

季節別出現頻度の違いを考慮した高波の確率過程モデルの提案

北 野 利 一*・喜 岡 渉**

季節によって海域の気象条件が異なるため、高波の出現率を一定とみなすことは難しい。しかしながら、波浪極値解析に用いられる高波出現の確率過程モデルには、従来の研究では、平均出現数のみを母数とした単純ポアソン過程が用いられてきた。本研究では、出現強度の概念を導入し、出現数の疎密の効果を考慮した確率過程モデルを検討した。その結果、出現数は負の 2 項分布に、出現周期はパレート分布に従うことがわかった。また、以上の検討に対する具体例も示した。

1. はじめに

極大値資料から確率波高を算定する際、単位時間における平均出現数が重要となる。毎年最大値資料とは異なり、高波の超過確率を時間量（例えば、再現期間）と関連づけるのに、単位時間あたりの平均出現数が不可欠な量となるためである。これまでの多くの極値波浪解析法においては、単位時間あたりの平均出現数を算出し、年間出現数の確率的変動特性をポアソン分布で与えている。確かに、ポアソン分布は、出現が稀な事象の出現数に対して、小数の法則を適用して得られる理論分布である。しかしながら、ポアソン分布と年間出現数の標本との適合性の検討が行われることは少ない。むしろ、出現数の確率分布をポアソン分布とする前提に対し、否定的に疑問視した研究は皆無といえる。このことは、波高の母分布に対して標本との適合性が入念に検討されることと対照的である。

本研究では、はじめに、高波出現数の標本とポアソン分布との適合性について 2 種類の検定法を示す。ポアソン分布であるという仮説が棄却される要因として、短期的には夏・冬期における高波の出現成因により、長期的には気候変動にตอบสนองして、高波出現数の疎密の傾向が存在することが考えられる。その傾向を表現するために、出現強度という概念を導入し、出現強度の変動を考慮した確率過程モデルを構築する。結果として、モデルから得られる統計的特性を理論的に検討し、高波データへの適用も試みる。

2. 単純ポアソン過程モデル

(1) 2 項過程とポアソン過程

ベルヌイ試行を独立に繰り返して出現と非出現の時系列を得る確率過程は 2 項過程と呼ばれる。出現確率が $p \in R[0,1]$ である時、全試行数 n 回のうち k 回出現する確率 $P_n(k)$ は、次式の 2 項分布で表される。

$$P_n(k) = {}_n C_k p^k (1-p)^{n-k} \dots\dots\dots (1)$$

上式の平均は np である。そこで、平均出現回数 np が一定値 λ となるように固定し、全試行数 n を無限大に増大させると、以下のように 2 項分布はポアソン分布に近似できる。

$$P(k) = P(k; \lambda) = \frac{\lambda^k}{k!} e^{-\lambda} \dots\dots\dots (2)$$

ポアソン分布は、低頻度に出現する現象の出現数を示す基本的な分布であることから、極値統計解析にも応用される。代表的な例として、1 年間の平均出現数を λ とし、極大波高の出現数がポアソン分布に従う場合を考える。この時、極大波高資料の累積確率関数 $F_*(x)$ にポアソン分布を重み関数として掛け合わせれば、年最大高波資料の累積確率関数 $F(x)$ は、 $F_*(x)$ を用いて

$$F(x) = \sum_{k=0}^{\infty} F_*(x) P(k) = e^{-\lambda(1-F_*(x))} \dots\dots\dots (3)$$

と表される (Goda, 2000)。

2 項分布とポアソン分布の重要な違いは、ポアソン分布の定義域は正の無限大を含んでおり、ポアソン分布には全試行数という概念がないことである。また、いずれの分布についても、非負の整数を定義域としており、平均出現数が十分大きく、出現数 k がゼロとなる確率が十分小さい時、式 (1) および (2) は次式の正規分布に近似できる。

$$Q(k) = \frac{1}{\sqrt{2\pi}\sigma} \int_{k-1}^k e^{-\frac{(z+0.5-\mu)^2}{2\sigma^2}} dz \dots\dots\dots (4)$$

ここで、 μ および σ^2 は、式 (1) の 2 項分布に対しては、 $\mu = np$; $\sigma^2 = np(1-p)$ $\dots\dots\dots (5)$

であり、式 (2) のポアソン分布に対しては、

$$\mu = \sigma^2 = \lambda \dots\dots\dots (6)$$

となる。式 (5) および (6) はそれぞれ元の分布の平均 μ および分散 σ^2 である。式 (6) をみるとおり、ポアソン分布の分散は平均に一致することも顕著な特徴といえる。

(2) カイ自乗適合度検定

標本から得られる経験分布と想定する理論分布との適合性を検討する際に最も代表的な検定として、カイ自乗適合度検定が知られる。特に、出現数を対象にポアソン

* 正会員 工博 名古屋工業大学助教授 工学研究科社会工学専攻
** 正会員 Ph.D. 名古屋工業大学教授 工学研究科社会工学専攻

分布が適合するか否かを判断するという問題は、多くのテキスト (例えば, Feller, 1968) にも取り上げられる典型的な問題である。高波の出現数に対して、以下のように定式化できる。

毎年の高波の出現数をいくつかに分し、そのそれぞれに対して、 n 年間の標本から得られる頻度 N_i とポアソン分布から得られる出現確率 θ_i を用いれば、次式の統計量 χ^2 が得られる。

$$\chi^2 = \sum_{i=1-l} (N_i - n\theta_i)^2 / n\theta_i \quad (7)$$

ここで、母分布から得られる出現確率 ϕ_i が出現確率 θ_i に一致する、すなわち、

$$H_0: \phi_i = \theta_i \quad (8)$$

であるという帰無仮説 H_0 が真の時、式 (7) で表される統計量 χ^2 は近似的に自由度 $(l-1)$ のカイ自乗分布に従う。以上は、頻度 $N_{i=1-l}$ の同時確率密度関数 f_M が、

$$f_M = n! \theta_1^{N_1} \theta_2^{N_2} \cdots \theta_l^{N_l} / (N_1! N_2! \cdots N_l!) \quad (9)$$

となる多項分布に従うことに基づくものである。また、式 (8) の帰無仮説 H_0 に対する対立仮説 H_1 は

$$H_1: \phi_i \neq \theta_i (1 \leq i \leq l) \quad (10)$$

となる。

カイ自乗適合度検定を実行する際に注意すべき点が主に2つある。その1つとして、ポアソン分布のように定義域に無限大を含む場合、ある出現数以上から無限大までを1つの出現数の区分としなければならないことである。また、各区分の出現数の期待値 $n\theta_i$ はできるかぎり等しいことが望ましい。ここで、表-1に示す秋田港の1991~2000年までの10年間における各年の高波出現数を例に考える。なお、高波の抽出については、波高に対して極値の漸近理論が適用できるように、図-1に示すとおり、超過平均波高が線形的に変化する波高 (4.5 m) を下限として高波を抽出した (方法の詳細は、Kitanoら (2003) を参照)。カイ自乗適合度検定を適用するのにあたり、表-1に縦線で付与するように、出現数を4つに分割することになる。この場合、自由度は3となり、式 (7) のカイ自乗統計量は、

$$\chi^2 = 7.1 < \chi^2(0.95, 3) = 7.8 \quad (11)$$

となる。したがって、ギリギリではあるが、式 (8) の帰無仮説を棄却できず、年間出現数の母分布をポアソン分布とすることを容認することになる。

自由度を大きくとるためには、出現数の区分数 l を十分多くとらなければならない。しかしながら、高波の出現頻度の観測年数は20~30年程度であり、上述の例からわかるように、自由度を十分に大きくとれず、適切な検定であるとはいえない。これが2つめの注意事項であり、カイ自乗適合度検定は出現数の頻度という情報量 (情報量は自由度に比例) の一般に少ない統計量で行うため、

表-1 秋田港における高波の出現数 (1991~2000年)

発生数	0~3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18~
年数	0	2	2	1	0	1	0	0	1	0	1	1	0	0	1	0
区分 i	1		2		3			4								

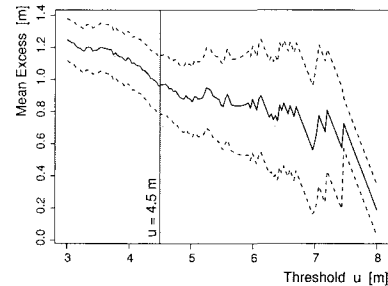


図-1 超過平均波高の閾値による変化

検定が適切であるかどうかには注意すべきである。図-2に示すポアソン分布と標本による経験分布は、上述の例に対する検定を視覚的に示したものであり、カイ自乗適合度検定は、この程度の“粗い”検討をしているに過ぎないことを注意すべきである。

(3) 平均と分散の比に基づく検定

ポアソン分布は平均が比較的大きければ正規分布に近似できること、ポアソン分布の平均と分散は等しいこと、その標本平均と標本分散は独立であること (証明は t 分布の誘導を参照) の以上の3点から、各年の観測出現数 $k_{i=1-N}$ がポアソン分布に従うなら、以下の統計量 χ^2_λ は、近似的に自由度 $(N-1)$ のカイ自乗分布に従うことが示される。

$$\chi^2_\lambda = \sum_{i=1-N} (k_i - \hat{\lambda})^2 / \hat{\lambda} \quad (12)$$

ここで、 k_i は各年の観測出現数であり、 $\hat{\lambda}$ は次式で得られる総観測年数 N 年間の標本平均である。

$$\hat{\lambda} = \sum_{i=1-N} k_i / N \quad (13)$$

また、各年の観測出現数の標本不偏分散 $\hat{\sigma}^2$ は、

$$\hat{\sigma}^2 = \sum_{i=1-N} (k_i - \hat{\lambda})^2 / (N-1) \quad (14)$$

である。したがって、標本平均と標本分散の比

$$\rho = \hat{\sigma}^2 / \hat{\lambda} = \chi^2_\lambda / (N-1) \quad (15)$$

は、自由度で規格化したカイ自乗分布に従う。カイ自乗分布の平均は自由度と一致するため、図-3に示すように、自由度を増やせば、自由度で規格化したカイ自乗分布の確率密度は1の周辺に集中する。また、ここでの自由度は、観測年数から1をひいたものであるため、この検定における情報量は2.2節の適合度検定における情報量よりも常に大きく、よりの確な検定といえる。なお、本節の帰無仮説 H_0 および対立仮説 H_1 は、母平均 λ およ

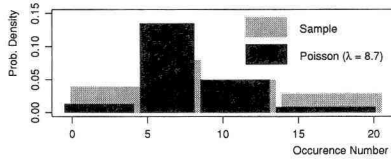


図-2 ポアソン分布と経験分布

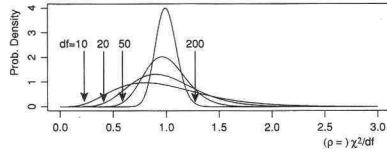


図-3 自由度で規格化したカイ自乗分布

び母分散 σ^2 を用いて、以下のように表される。

$$H_0: \sigma^2/\lambda = 1; H_1: \sigma^2/\lambda \neq 1 \dots\dots\dots (16)$$

表-1 の高波出現数に対して、式 (12) のカイ自乗統計量は、

$$\chi^2(0.95, 9) = 16.9 < \chi^2 = 23.0 \dots\dots\dots (17)$$

となる。この場合、有意水準 5% で帰無仮説は棄却される。したがって、平均と分散が等しいという性質をもつポアソン分布を母分布とすることが疑わしい。

ここでは、式 (12) の統計量 χ^2 が近似的にカイ自乗分布に従うことを利用した検討であるが、統計量 χ^2 の精密な分布については、Fisher (1950) で論じられていることを付記する。

(4) 出現周期の分布

2 項過程において、連続して k 回の非出現の後 ($k+1$) 回めにはじめて出現する確率は、

$$p_c(k) = p(1-p)^k \dots\dots\dots (18)$$

と表される幾何分布により得られる。2 項過程における時間ステップを微小時間間隔 dt とすることにより離散化されている時間軸を連続化すれば、ある出現時刻から引き続く出現時刻の時間間隔、すなわち出現周期 τ の分布は連続分布となり、その確率密度関数は、

$$p(\tau; \lambda) = \lambda e^{-\lambda\tau} \dots\dots\dots (19)$$

と表される指数分布となる。単位期間における平均出現数 λ とし、微小時間間隔 dt における出現確率

$$p = \lambda dt = \lambda\tau/k \dots\dots\dots (20)$$

を式 (18) の幾何分布に代入し、 k を無限大に増大させれば、式 (19) は得られる。なお、微小時間間隔 dt に複数の出現が生じない程度に dt が十分小さければ、各時刻でベルヌイ試行を行うことに相当する。また、ここでの極限 ($k \rightarrow \infty$) は、2(1) 節での全試行回数の極限 ($n \rightarrow \infty$) と同義であるため、単純ポアソン過程における出現周期の分布は、指数分布であるといえる。

表-1 に示す高波出現について、各出現の周期を指数確率紙に図示したものが図-4 である。指数分布に従う

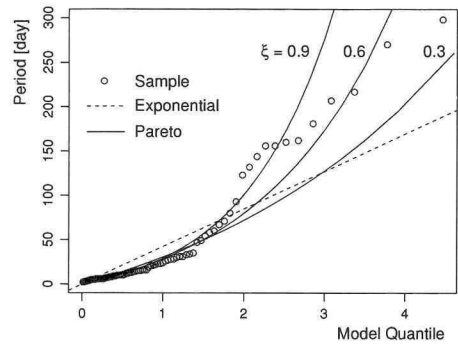


図-4 出現周期の確率点表示

標本は指数確率紙上でほぼ直線上に配置されるのに対し、図-4 の標本は、下に凸の曲線を示唆する。したがって、高波の確率過程モデルを単純ポアソン過程とすることは疑わしく、出現数がポアソン分布とすることを棄却した 2(3) 節での検定結果を裏付けるものである。図-4 の作成には表-1 に直接現れない量を用いるため、表-1 に現われる量のみを用いた 2(2) 節および 2(3) 節よりも情報量が明らかに多い。そのため、出現数の資料を用いるよりも、出現周期の資料を用いて検定する方がよいように思える。しかし、出現周期に着目する場合には、図-4 のように、視覚に頼らざるを得ないので、数量的に検定することは難しいといえる。

3. 出現強度の違いを考慮したポアソン過程モデル

(1) 出現強度

図-4 に示した高波の発生周期分布が下に凸となる傾向を説明するために、より広いクラスの確率過程モデルが必要となる。日本海沿岸域と太平洋沿岸域とで出現の物理的要因やその量的傾向は異なるものの、一般に、高波の出現数の傾向は夏期と冬期で異なることが多い。また、エルニーニョなどの地球規模での気候変動の影響により、同じ夏期 (冬期) であっても、その期待される出現数は異なるものと考えられる。ところで、式 (19) に示すポアソン分布は唯一の母数 λ で支配され、その母数の値は期待される出現数を意味する。参考として、表-1 の高波出現を時間軸上に示したものが図-5 である。以下では、ポアソン分布の母数を出現強度とよび、確率変量として扱い、出現強度 λ が時間的に変動する確率過程モデルを検討する。

図-6 は、確率過程モデルを模式的に表したものである。この図では 1 年を 2 つに区分し、半期毎に出現強度が変動し、半期の区間内では、高波の出現は単純ポアソン過程と考え、各出現強度に応じて実現値が確率的に与えられることを示している。なお、半期毎に変動する出

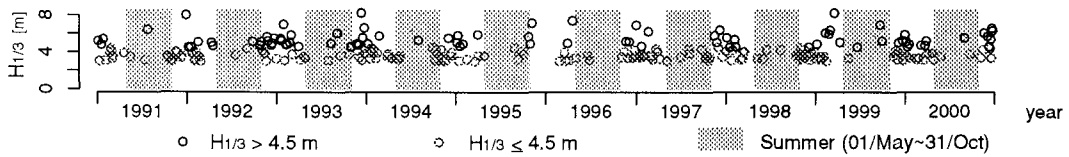


図-5 秋田港 (1991-2000 年) における高波出現の時系列

現強度については、例えば、日本海沿岸における高波の発生は、冬期に密で夏期に疎であるという時間軸に沿って現われる特性があるものの、単一の分布関数で変動させても差し支えない。なぜなら、疎密の周期毎に、副期間の数（この場合、夏と冬の2個）の出現強度の値を単一の分布関数から独立に抽出し、疎密特性にあわせて割り振ればよい（この場合、抽出した大きい値の方を冬、他方を夏とすればよい）ためである。

出現強度が変動するモデルについては、Greenwood・Yule (1920) が、作業者の事故数を対象にすでに検討している。しかし、それは作業者の個体変動を考慮したモデルである。本研究は、それを確率過程における時間的変動に適用したものであり、確率過程に固有な出現周期についても検討する点が大きく異なる。

(2) 出現数の分布

モデルが複雑化することを避けるために、出現強度 Λ の変動性を表す分布関数を以下のガンマ分布で与える。

$$g(\lambda) = g(\lambda; \alpha, \beta) = \beta^\alpha \lambda^{\alpha-1} e^{-\beta\lambda} / \Gamma(\alpha); \quad \alpha, \beta > 0 \quad (21)$$

ガンマ分布の平均 μ_g と分散 σ_g^2 は、

$$\mu_g = \alpha/\beta; \quad \sigma_g^2 = \mu_g/\beta \quad (22)$$

である。平均を一定値にとり、分散が異なるガンマ分布の例を図-7に示す。出現強度の分布形として、顕著な不適切さは見受けられない。 β が無限大の場合、確率1で出現強度は一定値 μ_g をとる。

着目する期間（例えば、1年間）に疎密の周期に応じた副期間の数を M （1年を夏と冬の2期に分割する場合は、 $M=2$ ）とし、各副期間における出現強度の変動特性を式 (21) のガンマ分布 g で与えたとすれば、副期間における出現数の確率分布 $P_{\text{sub}}(k)$ は、以下のように負の2項分布として得られる。

$$P_{\text{sub}}(k) = \int_0^\infty g(\lambda) P(k; \lambda/M) d\lambda \\ = \frac{\Gamma(k+\alpha)}{k! \Gamma(\alpha)} \left(\frac{\beta M}{1+\beta M} \right)^\alpha \left(\frac{1}{1+\beta M} \right)^k \quad (23)$$

負の2項分布の再生性により α を αM に置き換えると、着目する期間における出現数の確率分布 $P(k)$ は、

$$P(k) = \frac{\Gamma(k+\alpha M)}{k! \Gamma(\alpha M)} \left(1 + \frac{1}{\beta M} \right)^{-\alpha M} \left(\frac{1}{1+\beta M} \right)^k \quad (24)$$

と得られる。その平均 μ_{NB} と分散 μ_g は

$$\mu_{NB} = \alpha/\beta; \quad \sigma_{NB}^2 = \{1 + (\beta M)^{-1}\} \mu_{NB} \geq \mu_{NB} \quad (25)$$

である。出現数の平均 μ_{NB} は出現強度の平均 μ_g に一致し、式 (2) との比較のためと記す。出現数の分散 σ_{NB}^2 はその平均 $\mu_{NB} (= \lambda)$ より常に大きい。出現強度の分散 σ_g^2 が小さくなる ($\beta \rightarrow \infty$) につれ、出現数の分散 σ_{NB}^2 はその平均 λ に近づく。それと同時に、

$$\left(1 + \frac{1}{\beta M} \right)^{-\alpha M} = \left\{ \left(1 + \frac{1}{\beta M} \right)^{\beta M} \right\}^{-\alpha} \simeq e^{-\lambda} \\ \frac{\Gamma(k+\alpha M)}{\Gamma(\alpha M) (1+\beta M)^k} = \omega_k \omega_{k-1} \dots \omega_1 \simeq \lambda^k \quad (26)$$

であるので、出現数の分布 $P(k)$ は、式 (2) のポアソン分布 $P(k)$ に近似される。なお、式 (26) における $\omega_{j=1-k}$ は、

$$\omega_j = \frac{\lambda + j/(\beta M)}{1 + 1/(\beta M)} \quad (27)$$

である。このことは、本章で検討する確率過程が単純ポアソン過程を含むことを示している。

副期間数 M を増大させる場合も、式 (26) と同様のことが成立し、出現数の分布 $P(k)$ は、ポアソン分布に近似される。これは、着目する期間に対し、出現数の疎密の周期が十分に短ければ、疎密による影響が小さいことを示している。したがって、極値波浪解析で検討するよう

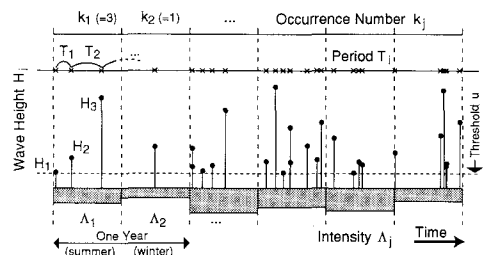


図-6 出現強度の変化するポアソン過程の模式図

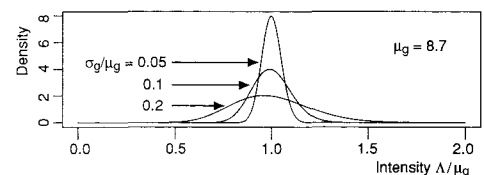


図-7 ガンマ分布

な数十年から数百年のオーダーの期間の出現数に着目する場合は、1 年末満の出現強度の変化ではなく、むしろ、数年以上のオーダーの変化の有無に注意しなければならないことを示唆している。

(3) 出現周期の分布

出現強度は単位時間当たりの出現数の期待値である。出現周期を考える場合には、その表現を出現周期あたりの出現数に改める必要がある。なぜなら、出現周期の分布は 1 つの出現に対するものであり、単位時間に対するものではないからである。そこで、周期と周波数の関係より、出現周期に逆比例させた出現強度を考えれば、1 つの出現に対する出現強度（すなわち、出現周期あたりの出現強度）となる。したがって、ここで扱う出現強度の分布 $g'(\lambda)$ は、

$$g'(\lambda) \propto g(\lambda) \cdot \lambda \\ = g(\lambda; \alpha+1, \beta) \dots\dots\dots (28)$$

となる。式 (19) の指数分布と式 (28) のガンマ分布を用いれば、出現強度を考慮した出現周期の分布は、

$$p'(\tau) = \int_0^\infty g'(\lambda) p(\tau; \lambda) d\lambda = \frac{1}{\sigma} \left(1 + \xi \frac{1}{\sigma}\right)^{-(1+1/\xi)} \dots\dots\dots (29)$$

と得られる。ここで、

$$\sigma = \beta/(\alpha+1); \quad \xi = 1/(\alpha+1) \dots\dots\dots (30)$$

となり、 ξ の値は常に正であることに注意する。また、式 (29) はパレート分布として知られ、指数確率紙上で下に凸となる曲線を描く ($\xi=0$ の場合は、指数分布)。図 4 における点線は $\xi=0.3, 0.6$ および 0.9 の場合のパレート分布を描いたものである。

出現周期の平均は、以下のように得られ、平均出現数の逆数に一致する。

$$T = \sigma/(1-\xi) = \beta/\alpha = 1/\lambda \quad (\xi < 1) \dots\dots\dots (31)$$

高波の出現を単純ポアソン過程とする場合、指数分布の母数 $\hat{\lambda}$ の最尤推定は、単純に、

$$\hat{\lambda}^{-1} = \sum_{i=1}^K \tau_i / K \dots\dots\dots (32)$$

となる ($\tau_{i=1-K}$ は出現周期の標本値) ので、平均周期の推定値 $\hat{T}_0$ は $\hat{\lambda}$ の逆数として得られる。他方、出現強度を考慮する場合、パレート分布の母数 α および ξ の最尤推定を数値的に行い (例えば、Smith (1985) や Kitano ら (2003) 等を参照)、式 (31) に代入することにより平均周期の推定値 $\hat{T}'$ が得られる。表 1 の高波標本に対し、それぞれの推定値は、

$$\left. \begin{aligned} \hat{T}_0 &= 42.4 \pm 9.0 \text{ [day]} \\ \hat{T}' &= 49.0 \pm 31.1 \text{ [day]} \end{aligned} \right\} \dots\dots\dots (33)$$

と得られる。指数分布を母分布とした推定値 $\hat{T}_0$ に比して、パレート分布を母分布とした推定値 $\hat{T}'$ は、そのとりうる範囲が広い。このことを逆に見て、パレート分布で推定するより、指数分布で推定する方が誤差は少ないと考えることは誤りである。指数分布を母分布とすれば、平均出現周期を 25 日あるいは 60 日とすることは棄却される。しかし、この判断は、パレート分布を母分布とする立場から見れば、第 2 種の誤りであるといえる。

4. 結 語

従来の波浪極値解析における高波出現のモデルは、単純ポアソン過程モデルである。しかしながら、出現数の疎密の周期的な傾向が見られる場合には、出現強度の変化を考慮しなければならない。そこで、単純ポアソン過程を帰無仮説とする検定を提示し、棄却される場合に検討すべき確率過程モデルを提案した。

なお、検定には検出力の議論が不可欠である。本論文で検討した検定に対する検出力の議論は割愛した。また、出現強度の変化を考慮した場合の検定およびその検出力についても、紙面の都合で割愛した。これらについては別の機会に報告したい。

謝辞: 本研究は科学研究費 若手 (B) 15760377 による成果の一部であることを付記する。論文推敲に際し、京都大学防災研究所 間瀬 肇 助教授にアドバイスをいただいた。また、高波資料の整理や図面の作成にあたり、名古屋工業大学 大学院 河井 敦君に協力いただいた。ここに謝意を表する。

参 考 文 献

- Feller, W. (1968): An introduction to probability theory and its applications, Wiley, 509 p.
- Fisher, R. A. (1950): The significance of deviations from expectations in a Poisson series, Biometrics, Vol. 6, pp. 17-24.
- Goda, Y. (2000): Random seas and design of maritime structures, World Scientific, 443 p.
- Greenwood, M. and G. U. Yule (1920): An enquiry into the nature of frequency distributions of multiple happenings, with particular reference to the occurrence of multiple attacks of disease or repeated accidents, J. Roy. Statist. Soc. Series A 89, pp. 255-279.
- Kitano, T., W. Kioka and H. Mase (2004): A new analysis for extreme wave statistics by Poisson-GPD Model, APAC 2003, paper no. 034.
- Smith, R. L. (1987): Estimating tails of probability distributions, Ann. Statist., Vol. 15, No. 8, pp. 1174-1207.