

極値波浪解析における記録更新値の統計的特性

北野利一*

再現確率波高は、極値の平均発生間隔として再現期間を定義するに必要な基準値である。したがって、具体的な標本においては、再現期間の十数倍程度の長時間に対し、はじめて現実的な平均間隔という意味を持つものである。本研究では、再現確率波高および再現期間を拡張した概念として、与えられた将来期間における記録更新値の漸近理論を誘導した。記録更新値の統計特性として、超過確率が十分小さくなる程度に既往最大値が十分に大きい時には、漸近理論が適用でき、記録更新値の確率分布は一般化極値分布や一般化パレート分布に密接に関係することを示した。

1. はじめに

極値波浪解析においては、解析対象とする波浪資料に対し、適合分布の採択、位置および尺度母数の推定、所与の再現期間に対応する確率波高の推定および信頼区間を求めるなどを主な目的としている。しかし、推定すべき量の一つである確率波高は、本来の定義から考えれば、予め再現期間を与えて定まる量ではなく、極値の平均発生間隔として再現期間を定義するに必要な基準値である。したがって、確率波高は来襲外力の規模を表わす代表量の一つであるが、誤解が生じやすい概念量である。それは、唯一の標本としての現実に対し、再現期間の数倍あるいは十数倍の長期間に対して、はじめて現実的な平均値としての意味を持つものである。

本研究では、対象とする波浪資料の既往最大値に対し、再現統計量と再現期間を拡張させた概念として、想定する将来の期間を与え、その期間内に既往最大値を越える記録更新値の統計的特性を理論的に検討するものである。なお、統計数学の分野で検討されている新記録の理論では、引き続くな記録を定めるための既往最大値として新記録を定義して、常に更新される既往最大値の確率過程の特性を考察するのに対して、本研究の記録更新値は定義が異なる。

2. 再現統計量と再現期間

再現統計量、すなわち、確率波高は、本来の定義にもどって考えれば、極値の発生間隔の平均量として再現期間を定義するに必要な基準値である。

ある基準値としての波高 h を予め与え、累積確率関数 F とする毎年最大値資料に対して、基準波高 h_R を越える事象の発生間隔が r 年となる確率分布 $P(r)$ は、

$$P(r) = \{F(h)\}^{r-1} \{1 - F(h)\} \quad (1)$$

である。式 (1) で表わされる確率分布の平均値、すなわち、平均発生間隔としての再現期間 R は、

$$R = E(r) = \sum_{r=1}^{\infty} r P(r) = \{1 - F(h)\}^{-1} \quad (2)$$

となる。この結果をもとに、再現期間 R を与えた時に、

再現期間の逆数 R^{-1} が超過確率 $\{1 - F(h_R)\}$ となるようにして、確率波高 h_R を得るのである。また、式 (2) より、波高 h_R を下限値とし、それを上回る波高が発生する割合が R^{-1} であるとも解釈できる。

基準値を越える事象の発生回数 k に着目すれば、各年に基準値 h_R を越える確率 $a(=1 - F(h_R))$ を用いて、 $L(\gg 1)$ 年間に確率波高 h_R を越える波高の発生回数 k の確率分布 $P(k; aL)$ は、近似的に、Poisson 分布

$$P(k; aL) = \frac{(aL)^k}{k!} e^{-aL} \quad (3)$$

従う。本来とりうるべき発生回数は範囲 $[0, L]$ の整数であるのに対し、近似式である式 (3) の定義域は非負の整数全てとなる点に注意する。発生間隔の平均が再現期間であるため、再現期間 R 年間に平均して 1 回発生することになる。したがって、再現期間内に確率波高 h_R を越える波高の発生回数 k の確率分布 $P(k)$ は、

$$P(k) = P(k; 1) = e^{-1}/k! \quad (4)$$

となる。図-1 に示すとおり、再現期間内に確率波高を越える波高が発生しない確率 $P(k=0)$ は十分に大きく、発生する確率 $\{1 - P(k=0)\}$ の約半分である。

期間 L に生じる毎年最大値を $h_{i=1-L}$ とする資料に基づけば、i) 確率波高 h_R を与える場合には、波高 h_R を越える発生時間間隔 τ_{ij} ($=j-i; h_i, h_j > h_R$ かつ $h_{i+1, j} \leq h_R$) の平均値が再現期間 R に収束し、ii) 再現期間 R を与える場合には、波高 $h_{i=1-L}$ を大きさの順に並びかえて、上位から R/L の割合にある波高の下限値が確率波高 h_R に収束する。なお、図-1 を見るとおり、再現期間内に基準値を越える事象の発生回数は 0 ~ 3 回程度に集中するため、上述の収束値として、再現期間および確率波高が現

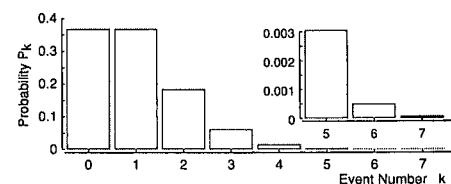


図-1 再現期間内に確率波高を越える波高の発生確率

実的に意味をもつためには、想定する期間 L を再現期間 R の十数倍～数十倍程度の期間とする必要がある。

以上の考察により、想定する再現期間と同程度か、より短い期間で与えられる現実的な標本に対して、再現期間および確率波高は、想定する再現期間以上の長い時間が経過して初めて実質的な意味をもつ量であることがわかる。現実的な標本に対する問題としては、むしろ、想定する期間を与え、その期間内で発生する高波がどのような特性をあるのか、という点が重要である。本研究では、対象とする資料の既往最大値に対し、想定する将来の期間内に、既往最大値を越える記録更新値、特に、工学的に重要となる、記録更新の最大値および最小値について、統計的特性を検討する。

3. 順序統計量と極値の基礎理論

(1) 順序統計量およびその確率分布

母分布から抽出される、標本サイズ N の標本について、最小値を1番とし、以下の関係を満たすように、昇順に順位 $i (=1 \sim N)$ を付けて並びかえた標本の値を順序統計量 x_i とよぶ。

$$x_{i=1} \leq x_{i=2} \leq \cdots \leq x_i \leq \cdots \leq x_{i=N} \quad (5)$$

各順序統計量 x_i の確率密度関数 $f_i(x_i)$ は、母分布の累積確率関数 $F(x)$ およびその密度関数 $f(x)$ を用いて、

$$f_i(x_i) = N_{N-i} C_{i-1} F^{i-1}(x_i) \{1 - F(x_i)\}^{N-i} f(x_i) \quad (6)$$

である(合田1990)。上式は、標本サイズ N 個の値の内、範囲 $x_i \leq x < x_i + dx_i$ の値をとる1つを除いた($N-1$)個に対し、($i-1$)個が x_i を越えない値をとり、($N-i$)($= (N-1)-(i-1)$)個が x_i を越える値をとる組み合わせの確率として2項分布より得られる確率

$$N_{N-i} C_{i-1} F^{i-1}(x_i) \{1 - F(x_i)\}^{N-i} \quad (7)$$

と、標本サイズ N の標本から範囲 $x_i \leq x < x_i + dx_i$ の値が1つ抽出される確率

$$N f_i(x_i) dx_i \quad (8)$$

の合成確率として導かれる。同様な考え方から、2つの順序統計量 x_i と $x_j (> x_i)$ の結合確率密度 $f_{i,j}$ も得られ、

$$\begin{aligned} f_{i,j}(x_i, x_j) &= \{N! f(x_i) [F(x_i)]_{i(i-1)}\} \\ &= [F(x_j) - F(x_i)]_{j(j-i-1)} [1 - F(x_j)]_{j(N-j)} f(x_j) \end{aligned} \quad (9)$$

となる。上式において、

$$[y]_{ik} = \frac{y^k}{k!} \quad (10)$$

という表記を用いている。

式(6)および(9)から、順序統計量 x_i を条件とする順序統計量 $x_j (> x_i)$ の条件付確率密度 $g(x_j|x_i)$ は、

$$g_{j,i}(x_j|x_i) = \frac{[1 - F(x_i)]_{j(N-j)} [F(x_j) - F(x_i)]_{j(j-i-1)}}{[1 - F(x_i)]_{j(N-i)}} f(x_j) \quad (11)$$

と得られる。上式の特別な場合として、i) 2つの順序統計量 x_i および x_j の順位が隣合う場合($j=i+1$)とii)大きい方の順序統計量 x_j の順位が標本内の最大順位となる場合($j=N$)には、順序統計量 $x_j (> x_i)$ の条件付確率密度 $g_{i+1,i}(x_j|x_i)$ および $g_{N,i}(x_j|x_i)$ は、

$$g^{(a)}(s) = as^{a-1} \frac{f(x_i)}{1 - F(x_i)} \quad (12)$$

とおいて、

$$g_{i+1,i}(x_j|x_i) = g^{(N-i)} \left(\frac{1 - F(x_j)}{1 - F(x_i)} \right) \quad (13)$$

および

$$g_{N,i}(x_j|x_i) = g^{(N-i)} \left(\frac{F(x_j) - F(x_i)}{1 - F(x_i)} \right) \quad (14)$$

となる。式(13)および(14)を区間 $[x_i, x_j]$ で積分することにより、特別な場合i)およびii)における順序統計量 x_j の条件付累積確率関数は、

$$G_{i+1,i}(x_j|x_i) = \int_{x_i}^{x_j} g_{i+1,i}(x|x_i) dx = 1 - \left\{ \frac{1 - F(x_j)}{1 - F(x_i)} \right\}^{N-i} \quad (15)$$

および

$$G_{N,i}(x_j|x_i) = \int_{x_i}^{x_j} g_{N,i}(x|x_i) dx = \left\{ \frac{F(x_j) - F(x_i)}{1 - F(x_i)} \right\}^{N-i} \quad (16)$$

となる。

(2) 極 値 理 論

母分布から抽出された標本の“平均値”が正規分布に従うように、母分布から抽出された標本の“最大値”的確率分布は、Fisher-Tippettの3種類の極値分布のいずれかに漸近する。3種類の極値分布は、一般化極値分布(GEV)として、次式に示される1つの関数系で表現することができる。

$$P_{GEV}(x; \mu, \sigma, \xi) = \exp \left\{ - \left[1 + \xi \left(\frac{x-\mu}{\sigma} \right) \right]^{-1/\xi} \right\} \quad (17)$$

上式の形状母数 ξ がゼロ、正および負をとることにより、性質の異なる3種類のFisher-Tippett分布に対応する。漸近する極値分布の種類(したがって、形状母数 ξ)を極値の吸引域(Domain of Attraction)とよび、母分布の非超過確率関数 $F(+\infty$ を含め、確率変数の定義域の上限値を x_∞)およびその密度関数 f に対し、

$$\xi = \lim_{x \rightarrow x_\infty} \frac{d}{dx} \left\{ \frac{1 - F(x)}{f(x)} \right\} \quad (18)$$

により形状母数が判定される(von Mises, 1954)。また、定数 μ および σ は、極値分布における位置および尺度母数であり、母分布の関数および最大値を選出するために抽出される数(いくつの内の最大値か)に依存する。

極値分布に密接に関連する分布として、閾値を越える量に関する漸近分布として、一般化パレート分布(GPD)

が知られる(Pickands, 1975)。すなわち、閾値 u とし、確率変数 $x(>u)$ に対して、

$$P_{GPD}(x-u; \sigma, \xi) = 1 - \left(1 + \xi \frac{x-u}{\sigma}\right)^{-1/\xi} \quad \dots \dots \dots (19)$$

と表わされる累積確率分布である。閾値を越える部分極大値資料にGPDを適用する方が、毎年最大値資料などの期間最大値資料に、最大値に関する漸近分布であるGEVを適用した推定よりも、標本数が増える分、推定精度が一般に向かうことが知られている。水文学分野では、Coles・Tawn(1996)が日降雨量資料に、海岸工学分野では、間瀬ら(2000)および北野ら(2002)が波浪の極大値資料に適用した検討を行っている。

式(19)におけるGPDの形状母数 ξ も式(18)より得られる値に一致し、尺度母数 σ は閾値および母分布に依存する。さらに、部分極大値資料を期間最大値資料として扱って得られるGEV分布の母数は、部分極大値資料に基づいて得られるGPDの母数および極大値の発生率により表現が可能である(Davison・Smith, 1990)。

閾値 u が十分に大きく、閾値を超過する確率が非常に小さくなり、累積確率分布 F に対して、

$$\varepsilon = 1 - F(u) \ll 1 \quad \dots \dots \dots (20)$$

となるならば、式(19)より、次式が得られる。

$$1 - F(x) = \varepsilon \left(1 + \xi \frac{x-u}{\sigma}\right)^{-1/\xi} \quad (x > u) \quad \dots \dots \dots (21)$$

この結果は、式(17)のGEVで表わされる毎年最大値資料を母分布とする場合にも当てはまる。すなわち、

$$\varepsilon_u = \left(1 + \xi \frac{u-\mu}{\sigma}\right)^{-1/\xi} \ll 1 \quad \dots \dots \dots (22)$$

を満たす閾値 u に対して、

$$\left(1 + \xi \frac{x-u}{\sigma}\right)^{-1/\xi} = \varepsilon_u \left(1 + \xi \frac{x-u}{\sigma_u}\right)^{-1/\xi} \ll 1 \quad \dots \dots \dots (23)$$

となり、式(21)の特別な場合として、

$$1 - P_{GEV}(x) = \varepsilon_u \left(1 + \xi \frac{x-u}{\sigma_u}\right)^{-1/\xi} \quad (x > u) \quad \dots \dots \dots (24)$$

と得られる。上式に含まれる定数 ε_u および σ_u は、閾値 u に依存する。以上の性質は次章で用いる。

4. 記録更新値の理論

(1) 記録更新数を既知とする(与える)場合

過去 $(N-L)$ 年と将来期間 L 年の期間に発生する合計 N 個の各年最大値を順序統計量として扱い、過去 M 年間の既往最大値 x_{PM} を順序統計量 x_i とすれば、将来 L 年間に $K=(N-i)$ 年分の各年最大値が既往最大値を越えていくことになる(図-2を参照)。以下では、記録更新値の発生数 K を記録更新数とよぶ。

記録更新数を“既知”とする場合とは、すでに記録された標本を対象に、過去のある時点(図-2では $y=10$)を境に、それ以前の過去における既往最大値に対し、そ

の時点以後に発生した記録更新値(すでに記録済)の統計特性の検討を想定する特殊な扱い(図-2B)である。

これに対し、次節での、記録更新数が“未知”であるとは、記録された標本全ての期間における既往最大値とし、これから記録されるべき未来の期間における記録更新値を対象とする、記録更新値の問題としての一般的な扱い方(図-2A)となる。

各年最大値 x の母分布 $F(x)$ が与えられている場合には、式(15)および(16)より、 K 年分の記録更新値における最大 x_M および最小値 x_m の累積確率分布 $G_{M,m}^{(K)}$ は、

$$G_{M,m}^{(K)}(x_M|x_{PM}) = \left\{1 - \frac{1 - F(x_M)}{1 - F(x_{PM})}\right\}^K \quad \dots \dots \dots (25)$$

および

$$G_m^{(K)}(x_m|x_{PM}) = 1 - \left\{\frac{1 - F(x_m)}{1 - F(x_{PM})}\right\}^K \quad \dots \dots \dots (26)$$

と得られる。さらに、既往最大値 x_{PM} が十分に大きく、

$$1 - F(x_{PM}) \ll 1 \quad \dots \dots \dots (27)$$

を満たす時には、

$$\frac{1 - F(x)}{1 - F(x_{PM})} = \left(1 + \xi \frac{x - x_{PM}}{\sigma_{PM}}\right)^{-1/\xi} \quad (x \geq x_{PM}) \quad \dots \dots \dots (28)$$

となる。上式において、 ξ は母分布の関数 F に依存し、式(18)により決まる定数、 σ_{PM} は母分布および既往最大値に依存する定数である。式(25)および(26)に式(28)を適用し、 K 年分の記録更新値における最大値 x_M および最小値 x_m の累積確率分布 $P_{M,m}^{(K)}$ は、漸近的に、

$$P_{M,m}^{(K)}(x_M|x_{PM}) = \left\{1 - \left(1 + \xi \frac{x - x_{PM}}{\sigma_{PM}}\right)^{-1/\xi}\right\}^K \quad \dots \dots \dots (29)$$

および

$$1 - P_m^{(K)}(x_m|x_{PM}) = \left(1 + \xi \frac{x - x_{PM}}{\sigma_{PM}}\right)^{-1/\xi} \quad \dots \dots \dots (30)$$

に従う。なお、式(30)に現れる定数 σ' および ξ' は、式(29)における σ_{PM} および ξ と以下の関係がある。

$$\left(\frac{\sigma'}{\xi'}\right) = \frac{1}{K} \left(\frac{\sigma_{PM}}{\xi}\right) \quad \dots \dots \dots (31)$$

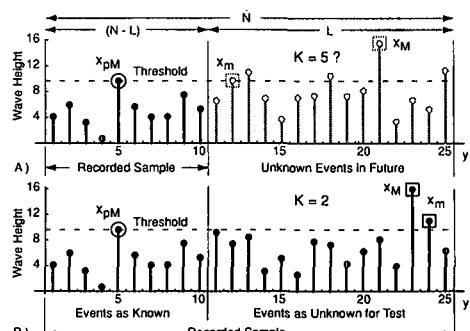


図-2 記録更新値の定義(2とおりの標本の扱い方)

式(29)～(31)を見るとおり、過去の期間や将来期間 L は、直接的には、記録更新値の確率分布 $P_{M,m}^{(K)}$ に無関係であり、 $P_{M,m}^{(K)}$ に直接関係するのは、記録更新数 K のみとなることは明らかである。また、式(30)より、毎年最大値の母分布 F がどのようなものであっても、記録更新の最小値 x_m は、漸近的にGPDに従う。また、その母数 (ϕ, ξ) が式(31)で表わされることから、記録更新数 K が多ければ多いほど、GPDの広がりのスケールは小さくなり、記録更新の最小値 x_m は、既往最大値 x_{PM} のより近傍値をとる傾向となる。記録更新数 $K=1$ の時、唯一の記録更新値は最大かつ最小があるので、確率分布 $P_{M,m}^{(K=1)}$ は確率分布 $P_{m,m}^{(K=1)}$ に一致している。

また、特別な場合として、各年最大値が GEV に従う場合、すなわち、各年最大値が極値となっている場合、GEV の母数 (μ, σ) および既往最大値 x_{PM} を用いて、式 (29) および (31) にあらわれる σ_{PM} は、

$$\frac{\sigma_{PM}}{\sigma} = 1 + \xi \frac{x_{PM} - \mu}{\sigma} \quad \dots \dots \dots \quad (32)$$

と表せる。

(2) 記録更正值数を未知とする場合

想定する将来期間 L 年の間に生じる、記録更新数 K の確率分布 $P_{(K)}$ は、式 (3) より、

$$P_{\{K\}}(x_{pM}) = P(K; l) = [l]_{\{K\}} e^{-l} \quad \dots \dots \dots \quad (33)$$

と表わされる Poisson 分布に従う。ただし、式(10)の表記法を用いており、定数 l は、既往最大値 x_{PM} に依存し、次式で示す平均更新数である。

$$l = L\{1 - F(x_{pM})\} \quad \dots \dots \dots \quad (34)$$

式(29)および(30)で得られた記録更新値の累積確率分布 $P_{M,m}^{(K)}$ に、式(33)による更新数 K の確率 $P_{(K)}$ を掛けて総和をとることにより、本質的に未知である、未来の記録更新数を予め与えずに、記録更新値の累積確率分布を得ることができる。記録更新値の最大値 r_u および

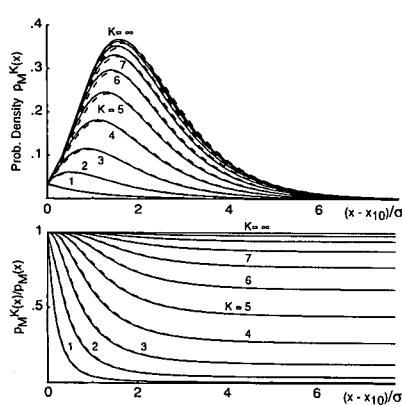


図-3 記録更新値 x_M の確率密度 p_M と部分確率密度 p_M^K の占める割合

最小値 x_m の確率密度関数 $p_{M,m}$ は、累積確率分布を確率変数で微分することにより、

$$p_M(x) = p_M^{K' \rightarrow \infty}(x) = \frac{P_{\text{GEV}}(x; \mu_l, \sigma_l, \xi) - e^{-l}}{1 - e^{-l}} \dots \dots (35)$$

および

$$p_m(x) = p_m'^{-\infty}(x) = \frac{1 - \exp(-LP_{\text{GPD}}(x - u; \sigma, \xi))}{1 - e^{-t}} \quad \dots \quad (36)$$

と得られる。ただし、式(35)および(36)における $p_{M,m}^{K'}$ は、後述の考察で必要となる部分確率密度であり、

$$p_{M,m}^{K'}(x) = e^{-\lambda} \frac{d}{dx} \sum_{K=1-K'} [l]_{\{K\}} P_{M,m}^{(K)}(x|x_{p_M}) \dots \dots \quad (37)$$

$$\mu_l = u + \sigma \xi^{e-1} (l^e - 1); \quad \sigma_l = \sigma l^e \quad \dots \dots \dots \quad (38)$$

である。将来期間が L 年であるので、 $K=1 \sim L$ の総和と考えねばならないが、発生回数の確率分布を式 (3) の Poisson 分布で近似しているため、 $K=1 \sim \infty$ の総和となることに注意する。式 (35), (36) および (38) に定数 l を含むため、式(34)を見るとおり、確率密度関数 $p_{M,m}$ は、いずれも、過去の期間(に依存する、既往最大値 x_{PM}) および将来期間 L に、間接的には関係する。また、式(35)で得られた記録更新値の最大値の確率分布 p_M は、 L 年最大統計量(合田, 1990)を既往最大値で切断した確率分布に一致している。

過去 10 年における既往最大値を 10 年確率波高で与え、将来期間を 50 年とする時、 $l=5 (=50/10)$ となる。さらに、各年最大値の分布を Gumbel 分布とした場合 ($\xi=0$) を例として、その確率密度 $p_{M,m}$ および部分確率密度 $p_{M,m}^k$ を確率密度 $p_{M,m}$ で除した割合 (p_s^k/p_s) $_{s=M,m}$ について、図-3 および 4 に示す。確率密度に関しては、記録更新最大値の分布 p_M が、Gumbel 分布 (この例での L 年最大統計量の分布) を切断したものとなること、記録更新最小値の部分確率密度 g_m^k が GPD を足しあわせたもの

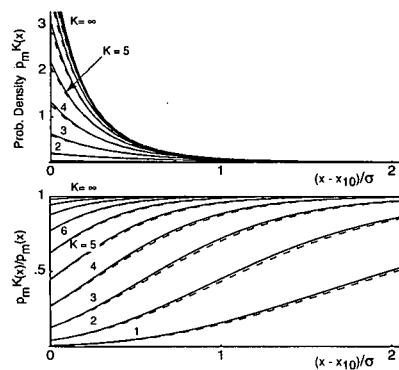


図-4 記録更新値 x_m の確率密度 p_m と部分確率密度 p_m^K の占める割合

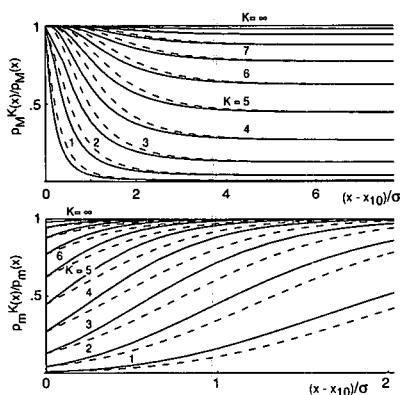


図-5 記録更新値 x_M および x_m の確率密度 $p_{M,m}^K$ の占める割合（既往最大値が十分に大きくなき場合）

になることが確認できる。図中の破線は、極値の漸近特性である式(28)を用いずに、式(35)および(36)の $P_{M,m}^{(K)}$ を、式(25)および(26)の $G_{M,m}^{(K)}$ に置き換えて得られる厳密分布である。既往最大値が10年確率波程度の大きさであっても、漸近理論による結果が妥当なものであることが確認できる。

また、図-3および4において、想定する将来期間に発生する更新数に着目する際には、各確率変数 x の断面での曲線の差を見ればよい。部分確率密度の占める割合 (p_s^K/p_s) _{$s=M,m$} については、記録更新最大値の場合には右下がりの曲線に、記録更新最小値の場合には右下がりの曲線になっている。このことは、記録更新値が大きい値をとるのは、最大値の場合には更新数が多い傾向を最小値の場合には更新数が少ない傾向を示している。

図-5は、既往最大値を2年確率波高とし、図-3およ

び4と同じ条件($l=5$)になるように、将来期間を10年としたものである。このように通常考査の対象としない程度にまで、既往最大値を小さくすれば、漸近理論による結果は、 $G_{M,m}^{(K)}$ を用いた厳密解からの乖離が大きい。

5. 結語

本研究では、再現確率波高および再現期間を拡張した概念として、与えられた将来期間における記録更新値の漸近理論を誘導し、その統計特性を検討した。

本理論の応用として、具体的な波浪データを用いた検討については、別の機会に改めて報告する。

一部の図面作成にあたっては、藤田裕嗣氏（中村石材工業（株））の協力に感謝する。

参考文献

- 北野利一・間瀬 登・喜岡 渉・矢野陽一郎 (2002): 一般化パレート分布による極値解析—拡張形状母数の推定—、海岸工学論文集、第49巻、(印刷中)。
- 合田良実 (1990): 港湾構造物の耐波設計、鹿島出版、333 p.
- 間瀬 登・北野利一・森岡智史 (2000): 波浪極値統計解析における一般化パレート分布の適用性に関する研究、海岸工学論文集、第47巻、pp. 221-225。
- Coles, S. G. and J. A. Tawn (1996): Modelling extremes of the areal rainfall process, Jour. Statist. Soc., ser. B, Vol. 58, pp. 329-348.
- Davison, A. C. and R. L. Smith (1990): Models for exceedances over high thresholds (with discussion), Jour. Roy. Statist. Soc., ser. B, Vol. 52, pp. 393-442.
- Pickands, J. (1975): Statistical inference using extreme order statistics, Ann. Statist., Vol. 3, pp. 119-131.
- von Mises, R. (1954): La distribution de la plus grande de n valeurs, Rev. Math. Union Interbalkanique, Vol. 1, pp. 141-160.