

一般化パレート分布の最大エントロピー法による 特徴付けに基づく推定量の構成

河村 敏彦[†]・岩瀬 晃盛[†]

(受付 2003 年 7 月 28 日; 改訂 2004 年 4 月 1 日)

要 旨

一般化パレート(GP)分布は Pickands によって導入されて以来, 様々な分野において多くの研究者によって研究されてきた. しかしながら尺度パラメータ ($\sigma > 0$) および形状パラメータ ($-\infty < k < \infty$) をもつ一般化パレート(GP)分布のパラメータの推定は一般的には容易ではない. Castillo and Hadi は $k \leq -1$ のとき最尤推定量は存在せず, また $k > 1/2$ のときには 2 次以上のモーメントは存在しないためモーメント法や確率重み付きモーメント法による推定量は存在しないと述べている. 本稿では最大エントロピー法を用いてパラメータのすべての値に対して推定量を構成する. さらに提案した方法を実データに適用する.

キーワード: 一般化パレート分布, 最大エントロピー法, 点推定, 分布の特徴付け.

1. はじめに

確率密度関数:

$$(1.1) \quad g(x; \sigma, k) = \begin{cases} \frac{1}{\sigma} \left(1 + \frac{kx}{\sigma}\right)^{-\frac{1}{k}-1}, & k \neq 0, \quad \sigma > 0 \\ \frac{1}{\sigma} \exp\left(-\frac{x}{\sigma}\right), & k = 0, \quad \sigma > 0 \end{cases}$$
$$0 < x < 1/\max(0, -k/\sigma)$$

をもつ一般化パレート(GP)分布は Pickands (1975) によって導入されて以来, さまざまな自然現象や工学における諸問題(例えば波浪解析, 地震, 信頼性解析など)に対し研究されてきた. ここで GP 分布の確率密度関数(1.1)のグラフを $\sigma = 1$ を固定した(X/σ を無次元量 X にすることに相当)ときの k の変化による様相だけ図 1, 2 に与えておく.

Baxter (1980) はパレート分布のパラメータの最小分散不偏推定量を構成し, Abdel-Ghaly et al. (1998) はある装置の故障時間分布をパレート分布と仮定し, 数値計算によるパラメータの最尤推定量を議論している. なおこの拡張されていないパレート分布については Johnson et al. (1994), Chapter 20 を参考されたい.

しかしながら拡張されたパレート分布, すなわち尺度パラメータ σ ($0 < \sigma < \infty$), 形状パラメータ k ($-\infty < k < \infty$) をもつ GP 分布のパラメータ推定は一般的には容易ではない. Castillo and Hadi (1997) は $k \leq -1$ のとき最尤推定量は存在せず, また $k > 1/2$ のときには 2 次以上の

[†] 広島大学 工学研究科: 〒739-8527 広島県東広島市鏡山 1-4-1

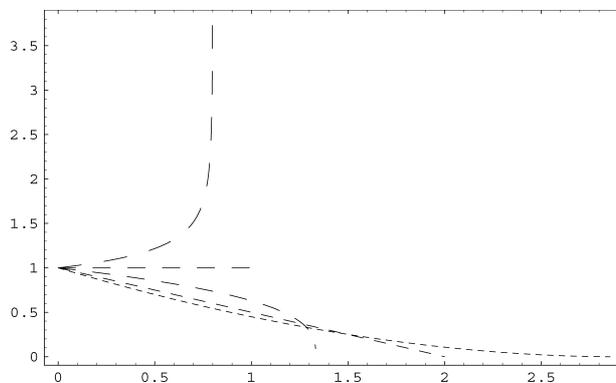


図 1. $\sigma = 1$ を固定したときの $k = -1.25, -1, -0.75, -0.5, -0.35$ に対する GP 分布の密度関数：上から点線が粗い順に $k = -1.25, -1, -0.75, -0.5, -0.35$ に対応する．

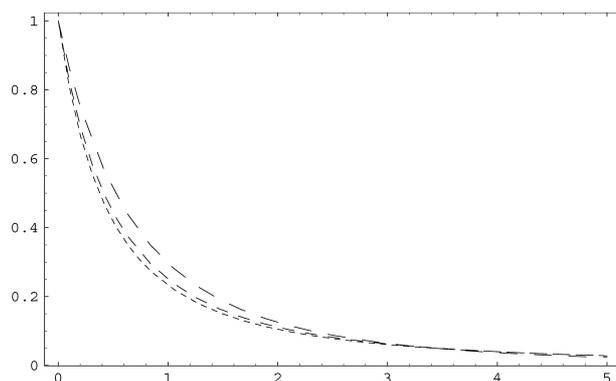


図 2. $\sigma = 1$ を固定したときの $k = 0.5, 1, 1.25$ に対する GP 分布の密度関数：上から点線が粗い順に $k = 0.5, 1, 1.25$ に対応する．

モーメントは存在しないためモーメント法や確率重み付きモーメント法による推定量は存在しないと述べている．Hosking and Wallis (1987) は GP 分布の $-1/2 < k < 1/2$ の場合のみパラメータの推定量の性質を考察している．Singh and Guo (1995) は最大エントロピー法を用いて 3 パラメータの GP 分布の形状パラメータが $0 < k < \infty$ において推定量の構成を行った．そこで本研究の目的は最大エントロピー法によって、GP 分布のすべてのパラメータ値に対する推定量を構成することである．

我々は次節で最大エントロピー法に基づいて 2 パラメータの GP 分布の特徴付けおよびパラメータの推定量を明示する．最後に第 3 節では第 2 節で構成したパラメータの推定量を求めるアルゴリズムを示し、実データに適用する．

2. 最大エントロピー法に基づく GP 分布の特徴付けと推定量の構成

この節ではまず Jaynes (1957) によって導入された最大エントロピー法を簡単に述べる． $-\infty \leq$

$a < b \leq \infty$ とし, 確率密度関数 $f(x)$ は

$$f(x) > 0 \quad (a < x < b), \quad f(x) = 0 \quad (\text{その他})$$

を満足するものとする. さらに, $y_i(x)$ ($i = 1, 2, \dots, m$) は (a, b) 上の可積分関数で, 与えられた定数の組 c_1, c_2, \dots, c_m に対して制約条件

$$E_f[y_i(X)] = \int_a^b y_i(x)f(x)dx = c_i, \quad i = 1, 2, \dots, m$$

を満足するものとする. このとき分布のエントロピー:

$$(2.1) \quad H(X) = - \int_a^b f(x) \log f(x) dx$$

を最大にする分布の確率密度関数は適当な定数 $\lambda_0, \lambda_1, \dots, \lambda_m$ が存在し

$$f(x) = \exp \left\{ \lambda_0 + \sum_{i=1}^m \lambda_i y_i(x) \right\}$$

で与えられる. ただし $\log(\cdot)$ は自然対数である. この最大エントロピー法によるべき逆ガウス型分布および一般化 Gumbel 分布などの特徴付けは最近 Kawamura and Iwase (2003) によって考察されている.

最初に本稿において考察の対象である GP 分布を最大エントロピー法によって特徴付けをする.

定理 1. 確率変数 X は $0 < x < 1/\max(0, -k/\sigma)$ 上で定義された確率密度関数 $f(x)$ をもち, これは制約条件:

$k \neq 0$ のとき,

$$E_f \left[\log \left(1 + \frac{kX}{\sigma} \right)^{-1/k} \right] = -1,$$

$k = 0$ のとき,

$$E_f \left[-\frac{X}{\sigma} \right] = -1$$

を満たすものとする. ただし $0 < \sigma < \infty, -\infty < k < \infty$ である. この条件のもとで $a = 0, b = 1/\max(0, -k/\sigma)$ としたエントロピー(2.1)を最大にする X の分布は GP 分布(1.1)である.

証明. $k \neq 0$ のとき, X_{GP} が GP 分布に従う確率変数とすると,

$$1 - \left(1 + \frac{kX_{GP}}{\sigma} \right)^{-1/k} \sim U(0, 1) : \text{一様分布}$$

より,

$$\left(1 + \frac{kX_{GP}}{\sigma} \right)^{-1/k} = U \sim U(0, 1).$$

ゆえに,

$$\log \left(1 + \frac{kX_{GP}}{\sigma} \right)^{-1/k} = \log U.$$

よって,

$$E_g \left[\log \left(1 + \frac{kX_{GP}}{\sigma} \right)^{-1/k} \right] = \int_0^1 \log u \, du = -1.$$

これより

$$\begin{aligned} - \int_0^b f(x) \log g(x) \, dx &= - \int_0^b f(x) \left\{ \log \frac{1}{\sigma} - \left(\frac{1}{k} + 1 \right) \log \left(1 + \frac{kx}{\sigma} \right) \right\} \, dx \\ &= \log \sigma + \left(\frac{1}{k} + 1 \right) E_f \left[\log \left(1 + \frac{kX}{\sigma} \right) \right] \\ &= \log \sigma + 1 + k \\ &= - \int_0^b g(x) \log g(x) \, dx = H(X_{GP}) \end{aligned}$$

となる. 一方, 情報不等式により

$$\begin{aligned} H(X) &\leq - \int_0^b f(x) \log g(x) \, dx = - \int_0^b g(x) \log g(x) \, dx \\ &= H(X_{GP}) \end{aligned}$$

となり, $k \neq 0$ のとき GP 分布は制約条件の下でエントロピーを最大にする.

次に $k = 0$ のときを示す. このとき GP 分布の密度関数 $g(x)$ が制約条件を満たしていることは明らか. これより

$$\begin{aligned} - \int_0^\infty f(x) \log g(x) \, dx &= - \int_0^\infty f(x) \left\{ \log \frac{1}{\sigma} - \frac{x}{\sigma} \right\} \, dx \\ &= \log \sigma + \frac{1}{\sigma} E_f[X] = \log \sigma + 1 \\ &= - \int_0^\infty g(x) \log g(x) \, dx = H(X_{GP}) \end{aligned}$$

となる. 一方, 情報不等式から

$$\begin{aligned} H(X) &\leq - \int_0^\infty f(x) \log g(x) \, dx = - \int_0^\infty g(x) \log g(x) \, dx \\ &= H(X_{GP}) \end{aligned}$$

となり, $k = 0$ のときも GP 分布は制約条件のもとでのエントロピーを最大にする. □

注意. $k > 0$ のとき GP 分布の密度関数 $g(x)$ が制約条件を満たすことは Gradshteyn and Ryzhik (2000), p. 555 の公式を用いることによって以下のように示すこともできる.

$$\begin{aligned} E_g \left[\log \left(1 + \frac{kX}{\sigma} \right) \right] &= \int_0^\infty \log \left(1 + \frac{kx}{\sigma} \right) \left(1 + \frac{kx}{\sigma} \right)^{-\frac{1}{k}-1} \frac{dx}{\sigma} \\ &= \frac{B(1, \frac{1}{k})}{k} \left\{ \psi \left(1 + \frac{1}{k} \right) - \psi \left(\frac{1}{k} \right) \right\} = k. \end{aligned}$$

ただし $B(\cdot, \cdot)$, $\psi(\cdot)$ はそれぞれベータ関数, ディガンマ関数を表す. また $k < 0$ のとき GP 分布の密度関数 $g(x)$ が制約条件を満たしていることは Gradshteyn and Ryzhik (2000), p. 554 の公式を用いることにより以下のように示すこともできる.

$$E_g \left[\log \left(1 + \frac{kX}{\sigma} \right) \right] = \int_0^{-\frac{\sigma}{k}} \log \left(1 + \frac{kx}{\sigma} \right) \left(1 + \frac{kx}{\sigma} \right)^{-\frac{1}{k}-1} \frac{dx}{\sigma}$$

$$= \frac{B(1, -\frac{1}{k})}{-k} \left\{ \psi\left(-\frac{1}{k}\right) - \psi\left(1 - \frac{1}{k}\right) \right\} = k.$$

次に Singh and Rajagopal (1986) の結果を上記の定理 1 で得られた特徴付けの結果に適用して次の結果を得る.

定理 2. ランダムサンプリングに基づく標本 $\{x_1, x_2, \dots, x_n\}$, $n \geq 2$, の下で最大エントロピー法によるパラメータ (σ, k) の推定量は $k \neq 0$ のとき,

$$\begin{cases} \frac{1}{n} \sum_{i=1}^n \log \left(1 + \frac{\tilde{k}x_i}{\tilde{\sigma}} \right) = \tilde{k} \\ \frac{1}{n} \sum_{i=1}^n \frac{1}{1 + \tilde{k}x_i/\tilde{\sigma}} = \frac{1}{1 + \tilde{k}} \end{cases}$$

を満たす $(\tilde{\sigma}, \tilde{k})$ で与えられる. 特に事前情報により $k = 0$ が既知のときは, σ の推定量は

$$\tilde{\sigma} = \frac{1}{n} \sum_{i=1}^n x_i$$

で与えられる.

証明. $k \neq 0$ のとき, $b = 1/\max(0, -k/\sigma)$ とすれば, 定理 1 の制約条件より $f(x)$ は

$$f(x) = \exp \left\{ \lambda_0 + \lambda_1 \log \left(1 + \frac{kx}{\sigma} \right)^{-1/k} \right\}$$

となる. ただし, λ_0, λ_1 は Lagrange の未定乗数である. また $\int_0^b f(x)dx = 1$ より

$$e^{\lambda_0} = \frac{\lambda_1 - k}{\sigma}$$

を得る. これより

$$\begin{aligned} H(X) &= - \int_0^b f(x) \log f(x) dx \\ &= - \log(\lambda_1 - k) + \log \sigma - \lambda_1 \int_0^b \log \left(1 + \frac{kx}{\sigma} \right)^{-1/k} f(x) dx \end{aligned}$$

となる. この $H(X)$ を λ_1, σ で偏微分をしそれぞれ 0 とおくと

$$\begin{aligned} \frac{\partial H(X)}{\partial \lambda_1} &= \frac{-1}{\lambda_1 - k} - E \left[\log \left(1 + \frac{kX}{\sigma} \right)^{-1/k} \right] = 0 \\ \frac{\partial H(X)}{\partial \sigma} &= \frac{1}{\sigma} - \left(\frac{\lambda_1}{\sigma^2} \right) E \left[\frac{X}{1 + kX/\sigma} \right] = 0 \end{aligned}$$

となる. 一方 (1.1) から $\lambda_1 = k + 1$ である. これより

$$\begin{cases} E \left[\log \left(1 + \frac{kX}{\sigma} \right) \right] = k \\ E \left[\frac{1}{1 + kX/\sigma} \right] = \frac{1}{1 + k} \end{cases}$$

を得る．ここで上式左辺の確率変数 X に対する期待値を標本 $\{x_1, x_2, \dots, x_n\}$ に関する算術平均に置き換えることにより定理 2 が示される．

同様に $k = 0$ のときを示す．定理 1 の制約条件より $f(x)$ は

$$(2.2) \quad f(x) = \exp \left\{ \lambda_0 + \lambda_1 \left(-\frac{x}{\sigma} \right) \right\}$$

となる．また $\int_0^\infty f(x) dx = 1$ より

$$e^{\lambda_0} = \frac{\lambda_1}{\sigma}$$

を得る．これより

$$H(X) = - \int_0^\infty f(x) \log f(x) dx = \log \sigma - \log \lambda_1 + \frac{\lambda_1}{\sigma} E[X]$$

となる．上式を λ_1 で偏微分をし，0 とおくと

$$(2.3) \quad \frac{\partial H(X)}{\partial \lambda_1} = \frac{E[X]}{\sigma} - \frac{1}{\lambda_1} = 0$$

となる．一方 (1.1) から $\lambda_1 = 1$ である．これより

$$E[X] = \sigma$$

を得る．ここで上式左辺の確率変数 X に対する期待値を標本 $\{x_1, x_2, \dots, x_n\}$ に関する算術平均に置き換えることにより定理 2 が示される．□

この定理により GP 分布のすべてのパラメータ値に対して推定量を構成することができ，次節でこの推定法によるデータ解析を行う．

3. データ解析

この節では，前節の定理 2 で構成した最大エントロピー法によるパラメータの推定法を Choulakian and Stephens (2001) によって解析されたデータ(ある未知の閾値以上の超過データが GP 分布に従うとみなせるデータ)に適用する．ここで，GP 分布のパラメータ (σ, k) の最大エントロピー法による推定値(maximum entropy method estimate, MEME) $(\tilde{k}, \tilde{\sigma})$ は次のアルゴリズムで求める(ただし存在および一意性については解析的には未解決である)．

- Step 1. $\xi := k/\sigma$ とおくと

$$\frac{1}{n} \sum_{i=1}^n \log(1 + \xi x_i) = k$$

$$\frac{1}{n} \sum_{i=1}^n \frac{1}{1 + \xi x_i} = \frac{1}{1 + k}$$

となり，第 1 式を第 2 式に代入すると

$$\frac{1}{n} \sum_{i=1}^n \frac{1}{1 + \xi x_i} = \frac{1}{1 + \frac{1}{n} \sum_{i=1}^n \log(1 + \xi x_i)}$$

表 1. Wheaton River のデータセットの最小値を 1 つずつ落としたときの k と σ に対する MEME の値.

Deleted observation	n	\tilde{k}	$\tilde{\sigma}$ [m^3/s]	Deleted observation	n	\tilde{k}	$\tilde{\sigma}$
None	72	-0.194	15.19	1.9	55	-0.251	17.34
0.1	71	-0.196	15.31	2.2	54	-0.260	17.99
0.3	70	-0.196	15.33	2.2	53	-0.258	18.14
0.4	69	-0.198	15.47	2.5	52	-0.273	18.51
0.4	68	-0.180	16.25	2.5	51	-0.273	18.72
0.6	67	-0.203	16.76	2.5	50	-0.271	18.90
0.6	66	-0.185	16.55	2.7	49	-0.263	18.77
0.7	65	-0.189	15.59	2.8	48	-0.269	19.07
1	64	-0.209	15.59	3.4	47	-0.266	18.70
1.1	63	-0.224	16.35	3.6	46	-0.270	18.88
1.1	62	-0.217	16.35	4.2	45	-0.263	18.52
1.4	61	-0.229	16.53	5.3	44	-0.271	17.81
1.5	60	-0.232	16.78	5.3	43	-0.267	17.97
1.7	59	-0.253	16.85	7	42	-0.251	16.56
1.7	58	-0.252	17.45	7.3	41	-0.253	16.61
1.7	57	-0.251	17.18	8.5	40	-0.238	15.46
1.7	56	-0.249	17.34				

を得る . ここで

$$(3.1) \quad d := \frac{1}{n} \sum_{i=1}^n \frac{1}{1 + \xi x_i} - \frac{1}{1 + \frac{1}{n} \sum_{i=1}^n \log(1 + \xi x_i)}$$

とする . ただし , $\xi = 0$ をとりうる場合があることを注意しておく .

• Step 2. ある十分小さな ε に対し $|d| < \varepsilon$ をみたま $\tilde{\xi}$ を求め , この $\tilde{\xi}$ を Step 1 の第 1 式に代入して \tilde{k} :

$$\tilde{k} = \frac{1}{n} \sum_{i=1}^n \log(1 + \tilde{\xi} x_i)$$

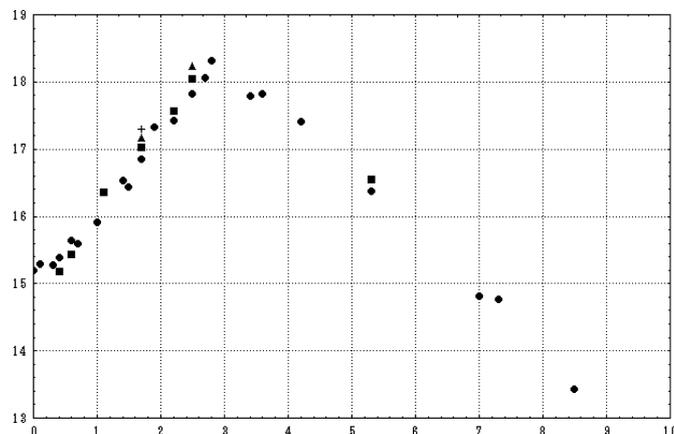


図3. 閾値 u に対する $\sigma + ku$ の推定値(横軸: u , 縦軸: $\tilde{\sigma} + \tilde{k}u$). ここで同一の u 上にある印 \square , \triangle , \circ および $+$ はそれぞれ n が1つずつ少なくなる点を意味する.

を求める. さらに

$$\tilde{\sigma} = \frac{\tilde{k}}{\tilde{\xi}}$$

により $\tilde{\sigma}$ が求まる.

Wheaton River のデータセットの最小値を1つずつ落としたときの k と σ の最大エントロピー法によるパラメータ推定値(MEME) $\tilde{k}, \tilde{\sigma}$ を表1に与える. この結果を用いてさらに図3で横軸を u とした $\tilde{\sigma} + \tilde{k}u$ の打点結果を与える. 確率変数 X が $GP(k, \sigma)$ に従っているとき, 任意の閾値 u に対して $X > u$ であるという条件の下での $X - u$ の分布は $GP(k, \sigma + ku)$ に従うという性質が知られている. この図3を見ると, $u = 2.8$ 以降は u に関して k の推定値が負であることより概ね線形に減少していると思われる. このことは少なくとも $u = 2.7$ までは GP 分布であることが否定され, また $u = 2.8$ 以降は GP 分布であることが否定されないことを意味している. 今の場合, k の正負に関する事前情報を必要とする最尤推定値を用いた上の条件付分布の性質の利用による検討は同様に可能であろうが, ここで提示した MEME は k の正負に関する事前情報を必要とせず, 事実上は特に問題にならないとしても機械的に計算を行うことが可能となるものであり閾値の決定の一つの結果を与えている.

謝 辞

本論文の執筆にあたって二人の査読者の有益なコメントのおかげで内容は大幅に改善されました. 特に査読者には具体的な多くの計算や提案などを頂き, 本稿に反映させて頂きました. 例えば定理1の証明において GP 分布 ($k \neq 0$) が制約条件を満たすことにより簡明な証明, またデータの解析結果においての貴重なコメントを査読者に指摘して頂きました. ここに深く感謝致します.

参 考 文 献

Abdel-Ghaly, A. A., Attia, A. F. and Aly, H. M. (1998) Estimation of the parameters of Pareto

- distribution and the reliability function using accelerated life testing with censoring, *Communications in Statistics. Simulation and Computation*, **27**, 469–484.
- Baxter, M. A. (1980). Minimum variance unbiased estimation of the parameters of the Pareto distribution, *Metrika*, **27**, 133–138.
- Castillo, E. and Hadi, A. S. (1997). Fitting the generalized Pareto distribution to data, *Journal of the American Statistical Association*, **92**, 1609–1620.
- Choulakian, V. and Stephens, M. A. (2001). Goodness-of-fit tests for the generalized Pareto distribution, *Technometrics*, **43**, 478–484.
- Gradshteyn, I. S. and Ryzhik, I. M. (2000). *Table of Integrals, Series, and Products*, 6th ed., Academic Press, San Diego, California.
- Hosking, J. R. M. and Wallis, J. R. (1987). Parameter and quantile estimation for the generalized Pareto distribution, *Technometrics*, **29**, 339–349.
- Jaynes, E. T. (1957). Information theory and statistical mechanics, *Physical Review B*, **106**, 620–630.
- Johnson, N. L., Kotz, S. and Balakrishnan, N. (1994). *Continuous Univariate Distributions*, Vol. 1, 2nd ed., John Wiley & Sons, New York.
- Kawamura, T. and Iwase, K. (2003). Characterizations of the distributions of power inverse Gaussian and others based on the entropy maximization principle, *Journal of the Japan Statistical Society*, **33**, 95–104.
- Pickands, J. (1975). Statistical inference using extreme order statistics, *The Annals of Statistics*, **3**, 119–131.
- Singh, V. P. and Guo, H. (1995). Parameter estimation for 3-parameter generalized Pareto distribution by the principle of maximum entropy (POME), *Hydrological Sciences Journal*, **40**, 165–182.
- Singh, V. P. and Rajagopal, A. K. (1986). A new method of parameter estimation for hydrologic frequency analysis, *Hydrological Science and Technology*, **2**, 33–40.

Composition of Estimators Based on Characterization by Maximum Entropy Method of a Generalized Pareto Distribution

Toshihiko Kawamura and Kōsei Iwase

(Artificial Complex Systems Engineering, Hiroshima University)

Since a generalized Pareto (GP) distribution was introduced by Pickands, it has been studied by many authors in various fields. However, generally estimation of the parameters of the GP distribution with scale parameter ($\sigma > 0$) and shape parameter ($-\infty < k < \infty$) is not easy. Castillo and Hadi stated that when $k \leq -1$, the maximum likelihood estimates do not exist, and also for $k > 1/2$, second and higher moments do not exist, and hence both estimates based on the method of moments and the probability-weighted moment estimates do not exist. This paper propose, for any real number k , a method for estimating parameters based on the maximum entropy method. Furthermore, it is applied for reference data in the proposed method, and numerical computation is carried out.