

COMPARACIÓN DE ANÁLISIS EXTREMO UNI Y BIVARIADO PARA MAREJADAS RECIENTES EN CHILE

Marco Antonio Álvarez Rivera, José Beyá Marshall y Ariel Andrés Gallardo Yáñez
Escuela de Ingeniería Civil Oceánica, Universidad de Valparaíso, Chile
marco.alvarez.24@gmail.com, jose.beya@uv.cl, ariel.gallardo@uv.cl

RESUMEN:

Existen muchas técnicas para estimar valores de retorno basados en la altura significativa espectral del oleaje, sin embargo, es importante conocer la incidencia de otras variables en la recurrencia de los eventos extremos. En este estudio se incorpora el periodo pico del oleaje en el cálculo de periodos de retorno mediante el empleo de funciones cópulas, las que permiten calcular la probabilidad conjunta de dos o más variables en función de sus distribuciones marginales y su grado de dependencia. Se analizan 6 bases de datos de oleaje ampliamente utilizadas para un punto ubicado costa afuera de la bahía de Valparaíso y se comparan con el enfoque tradicional que contempla sólo una variable. Finalmente se analizan y comparan distintas estimaciones del periodo de retorno para dos marejadas importantes ocurridas recientemente en Valparaíso.

ABSTRACT:

Currently there are many techniques for estimating returns periods based on the spectral significant height, however sometimes it is important to understand the influence of other variables on recurrence of extreme events. In this study, the peak period is incorporated in the calculation of return periods by using copula functions, which allow calculating the joint probability of two or more variables depending on their marginal distributions and its level of dependency. Six widely used waves databases are analyzed at a point offshore Valparaiso. The results obtained were compared with the conventional approach that considers only one variable. Finally different estimates of the return period for two important events occurred recently in Valparaiso are compared and analyzed.

PALABRAS CLAVES: periodo, retorno, bivariado

INTRODUCCIÓN

Uno de los aspectos más importantes que determina el tipo, costo y factibilidad de una obra marítima es la altura de la ola de diseño. Este valor se define en base a la recurrencia de tormentas o marejadas; eventos extremos que generan solicitaciones críticas para la estabilidad de la estructura. Conociendo las características de estas marejadas es posible estimar el periodo de retorno de cualquier tormenta, donde la altura significativa máxima es la variable más utilizada para medir la intensidad de estos eventos.

A pesar de la necesidad de conocer con precisión el clima de oleaje, en la actualidad existe una gran incertidumbre debido a la escasa cantidad de mediciones continuas y de largo plazo disponibles. Valdivia et al. (2015) mostró que, comparando diversas bases de datos de modelos de re-análisis ampliamente utilizados, existe una gran variabilidad en la determinación de periodos de retorno y la consecuente elección de la altura de diseño. Adicionalmente mostró que las bases de datos presentan ajustes deficientes en relación a las mediciones para los valores más altos de las alturas de oleaje, evidenciando el bajo nivel de confianza de los valores extremos obtenidos a partir de los análisis de dichas bases de datos.

Por otra parte, la práctica común de sólo considerar la altura para determinar el periodo de retorno puede inducir errores importantes en las solicitaciones de diseño. En el presente trabajo se realiza un esfuerzo por integrar el período pico al cálculo de los valores de retorno a partir de un análisis de valores extremos bi-variado que se compara con el enfoque actual, uni-variado, que sólo considera alturas de oleaje. Se identifican las diferencias en el cálculo de los períodos de retorno de dos eventos de gran magnitud ocurridos recientemente en Chile utilizando los métodos uni- y bi-variados y las distintas bases de datos.

Para el análisis bi-variado se utiliza un modelo de cópula de dos variables similar al utilizado por Requena et al. (2013), quien trabajó con variables hidrológicas de caudal y volumen para estimar la probabilidad de ocurrencia de crecidas en el diseño de represas. En este estudio se utiliza una metodología similar con los correspondientes ajustes al ámbito marítimo.

METODOLOGÍA

Se analizaron 6 bases de datos en un punto ubicado costa afuera y frente a la Bahía de Valparaíso, Chile (-33° , -73.5° aprox.). Para cada base se hizo un análisis de clima extremo; identificando tormentas y calculando periodos de retorno en función de una y dos variables. Además se incorporaron al análisis mediciones de boyas cercanas de las marejadas del 3-7-2013 y 8-8-2015, siendo esta última la que generó mayores daños y costos de reparación en el borde costero (Campos et al., 2015; Winckler et al., 2015).

Las bases de datos, que contienen entre 20 y 34 años de información, fueron: NOAA WW3 (NOAA, 2015), NOAA CFSR (Saha et al., 2010), FUGRO-OCEANOR obtenida del proyecto Corfo-INNOVA 09CN14-5718 “Catastro del recurso energético asociado a oleaje para el apoyo a la evaluación de proyectos de generación de energía undimotriz”, la base de datos de la Dirección de Obras Portuarias de Chile (DOP), y dos bases generadas con el modelo numérico Wavewatch III v.418 para el proyecto FONDEF IDeA IT3I20006 “Un Atlas de Oleaje para Chile”, la primera (AOC1) calibrada según recomendaciones de Ardhuin et al., (2011) y la segunda (AOC2) calibrada específicamente para clima extremo. Las mediciones de las marejadas fueron obtenidas del Servicio Hidrográfico y Oceanográfico de la Armada. La tabla 1 muestra los parámetros de cada marejada.

Tabla 1.- Parámetros característicos medidos de las marejadas de 2013 y 2015 en Valparaíso

Fecha Tormenta	Hs	Tp	Dp
03-07-2013	6.40 m	20.0 s	233°
08-08-2015	7.23 m	13.3 s	310°

Para cada base de datos se identificaron las 35 mayores tormentas mediante el método POT (PeakOverThreshold; Goda, 2000) y se fijó un espaciamiento temporal entre ellas de por lo menos 4

días para asegurar su independencia (Valdivia et al., 2015). A cada una de las bases de datos se le aplicó el análisis uni- y bi-variado, incorporando las marejadas del 3-7-2013 y 8-8-2015 para observar el nivel de ajuste de los modelos generados y comparar las diferencias en los periodos de retorno estimados a partir de las dos metodologías utilizadas.

Para el cálculo de períodos de retorno basado sólo en alturas existen muchas técnicas que permiten estimar la recurrencia de tormentas. En este estudio se utilizó el método de mínimos cuadrados. A cada serie de alturas se le ajustó 3 distribuciones de probabilidad típicas (Gumbel, Fréchet y Weibull) según sus fórmulas de ploteo de posición, se obtuvo la variable reducida y se escogió la que tuviera mayor correlación con la serie real de datos. (Valdivia et al., 2015).

Para el caso bi-variado, el periodo de retorno conjunto se obtuvo a través de funciones cópulas, las cuales tienen la propiedad de representar la relación de dependencia entre dos o más variables y permiten definir las funciones marginales de cada variable de forma independiente. La distribución de probabilidad conjunta queda definida entonces por las frecuencias marginales de cada una de las variables y la dependencia existente entre ambas (Sklar, 1959):

$$F_{X,Y}(x,y) = C(F_X(x), F_Y(y)), x, y \in R \quad [1]$$

Donde $F_X(x)$ y $F_Y(y)$ son las funciones de distribución de probabilidad acumulada de las variables y $C: [0,1]^2 \rightarrow [0,1]$ es la función cópula utilizada para representar la ley de dependencia entre ambas. Esta puede ser de tipo paramétrico o empírico dependiendo de si su expresión responde o no a una ecuación paramétrica. En el primer caso, existe un parámetro (θ) que cuantifica la relación de dependencia entre las variables, mientras que en el segundo no participa ningún parámetro y es la función cópula la que se ajusta de forma local a los datos.

La familia de cópulas más utilizada es la Arquimediana, dentro de la cual destacan las cópulas Frank, Clayton y Gumbel, siendo esta última la que mejor representa la dependencia entre valores extremos. Las variables utilizadas corresponden a la altura significativa (X) y el periodo pico (Y) de las tormentas definidas anteriormente.

Se ajustó una distribución a la serie de alturas y otra a la serie de períodos y se encontró la función cópula que mejor las relacionaba, según la metodología descrita por Requena et al. (2013). Las distribuciones consideradas para la serie de alturas y periodos fueron las siguientes: Normal, LogNormal, Gamma, Weibull, Gumbel, Logistic, LogLogistic y Valores Extremos Generalizada. Para la serie de alturas además se ajustó la distribución de Pareto Generalizada utilizando el valor umbral obtenido del método POT como parámetro de localización; los demás parámetros de ésta y el resto de las distribuciones se definieron a partir de los estimadores de máxima verosimilitud. La elección de la mejor distribución se hizo en base al test de Kolmogorov-Smirnov, escogiéndose aquella que tuviera el mayor p-valor para un nivel de significancia de 0.05.

Una vez definidas las distribuciones de altura y periodo se testearon 4 funciones cópula y se eligió la de mejor ajuste según el estadístico de Cramér-von Mises (S_n ; Requena et al., 2013; [2]), el cual cuantifica la diferencia entre la cópula empírica y la paramétrica testeada. Se consideraron las cópulas Frank, Clayton, Gumbel y la Gaussiana como posibles candidatas. Para generar el ajuste se calculó el parámetro θ de cada cópula a partir del coeficiente de Kendall, un indicador que mide el nivel de dependencia entre las variables. También se consideró la posibilidad de que las variables tuvieran cero dependencia en cuyo caso la probabilidad acumulada se calculó como $F_X(x) \cdot F_Y(y)$. Se seleccionó la cópula que tenía el menor valor del estadístico de Cramér-von-Mises, definido como:

$$S_n = \sum_{i=1}^n \left\{ C_n \left(\frac{R_i}{n+1}, \frac{S_i}{n+1} \right) - C_{\theta_n} \left(\frac{R_i}{n+1}, \frac{S_i}{n+1} \right) \right\}^2 \quad [2]$$

Donde C_n es la cópula empírica, C_{θ_n} es la cópula paramétrica estimada a partir de las observaciones, R_i y S_i son los rankings asociados a las series observadas de las variables y n es el número de tormentas.

Una vez identificada la función de distribución conjunta formada por las leyes de frecuencia marginales y la cópula, se calculó la probabilidad acumulada conjunta de acuerdo a la ecuación [1] para un conjunto de valores de altura y periodo, y luego se calculó el periodo de retorno a esa probabilidad.

Dado que la serie de periodos (Y) no corresponde a una serie de valores extremos, se definió el periodo de retorno bi-variado de forma distinta dependiendo de si el valor de la variable Y estaba en la cola inferior o superior de la distribución. Si el valor se encontraba sobre el punto máximo de la función densidad (en la cola superior), el periodo de retorno conjunto se definió de la forma clásica; a partir de la probabilidad de excedencia de alturas y de periodos (Requena et al., 2013):

$$T_{X,Y+} = \frac{u_T}{P(X > x \wedge Y > y)} = \frac{u_T}{1 - F_X(x) - F_Y(y) + C(F_X(x), F_Y(y))} \quad [3]$$

Por otro lado, si el valor de la variable Y se encontraba bajo el punto máximo de la función densidad (en la cola inferior), el periodo de retorno conjunto se definió a partir de la probabilidad de excedencia de altura y la probabilidad de ocurrencia de periodos, según la ecuación:

$$T_{X,Y-} = \frac{u_T}{P(X > x \wedge Y \leq y)} = \frac{u_T}{F_Y(y) - F_X(x) + C(F_X(x), (1 - F_Y(y)))} \quad [4]$$

Donde $C(F_X(x), F_Y(y)) = P(X \leq x \wedge Y \leq y)$ y u_T es el tiempo promedio de arribo entre dos eventos sucesivos (N° años/ N° eventos).

RESULTADOS

Del análisis uni-variado se obtienen curvas de altura significativa versus periodo de retorno para cada base de datos. En la figura 1 se puede ver el nivel de ajuste de cada distribución y las tormentas identificadas según el método POT. Además se agregaron líneas horizontales de las dos marejadas más importantes ocurridas recientemente en Valparaíso para ver las diferencias en los periodos de retorno teóricos predichos por las distintas bases de datos. Los resultados de este análisis se presentan en la tabla 2.

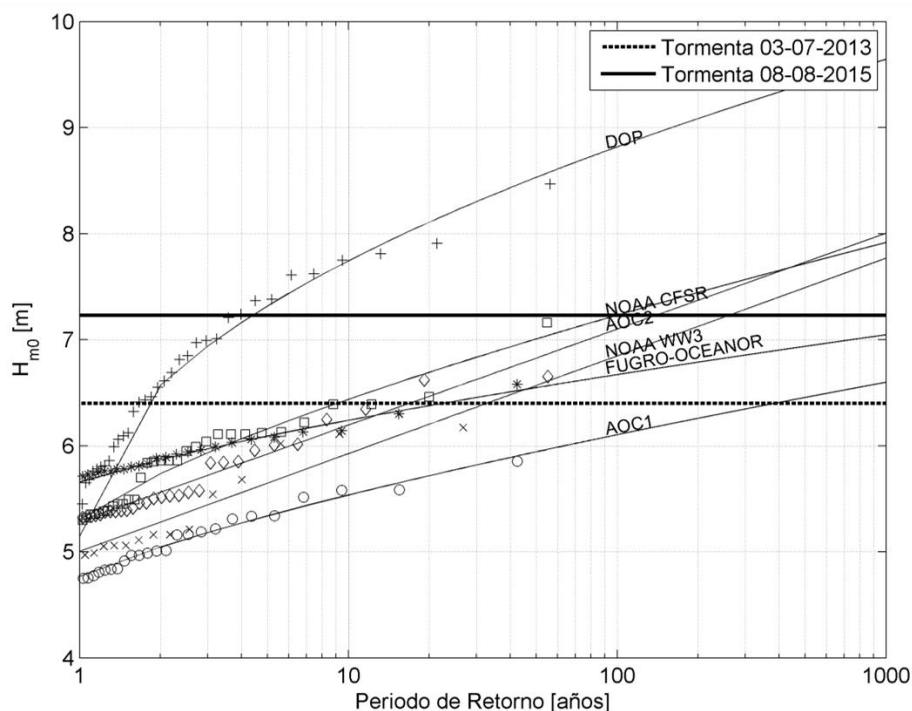


Figura 1.- Altura significativa v/s periodo de retorno para diversas bases de datos incluyendo las alturas del análisis extremo (símbolos). Líneas horizontales indican dos marejadas recientes en Valparaíso.

El análisis bi-variado, si bien considera un método diferente y más alternativas para el ajuste de la serie de alturas, arroja resultados similares a los del uni-variado respecto a esta variable, siendo la distribución Generalizada de Pareto la función con la que se obtiene el mejor ajuste para casi todas las bases. A modo de ejemplo, en la figura 2a, se presenta el ajuste de dicha distribución a la serie de alturas de la base AOC1. Para el caso de la serie de periodos, el ajuste nunca fue satisfactorio ya que en casi todas las bases se encontró una gran aleatoriedad en los datos. A pesar de esto, la distribución Gumbel resultó ser la función de mejor ajuste para 3 de las 6 bases (figura 2b). Respecto a la distribución conjunta se observó un grado de dependencia casi nulo, lo que se ve reflejado en el grado de dispersión de los datos (figura 2c), donde se grafica $R_i/(n+1)$ vs $S_i/(n+1)$ siendo R_i el ranking de X_i dentro de la serie de datos de altura y S_i el ranking de Y_i dentro de la serie de datos de periodo. La poca dependencia entre las variables hizo que no se viera una tendencia clara durante la selección de la cópula, lo que demuestra la poca relación existente entre la altura significativa y el periodo pico de las tormentas presentes en esta zona de Chile (figura 2d).

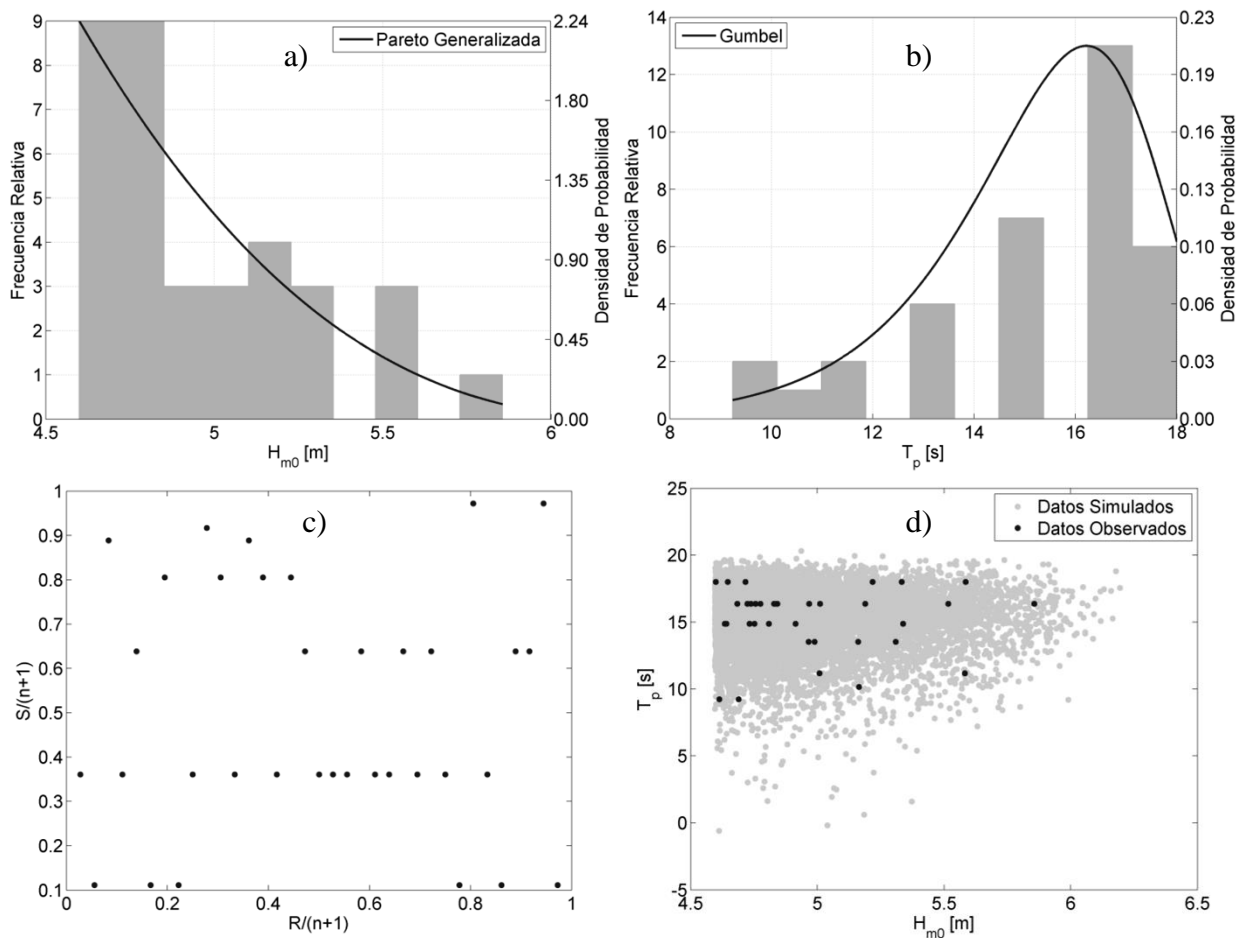


Figura 2.-Ajuste de la base de datos AOC1. a) Ajuste de la serie de alturas. b) Ajuste de la serie de periodos. c) Gráfico de ranking. d) Ajuste de la función cópula.

Las distribuciones de mejor ajuste para cada variable y la cópula que mejor representa la dependencia entre éstas se presentan en la tabla 2. En la figura 3 se muestra el gráfico final con los periodos de retorno más importantes para cada base de datos.

Tabla 2.-Distribución y cópula de mejor ajuste para cada base de datos.

Base de Datos	Función de Distribución de Mejor Ajuste		
	Altura	Periodo	Cópula
AOC1	Pareto	Gumbel	Independiente
AOC2	Pareto	Gumbel	Clayton
FUGRO-OCEANOR	LogLogistic	Logistic	Independiente
NOAA CFSR	Pareto	Normal	Clayton
NOAA WW3	GEV	Gumbel	Gumbel
DOP	Pareto	Weibull	Frank

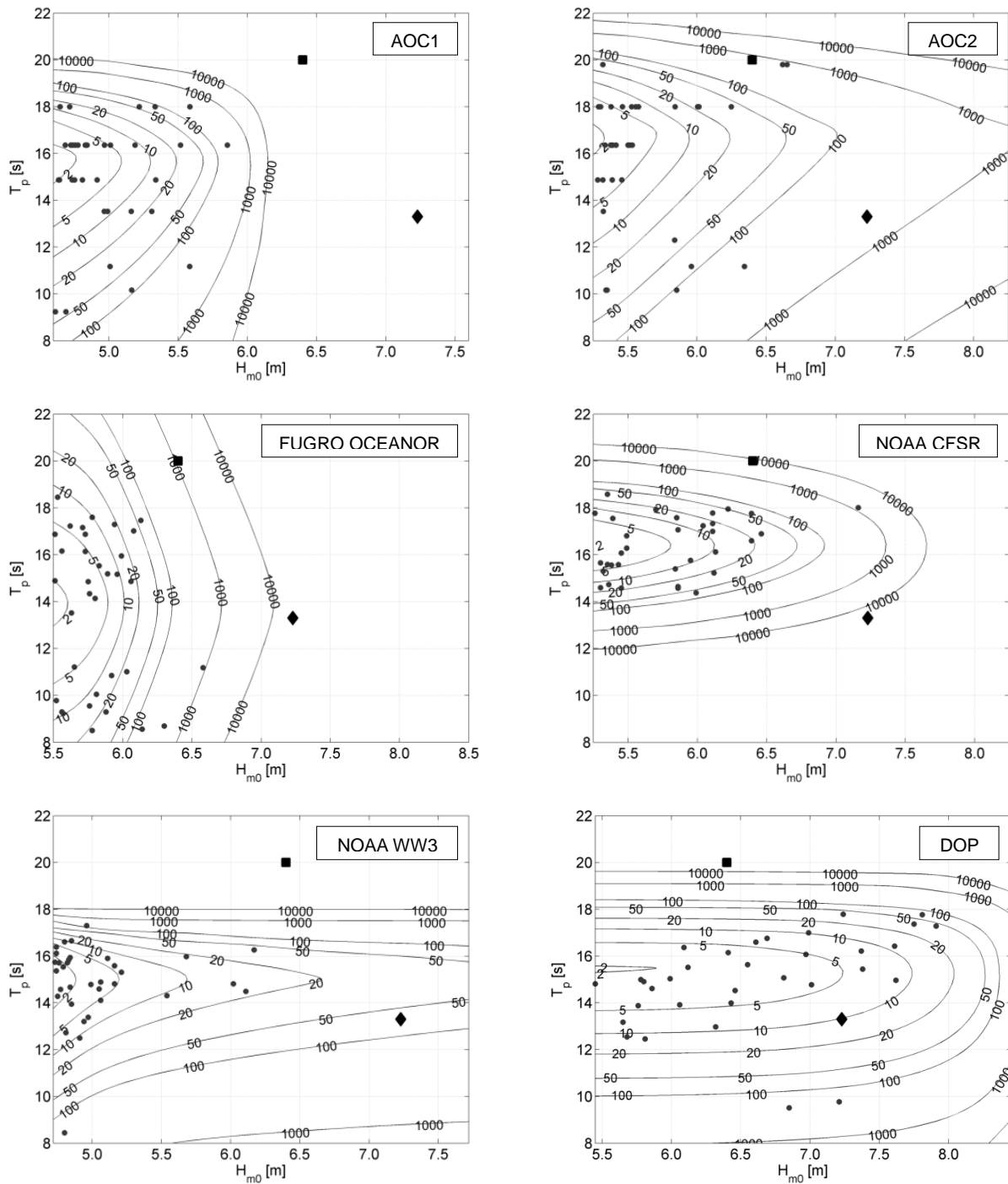


Figura 3.-Curvas de periodo de retorno para altura significativa espectral y periodo pico para cada base de datos. Los puntos corresponden a las tormentas seleccionadas mediante el método POT, el cuadrado corresponde a la tormenta del 3-7-2013 y el diamante a la tormenta del 8-8-2015.

Tabla 3.- Periodos de retorno uni- y bi-variados para dos marejadas y diversas bases de datos. Se incluye el error medio (H_{m0} modelo – H_{m0} medición) y la banda de confianza del 95% (según distribución normal) como indicador de la bondad de ajuste entre las 20 mayores mediciones de boya coincidentes con las bases de datos.

Período de retorno (años)	03-07-2013		08-08-2015		Error medio $H_{m0} \pm$ Banda de Confianza 95% [m]
	Uni-variado	Bi-variado	Uni-variado	Bi-variado	
AOC1	385	>10000	>10000	>10000	-0,77 \pm 0,74
AOC2	17	595	139	508	-0,27 \pm 1,18
FUGRO-OCEANOR	22	1115	3361	>10000	0,11 \pm 1,05
NOAA CFSR	9	6230	100	>10000	-0,41 \pm 0,95
NOAA WW3	33	>10000	260	66	-0,77 \pm 1,04
DOP	2	>10000	4	10	-

CONCLUSIONES

El método bi-variado presentado en este trabajo considera más alternativas de distribución para el ajuste de las variables que el uni-variado tradicional y permite evaluar la frecuencia de ocurrencia de eventos extremos considerando también su período pico. Esto es importante ya que la marejada del 3-7-2013 presentó un periodo pico muy elevado y fue evidentemente un evento atípico.

El análisis bi-variado presenta un mayor grado de incertidumbre debido a la incorporación de la función cópula, agrega supuestos y ajustes al análisis uni-variado convencional. En el análisis bi-variado además, para la obtención de las cópulas, es necesario incorporar un mayor número de datos que no son eventos extremos según el análisis uni-variado. El grado de dependencia entre las variables tiene una influencia importante en los resultados, obteniéndose estimaciones de mayor confiabilidad cuando existe algún tipo de tendencia. En este caso, sin embargo, la dependencia resultó ser casi nula lo que llevó a un ajuste deficiente de las funciones cópulas.

Se observan diferencias superiores a los 10.000 años en los períodos de retorno estimados entre bases de datos y métodos uni- y bi-variados para ambas marejadas. La diferencia entre las bases de datos puede explicarse por la utilización de distintos modelos de generación de oleaje mientras que la diferencia entre los métodos se debe a la incorporación del periodo pico como parámetro adicional, lo cual produce estimaciones de periodos de retorno generalmente más altas. Esto se ve reflejado claramente en la marejada del 2013, la cual, al presentar un periodo pico atípico para la mayoría de las bases de datos, genera una estimación de recurrencia muy baja para tormentas de este tipo en el futuro.

La única excepción a lo anteriormente descrito es el caso de NOAA WW3; base que arroja un periodo de retorno menor según el análisis bi-variado para la tormenta del 2015. Esto se debe principalmente a que se usaron distribuciones diferentes para el ajuste de la serie de altura de cada metodología, lo que arrojó probabilidades de excedencia mayores (periodos de retorno menores) para el caso bi-variado. Se recomienda entonces elegir la misma metodología de ajuste para futuros estudios con el fin de obtener resultados más fáciles de contrastar.

La confiabilidad de los períodos de retorno obtenidos es baja debido a que ninguna base de datos, incluyendo las comerciales FUGRO-OCEANOR y DOP, son capaces de reproducir las marejadas medidas con exactitud. Esto se refleja en que los errores absolutos de todas las bases de datos son superiores a 1 m para un nivel de confianza del 95%. Todas las bases de datos presentan importantes sesgos hacia la subestimación a excepción de FUGRO-OCEANOR que tiene un error medio (sesgo) positivo. Esta baja confiabilidad pone en evidencia la validez de los estudios de oleaje de los diseños de las obras marítimas.

El análisis bi-variado tiene el potencial de entregar información más detallada para el diseño, sin embargo, se requiere contar con estadísticas de oleaje de mejor calidad para obtener resultados confiables. Por esta razón es urgente contar con una mayor cantidad de mediciones y continuar con los esfuerzos por lograr modelos de re-análisis de oleaje más precisos. Se recomienda para futuras investigaciones integrar otros parámetros como la dirección del oleaje en un análisis multivariado.

REFERENCIAS

- Ardhuin, F., J. Hanafin, Y. Quilfen, B. Chapron, P. Queffelec, M. Obrebski, y otros.** (2011). "Calibration of the "IOWAGA" global Wave Hindcast (1991-2011) Using ECMWF and CFSR Winds". *12th International Workshop on Wave Hindcasting and Forecasting, Kohala Coast, Hawai'i*. Hawái, Estados Unidos.
- Campos, R., Beyá, J. y Mena, M.** (2015). "Cuantificación de los daños históricos a infraestructura costera por marejadas en las costas de Chile". *XXII Congreso Chileno de Ingeniería Hidráulica*. Santiago, Chile.
- Goda, Y.** (2000). "Random seas and design of maritime structures (2nd ed.)". *Yokohama National University*. Yokohama, Japón.
- González, M. y Córdova, J.** (1997). "Aplicación de distribuciones de probabilidades bivariadas en la evaluación de daños ambientales causados por ocurrencia de eventos extremos simultáneos". *IV Congreso Interamericano sobre Medio Ambiente*. Caracas, Venezuela.
- Liu, Z. y Frigaard, P.** (2001). "Generation and Analysis of Random Waves". *Laboratory for Hydraulik og Havnebygning, Institutet for Vand, Aalborg Universitet*. Aalborg, Dinamarca.
- Mediero O., Garrote de Marcos, L.** (2015). "Comprobación de la seguridad hidrológica de aliviaderos mediante un análisis multivariado con la técnica de las cópulas". *IV Jornadas de Ingeniería de Agua*. Córdoba, España.
- NOAA** (2015). National Oceanographic and Atmospheric Administration. Recuperado May-2015, NOAA Wavewatch III (NWW3) Global wave forecast model data archive. <http://polar.ncep.noaa.gov/waves/download.shtml>.
- Requena, A. I., Mediero, L., y Garrote, L.** (2013). "A bivariate return period based on copulas for hydrologic dam design: accounting for reservoir routing in risk estimation". *Hydrology and Earth System Sciences*. Madrid, España.
- Saha, S.** (2010). NCEP Climate Forecast System Reanalysis (CFSR) Selected Hourly Time-Series Products, January 1979 to December 2010. Recuperado el 03 de 2015, de Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory.: <http://dx.doi.org/10.5065/D6513W89>.
- Sklar, A.** (1959). "Fonctions de répartition à n dimensions et leurs marges". *Publications de l'Institut de Statistique de L'Université de Paris*. París, Francia.
- Valdivia, J., Gallardo, A. y Beyá, J.** (2015). "Incertidumbre en los valores extremos del oleaje en Chile". *IV Congreso de Oceanografía Física, Meteorología y Clima (Presentación Oral)*. Valparaíso, Chile.
- Winckler, P., Contreras, M., Beyá, J. y Molina, M.** (2015). "El temporal del 8 de agosto de 2015 en la bahía de Valparaíso". *XXII Congreso Chileno de Ingeniería Hidráulica*. Santiago, Chile.