

II-4 上下限censoringを考慮した極値統計解析モデルの精度の検討

愛媛大学工学部 正員 山口正隆 愛媛大学工学部 正員 畑田佳男
愛媛大学工学部 大福 学 愛媛大学大学院 学生員 ○野中浩一

1. はじめに：母数推定に最小2乗法を用いる極値統計解析モデルは、欠落を伴うcensored dataに対して、最尤法に基づく方法のように一般的な形では適用できないが、順位の明らかな資料を欠く場合には、相当する順位を空位とする簡単な操作によって適用可能になる。しかし、その精度は明らかでない。そこで、本研究では広範な入力条件のもとにモンテカルロシミュレーションで発生したcensored dataの極値統計解析に基づいて、最小2乗法を用いた拡張型モデルの精度を種々の観点から明らかにする。また、欠落を伴う風速資料などの観測資料に拡張型モデルを適用して再現確率統計量とその分散（標準偏差）を推定し、最尤法に基づく結果との比較・検討を行う。

2. モンテカルロシミュレーション：censoringは、censoring値が固定され標本数がランダムである第1種censoringと、逆に標本数が固定され基準値がランダムである第2種censoringに分類され、それぞれが上限（基準値以上の値）あるいは下限（基準値以下の値）あるいは上下限（両者の値）をもつ資料がcensoringを受ける3種類に細分され、合計6種類に分けられる。本研究ではGumbel分布、Weibull分布、FT-II型分布を対象として、6種類のcensored dataをモンテカルロシミュレーションにより発生する。そして、最小2乗法による母数推定と相関係数 ρ 、再現確率統計量 H_n 、jackknife法による分散 σ^2 の計算を各標本ごとに行い、諸量の平均値、再現確率統計量のbiasおよび再現確率統計量の分散 Var を求める。シミュレーションは年最大値資料および極大値資料の場合に、censoring条件、資料年数 K 、候補分布の種類や形状母数および事象の年平均発生数を変化させて行い、これらの影響を検討する。

3. シミュレーション結果：図-1は形状母数 $k=1.4$ のWeibull分布に基づく第2種censoring（上限・下限censoring比 $\mu_v=0.1$, $\mu_l=0.1$ ）を受けた年最大値資料に対する相関係数 ρ 、無次元bias $\Delta \tilde{H}$ 、無次元標準偏差 $\tilde{\text{Var}}^{1/2}$ 、無次元標準偏差推定値 $\tilde{\sigma}_i$ の%値と資料年数 K との関係をcensoringの種類あるいは形状母数 k をパラメータとして示す。標本ごとの変動性が高い上位値資料の削除により上限・上下限censoringの場合に相関係数が大きい。また、 $\text{Var}^{1/2}$ はcensoringの有無にあまり依存しない。これは第1種censoringの場合も同様である。また、biasはcensoringの有無および形状母数によらずほとんど生じず、一方 $\text{Var}^{1/2}$ の推定誤差 $(\tilde{\sigma}_i - 1)$ は最大20%に達するが、資料

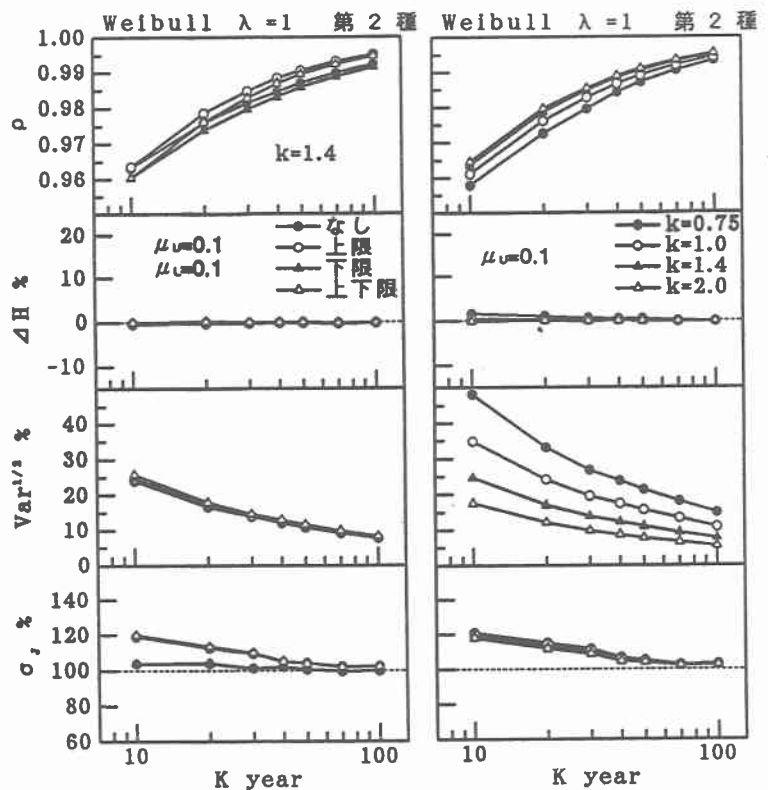


図-1

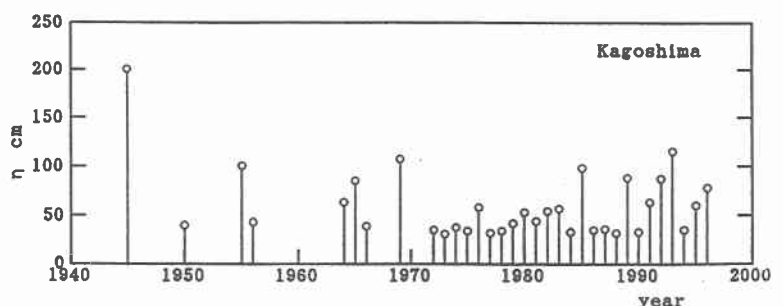


図-2

年数の増加とともに減少する。

4. 観測資料への適用：鹿児島での52年間の年最大高潮偏差資料を第1種下限censored data, 大王崎での34年間の年最大風速資料を第1位資料を欠く第2種上限censored data, 油津での16年間の年最大波高資料を最下位資料を欠く第2種censored dataとみなして、候補分布をGumbel分布, Weibull分布あるいはFT-II型分布, 最適分布選択基準を相関係数基準とする極値統計解析を行い、再現確率統計量とその標準偏差を推定した。ここでは紙数の都合上、鹿児島および大王崎の結果を示す。鹿児島の年最大高潮偏差資料（図-2）は1945～1996年のうち1972年以降毎年得られているだけである。また、100cm以上の高潮偏差は52年間で4回生起しているが、そのうち3回は1972年以前であり、しかも第1位高潮偏差200cmは1945年の台風4516号（枕崎台風）により発生している。解析は下限censoring値を $\eta_c=50, 60, 80\text{cm}$ の3種類に設定し、第1種下限censoring問題として行った。表-1は鹿児島における結果の一覧である。

第1位高潮偏差が突出した値をとるので、100年確率高潮偏差 η_{100} の標準偏差 $\sigma_{\eta_{100}}$ がかなり大きく、従って推定値の信頼度は高くない。一方、最尤法に基づく結果によれば、最適分布は $k=4.34$ のFT-II型分布であり、最小2乗法に基づく最適分布より分布幅が狭い。また、確率高潮偏差は最小2乗法に基づく結果とほぼ一致し、しかも標準偏差は約1/3となって小さい。したがって、鹿児島における確率高潮偏差に関しては、最尤法に基づく推定値の信頼性がより高いと考えられる。一方、志摩半島先端に位置する大王崎では1953～1986年に得られた年最大風速資料（図-3）を用いる。この資料では、第1位風速は周辺の風観測資料の調査から明らかなように、観測期間中の最大風速は台風5915号によって生起していると推定されるが、欠測となっているので、第1位資料を欠いた第2種上限censoring問題として解析を行う。表-2は大王崎における結果であり、第1位資料を無視した場合の結果および最尤法に基づく結果も与えられる。この場合の適合度は非常に高い。censored dataとしての解析によれば、100年確率風速 U_{100} は最尤法に基づく値と一致するが、標準偏差 $\sigma_{U_{100}}$ はより大きい。またuncensored dataとしての解析では、標準偏差のみならず確率風速も小さい。しかし、第2位風速が36.6m/sであることを考慮すると、確率風速の推定値は過小である。

5. まとめ：①本モデルはuncensored dataや下限censored dataに対して、候補分布の種類やその形状母数および下限censoringの種類や割合によらず、biasをほとんど生じず、標準偏差を適正に評価する。②本モデルは候補分布の種類やその形状母数によらず、第1種上限・上下限censored dataに対して負のbiasを生じるが、標準偏差を適正に評価し、逆に第2種上限・上下限censored dataに対してbiasをほとんど生じないが、上限censoring比が小さいほど標準偏差を過大評価する。ただし、biasおよび標準偏差の推定誤差は標本数の増大とともに減少する。③標本数を大きくとれる極大値資料に対するbiasおよび標準偏差の推定誤差が年最大値資料に対するものと比べて小さい。④欠落を伴う種々の観測資料への適用結果によれば、本モデルは最尤法に基づくモデルと同程度の再現確率統計量の推定値を与えるが、より大きい標準偏差を生じることから、推定値の有効性の点で最尤法に基づくモデルより劣る。

表-1

data(AM)	N	K	k	ρ	$\eta_{100}\text{ cm}$	$\sigma_{\eta_{100}\text{ cm}}$
$\eta_c \geq 50\text{cm}$	19	52	0.60	0.9747	183	52
$\eta_c \geq 60\text{cm}$	14	52	0.50	0.9733	181	55
$\eta_c \geq 80\text{cm}$	10	52	0.50	0.9743	178	56
MLM		52	F4.34		182	19

K : years k : Weibull dist., F4.34:FT-II ($k=4.34$)

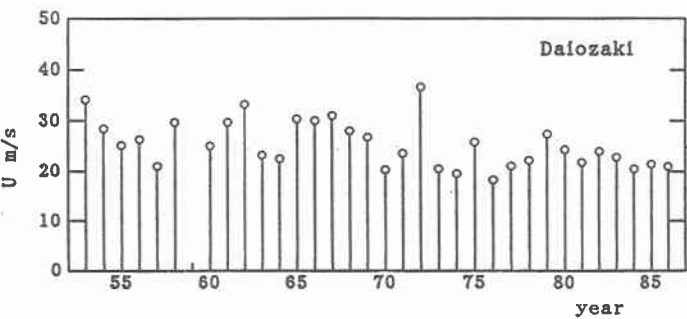


図-3

表-2

data(AM)	N	K	k	ρ	U_{100}	$\sigma_{U_{100}}$
censored	33	34	1.40	0.9966	42.7	3.6
uncensored	33	34	1.70	0.9954	38.9	2.1
MLM	33	34	1.51		42.7	2.9

K : years k : Weibull dist. $U_{100}, \sigma_{U_{100}}$: m/s