SEVECLIM: SOFTWARE PARA ANALIZAR MODELOS PROBABILÍSTICOS PARA CARACTERIZAR VARIABLES CLIMÁTICAS EXTREMAS

Pedro Roura-Pérez*, Vivian Sistachs-Vega**, Jorge Alfonso Rodríguez Méndez*** y Raimundo Vega-González*

*Centro del Clima, Instituto de Meteorología, Cuba. E-mail: pedro.roura@insmet.cu

**Facultad de Matemática y Computación, Universidad de La Habana, Cuba.

***Instituto de Medicina Tropical: Pedro Kourí (IPK), Cuba.

ABSTRACT

Extreme Value Theory is a discipline that develops modern techniques to describe the least common events, which in itself makes it a “unique” discipline. The goal of this theory is to analyze observed extreme values and to estimate values for different return periods. This discipline is of high importance for the National Meteorological Institute and especially to the Clime Center due to the characterization of the extreme climatological variables, this is the reason for the development of a software to provide these services obtaining relevant information as well as an economical gain. The knowledge of the principal climatic variables associated with extreme events, as well as the historical data in series with a observational time period of the highest temporal longitude is also important, for example in construction. In this line of work is required a specific safety margin to elaborate measures to prevent social and economic affectations as well as to maintain the safety of the population. SEVECLIM was developed with the purpose to facilitate the work and to reduce calculation errors. The software allows a user to introduce the extreme variables data. The goal is to characterize through different probability distributions the different extreme climatological variables to obtain the model that best fit the data, to estimate the return periods of 1%, 2 %, 5 % y 10 %, as well as the main statistics. It is show the proposed work methodology to a better understanding in the software utilization. With the model that best fit the data, it is possible to be closer to reality thus providing a better knowledge of the regime of the studied variable; it will also allow us to offer these services to other institutions to obtain economic gains.

KEYWORDS: Extreme Value, probability distributions, return periods

MSC: 62P12

RESUMEN

La Teoría de Valores Extremos es una disciplina que desarrolla técnicas y modelos para describir los sucesos menos comunes, lo cual hace que sea una disciplina “única”. El objetivo de esta teoría es analizar valores extremos observados y estimar valores extremos para distintos períodos de retornos. Es muy importante para el Instituto Nacional de Meteorología y en especial para el Centro del Clima la caracterización de las variables climatológicas extremas por lo que se ha desarrollado un software con el cual podemos brindar servicios y obtener beneficios económicos. Es necesario conocer las principales características de las variables climatológicas asociadas a eventos extremos, así como datos históricos de interés en series con un período observacional de la mayor longitud temporal admisible, por ejemplo, en la esfera de la construcción de obras se requiere un margen de seguridad específico para la elaboración de medidas relacionadas con las posibles afectaciones socio-económicas y la protección a la población. SEVECLIM se desarrolló con la idea de facilitar el trabajo y eliminar errores a la hora de realizar los cálculos. El software permite a un usuario introducir los datos de las variables extremas, el objetivo es caracterizar a través de diferentes distribuciones de probabilidad las diferentes variables climatológicas extremas para obtener el mejor modelo de ajuste a los datos y estimar los valores esperados de los períodos de retorno 1 %, 2 %, 5 % y 10 %, así como los principales estadígrafos. Se brinda la metodología propuesta para una mejor comprensión a la hora de ser utilizado. Con el modelo más representativo en el ajuste de los datos, se permite dar estimados más cercanos a la realidad y brindando un mayor conocimiento del régimen de la variable en cuestión. Además permitirá brindar servicios a distintas instituciones para obtener beneficios económicos.

PALABRAS CLAVE: Valores extremos, distribuciones de probabilidad y período de retorno

1. INTRODUCCIÓN

La Teoría de Valores Extremos (TVE) ha emergido como una de las más importantes ramas de la estadística que se aplica a la ciencia en los últimos 50 años; en particular en las últimas dos décadas se ha desarrollado rápidamente tanto desde el punto de vista metodológico como de las aplicaciones. La mayoría de los estudios estadísticos tratan de la modelación del promedio de la distribución de la variable de interés, el cual se estima con la media muestral. En la TVE el interés principal no está en el promedio, sino en los valores más bajos o más altos de la variable bajo estudio, es decir, el interés está en los eventos asociados a la cola de la distribución.

Según plantea Coles, "la Teoría de Valores Extremos es una disciplina que desarrolla técnicas y modelos para describir los sucesos menos comunes, lo cual, para él, hace que sea una disciplina única" [2]. Para Gumbel (autor del considerado libro de referencia para el estudio de valores extremos),"el objetivo de la TVE es analizar valores extremos observados y predecir valores extremos en el futuro" [4]. Esta, en particular, interesa mucho pues describe los objetivos con las variables climatológicas.

Los TVE tienen muchas aplicaciones en la práctica, una muy importante es en el área de la meteorología. Las variables climatológicas (viento máximo anual, temperatura mínima absoluta anual entre otras), es común la necesidad del conocimiento de los valores extremos en diferentes renglones de la sociedad como son la agricultura, ingeniería, la defensa civil, etc. Debido a la naturaleza de los algoritmos asociado al modelo de probabilidades, se hace necesario la elaboración efectiva de una herramienta computacional que facilite el cálculo, por ejemplo de los valores críticos de la variable que corresponden a distintos períodos de retorno de interés socio-económico [6].

En este trabajo se describe un software y las herramientas empleadas para su desarrollo. Se especifica su funcionamiento y la metodología a utilizar en datos de variables climáticas extremas. Para esto desarrollamos tres epígrafes, uno referido a la teoría estadística, otro referido a la metodología empleada y por último ilustramos con un ejemplo de viento máximo.

El objetivo general es realizar un software para el análisis de diferentes eventos climatológicos extremos a través de modelos probabilísticos. Dentro de los objetivos específicos tenemos: analizar diferentes distribuciones de probabilidad para obtener el mejor modelo de ajuste a los datos. Realizar el análisis del evento extremo asociado al modelo ajustado. Estimar los valores esperados de los períodos de retorno 1 %, 2 %, 5 % y 10 % del evento extremo asociado al modelo ajustado. Se entiende por período de retorno del 1%, al evento que ocurre una vez cada 100 años, análogamente es lo que ocurre con el 2% (50 años), 5 % (20 años) y el 10 % (10 años):

La modelación de datos observacionales de las variables extremas tiene gran importancia económica relativa a los posibles efectos destructivos de estos eventos en Cuba, sabiendo que está ubicada en el lecho de las trayectorias de organismos ciclónicos tropicales del Atlántico Norte. Un ejemplo importante de lo anterior lo constituye el desastre provocado por el huracán Flora (categoría Saffir-Simpson dos y trayectoria muy irregular) en la región oriental de Cuba del cuatro al ocho de octubre de 1963, donde los totales de lluvia acumulados fueron de 1500 a 1600 milímetros (El Instituto Nacional de Recursos Hidráulicos (INRH) reportó un máximo de 2550 mm), totalizando en la ciudad de Santiago de Cuba, 735.1 mm en 24 horas el día cinco; los muertos por inundaciones pluviales fueron más de 1200, además de una enorme cantidad de damnificados y cuantiosas pérdidas materiales en la agricultura, las viviendas y la infraestructura en general [8].

Este software es muy importante para el Instituto de Meteorología (INSMET) y en especial para el Centro del Clima (CENCLIM) para brindar servicios sobre la caracterización de las variables extremas en diferentes esferas como la construcción de obras, la planificación urbana, la agricultura, para la hidrología. Además para la Directiva #1, debido a que ellos son los encargados de la toma de decisiones ante la ocurrencia de un evento extremo, previniendo a la población evitando así las pérdidas de vidas humanas y la disminución de las pérdidas materiales.

2. DISTRIBUCIONES UTILIZADAS EN EL SOFTWARE

En este epígrafe se comentan de la Teoría de Valores Extremos Generalizada los aspectos fundamentales de las distribuciones que se utilizan en el software. La distribución de Fisher Tippett de tipo I o distribución de Gumbel llamada así en honor de Emil Julius Gumbel. Fue creada en la década del 20 con la teoría de los valores extremos [5, 8, 10].

a. Distribución de Gumbel para máximos (Gumbel I) de parámetros $\boldsymbol{\alpha}, \boldsymbol{\beta} > \mathbf{0}$

Función de distribución: $F(x) = e^{-e^{\left[-\left(\frac{x-\alpha}{\beta}\right)\right]}} = P(X \leq x), donde\ (-\infty < x < +\infty)$ (1)

Función de densidad: $f(x) = \frac{1}{\beta} e^{-e^{\left[-\left(\frac{x-\alpha}{\beta}\right)\right]}} e^{-\left(\frac{x-\alpha}{\beta}\right)}, donde\ (-\infty < x < +\infty)$ (2)

b. Distribución de Gumbel para mínimos (Gumbel II) de parámetros $\boldsymbol{\alpha}, \boldsymbol{\beta} > \mathbf{0}$

Función de distribución: $F(x) = 1 - e^{-e^{\left[\left(\frac{x-\alpha}{\beta}\right)\right]}} = P(X \le x), donde\ -\infty < x < +\infty$ (3)

Función de densidad: $f(x) = \frac{1}{\beta} e^{-e^{\left[\left(\frac{x-\alpha}{\beta}\right)\right]}} e^{\left(\frac{x-\alpha}{\beta}\right)}, donde\ (-\infty < x < +\infty)$ (4)

La distribución de Fisher Tippett de tipo II o distribución de Frechet recibe su nombre de Maurice Frechet, que escribió un artículo relacionado con ella en 1927. También trabajaron con ella Fisher y Tippett en 1928 y Gumbel en 1958 [5, 8].

c. Distribución de Frechet para máximos (Frechet I) de parámetros $\boldsymbol{\alpha}, \boldsymbol{\beta} > \mathbf{0}$

Función de distribución: $F(x) = e^{-\left(\frac{x}{\alpha}\right)^{-\beta}} = P(X \le x), tal\ que\ (x \ge 0)$ (5)

Función de densidad: $f(x) = \frac{\beta}{a}\left(\frac{x}{a}\right)^{-1-\beta} e^{-\left(\frac{x}{\alpha}\right)^{-\beta}}, tal\ que (x \ge 0)$ (6)

d. Distribución de Frechet para mínimos (Frechet II) de parámetros $\boldsymbol{\alpha}, \boldsymbol{\beta} > \mathbf{0}$

Función de distribución: $F(x) = 1 - e^{-\left(\frac{x}{\alpha}\right)^{\beta}} = P(X \le x)$, tal que $(x \ge 0)$ (7)

Función de densidad: $f(x) = \beta\alpha^{-\beta}x^{\beta-1}(1 - F(x))$, tal que $(x \ge 0)$ (8)

En el caso que se cuente con muestras pequeñas de datos, existe el fenómeno de la ocurrencia inestabilidad en la modelación, siendo posible que el ajuste sea por la distribución Normal o Log-Normal.

e. Distribución Normal de parámetros $-\infty < \boldsymbol{\alpha} < +\infty,\ \boldsymbol{\beta} > \mathbf{0}$

La gráfica (campana de Gauss) describe de forma aproximada muchos fenómenos de la naturaleza de mediciones físicas en diversas áreas. Fue presentada por primera vez por Abraham de Moivre en un artículo del año 1733, sin embargo el nombre de Gauss se ha asociado a esta distribución porque también derivo su ecuación a partir de un estudio de errores en mediciones repetidas de la misma cantidad y la usó ampliamente en estudios de datos astronómicos.

La principal desventaja cuando se trata de describir variables climatológicas es que esta varia en un rango continuo de $(-\infty, +\infty)$ mientras algunas son no negativas como la precipitación y el viento. Además que es simétrica alrededor de la media, mientras los valores extremos tienden a ser asimétricos.

La función de distribución:

$F(x) = \int_{-\infty}^{x} f(x)dx = P(X \le x), tal\ que\ 0 < F(x) < 1$ (9)

Su función de densidad: $f(x) = \frac{1}{\beta\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\alpha}{\beta}\right)^2}$ (10)

Donde x representa la variable, α es la media y β^2 es la varianza.

f) **Distribución Log-Normal de parámetros $\boldsymbol{\alpha}, \boldsymbol{\beta} > \mathbf{0}\ \boldsymbol{y}\ \boldsymbol{x} > \mathbf{0}$**

La distribución Log-Normal es una distribución de probabilidad de una variable aleatoria cuyo logaritmo está normalmente distribuido. Es decir, si X es una variable aleatoria con una distribución normal, entonces e^X tiene una distribución Log-Normal. Está, a diferencia de la distribución Normal está limitada a $X > 0$.

La función de distribución: $F(x) = \int_{-\infty}^{x} f(t)dt = P(X \le x) = \bar{N}\left(\frac{\ln x - \alpha}{\beta}\right)$ (11)

Y la función de densidad: $f(x) = \frac{1}{\beta x\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln x - \alpha}{\beta}\right)^2}$ (12)

3. FÓRMULAS DE PLOTEO

Dada una muestra de n datos observacionales x_i, existen varias fórmulas para el cálculo de las probabilidades muestrales acumuladas de ocurrencia de sucesos o Distribuciones Acumulativas Empíricas (DAE): Estas fórmulas de ploteo son utilizadas en el procedimiento de ajuste de una ley de probabilidad teórica a datos observacionales de variables extremas, mediante el método de Variable Reducida, representadas por las siguientes expresiones:

$\widehat{F}_1(x_i') = \frac{m_i'}{n+1}$ Distribución Acumulativa Empírica por Weibull (1939); (13)

$\widehat{F}_2(x_i') = \frac{m_i'-0,31}{n+0,38}$ Distribución Acumulativa Empírica por Beard (1943); (14)

$\widehat{F}_3(x_i') = \frac{m_i'-0,4}{n+0,2}$ Distribución Acumulativa Empírica por Cunnane (1978); (15)

$\widehat{F}_4(x_i') = \frac{m_i'-0,5}{n}$ Distribución Acumulativa Empírica por Hazen (1930); (16)

donde x_i son los datos cronológicos; x_i' son los estadígrafos ordinales de la serie, o sea, los datos cronológicos x_i ordenados en forma ascendente (de menor a mayor); n es el tamaño de la muestra y m_i' es el rango de la observación x_i'. El rango m_i' es igual al número de orden, si el dato no se repite; si el dato se repite; m_i' es igual a la media aritmética de los números de orden correspondientes, y este es el rango que se le asigna a cada valor repetido. Todas estas expresiones son valores aproximados de las distribuciones teóricas [8, 12, 13].

3.1. Análisis para la selección del modelo

Se aplica el método conocido de Variable Reducida con el objetivo de seleccionar la distribución adecuada para el cálculo del valor extremo de interés. Este método consiste, para cada muestra $\{x_i\}$ de tamaño n, en la obtención de los estadígrafos ordinales $\{x_i'\}$ correspondientes (ordenación ascendente) y en la transformación por logaritmación neperiana simple o reiterada de los modelos exponenciales teóricos en formas lineales de los tipos:

$y = -\mathrm{Ln}[-\mathrm{Ln}\, F(x)] = Ax' + B$, para la distribución de Gumbel I; (17)

$y = \mathrm{Ln}[-\mathrm{Ln}(1 - F(x))] = Ax' + B$, para la distribución de Gumbel II; (18)

$y = -\mathrm{Ln}[-\mathrm{Ln}\, F(x)] = A\,\mathrm{Ln}\, x' + B$, para la distribución de Frechet I; (19)

$y = \mathrm{Ln}[-\mathrm{Ln}(1 - F(x))] = A\,\mathrm{Ln}\, x' + B$, para la distribución de Frechet II; (20)

$y = Ax' + B$, para la distribución Normal; (21)

$\mathrm{y} = A\,\mathrm{Ln}\, x' + B$, para la distribución Log-Normal; (22)

Donde y es la variable reducida y $F(x)$ se estima mediante $\hat{F}(x')$; en el caso particular de las distribuciones Log-Normal y Normal se tiene que los valores y_i de y son las abscisas de la distribución Normal (canónica) $N(y; 0; 1)$ tales que $N(y_i; 0; 1) = \hat{F}(x_i'), i = 1, 2, \dots, n$. De esta manera, ploteando los puntos (x_i', y_i) o $(\ln x_i', y_i)$, según el caso, se procede a la determinación unívoca de los parámetros A y B del modelo lineal equivalente (estimadores mínimo-cuadráticos), así como del valor adquirido por el coeficiente de correlación γ, seleccionándose conjuntamente el modelo de mejor ajuste y la fórmula de cálculo $\hat{F}$ que corresponden al mayor valor de γ (existen $6 * 4 = 24$ valores de γ, pero para valores máximos solo $4 * 4 = 16$ posibles para cada conjunto de datos y los otro 8 son para valores mínimos): En aquellos casos en que el mayor valor de γ coincide prácticamente para dos modelos distintos se toma, en principio, como modelo de mejor ajuste el que tiene el menor valor del error standard del estimado $\hat{e}$ y cuando este modelo resulta inaceptable para las estimaciones se selecciona el modelo que le sigue inmediatamente (según la gradación de γ) para caracterizar la serie en cuestión [8, 13].

4. MÉTODOS PARA EL CÁLCULO DE LOS PARÁMETROS DEL MODELO

Analizando las distribuciones para valores extremos nos damos cuenta que estas dependen de dos parámetros α y β donde el primero es un parámetro de localización y el segundo es un parámetro de escala. Para calcularlos se han implementado tres métodos, el método de Sneyers, el método de los Momentos y un método por los Mínimos Cuadrados.

a. Método de Sneyers

Este método fue planteado por Sneyers en la nota técnica 143 de la OMM (en francés) [9]. La eficiencia de estos estimadores se considera como excelente porque si $n > 10$, la eficiencia $\hat{\alpha}$ está muy cerca de uno, mientras que la eficiencia de $\hat{\beta}$ es mayor que 0.80. Cuando se tiene una muestra pequeña de datos de una

variable extrema, los resultados obtenidos pueden ser inestables debido a una subestimación o sobrestimación de los valores críticos en las extrapolaciones [8, 11].
La fórmula para Gumbel sería:

- Gumbel I: $\hat{\beta} = \frac{b_n}{n}(F_n)x_i^{'}, \ \ \hat{\alpha} = \bar{x}_i - 0.577216\hat{\beta}$ (23)
- Gumbel II: $\hat{\beta} = \frac{b_n}{n}(F_n)x_i^{'}, \ \hat{\alpha} = \bar{x}_i + 0.577216\hat{\beta}$ (24)

Mientras para Frechet sería:

- Frechet I: $\hat{\beta} = \frac{1}{\frac{b_n}{n}(F_n)lnx_i^{'}}\ , \ \hat{\alpha} = e^{\overline{lnx}_i - \frac{0.577216}{\hat{\beta}}}$ (25)
- Frechet II: $\hat{\beta} = \frac{1}{\frac{b_n}{n}(F_n)lnx_i^{'}}\ , \hat{\alpha} = e^{\overline{lnx}_i + \frac{0.577216}{\hat{\beta}}}$ (26)

Donde $F_n = \sum_{i=1}^{n}\left(1 - \sum_{j=i}^{n}\frac{1}{j}\right)$ (27)

y $b_n = 1 + e^{-0.975652 * L_n(n-1) + 1.043532 - \frac{1.950309}{L_n(n-1)} + \frac{3.574231}{L_n^2(n-1)}}$ (28)

El coeficiente b_n es un factor que se utiliza para eliminar el sesgo en la estimación.

b. Método de los Momentos

Uno de los métodos más conocido en la estimación de parámetros de una distribución de probabilidad es el de los Momentos, desarrollado por primera vez por Karl Pearson en 1902. Considera que buenos estimadores de los parámetros de una función de probabilidad son aquellos para los cuales los momentos de la función de densidad de probabilidad son iguales a los momentos correspondientes de la información de la muestra [5]. Este método es bastante sencillo y brinda estimadores bastante confiables cuando la muestra es grande.
La fórmula para Gumbel sería:

- Gumbel I: $\hat{\beta} = \frac{\sqrt{6}}{\pi}\sigma_x, \ \hat{\alpha} = \bar{x}_i - 0.577216\hat{\beta}$ (29)
- Gumbel II: $\hat{\beta} = \frac{\sqrt{6}}{\pi}\sigma_x, \ \hat{\alpha} = \bar{x}_i + 0.577216\hat{\beta}$ (30)

Donde σ_x es la desviación típica muestral de los datos.
Mientras para Frechet sería:

- Frechet I: $\hat{\beta} = \frac{\pi}{\sqrt{6}\sigma_{\ln x}}\ , \ \hat{\alpha} = e^{\overline{lnx}_i - \frac{0.577216}{\hat{\beta}}}$ (31)
- Frechet II: $\hat{\beta} = \frac{\pi}{\sqrt{6}\sigma_{\ln x}}\ , \ \hat{\alpha} = e^{\overline{lnx}_i + \frac{0.577216}{\hat{\beta}}}$ (32)

Donde $\sigma_{\ln x}$ es la desviación típica muestral del logaritmo neperiano de los datos.

c. Método de los Mínimos Cuadrados

Otra forma de calcular los estimados consiste en tomar las rectas aproximadas por el método de Mínimos Cuadrados ajustada a los datos experimentales $\langle x', Y' \rangle$ para Gumbel o $\langle \ln x', Y' \rangle$ para Frechet donde Y' es el valor de la expresión de la variable reducida correspondiente, siendo A la pendiente de la recta y B el punto donde corta el eje Y [1].
Mientras para Gumbel I y Gumbel II sería:

$\hat{\beta} = \frac{1}{A}, \ \hat{\alpha} = -B\hat{\beta}$ (33)

Para Frechet I y Frechet II sería:

$\hat{\beta} = A\ , \ \hat{\alpha} = e^{-\frac{B}{\hat{\beta}}}$ (34)

d. Método de Máxima Verosimilitud

El método de Máxima verosimilitud, selecciona como estimador al valor del parámetro que tiene la propiedad de maximizar el valor de la probabilidad de la muestra aleatoria observada, es decir, encuentra el valor de los parámetros que maximizan la función de verosimilitud. La verosimilitud es la función de densidad de probabilidad conjunta de las variables independientes [9].

Las fórmulas para los estimadores de la Normal:

$$\boldsymbol{\hat{\alpha} = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}} \qquad (35)$$

$$\boldsymbol{\hat{\beta} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n}} = s} \qquad (36)$$

Mientras para los estimadores de la Log-Normal:

$$\boldsymbol{\hat{\alpha} = \frac{1}{n}\sum_{i=1}^{n} Lnx_i = \overline{Lnx}} \qquad (37)$$

$$\boldsymbol{\hat{\beta} = \sqrt{\frac{\sum_{i=1}^{n}(Lnx_i - \overline{Lnx})^2}{n}} = sLnx} \qquad (38)$$

5. VALORES CRÍTICOS QUE CORRESPONDEN A PERÍODOS DE RETORNO O PROBABILIDADES PREFIJADAS

Una definición importante es la de los valores críticos o valores de retorno, que en el campo de la climatología tienen mucha importancia. Son considerados como los cuantiles de la distribución de valores extremos. Así, si T_p (el período de retorno) es el cuantil de orden p de una variable con una distribución teórica empírica, entonces p es la probabilidad de que T_p sea superado una vez al año; y el período de retorno $\frac{1}{p}$, es el número de unidades de tiempo que transcurrirán en media entre dos veces en los que la variable supere el valor de T_p.

Dada la probabilidad p, el período de retorno T_p de un suceso se calcula mediante la fórmula:

$$T_p = \begin{cases} \frac{1}{p} \; si \; p < 0.50 \\ \frac{1}{1-p} \; si \; p > 0.50 \end{cases} \qquad (39)$$

Por ejemplo, si X es la variable aleatoria continua del viento máximo o racha máxima anual, la cual sabemos que se distribuye según cierta ley de probabilidad con distribución acumulativa $F(x)$, la probabilidad prefijada $p = p(X > x_0) = 0.01 = 1\%$, equivale a un período de retorno de $1/0.01 = 100\ años$, o sea para $100\ años$ podemos esperar q la variable sea igual o mayor que x_0 [5].

El problema consiste en calcular, a partir de la expresión de $F(x)$ donde los parámetros han sido estimados previamente, la abscisa crítica x_p ó x_T de la variable aleatoria X que corresponde a la probabilidad o período de retorno prefijado. Despejando la x se obtienen los siguientes resultados relativos a la probabilidad prefijada p del suceso $X > x$ (si X es continua, $p(X \le x) = p(X < x)$ y $p(X \ge x) = p(X > x)$ porque las probabilidades puntuales $p(X = x)$ son necesariamente nulas):

Distribución de Gumbel I: $x_p = \hat{\alpha} - \hat{\beta} * Ln[-Ln(p)]$ (40)

Distribución de Gumbel II: $x_p = \hat{\alpha} + \hat{\beta} * Ln[-Ln(1-p)]$ (41)

Distribución de Frechet I: $x_p = \hat{\alpha} * exp\{(-\frac{1}{\hat{\beta}}) * Ln[-Ln(p)]\}$ (42)

Distribución de Frechet II: $x_p = \hat{\alpha} * exp\{\left(-\frac{1}{\hat{\beta}}\right) * Ln[-Ln(1-p)]\}$ (43)

Distribución Normal: $x_p = \hat{\alpha} + \hat{\beta} * Z_p$ (44)

Donde $\overline{N}(Z_p;\ 0,1) = 1 - p$ (Z_p es la abscisa de la distribución Normal canónica que corresponde a la probabilidad acumulada $1 - p$):

Distribución Log-Normal: $x_p = exp(ln\hat{\alpha} + \hat{\beta} * Z_p)$ (45)

Donde $\overline{N}(Z_p;\ 0,1) = 1 - p$ (Z_p es la abscisa de la distribución Normal canónica que corresponde a la probabilidad acumulada $1 - p$):

6. METODOLOGÍA

Con la idea de facilitar el trabajo y eliminar errores a la hora de realizar los cálculos, se desarrolló una aplicación visual que permite a un usuario introducir los datos de alguna variable extrema y este le brinda el mejor modelo que ajusta los datos, la gráfica que corresponde a dicho modelo, los valores esperados para $10, 20, 50\ y\ 100$ años así como los principales estadígrafos de las variables principales que pueden ser de interés. La herramienta computacional que se presenta en este documento es una aplicación de nombre

SEVECLIM.exe, desarrollada sobre la plataforma .NET y utilizando el lenguaje de programación C#. La metodología propuesta en el software para el análisis de la caracterización de variables extremas tendrá en cuenta los siguientes pasos:

PASO 1: Cargar los datos desde un fichero .csv o teclear los datos manualmente.
PASO 2: Seleccionar la opción según la variable extrema de interés (máxima o mínima):
PASO 3: Se muestra el esquema de cálculo en la aplicación del método de Variable Reducida con los datos ordenados de menor a mayor (x_i'), el rango de x_i', las cuatro DAE, el logaritmo neperiano de x_i' y las cuatros variables reducidas en la DAE correspondiente. En el caso de la variable extrema máxima los cuatro valores de la inversa de la función normal acumulada en la DAE correspondiente.
PASO 4: Se observa en la tabla, la combinación de las DAE F_i y DAT F, asociado al coeficiente de correlación, error y gráfico correspondiente.
PASO 5: Se muestran los principales estadígrafos de interés para el análisis de la variable de estudio.
PASO 6: Se muestran los valores asociados a un período de retorno calculado por la distribución con mayor coeficiente de correlación y menor error. La estimación de los parámetros de la ley ajustada por el método seleccionado, mostrando el gráfico de distribución y densidad.
PASO 7: Se guarda en un fichero .xls los datos del esquema de cálculo en la aplicación del método de Variable Reducida, la combinación de las DAE F_i y DAT F, los estadígrafos fundamentales, los valores asociados a un período de retorno para cada distribución y los parámetros de estimados por cada método.

Como resumen se ilustra gráficamente (Figura 1) la metodología utilizada en el software SEVECLIM.exe.

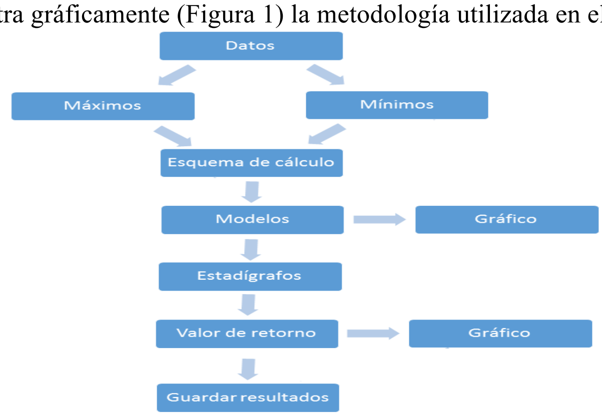

Figura 1. Metodología seguida en el software SEVECLIM.exe

6.1. Descripción del software a través del análisis del régimen del viento máximo anual en la estación meteorológica de Casablanca (1985-2014)

a. Entrada de datos

Parte principal del software es conocer los datos sobre los que se realizara el estudio y si son datos mínimos o máximos. Para facilitarles a los usuarios la entrada de los mismos al programa se le brindan las opciones de que ingresen los datos manualmente en la tabla habilitada para ello o la opción de llenarla con los datos de un fichero ".csv" (Figura 2):

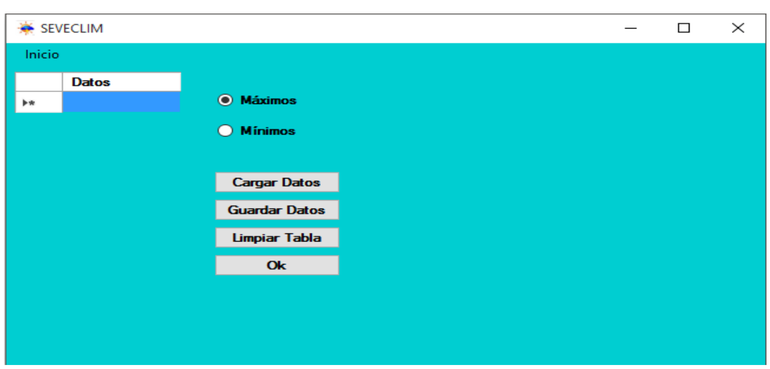

Figura 2. Pantalla inicial. Entrada de datos al sistema y selección de la variable

b. Esquema de cálculo en la aplicación del método de Variable Reducida