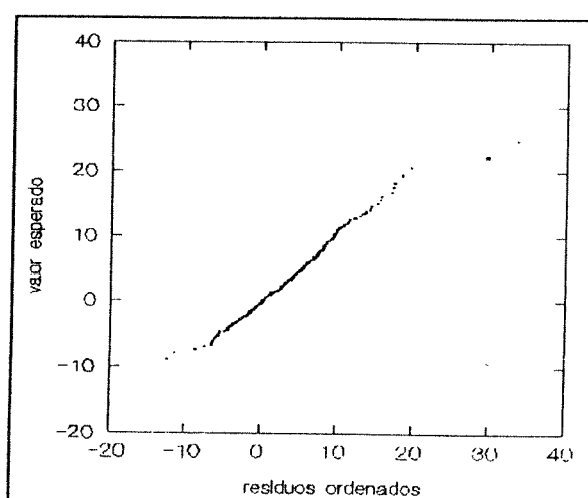


Figura 4: San Miguel

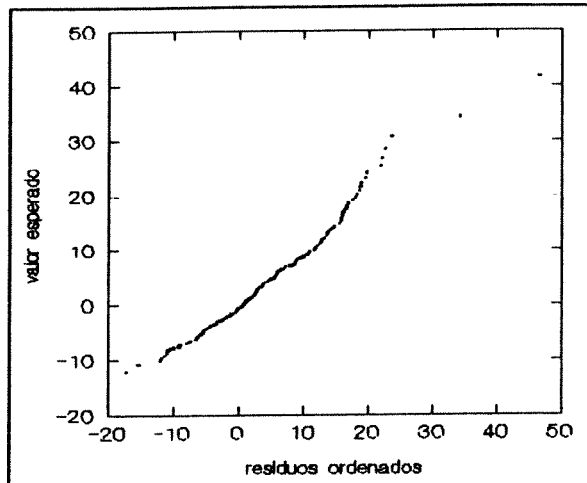


Figura 5: Aeroparque

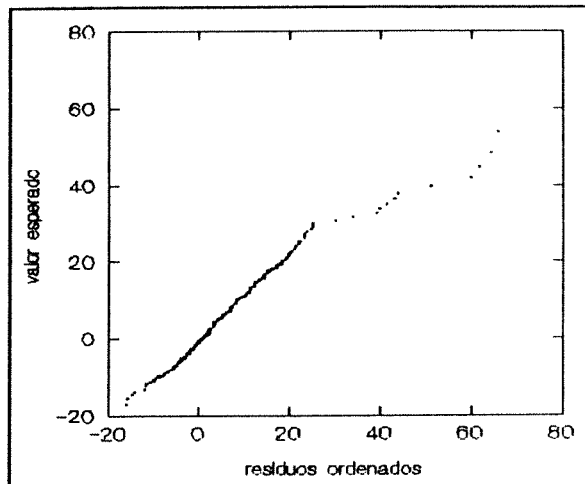


Figura 6: Villa Reynold

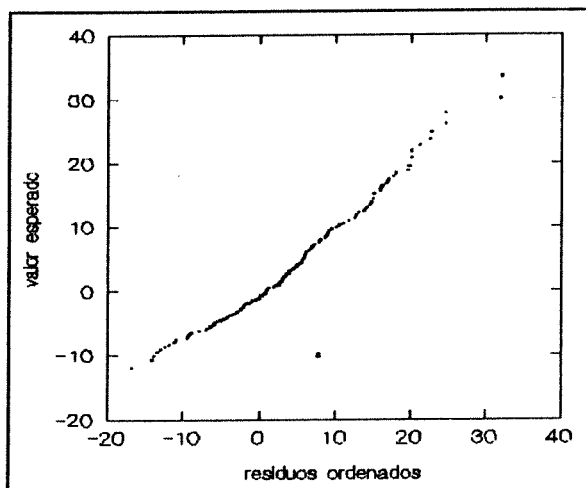


Figura 7: Junín

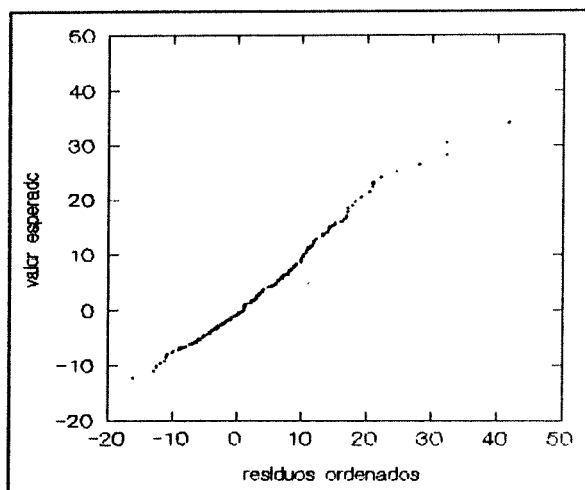


Figura 8: Bzeiza

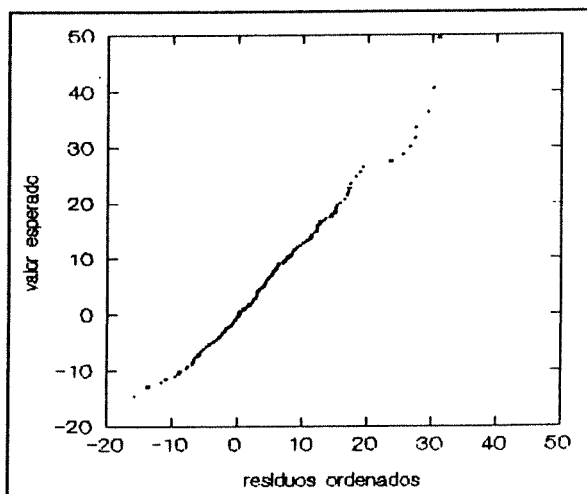


Figura 9: Resistencia

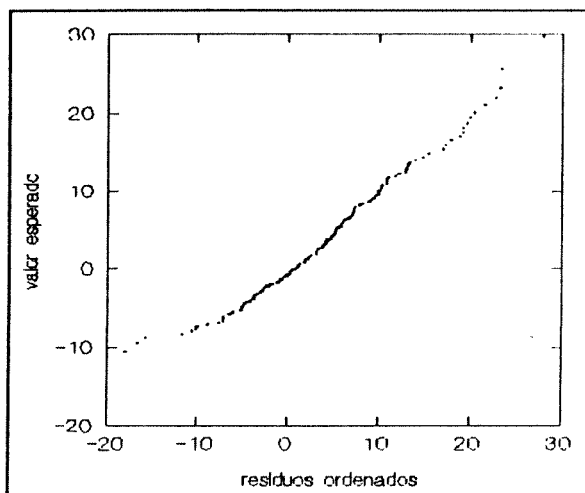


Figura 10: San Juan

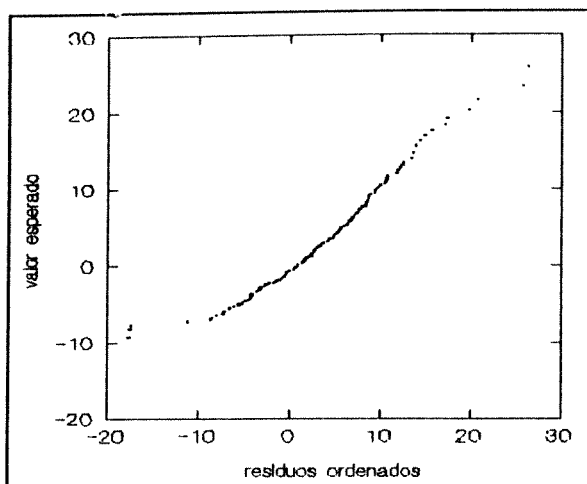


Figura 11: Córdoba

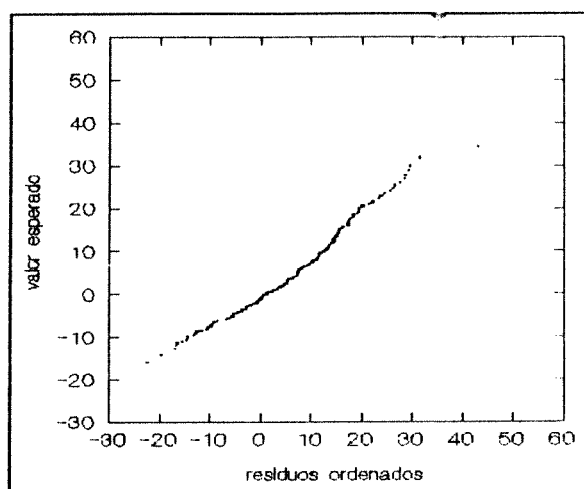


Figura 12: Trelew

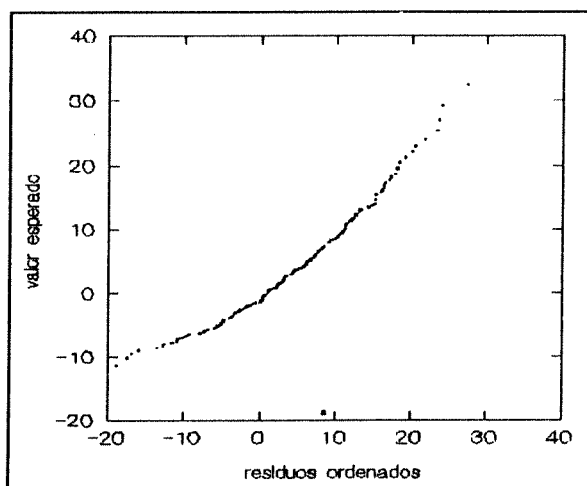


Figura 13: Villa Dolores

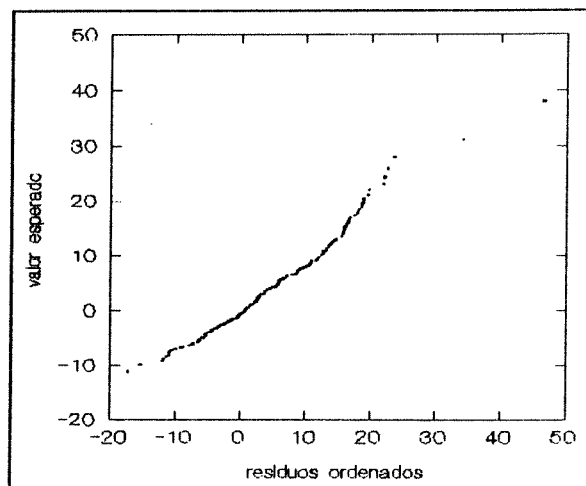


Figura 14: 9 de Julio

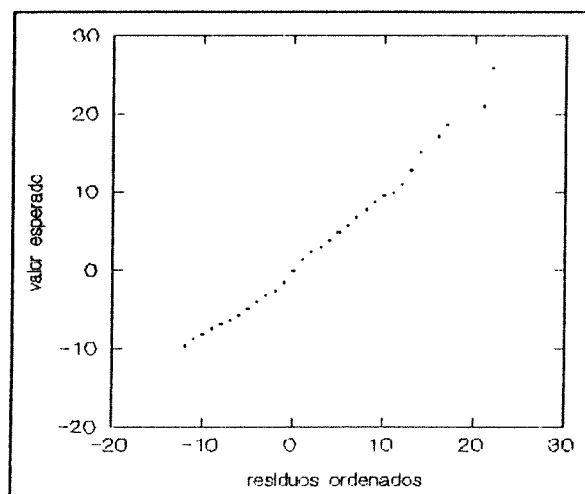


Figura 13: Lomitas

Para ver la independencia de los residuos se calculó su función de autocovarianzas. Si bien en varias estaciones algunas correlaciones son estadísticamente distintas de cero, esto no parece ser muy serio pues el tamaño de las autocorrelaciones es pequeño. Como un chequeo adicional, en las estaciones que parecían presentar residuos más correlacionados, se ajustó un modelo autorregresivo con el fin de extraer la parte estructural y comparar su "tamaño" con el tamaño de los residuos. Este análisis mostró que este tamaño, medido como el cociente de desvíos estándar, fue siempre poco significativo. Los autores están estudiando un modelo que tiene en cuenta la correlación de los residuos.

También se analizó la hipótesis de equi-distribución comparando las varianzas de los datos correspondientes a cada uno de los doce meses. Se encuentra que hay una tendencia a encontrar varianzas mayores asociadas con valores medios mayores. Este es un punto que podría perfeccionarse en un futuro estudio del problema.

6.- REFERENCIAS

- Salvatierra, S.M. y Viollaz A.J. (1993). Estimación del máximo de una serie de observaciones con función de medias periódica. Comunicación presentada en el Segundo Congreso Latinoamericano de Sociedades de Estadística.
- Salvatierra, Stélla (1994). Estimación del máximo de un proceso estocástico no estacionario. Tesis de Magister en Estadística Aplicada. Instituto de Investigaciones Estadísticas. Universidad Nacional de Tucumán.
- Tiago de Oliveira (1963). Decision results for the parameters of the extreme value (Gumbel) distribution based on the mean and the standard deviation. Trabajos de Estadística, 14, p.61-81.