

年最大日雨量データの「外れ値」の統計的評価 Statistical Evaluation of Outliers in Annual Maximum Daily Rainfall Records

近森秀高*・工藤亮治*

*岡山大学大学院環境生命科学研究科 (〒700-8530 岡山市北区津島中 3-1-1 岡山大学環境理工学部)

CHIKAMORI Hidetaka*, KUDO Ryoji*

*Graduate School of Environmental and Life Science, Okayama University
(3-1-3 Tsushima-naka, Kita-ku, Okayama-shi Okayama, 700-8530)

(Received 4 December 2020)

Abstract

Extremely large values of observed rainfall depth in record rainfall are often plotted as “outliers” on an extreme-value probability paper. Outliers are often treated as measurement errors and excluded from data for analysis although their existence should have important statistical implications if they are correct data. In this study, we investigated influence of the presence of “outliers” by fitting of generalized extreme value distributions to bootstrap samples of annual maximum daily rainfall data observed at each of 10 rainfall stations in Japan. The results showed that the bootstrap distributions can be classified by existence of the maximum of annual daily rainfall maxima in each sample, and that the bootstrap sample is combination of these samples. This suggests that the maximum value, which is an “outlier”, may follow a different distribution from other daily rainfall data.

Key words: extreme events, statistical evaluation, outlier, heavy rainfall, flood

要 旨

記録的な降雨で観測される非常に大きな降雨強度は、しばしば極値確率紙に「外れ値」としてプロットされる。外れ値は「異常値」として分析用のデータから除外されることがあるが、極端に大きなデータは、極端現象の定量的評価には貴重な情報となりうる。本研究では、全国 10 カ所の雨量観測点における「外れ値」を持つ年最大日雨量データを対象とし、この bootstrap 標本から得られる一般化極値分布のパラメータと確率日雨量の分布を調べ、「外れ値」の存在がこれらの分布に及ぼす影響を調べた。その結果、各 bootstrap 標本から得られる 100 年確率日雨量の分布は、年最大日雨量の最大値の有無によって異なる 2 個の分布の混合分布となった。これは、「外れ値」である最大値が他のデータと異なる分布に従っている可能性を示唆している。

キーワード: 極端現象, 統計的評価, 外れ値, 豪雨, 洪水

1. はじめに

近年、台風や前線などの影響で夏季に発生する豪雨災害では、従来の観測値を大きく上回る規模の雨量が観測される事例が少なくない。例えば、令和 2 年 7 月豪雨では、西日本から東日本、東北地方の広い範囲で大雨が発生し、7 月 4 日から 7 日にかけて九州で記録的な大雨により球磨川など大河川での氾濫が相次いだ。また、令和元年東日本台風（台風第 19 号）による大雨、暴風等では、台風の上陸により、関東地方や甲信地方、東北地方などで記録的な大雨となり、甚大な被害がもたらされた。平成 30 年 7 月豪雨では、西日本を中心に北海道や中部地方を含む全国的に広い範囲で集中豪雨が発生し、6 月 28 日から 7 月 8 日（9 時）までの総雨量が四国地方で 1,800mm、東海地方で 1,200mm、九州北部地方で 900mm、近畿地方で 600mm、中国地方で 500mm を超えるところがあり、7 月の月雨量平年値の 2～4

倍となる地点もあった。このような豪雨の原因として、地球温暖化に伴う気候変動の影響が長年にわたって指摘されていることから、今後、想定を上回る規模の豪雨の増加が予想され、ダムや用排水路などの灌漑排水に関わる施設の維持管理に問題が生じることが懸念される。

想定を上回る規模の観測値は、過去のデータに適応した確率分布からは大きくかけ離れた「外れ値」であることが多い。年最大値などの極値の標本では、これらの発生頻度が低い「外れ値」は、極端現象の定量的評価のために貴重な情報となるはずであるが、「外れ値」を含むデータに従来の統計解析法の手順に従って確率分布を適応しても、ほとんどの場合「外れ値」には適合しないために、異常値と同様に扱わざるを得ない。

本研究では、全国 10 カ所の雨量観測点におけるこのような「外れ値」を持つと見られる年最大日雨量データに対して、L 積率法および最尤法のそれぞれを用いて一般化極値分布を適応し、パラメータの推定法による「外れ値」への適合性の相違を調べた。また、各地点の雨量データから、bootstrap 法によって得られるリサンプリング標本 (bootstrap 標本) に一般化極値分布を適応してそのパラメータと確率日雨量の分布を調べ、「外れ値」の存在が一般化極値分布のパラメータの分布と確率水文量の推定値の分布に及ぼす影響を調べた。

なお、「外れ値」は他の測定値からかけ離れた測定値と定義されており、必ずしも異常値とは限らない。「外れ値」の判定についてはいくつかの定量的方法が提案されているが、ここでは、一般化極値分布を適応した際に、極値確率紙上で、適応した確率分布を表す曲線から著しく離れた値を「外れ値」と見なすことにした。

2. 解析対象資料

近年、記録的な豪雨に見舞われた雨量観測点を 10 カ所選定し、1987 年から 2017 年までの 40 年間に観測された年最大日雨量のデータを用いて解析を行った (気象庁 1)。解析対象とした観測点の位置を図 1 に示す。各解析対象観測点および周辺地域における豪雨の発生概況は以下のとおりである (気象庁 2, 気象庁 3)。

(1) 川井

平成 28 年 (2016 年) 8 月 16 日～8 月 31 日に台風第 7 号、第 11 号、第 9 号、第 10 号および前線による大雨・暴風に見舞われた。これらの台風等の影響で、東日本から北日本を中心に大雨や暴風となり、特に北海道と岩手県では、記録的な大雨となった。

(2) 只見

平成 16 年 (2004 年) 年 7 月 27 日から 30 日にかけて、新潟県と福島県会津を中心に大雨となった。特に、28 日から 30 日にかけては、大気の状態が不安定となり、「平成 16 年 7 月新潟・福島豪雨」(平成 16 年 7 月 12 日～7 月 14 日) を上回る記録的な大雨となった。

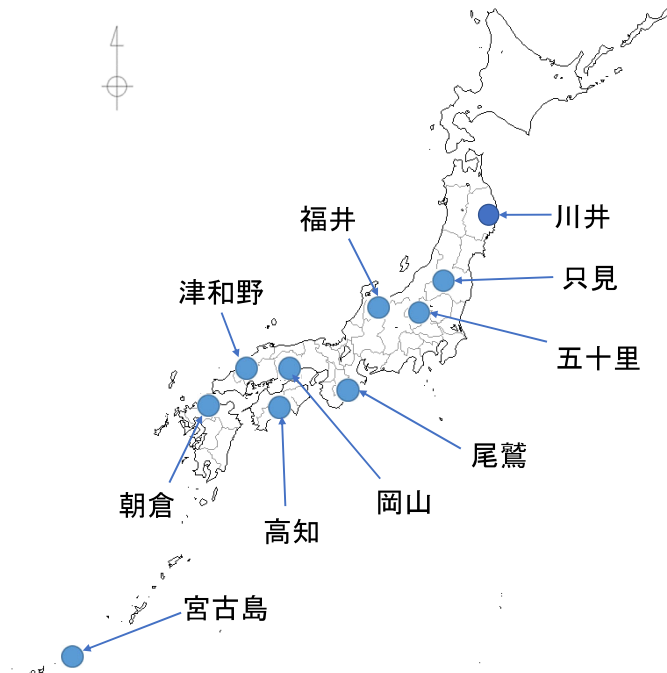


図1 解析対象地点

(3) 五十里 (いかり)

平成 27 年 (2015 年) 年 9 月 7 日～9 月 11 日にかけて、台風第 18 号等による大雨に見舞われた。台風第 18 号および台風から変わった低気圧に向かって南から湿った空気が流れ込んだ影響で、西日本から北日本にかけての広い範囲で大雨となり、特に関東地方と東北地方では記録的な大雨となった。

(4) 福井

平成 16 年 (2004 年) 年 7 月 17 日夜から 18 日にかけて、活発な梅雨前線が北陸地方をゆっくりと南下したのに伴い、福井県や岐阜県で大雨となった。福井市では 18 日の日雨量が 197.5mm を記録した。

(5) 岡山

平成 23 年 (2011 年)、紀伊半島に記録的豪雨をもたらした台風 23 号の接近により、岡山では 9 月 3 日に記録的豪雨に見舞われ 187mm/d の日雨量を記録した。これにより、近年 100 年間の観測最大日雨水量であった 159.0 mm/d (1990 年 9 月 18 日) を大幅に更新した。

(6) 津和野

平成 25 年 (2013 年) 8 月 23 日～25 日にかけて、島根県西部では、24 日の明け方と 25 日の明け方に猛烈な雨が降り、24 時間雨量が多いところで 400mm を超え、8 月の月雨量平年値の約 3 倍の雨量を記録した。

(7) 尾鷲

全国的にも有数の豪雨地帯に位置し、前述の、平成 23 年 9 月の紀伊半島豪雨など、記録的豪雨や水害の記録が数多く見られる地域である。

(6) 高知

平成 10 年 (1998 年) 9 月 23 日から 24 日にかけて、前線が四国の南海上から瀬戸内付近に北上し、25 日朝まで停滞した。この前線に向かって高気圧の縁を回る暖湿気流が流れ込んだため、高知市およびその周辺で 24 日を中心に長時間にわたって激しい雨が降り続いた。

(7) 朝倉

平成 29 年 (2017 年) 7 月 5 日から 6 日にかけて西日本では記録的な大雨となった。とくに、九州北部の朝倉地点では 7 月 5 日の日雨量が 500mm を超える未曾有の豪雨となり、九州北部豪雨と命名された。

(8) 宮古島

平成 29 年（2017 年）9 月 13 日から 9 月 18 日にかけて、台風第 18 号および前線による大雨・暴風等に見舞われた。台風や活発な前線の影響で、南西諸島や西日本、北海道を中心に大雨や暴風となった。また、台風の影響で、南西諸島や西日本を中心に猛烈なしけとなり、大潮の時期と重なったため西日本を中心に高潮となった。

3. 一般化極値分布の適応

3.1 一般化極値分布

一般化極値分布の確率分布関数 $F(x)$ および確率密度関数 $f(x)$ は以下のものである。

$$F(x) = \begin{cases} \exp\left\{-\left[1 - \frac{k(x-c)}{a}\right]^{1/k}\right\} & (k \neq 0) \\ \exp\left\{-\exp\left(-\frac{x-c}{a}\right)\right\} & (k = 0) \end{cases} \tag{1}$$

$$f(x) = \begin{cases} \frac{1}{a} \left\{1 - \frac{k(x-c)}{a}\right\}^{\frac{1-k}{k}} \exp\left\{-\left[1 - \frac{k(x-c)}{a}\right]^{1/k}\right\} & (k \neq 0) \\ \frac{1}{a} \exp\left[-\frac{x-c}{a} - \exp\left(-\frac{x-c}{a}\right)\right] & (k = 0) \end{cases} \tag{2}$$

ここに、 a, c および k はパラメータである。

3.2 L 積率法によるパラメータの推定

L 積率法（Hosking, 1997）による一般化極値分布のパラメータ推定の手順は以下のものである。

① 確率重みつき積率の推定

標本データの r 次の確率重みつき積率の推定値 b_r^* は、以下のようにして求められる。

$$b_r^* = \frac{1}{N} \sum_{j=1}^N x_j \{F(x_j)\}^r \tag{3}$$

ここに、 x_j は対象データであり、 j は昇順に並べたデータに付した番号、 N はデータの総数、 $F(x_j)$ はデータ x_j の非超過確率である。各データの非超過確率 $F(x_j)$ の推定にはプロットイング・ポジション公式を用いる。一般化極値分布の適応では以下の APL 公式がよく用いられる。

$$F(x_j) = \frac{j-0.35}{N} \tag{4}$$

② L 積率の計算

L 積率は、確率重みつき積率の推定値 b_r^* を下記の(5)~(7)の式の b_r に代入することにより、求められる。

$$\lambda_1 = b_0 \tag{5}$$

$$\lambda_2 = 2b_1 - b_0 \tag{6}$$

$$\lambda_3 = 6b_2 - 6b_1 + b_0 \tag{7}$$

③ パラメータの推定

まず、式(5)~(7)でそれぞれ表される 1~3 次の L 積率は以下のようになる。

$$\lambda_1 = c + \frac{a}{k} [1 - \Gamma(1+k)] \tag{8}$$

$$\lambda_2 = \frac{a}{k} (1 - 2^{-k}) \Gamma(1+k) \tag{9}$$

$$\frac{2\lambda_2}{\lambda_3 + 3\lambda_2} = \frac{1 - 2^{-k}}{1 - 3^{-k}} \tag{10}$$

ここで Γ はガンマ関数で以下のように定義される。

$$\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt \quad (11)$$

パラメータを求める手順としてはまず、式(10)より k を求め、これを用いて式(9)より a を求める。式(10)からは、 k を陽的に表すことはできないため、以下の近似式により求める。

$$k = 7.8590d + 2.9554d^2 \quad (12)$$

ここで、

$$d = \frac{2\lambda_2}{\lambda_3 + 3\lambda_2} - \frac{\ln 2}{\ln 3} \quad (13)$$

である。

式(9)および(8)より、 a, c はそれぞれ以下のように求める。

$$a = \frac{k\lambda_2}{(1-2^{-k})\Gamma(1+k)} \quad (13)$$

$$c = \lambda_1 - \frac{a}{k} [1 - \Gamma(1+k)] \quad (14)$$

3.3 最尤法によるパラメータの推定

確率密度関数が $f(x_i; \theta_1, \theta_2, \dots, \theta_l)$ ($\theta_1, \theta_2, \dots, \theta_l$ はパラメータ) で表される母集団から抽出した標本を x_1, x_2, \dots, x_n とするとき、パラメータ $\theta_1, \theta_2, \dots, \theta_l$ の関数となる尤度は以下のように表される (例えば、神田・藤田, 1982)。

$$L = \prod_{i=1}^n f(x_i; \theta_1, \theta_2, \dots, \theta_l) \quad (15)$$

一般化極値分布の確率密度関数を用いると、尤度関数は次式のように表される。

$$L = \prod_{i=1}^n \frac{1}{a} \left\{ 1 - \frac{k(x-c)}{a} \right\}^{\frac{1-k}{k}} \exp \left\{ - \left[1 - \frac{k(x-c)}{a} \right]^{\frac{1}{k}} \right\} \quad (16)$$

対数尤度関数 ℓ は、次式のようになる。

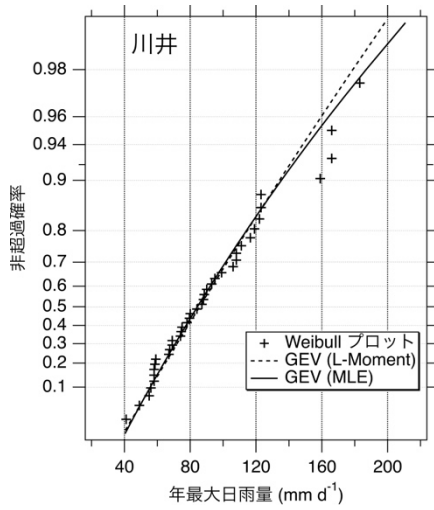
$$\begin{aligned} \ell = \ln L &= \sum_{i=1}^n \left(\ln \frac{1}{a} \left\{ 1 - \frac{k(x-c)}{a} \right\}^{\frac{1-k}{k}} \exp \left\{ - \left[1 - \frac{k(x-c)}{a} \right]^{\frac{1}{k}} \right\} \right) \\ &= \sum_{i=1}^n \ln \frac{1}{a} + \sum_{i=1}^n \frac{1-k}{k} \ln \left[1 - \frac{k(x-c)}{a} \right] - \sum_{i=1}^n \left[1 - \frac{k(x-c)}{a} \right]^{\frac{1}{k}} \end{aligned} \quad (17)$$

式(17)を最大とする a, c および k が最尤値となる。 a, c, k の最尤値は差分進化法 (Differential Evolution) (Storn and Price, 1997) を用いて推定した。

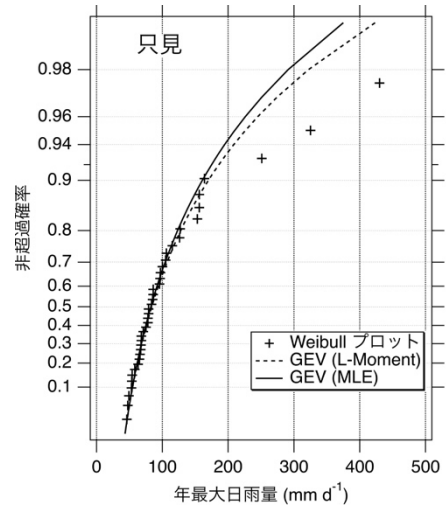
3.4 一般化極値分布の適応と「外れ値」

解析対象とした 10 カ所の雨量観測点における年最大日雨量データに一般化極値分布を適応した結果をそれぞれ図 2 (1)~(10) に示す。これらの図では、各地点における解析対象期間中の観測値の非超過確率を Weibull プロットで推定した結果と、一般化極値分布を L 積率法および最尤法により適応した結果を極値確率紙上に併示している。

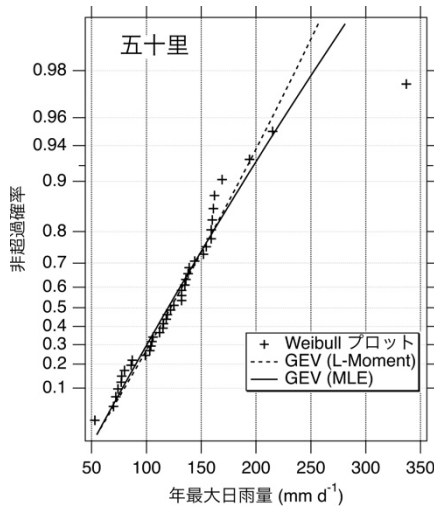
川井を除く 9 観測点で、適応した一般化極値分布を示す実線および破線に対し、年最大日雨量の上位 1~数個が右側に大幅に外れてプロットされ、いわゆる「外れ値」となっていることがわかる。これは、これらの年最大日雨量の非超過確率が、一般化極値分布では Weibull プロットよりも大きく評価されている、すなわち、確率年が長く評価されていることを示している。川井地点では、最大値および第 2 位の値は、L 積率法および最尤法によって適応した分布の曲線の近くに位置しているが、3, 4 位の値が若干右側へ外れている。



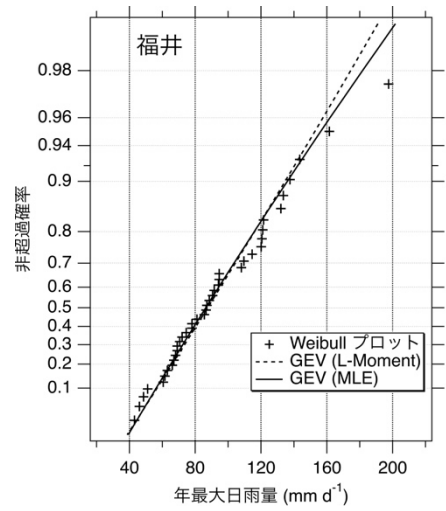
(1) 川井



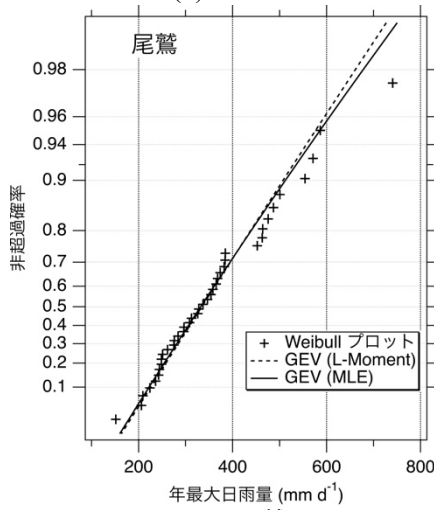
(2) 只見



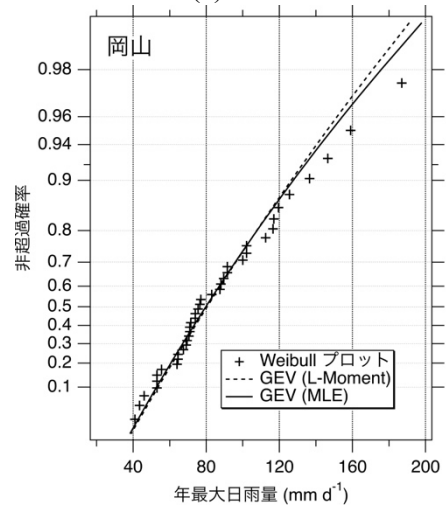
(3) 五十里



(4) 福井



(5) 尾鷲



(6) 岡山

図2 一般化極値分布の適応 (1 / 2)

パラメータ同定の手法 (L積率法および最尤法) による「外れ値」への適応度には、地点によって若干の違いが見られた。高い非超過確率の値に対する日雨量の値を比較すると、各地点における年最大日雨量の上位3個程度「外れ値」に対する非超過確率の推定値を見ると、最尤法による推定値がLモーメ

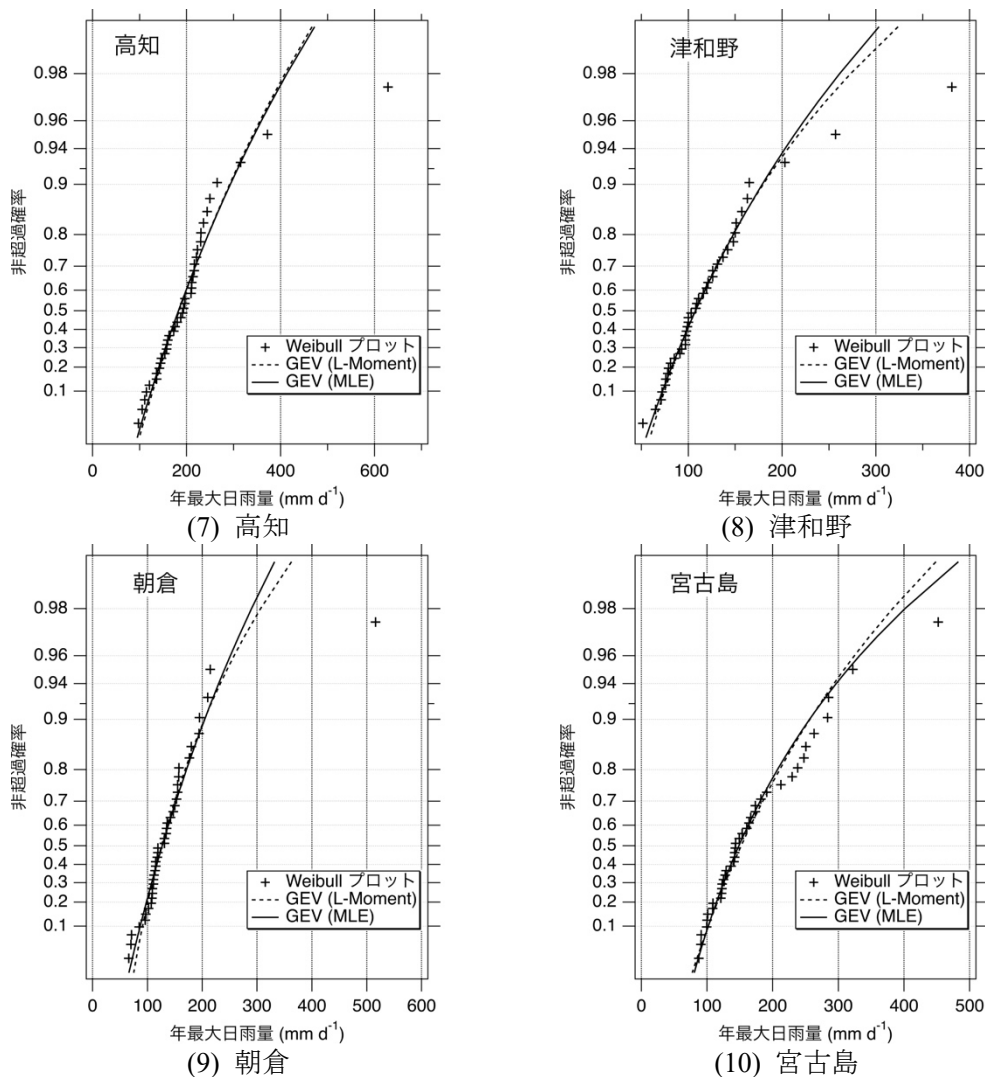


図2 一般化極値分布の適応 (2 / 2)

ント法による値を下回る地点が 10 箇所中 6 箇所であり、「外れ値」の発生頻度に明確な大小関係は見られなかった。

4. Bootstrap 法を用いた確率日雨量およびパラメータの分布の推定

4.1 Bootstrap 法の適用

ここでは、リサンプリング手法の一つである bootstrap 法を用いて年最大日雨量データの統計解析を行い、各 bootstrap 標本から得られる 100 年確率日雨量および一般化極値分布のパラメータの分布を調べた。具体的な手順は以下のとおりである。

- 1) 年最大日雨量の標本のサイズが N 個のとき、この標本から重複を許して N 個無作為に抽出する（復元抽出）。
- 2) 復元抽出を B 回繰り返す、 B 通りの bootstrap 標本を得る。
- 3) B 通りの標本それぞれに対して一般化極値分布を適応し、分位値を求める。
- 4) 求めた分位値の平均を求め、bootstrap 推定値とする。
- 5) 求めた分位値のヒストグラムを描き、分布の形状を調べる。ここでは、bootstrap 標本のサイズである B を 500 とした。

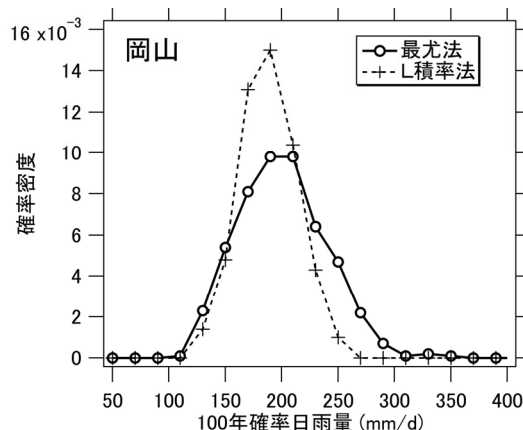


図3 100年確率日雨量の分布 (岡山)

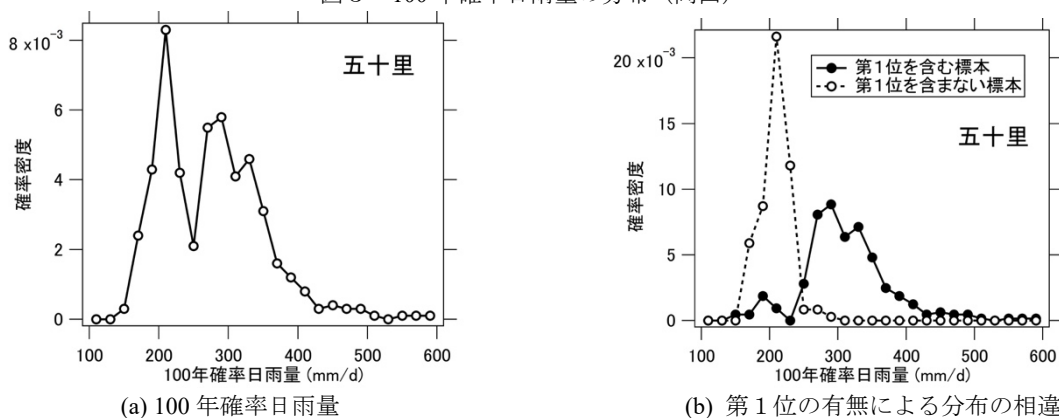


図4 100年確率日雨量の Bootstrap 推定値 (五十里)

4.2 Bootstrap 法による確率日雨量およびパラメータの分布推定

各 bootstrap 標本からそれぞれ 100 年確率日雨量を推定し、その分布を調べた。その結果、五十里、高知、朝倉の 3 地点では、L 積率法、最尤法のいずれを用いた場合でも 2 個以上のピークを持った分布が得られた。特に、朝倉地点では、L 積率法を用いた一般化極値分布によって求められた 100 年確率日雨量の分布には 3 個のピークが見られた。只見および宮古島では最尤法、津和野では L 積率法を用いた場合に 2 個のピークが見られた。これらの地点を除く 5 地点では、いずれの適応法を用いた場合もピークは 1 個となった。岡山における事例を図 3 に示す。

ピークが 2 個以上になった事例として、五十里および朝倉における 100 年確率日雨量および一般化極値分布のパラメータ k の分布を図 4～7 に示す。

図 4 に示した五十里地点における結果を見ると、図 4(a)では 100 年確率日雨量の分布が 2 個のピークを持つことが分かる。Bootstrap 標本を、第 1 位の値を含む標本と含まない標本とに分類し、それぞれの標本群から得られる 100 年確率日雨量の分布を調べた結果を図 4(b)に示す。この図を見ると、第 1 位の値の有無によって確率日雨量の分布が明確に分かれることが示されている。また、図 5(a)に示したパラメータ k の分布でも 2 個のピークが現れており、第 1 位の値を含む標本と含まない標本とに分けて示した図 5(b)では、2 つの分布が明瞭に分けられることが分かる。なお、 k 以外のパラメータでは、 c の分布には 2 個のピークが見られたが、第 1 位の有無による分布の違いは小さく、また、 a の分布には 1 個のピークしか見られなかった。このことから、これら 100 年確率日雨量の分布の相違は、主にパラメータ k の違いによって生じていると推察される。

図6および7に示した朝倉における結果でも五十里と同様の傾向が見られ、第1位の値の有無による分布の違いは五十里よりも明瞭であった。

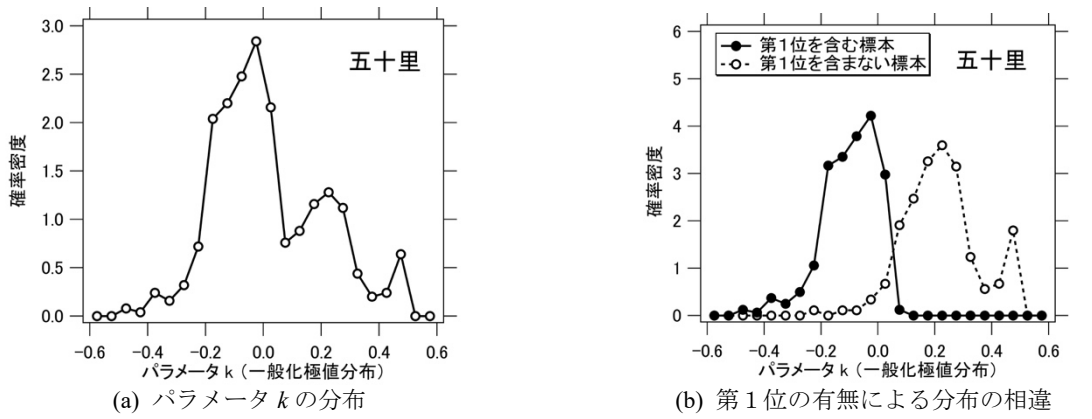


図5 一般化極値分布のパラメータ k の Bootstrap 推定値 (五十里)

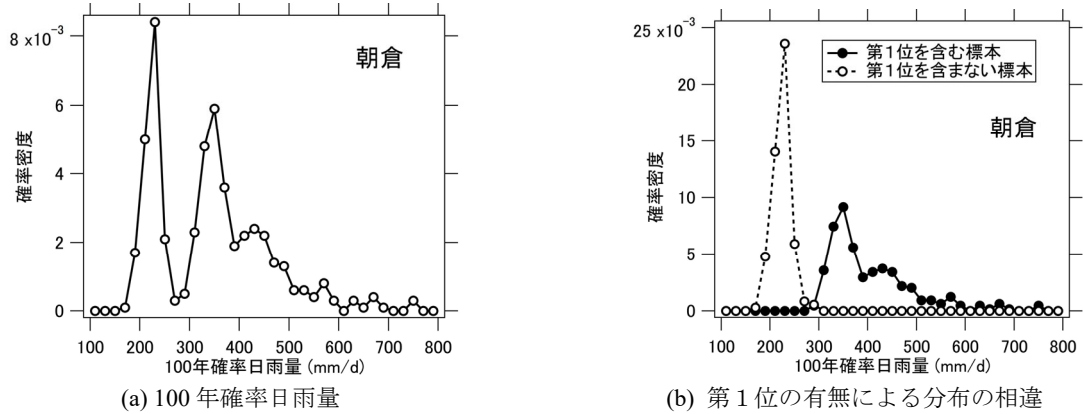


図6 100年確率日雨量の Bootstrap 推定値 (朝倉)

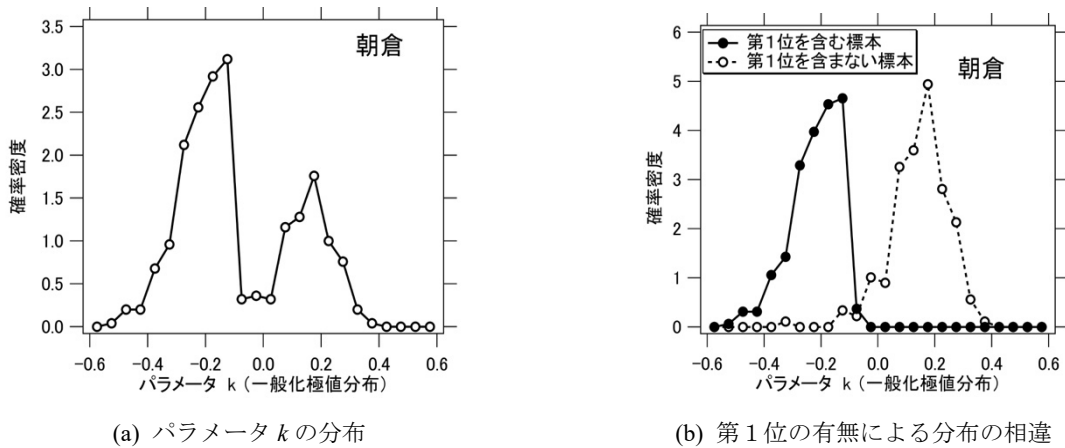


図7 一般化極値分布のパラメータ k の Bootstrap 推定値 (朝倉)

なお、100年確率日雨量およびパラメータの分布のピーク数は、いくつかの地点でパラメータ推定法による違いが見られた。例えば、朝倉地点における100年確率日雨量の分布には、一般化極値分布を最尤法により適応した場合にはピークが2個であったのに対し、L積率法を用いた場合にピークは3個となった。また、只見地点では、最尤法を用いて適応した場合はピークが2個であったのに対し、L積率法を用いた場合は1個であった。

5. おわりに

本研究では、川井、只見、福井、五十里、岡山、津和野、尾鷲、高知、朝倉、宮古島の10地点の1978年～2017年の年最大日雨量のデータを解析対象として、それぞれに一般化極値分布を適応し、分布形の比較検討を行った。また、各地点における解析対象データに、bootstrap法を適用し、100年確率日雨量およびパラメータの分布の形状を調べた。

得られた結果は以下のものである。

- (1) L積率法と最尤法を用いて解析対象データに適応した一般化極値分布の分布形と、プロットイングポジション公式によるプロットとを極値確率紙上で比較した結果、いずれの手法を用いた場合も、非超過確率が高い場合にプロットイングポジション公式によるプロットとの乖離が見られた。また、両手法で適合度に大差は見られなかった。
- (2) 解析対象データから得られるbootstrap標本にL積率法および最尤法を用いて一般化極値分布を適応し、標本から得られる確率日雨量のbootstrap平均によって描かれる分布の比較を行った。その結果、非超過確率が高い部分で最尤法の分布の方がプロットイングポジション公式によってプロットされた観測雨量に近づく傾向が見られた。
- (3) Bootstrap標本から得られた一般化極値分布のパラメータの分布と100年確率雨量の分布を比較すると、五十里と朝倉の2地点において、L積率法、最尤法のいずれを用いた場合も100年確率日雨量とパラメータ k の分布に2つのピークが認められた。
- (4) 2つのピークが認められた五十里、朝倉の2地点における100年確率日雨量の分布を各ピークが属する2つのグループに分割し、各グループが持つ100年確率日雨量の推定に用いられた年最大日雨量のbootstrap標本が1位の値を持つ割合を比較した。その結果、いずれの地点においても、1位の雨量の有無によって標本が所属するグループが異なった。同様に、パラメータ k の分布を2つのグループに分割し同様の検討を行った結果、いずれの地点においても、1位の雨量の有無によって所属するグループが異なった。これは、1位の年の年最大日雨量がそれ以外の年最大日雨量と異なる分布に属する可能性があるということを示唆している。

引用文献

神田 徹・藤田睦博 (1982) : 水文学 (新体系土木工学 26), 技法堂出版.

気象庁 1 (2018年2月26日参照) : 過去の気象データ検索,

<http://www.data.jma.go.jp/obd/stats/etrn/index.php>

気象庁 2 (2018年2月26日参照) : 災害をもたらした気象事例 (平成元年～本年),

http://www.data.jma.go.jp/obd/stats/data/bosai/report/index_1989.html

気象庁 3 (2018年2月26日参照) : 災害をもたらした気象事例 (昭和20～63年),

http://www.data.jma.go.jp/obd/stats/data/bosai/report/index_1945.html

Hosking, J.R.M. and Wallis, J.R. (1997): Regional Frequency Analysis: An Approach Based on L-Moments, Cambridge University Press.

Storn, R and K. Price (1997): Differential Evolution – A Simple and Efficient Heuristic for global Optimization over Continuous Spaces, Journal of Global Optimization.