DISEÑO HIDROLÓGICO

ANÁLISIS DE EXTREMOS

Edición 2024

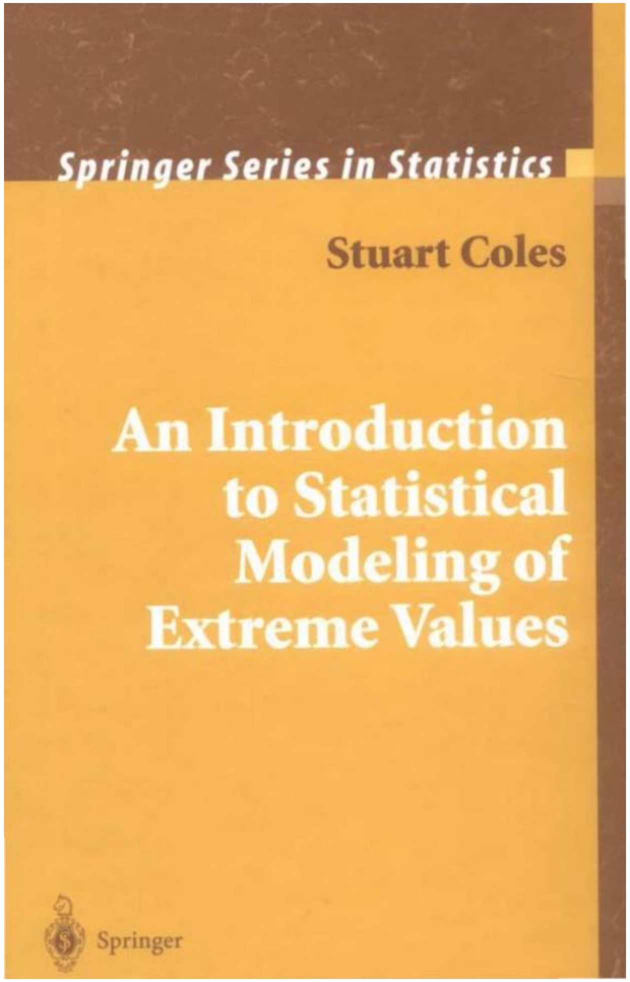
Rafael Terra (en base a notas de Sebastián Solari)

Instituto de Mecánica de los Fluidos e Ingeniería Ambiental (IMFIA)
Facultad de Ingeniería, Universidad de la República, Uruguay

rterra@fing.edu.uy

Bibliografía

- Coles (2001) “An Introduction to Statistical Modeling of Extreme Values”



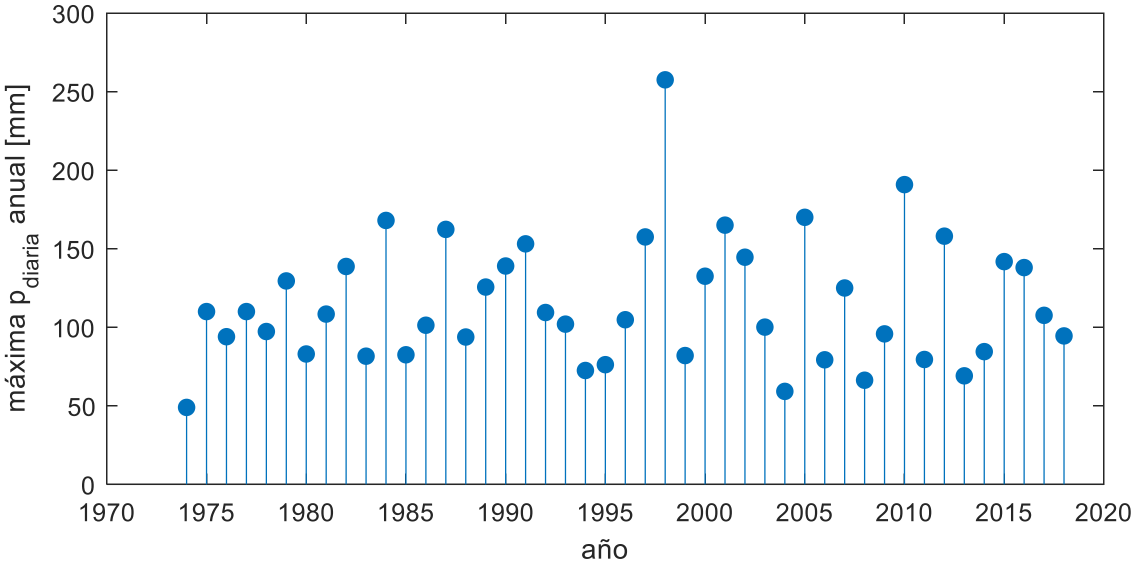
INTRODUCCIÓN

- Objetivo del análisis de extremos: Cuantificar el comportamiento de una variable para valores "atípicamente altos o bajos".
- Problema #1: Se intenta estimar la probabilidad de ocurrencia de eventos mayores (menores) a los observados.

Se necesitan **modelos** para **extrapolar** a partir de los datos disponibles

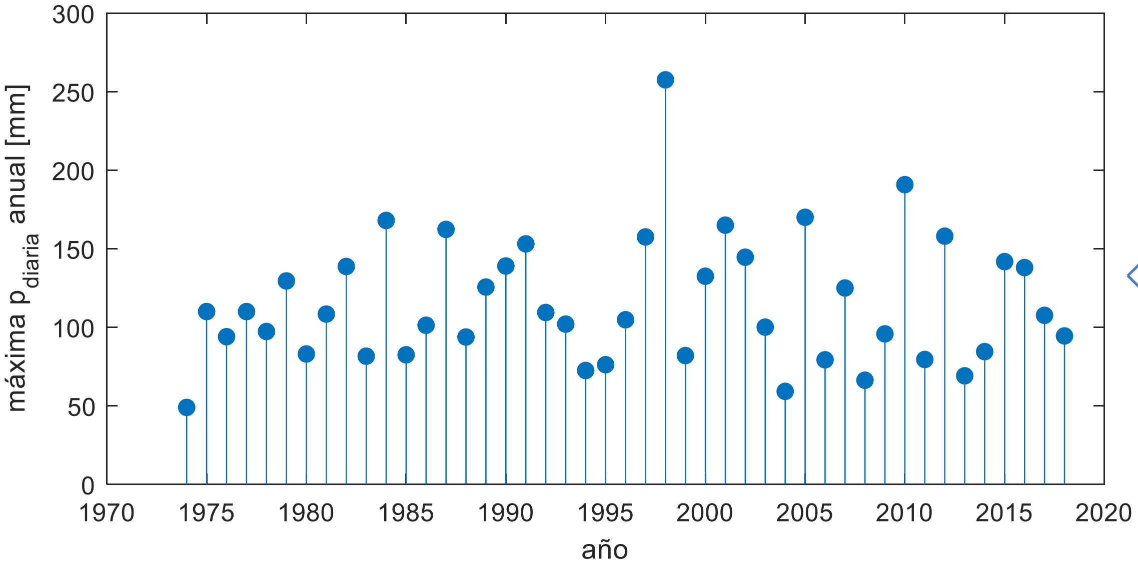
INTRODUCCIÓN

- Ejemplo: Variable aleatoria "máximo anual de la precipitación diaria"



INTRODUCCIÓN

- <u>Ejemplo</u>: Variable aleatoria "máximo anual de la precipitación diaria"



¿ Cuál es la probabilidad que en un año cualquiera tenga *max*(p_{diaria}) >= 300 mm ?

¿Cada cuántos años ocurre que *max*(p_{diaria}) >= 300 mm?

INTRODUCCIÓN

¿ Cuál es la probabilidad que en un año cualquiera tenga *max*(p_{diaria})>=300 mm ?

Estoy buscando el valor $p_{anual}(x) = 1 - F_{anual}(x)$ en donde $F_{anual}(x)$ es la **distribución de probabilidad acumulada** (CDF) de los máximos anuales.

$$F(x) = Prob\left[\max_{año}(p_{diaria}) \leq x\right] = 1 - p(x)$$

!

¿Cada cuántos años ocurre que *max*(p_{diaria})>=300 mm?

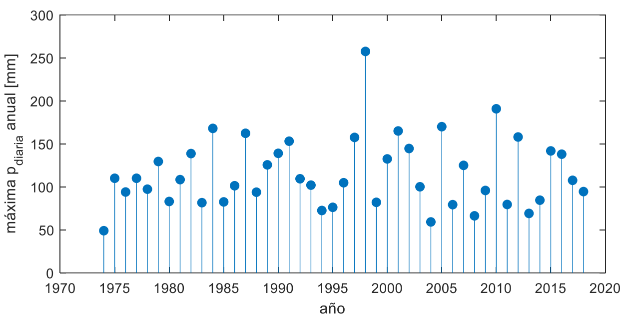
Estoy buscando el **período de retorno** $Tr(x)$, en años.

$$\boxed{\boldsymbol{Tr(x) = \frac{1}{p(x)}}}$$

Por ejemplo: p_{diaria}=300 mm se supera, en promedio, una vez cada 50 años (i.e. su período de retorno es 50 años), por lo que en un año cualquier su probabilidad de superación es P=1/50=0,02 o, lo que es lo mismo, la probabilidad de que la máxima p_{diaria} en un año sea igual o menor a 300 mm es F = 1-p=0,98.

INTRODUCCIÓN

- Ejemplo: Variable aleatoria “máximo anual de la precipitación diaria”



Puedo usar los datos para calcular la **distribución de probabilidad acumulada empírica** (ECDF):

Definition 2.4 Given an ordered sample of independent observations

$$x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$$

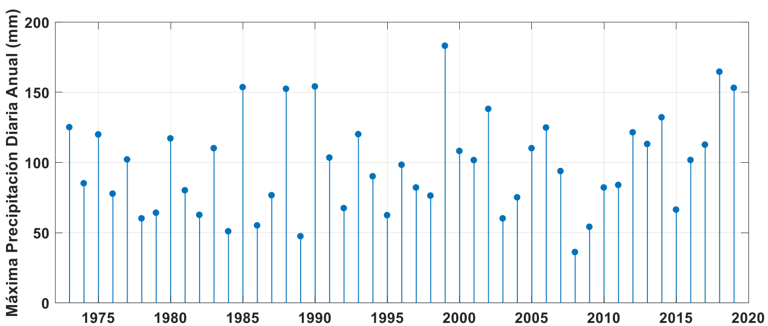
from a population with distribution function F, the **empirical distribution function** is defined by

$$\tilde{F}(x) = \frac{i}{n+1} \quad \text{for } x_{(i)} \leq x < x_{(i+1)}.$$

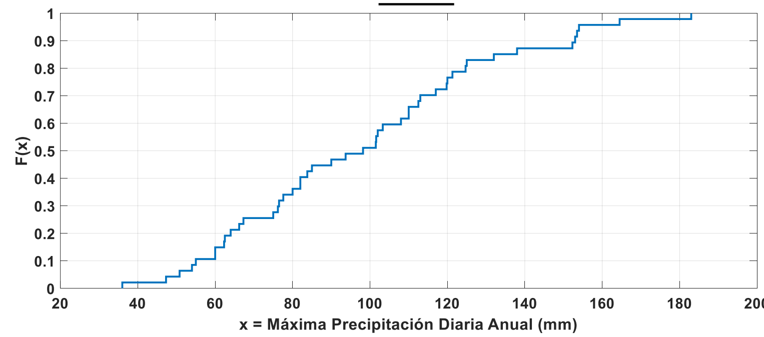
Δ

INTRODUCCIÓN

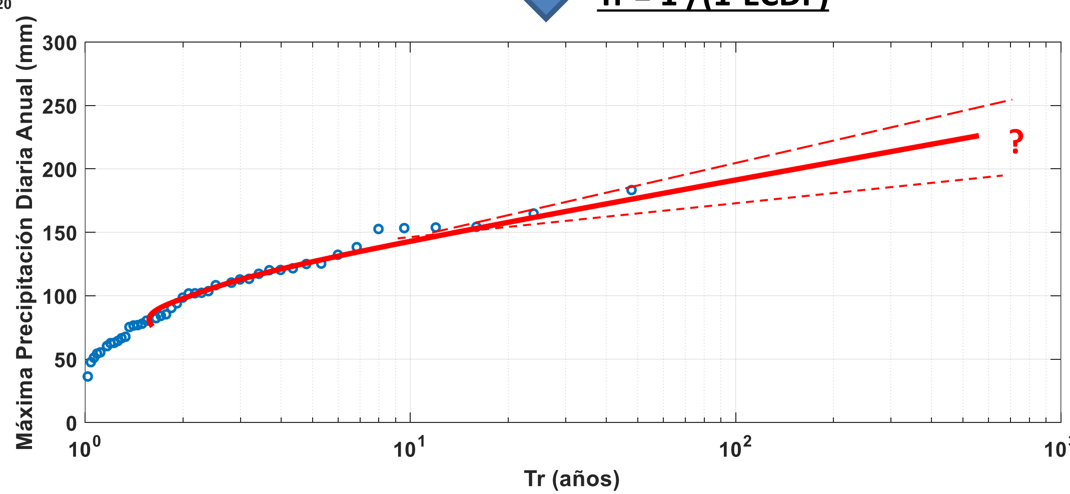
- Ejemplo: Variable aleatoria “máximo anual de la precipitación diaria”



ECDF



Tr = 1 /(1-ECDF)



Introducción

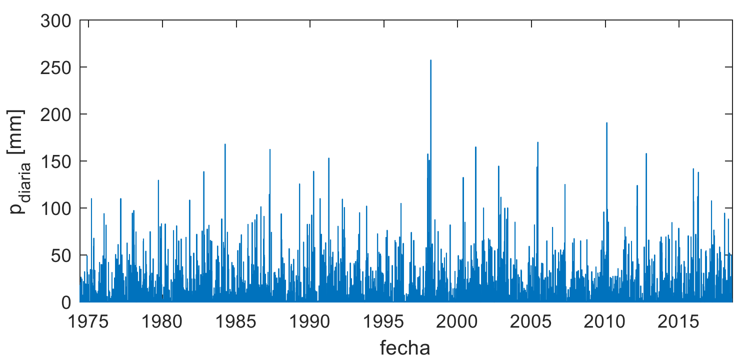
- Objetivo del análisis de extremos: Cuantificar el comportamiento de una variable para valores "atípicamente altos o bajos".

- Problema #2: La información disponible es, por definición, escasa (se está trabajando con valores atípicos; i.e. poco probables).

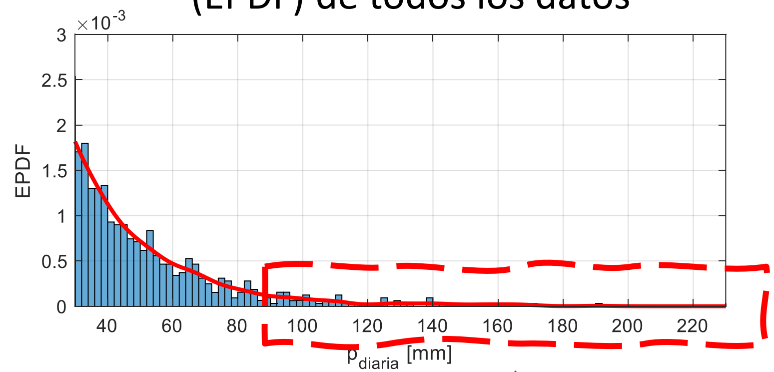
Es importante: (i) "maximizar" el uso de la información disponible y (ii) estimar la incertidumbre de las estimaciones.

INTRODUCCIÓN

Serie completa de datos



Función de densidad de probabilidad (PDF) empírica (EPDF) de todos los datos



Para el análisis de extremos solo nos interesa lo que pasa en la cola superior (inferior) de la distribución (upper/lower tail).

Teoría de valores extremos

Vamos a trabajar con la variable aleatoria M_n:

$$M_n = \max\{x_1, x_2, \dots, x_i, \dots, x_n\}$$

En donde $\{x_1, \dots, x_n\}$ es una secuencia de variables aleatorias independientes e idénticamente distribuidas (**i.i.d.**) con distribución $F(x)$.

Por ejemplo, x_i son los valores de precipitación diaria, o niveles horarios, etc., medidos a lo largo de un año con frecuencia fija.

Teoría de valores extremos

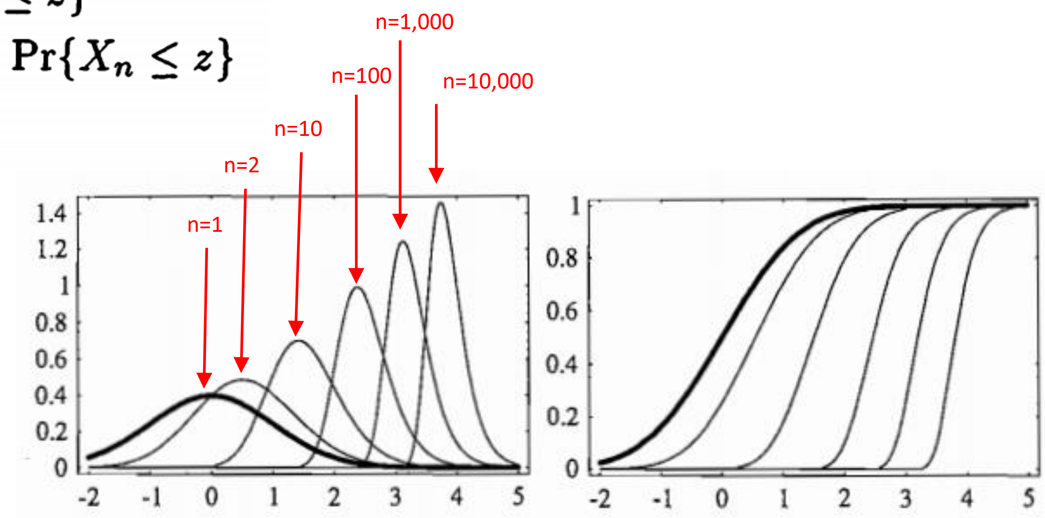
En principio la distribución de la variable M_n podría derivarse a partir de la distribución de x ($F(x)$) como:

$$\begin{aligned}\Pr\{M_n \leq z\} &= \Pr\{X_1 \leq z, \ldots, X_n \leq z\} \\ &= \Pr\{X_1 \leq z\} \times \cdots \times \Pr\{X_n \leq z\} \\ &= \{F(z)\}^n.\end{aligned}$$

TEORÍA DE VALORES EXTREMOS

En principio la distribución de la variable M_n podría derivarse a partir de la distribución de x ($F(x)$) como:

$$
\begin{aligned}
\Pr\{M_n \leq z\} &= \Pr\{X_1 \leq z, \ldots, X_n \leq z\} \\
&= \Pr\{X_1 \leq z\} \times \cdots \times \Pr\{X_n \leq z\} \\
&= \{F(z)\}^n.
\end{aligned}
$$



PDF y CDF obtenidas si F(x) es una distribución normal N(0,1)

Teoría de valores extremos

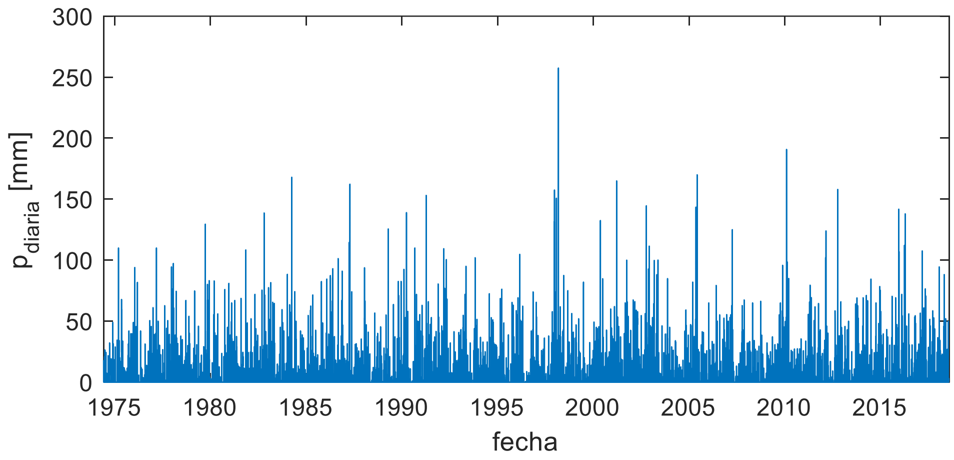
En principio la distribución de la variable M_n podría derivarse a partir de la distribución de x ($F(x)$) como:

$$\begin{aligned}\Pr\{M_n \le z\} &= \Pr\{X_1 \le z, \ldots, X_n \le z\} \\ &= \Pr\{X_1 \le z\} \times \cdots \times \Pr\{X_n \le z\} \\ &= \{F(z)\}^n.\end{aligned}$$

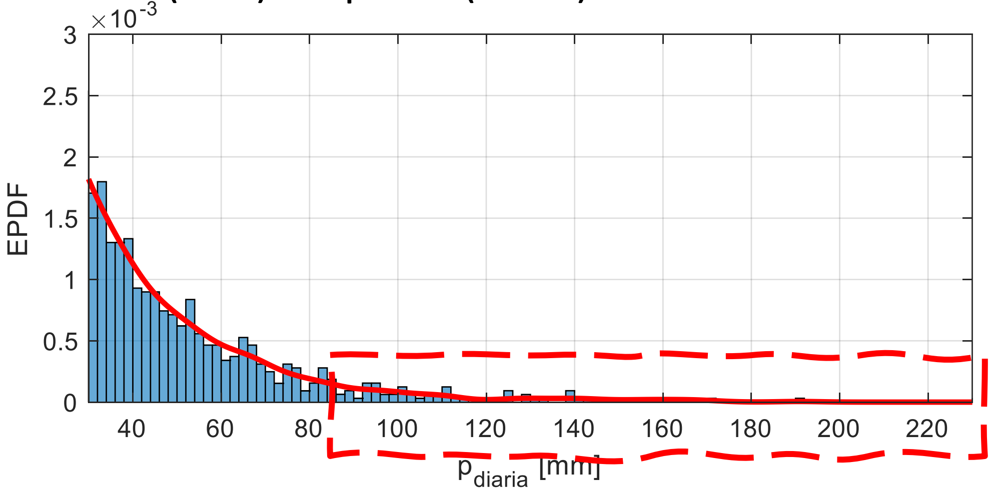
… sin embargo, pequeños errores en $F(x)$ pueden implicar grandes errores en $F(x)^n$.

TEORÍA DE VALORES EXTREMOS

Serie completa de datos



Función de densidad de probabilidad (PDF) empírica (EPDF) de todos los datos



¿Cuál es la distribución F(x) adecuada para la cola superior?

Teoría de valores extremos

¿Alternativa? → **Teorema de la teoría de los valores extremos**

El máximo M_n de n observaciones independientes e idénticamente distribuidas $\{x_i\}$ tiende a una distribución de probabilidad $G(z)$ cuando n tiende a infinito, independientemente de cuál sea la distribución $F(x)$ de $\{x_i\}$.

La distribución $G(z)$ se denomina **distribución generalizada de extremos** (***GEV*** distribution).

Teoría de valores extremos

- En las aplicaciones a casos reales:
  - ***n*** no tiende a infinito.
  - Las observaciones no son independientes.
  - La distribución F(x) no necesariamente es estacionaria.

A pesar de los cual el uso de la distribución GEV ha probado ser una herramienta valiosa para el análisis de extremos…

… siempre y cuando se tenga una comprensión suficiente de los procesos físicos que subyacen a los datos analizados!!

Teoría de valores extremos

- Distribución **GEV**: $$G(z) = \exp\left\{-\left[1+\xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\}$$

con $$1+\frac{\xi(z-\mu)}{\sigma} > 0$$

siendo:

μ → parámetro de posición

σ → parámetro de escala

ξ → parámetro de forma

TEORÍA DE VALORES EXTREMOS

- Según cuál sea el valor que toma el parámetro de forma ξ, la GEV se denomina:

Tipo I (EVI) o **Gumbel**, cuando $\xi = 0$ → $\mathrm{I}: G(z) = \exp\left\{-\exp\left[-\left(\frac{z-b}{a}\right)\right]\right\}, \quad -\infty < z < \infty;$

Tipo II (EVII) o **Fréchet**, cuando $\xi > 0$ → $\mathrm{II}: G(z) = \begin{cases} 0, & z \le b, \\ \exp\left\{-\left(\frac{z-b}{a}\right)^{-\alpha}\right\}, & z > b; \end{cases}$

Tipo III (EVIII) o **Weibull**, cuando $\xi < 0$ → $\mathrm{III}: G(z) = \begin{cases} \exp\left\{-\left[-\left(\frac{z-b}{a}\right)^{\alpha}\right]\right\}, & z < b, \\ 1, & z \ge b, \end{cases}$

Gumbel no tiene límite, Fréchet tiene límite inferior y Weibull superior

Teoría de valores extremos

- Según cuál sea el valor que toma el parámetro de forma ξ, la GEV se denomina:

Tipo I (EVI) o Gumbel, cuando $\xi = 0$ (sin límite inferior ni superior)

Tipo II (EVII) o Fréchet, cuando $\xi > 0$ (con límite inferior)

Tipo III (EVIII) o Weibull, cuando $\xi < 0$ (con límite superior)

¡Cuidado!

$$\boxed{\xi = -\gamma}$$

$$G(z) = \exp\left\{-\left[1+\xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\}$$

con $\quad 1+\frac{\xi(z-\mu)}{\sigma} > 0$

$$G(z) = \exp\left\{-\left[1-\gamma\left(\frac{z-\mu}{\sigma}\right)\right]^{\frac{1}{\gamma}}\right\}$$

con $\quad 1-\frac{\gamma(z-\mu)}{\sigma} > 0$

Tipo I (EVI) o Gumbel, cuando $\xi = 0$

Tipo II (EVII) o Fréchet, cuando $\xi > 0$

Tipo III (EVIII) o Weibull, cuando $\xi < 0$

Tipo I (EVI) o Gumbel, cuando $\gamma = 0$

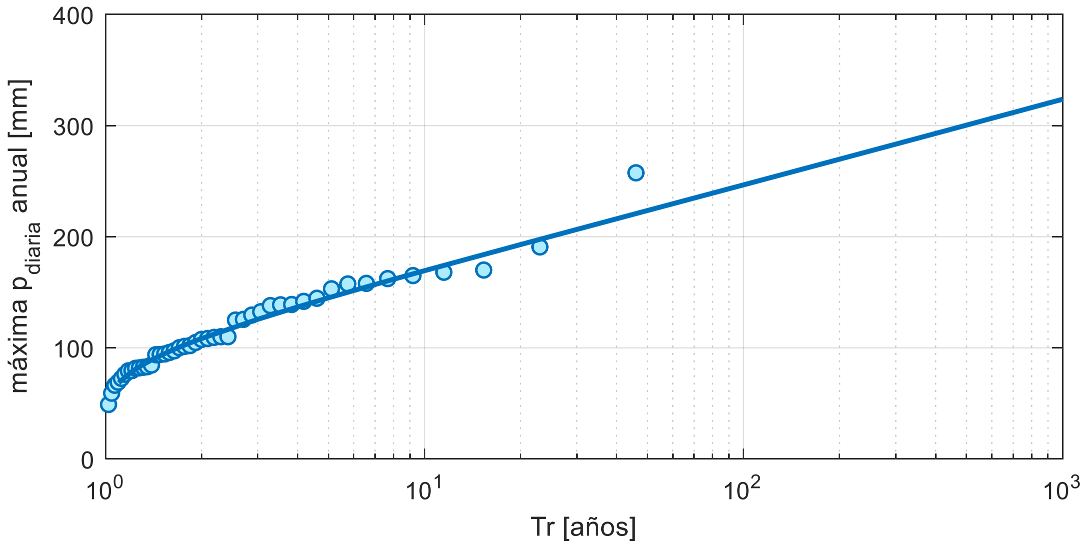
Tipo II (EVII) o Fréchet, cuando $\gamma < 0$

Tipo III (EVIII) o Weibull, cuando $\gamma > 0$

MÉTODO DE MÁXIMOS ANUALES

- Seleccionar máximos anuales a partir de la serie completa, "asegurando" i.i.d.
- Estimar parámetros (μ, σ, ξ) de la GEV a partir de los datos.
- Calcular cuantiles de distinto Tr (o probabilidad $p = 1/Tr$) con la GEV.

$$z_p = \begin{cases} \mu - \frac{\sigma}{\xi}\left[1 - \{-\log(1-p)\}^{-\xi}\right], & \text{for } \xi \neq 0, \\ \mu - \sigma \log\{-\log(1-p)\}, & \text{for } \xi = 0, \end{cases}$$



Método de Máximos Anuales

- Una crítica habitual al método de Máximos Anuales es que "desperdicia" información al no considerar todos los eventos extremos de la serie sino únicamente los máximos de cada año.
- En la medida que se dispone de series de registros medidos en alta frecuencia de forma continua, se suele recurrir a otra aproximación, basada en el uso de las excedencias sobre un umbral (Threshold Models).

MODELOS DE EXCEDENCIAS

De forma equivalente a lo planteado en el **teorema de la teoría de los valores extremos**:

Si M_n es el máximo de n observaciones iid $\{x_i\}$, cuya distribución de probabilidad $\Pr\{M_n \leq z\}$ tiende a una GEV cuando n tiende a infinito $\Pr\{\mathrm{M_n} \leq z\} \approx G(z)$.

Entonces para un valor de umbral u suficientemente grande, **la distribución de las excedencias sobre el umbral $\boldsymbol{y_i = (x_i - u)}$**, condicionada a $x > u$, seguirá una distribución de **Pareto Generalizada** (GP distribution):

$$H(y) = 1 - \left(1 + \frac{\xi y}{\tilde{\sigma}}\right)^{-\frac{1}{\xi}}$$

Válida en: $y > 0 \;;\; \left(1 + \frac{\xi y}{\tilde{\sigma}}\right) > 0$

“Suficientemente grande: $u > \mu$

¿ μ?

Con $\tilde{\sigma} = \sigma + \xi(u - \mu)$

MODELOS DE EXCEDENCIAS

- Distribución generalizada de Pareto (GP distribution)

Si el máximo de la serie Mn tiene distribución GEV: $G(z) = \exp\left\{-\left[1+\xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}$

Entonces las excedencias sobre un umbral suficientemente alto tienen una distribución GP:

$$H(y) = 1 - \left(1 + \frac{\xi y}{\tilde{\sigma}}\right)^{-1/\xi}$$

Con el mismo parámetro de forma ξ y con parámetro de escala $\tilde{\sigma}$ dado por:

$$\tilde{\sigma} = \sigma + \xi(u - \mu)$$

MODELOS DE EXCEDENCIAS

- Deducción de GP a partir de la GEV:

Let X have distribution function F. By the assumption of Theorem 3.1, for large enough n,

$$F^n(z) \approx \exp\left\{-\left[1+\xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}$$

for some parameters $\mu, \sigma > 0$ and ξ. Hence,

$$n \log F(z) \approx -\left[1+\xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\xi}.$$

But for large values of z, a Taylor expansion implies that

$$\log F(z) \approx -\{1 - F(z)\}.$$

Substitution into (4.5), followed by rearrangement, gives

$$1 - F(u) \approx \frac{1}{n}\left[1+\xi\left(\frac{u-\mu}{\sigma}\right)\right]^{-1/\xi}$$

for large u. Similarly, for $y > 0$,

$$1 - F(u+y) \approx \frac{1}{n}\left[1+\xi\left(\frac{u+y-\mu}{\sigma}\right)\right]^{-1/\xi}.$$

Hence,

$$\begin{aligned}
\Pr\{X > u+y \mid X > u\} &\approx \frac{n^{-1}\left[1+\xi(u+y-\mu)/\sigma\right]^{-1/\xi}}{n^{-1}\left[1+\xi(u-\mu)/\sigma\right]^{-1/\xi}} \\
&= \left[1 + \frac{\xi(u+y-\mu)/\sigma}{1+\xi(u-\mu)/\sigma}\right]^{-1/\xi} \\
&= \left[1 + \frac{\xi y}{\tilde{\sigma}}\right]^{-1/\xi},
\end{aligned}$$

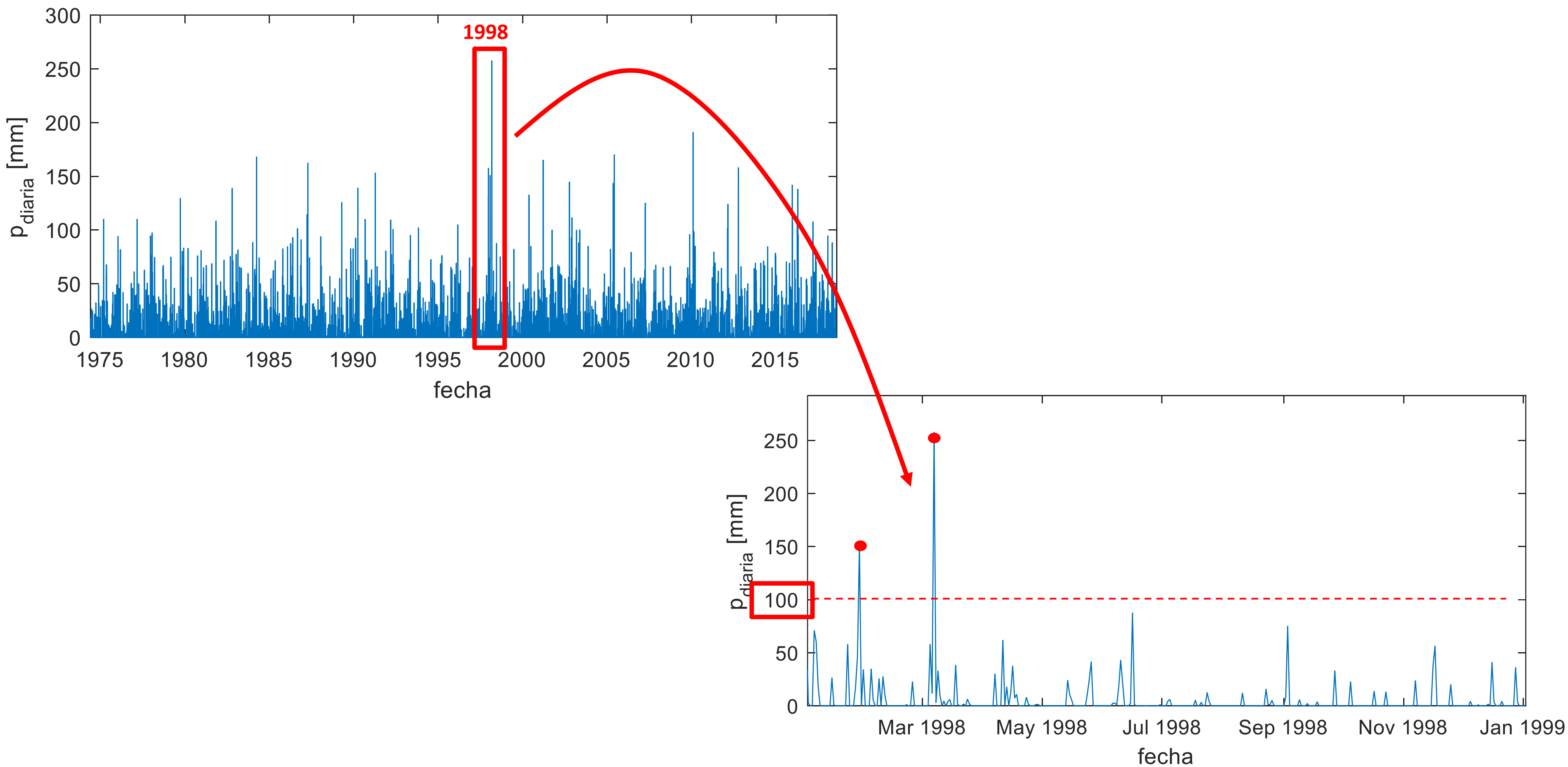
where

$$\tilde{\sigma} = \sigma + \xi(u-\mu),$$

as required.

MÉTODO DE PICOS SOBRE EL UMBRAL (POT)

- Selección de valores de excedencia de un determinado umbral "asegurando" que sean i.i.d.



MÉTODO DE PICOS SOBRE EL UMBRAL (POT)

Selección de valores de excedencia de un determinado umbral "asegurando" que sean i.i.d. → en general obtengo un número medio de picos por año $\boldsymbol{\nu} > \mathbf{1}$ (i.e. en promedio tengo más de un evento extremo por año).

La distribución de probabilidad empírica de la serie de excedencias $\{x_1, \dots, x_i, \dots, x_n\}$, siendo *n* el número de picos, se puede calcular como cualquier ECDF:

Definition 2.4 Given an ordered sample of independent observations

$$x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$$

from a population with distribution function F, the **empirical distribution function** is defined by

$$\tilde{F}(x) = \frac{i}{n+1} \quad \text{for } x_{(i)} \leq x < x_{(i+1)}.$$

MÉTODO DE PICOS SOBRE EL UMBRAL (POT)

Sin embargo, ahora p = 1-F(x) no es la probabilidad de que el máximo anual supere el valor x, sino que es la probabilidad de que el máximo de una tormenta (un evento) supere el valor x.

- **Cálculo de Tr_años**:

Valor esperado del número de eventos que ocurren entre dos eventos cuya magnitud es mayor o igual a x:

$$Tr_{eventos} = \frac{1}{1 - F_{eventos}(x)} = \frac{1}{p_{eventos}(x)} \text{ [eventos]}$$

Para pasar de Tr_eventos a Tr_años:

$$Tr_{\text{años}} = \frac{1}{\nu[1 - F_{eventos}(x)]} = \frac{1}{\nu \times p_{eventos}(x)} \text{ [años]}$$

MÉTODO DE PICOS SOBRE EL UMBRAL (POT)

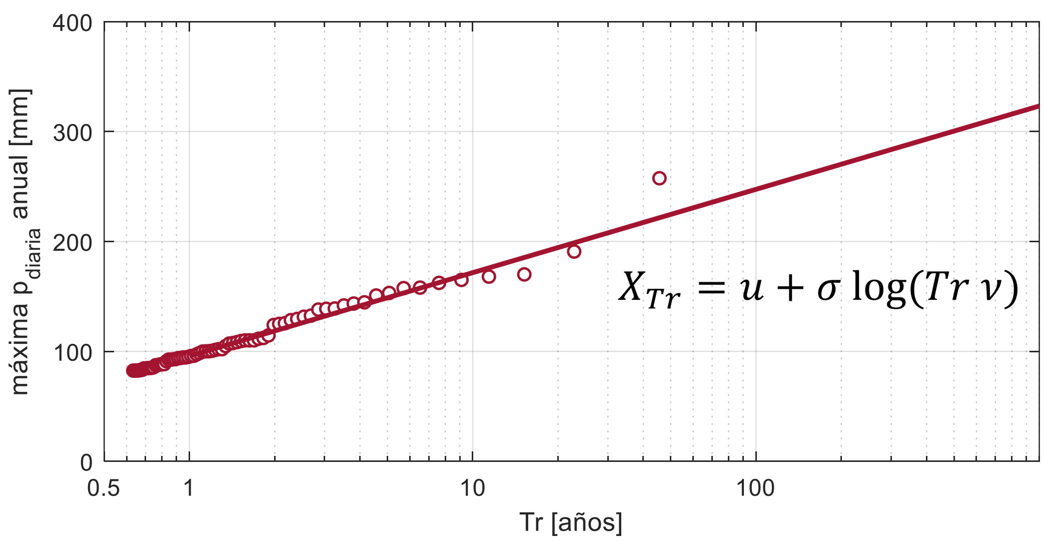
- Selección de valores de excedencia de un determinado umbral "asegurando" que sean i.i.d. (una media de ν eventos por año).
  - Estimar parámetros $(\tilde{\sigma}, \xi)$ de la GP a partir de los datos.
- Calcular cuantiles de distinto Tr (o probabilidad $p = 1/Tr$) con la GP y el número medio de eventos por año ν.

$$H(y) = 1 - \left(1 + \frac{\xi y}{\tilde{\sigma}}\right)^{-1/\xi}$$

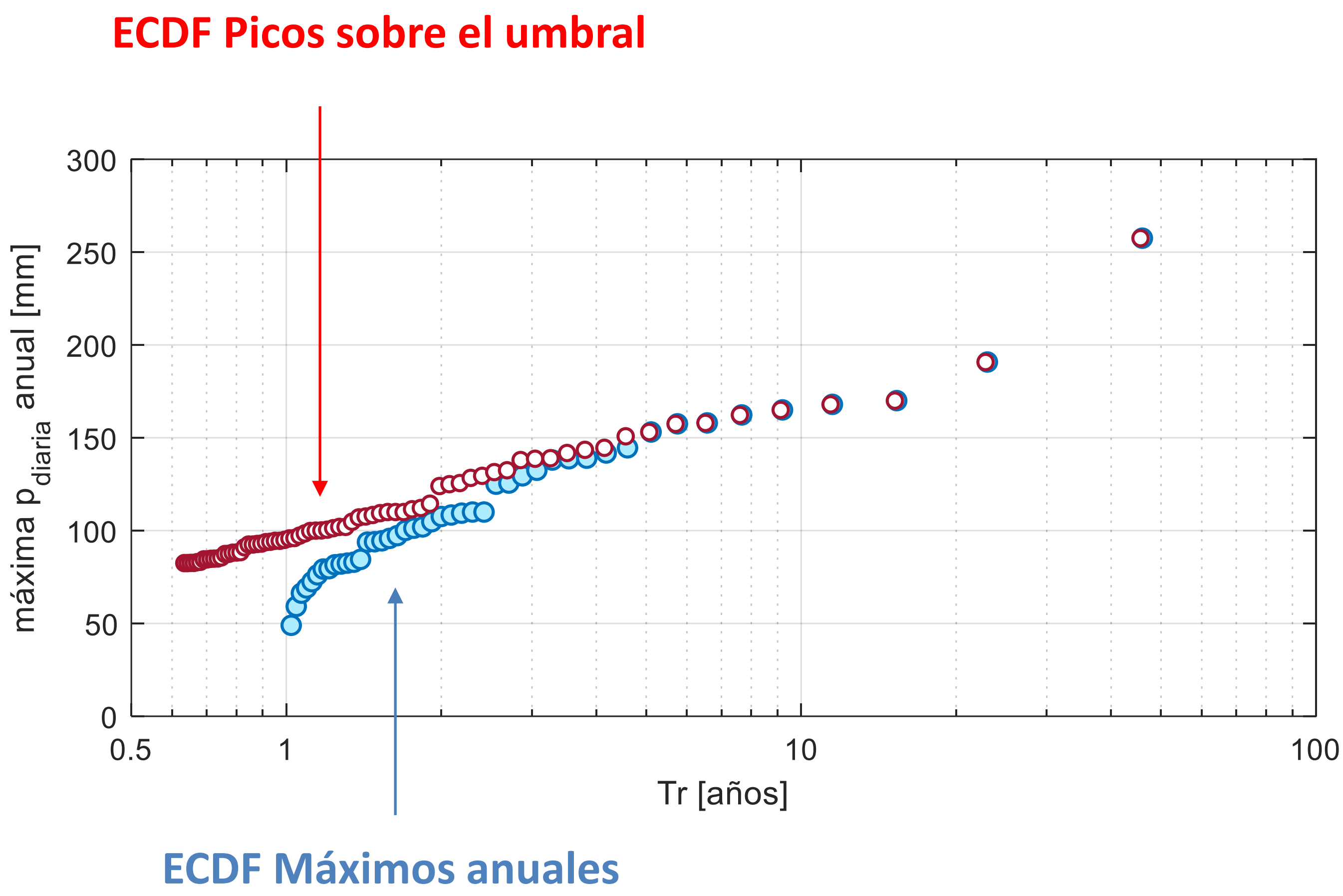
$$Tr = 1/\left(\nu.(1 - H)\right)$$

$$y = \frac{\sigma}{\xi}\left((Tr\,\nu)^{\xi} - 1\right)$$

$$X_{Tr} = u + \frac{\sigma}{\xi}\left((Tr\,\nu)^{\xi} - 1\right)$$



Método de picos sobre el umbral (POT)



¿Cómo se relaciona esto con las curvas IDF?

ISOYETAS DE LLUVIAS EXTREMAS EN URUGUAY

Precipitación de 3 horas y 10 años de período de retorno

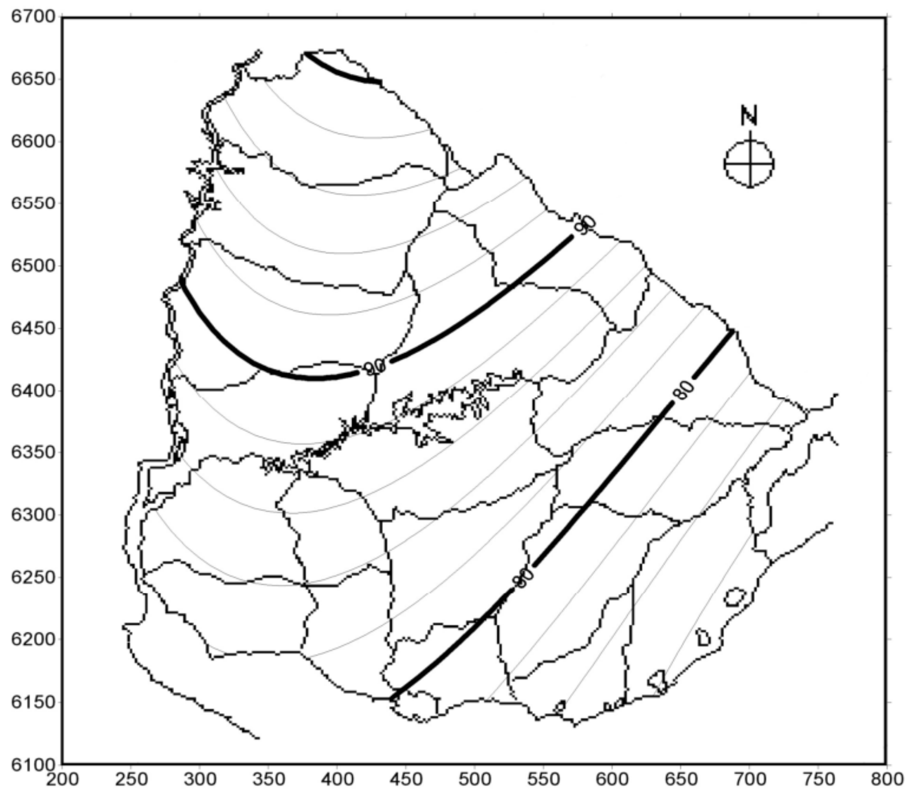


Figura 3.1.10: Isoyetas de precipitaciones extremas para Uruguay asociadas a una duración de 3 horas y 10 años de período de retorno.

$$\boxed{P_{(d,T_r,p)} = P_{(3,10,p)}.CT_{(T_r)}.CD_{(d)}}$$

Donde:

P(d,Tr,p) es la precipitación máxima para una duración d, período de retorno Tr y un punto del Uruguay p donde se quiere calcular.

T_r = Período de retorno en años

d = Duración en horas

$$CT_{(T_r)} = 0.5786.-0.4312.\log\left[Ln\left(\frac{T_r}{T_r-1}\right)\right]$$

$$\mathrm{CD}_{(\mathrm{d})} = \frac{0.6208.d}{(d+0.0137)^{0.5639}}$$ Para d< 3hs

$$\mathrm{CD}_{(\mathrm{d})} = \frac{1.0287.d}{(d+1.0293)^{0.8083}}$$ Para d> 3hs

Basado en Gumbel con otra cantidad de aproximaciones para proveer una función continua con Duración y Área, etc

Análisis de extremos

El ajuste (i.e. estimación de los parámetros) de una o más distribuciones de probabilidad no es suficientes…

… es necesario también:

- Verificar la **calidad del ajuste** y contrastar las distintas funciones obtenidas.

- Estimar **intervalos de confianza** de la extrapolación.

ANÁLISIS DE EXTREMOS

Edición 2024

Rafael Terra (en base a notas de Sebastián Solari)

Instituto de Mecánica de los Fluidos e Ingeniería Ambiental (IMFIA)
Facultad de Ingeniería, Universidad de la República, Uruguay

rterra@fing.edu.uy