

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

間 瀬 肇*・北 野 利 一**・森 岡 智 史***

1. はじめに

海岸海洋構造物の設計においては、まず第一に計画波浪の選定が必要である。そのためには、期間最大値資料や極大値資料を用い極値統計解析により、計画波浪を選定しなければならない。波浪極値統計解析の理論的基礎および選定手法については、合田 (1991) が詳しく解説している。山口ら (1997) は、モンテカルロシミュレーションにより種々の極値分布関数から年最大値資料を模擬し、確率波高のバイアスや標準偏差に及ぼす母数推定法、標本数、極値分布関数の形状母数の影響等を詳細に検討している。

年最大値資料は、i) 毎年 1 つのデータしか用いないので他のデータが無駄になる、ii) ある年の 2 番目に大きな事象が他の年の最大事象を上回っていても考慮されない、iii) 極値と考えられない小さな値でも年によっては最大値資料となるという欠点がある。田中・宝 (1999) は、河川流量の非毎年資料、すなわち、極大値資料を用いて確率水文量を評価する方法とその注意点を報告している。波浪観測資料は、流量観測資料に比べて観測データが少なく、波浪推算結果を援用することが多い。こうした資料は、ある閾値以上の大きな事象が抽出された極大値資料とみなされる。

本研究は、波浪の極値統計解析における一般化パレート分布の適用性を検討するものであり、特に、資料年数、平均発生率、母集団の極値分布関数、再現期間を変化させた場合の、確率波高に関する真値と推定値の誤差分布特性に注目する。

2. 一般化パレート分布の特徴

独立・等質な n 個の基準化されたデータからなる標本の極値の分布は、 $n \rightarrow \infty$ で極値 I 型分布、極値 II 型分布、極値 III 型分布のいずれかで表される。Jenkinson (1955) は、これら 3 種類の極値分布を一つの関数形表示、すなわち、一般化極値分布 (Generalized Extreme Value

Distribution, 略して GEV-D) として以下のように表した。

$$F(x) = \Pr\{X \leq x\} = \exp[-\{1 - \kappa(x - \mu)/\sigma\}^{1/\kappa}] \quad \dots\dots\dots (1)$$

ここで、 μ は位置母数、 σ は尺度母数、 κ は形状母数である。

観測データおよび波浪推算結果を有効に利用する方法として、ある値より大きな事象を抽出して標本とする閾値法 (Threshold Method) がある (Davison and Smith, 1990; 田中・宝, 1999)。分布関数が $F(x)$ で表される確率変数 X に対して、閾値 u が十分に大きい場合、 $X > u$ となる条件付確率分布関数

$$P(X \leq x | X > u) = \frac{F(x) - F(u)}{1 - F(u)}; x > u \quad \dots\dots\dots (2)$$

は、次式で表される一般化パレート分布 (Generalized Pareto Distribution, 略して GPD) で近似できる (Pickands, 1975)。

$$G(x) = 1 - \{1 - \kappa(x - u)/\sigma\}^{1/\kappa}; x > u, \kappa \neq 0 \quad \dots (3)$$

$\kappa = 0$ の場合には、次式で表される指数分布となる。

$$G(x) = 1 - \exp\{-(x - u)/\sigma\}; x > u \quad \dots\dots\dots (4)$$

GPD の特徴は、以下のようである。

i) 確率変数 $X - u$ が GPD (σ, κ) に従うとき、閾値 $\bar{u}$ (ただし、 $\bar{u} > u$) を越える $X - \bar{u}$ の分布関数は $\sigma^* = \sigma - \kappa(\bar{u} - u)$ とする GPD (σ^*, κ) となる。

ii) 一年に n 個の観測値があり、閾値 u を超過する確率が λ で、 u を越える観測値が GPD (σ, κ) で表される時、年最大値に対する分布関数は次式で表される (Davison and Smith, 1990)。

$$F(x) = \exp[-\lambda \{1 - \kappa(x - u)/\sigma\}^{1/\kappa}]; x > u \quad \dots (5)$$

これは位置母数、尺度母数、形状母数がそれぞれ、 $u - \sigma((\lambda n)^{-\kappa} - 1)/\kappa$, $\sigma(\lambda n)^{-\kappa}$, κ の GEV-D である。すなわち、GPD から推定されるパラメータは GEV-D におけるパラメータとなる。

GPD と GEV-D との関係については、Stedinger et al. (1993) がまとめており、田中・宝 (1999) が引用している。要点は以下のようである。閾値を越える事象の発生過程をポアソン過程と仮定する。極大値に対する分布関数を $G(x)$ 、年最大値に対する分布関数を $F(x)$ とする

* 正会員 工博 京都大学助教授 防災研究所
** 正会員 工博 名古屋工業大学助手 社会開発工学科
*** 京都大学工学部 地球工学科

と、2つの分布関数には以下の関係がある(合田, 1991)

$$F(x) = \exp[-\lambda\{1-G(x)\}] \quad \cdots \cdots \cdots (6)$$

ここで、 $\lambda(=n\lambda')$ は平均発生率である。式(6)の $G(x)$ に式(3)のGPDを代入すると、式(5)が得られる。

いま、再現期間を R および再現確率統計量を x_R とする。 R と年超過確率 $1-F(x_R)$ との関係は以下のようである。

$$\frac{1}{R} = 1 - F(x_R) = 1 - \exp[-\lambda\{1-G(x_R)\}] \quad \cdots \cdots (7)$$

また、

$$R' = 1/\{\lambda\{1-G(x_R)\}\} \quad \cdots \cdots \cdots (8)$$

として求まる R' は、極大値解析による x_R に対する再現期間である。発生頻度の低い事象については、 R と R' はほぼ等しくなる。

GPDにおける閾値の適切な選択法については、mean residual life plot (Davison and Smith, 1990) を利用する。先に示したGPDのi)の性質より、 $\sigma - \kappa(\bar{u} - u) > 0$ なる $\bar{u}$ に対して、 $\kappa > -1$ ならば

$$E(X - \bar{u} | X > \bar{u}) = \frac{\sigma - \kappa(\bar{u} - u)}{1 + \kappa} \quad \cdots \cdots \cdots (9)$$

となり、 $\bar{u}$ を越える X の平均値は $\bar{u}$ に比例することを意味する。この性質を利用して $\bar{u}$ の最小値を選定する。

例えば、2つの発生原因が異なる出現値が混じっている標本資料において、発生原因別に有意な出現値の偏りがある場合、適切に閾値を設定してどちらかの発生原因による出現値を抜き出せた極大値資料を作成できれば、GPDがうまく適用できることになる。

3. シミュレーションによる検討

シミュレーションによる検討は、以下の条件の下で行った。

- i) 資料年数 N : 10, 20, 30年
- ii) 平均発生率 λ : 15, 30
- iii) 母集団分布関数:

Weibull 分布

$$F(x) = 1 - \exp[-\{(x-B)/A\}^k]; B \leq x < \infty \quad \cdots (10)$$

において、a) $k=1.7$, $A=2.79$, $B=1.0$ およびb) $k=2.0$, $A=3.33$, $B=1.0$ としたもの

Gumbel 分布

$$F(x) = \exp[-\exp\{-(x-B)/A\}]; -\infty \leq x < \infty \quad \cdots \cdots \cdots (11)$$

において、 $A=1.0$, $B=3.0$ としたもの

- iv) 再現期間 R : 50, 100年

シミュレーションの手順は、以下のようである。

i) 母集団から標本資料を作成するため、 N , λ および極値分布関数を定め、 $N\lambda$ 個の(0~1)区間の一様乱数を非超過確率として与え、対応する $N\lambda$ 個の極値を求める。ま

た、 R 年確率波高の真値 H_{real} を求める。

ii) 作成された標本資料に対して最小2乗法を用いてWeibull分布の母数を推定し、それより R 年確率波高の推定値 H_{est} を計算する。手順は、合田(1991)に従った。

iii) 閾値を設定してそれより大きな極値データのみを対象とする場合にGPDをあてはめる。閾値の設定に関しては、式(9)で示した性質をここでは用いないで、簡単に標本平均値の0.5, 0.7, 1.0, 1.3, 1.5倍を閾値とした5種類の標本資料に対してGPDをあてはめ、そのうち宝・高棹(1988)が提案した最小2乗基準が最小で、かつ0.03以下の場合に R 年確率波高の推定値 H_{est} を求める。母数推定に当たっては、Lモーメント法を用いた(Stedinger et al., 1993)。プロットング公式としてはCunnane公式を用いた。

iv) H_{real} と H_{est} の差を求める。

v) i)~iv)を5000回繰り返して、確率波高に関する真値と推定値の誤差分布特性を調べる。

図-1は、母集団の分布関数がWeibull分布である標本資料に対して、50年確率波高に関する真値と推定値の誤差の分布を示したものである。標本と推定値の相関係数は0.99以上である。この図においては、 $\lambda=30$ と固定し、観測年数 N を10, 20, 30と変化させた。なお、横軸は誤差の標準偏差 σ で無次元化してある。○印がWeibull分布をあてはめて得られた結果、●印が閾値資料にGPDをあてはめて得られた結果である。母数は、50年確率波高が約10mになるように設定した。(b)図のみ母集団分布として与えるWeibull分布の母数が異なっている。(b)図のケースは $k=2.0$ であり、最小2乗法において固定値として扱われる0.75, 1.0, 1.4, 2.0のうちの1つであるが、(c)図では $k=1.7$ となっている。(b)図のケースでは、 $k=2.0$ となって尺度母数 A と位置母数 B が推定される。このような場合誤差分布の最頻値は0となり、分布形は左右対称になる。(c)図のケースにおいては、 $k=1.4$ あるいは2.0のもので A および B が推定される。誤差分布は2山型分布になるが、これは k が1.4か2.0のどちらかが選ばれたことにより、確率波高が大きくなる場合と小さくなる場合に分かれることによるものと考えられる。(a), (c), (d)図のWeibull分布をあてはめた結果は、観測年数が大きくなるにつれて2山のピークが鋭くなる。

一方、(a)~(d)図のすべての図において、GPDをあてはめた場合の誤差分布は最頻値が負値となり、正方向に裾をひく分布形になる。また、GPDを用いた場合の誤差分布形はほとんど同じになる。

図-2には、50年確率波高の結果に加え、100年確率波高のそれも含めて、GPDを用いた場合の誤差分布を示した。それらの結果は点線で示してある。図中には、これ

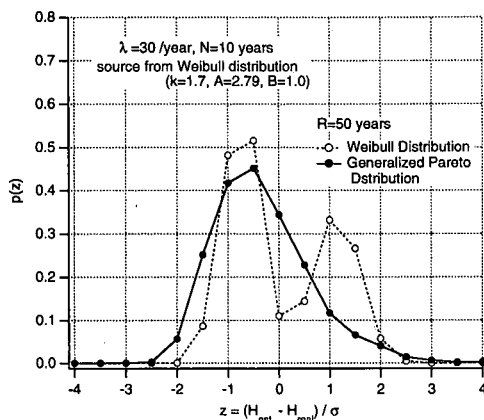
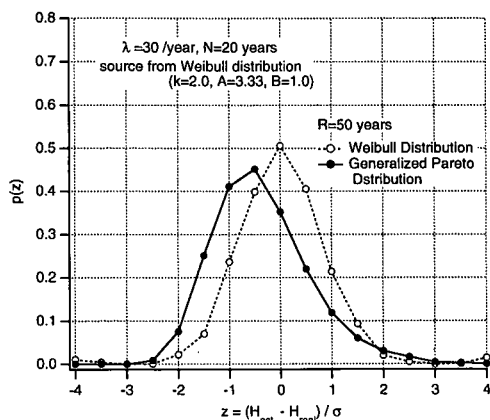
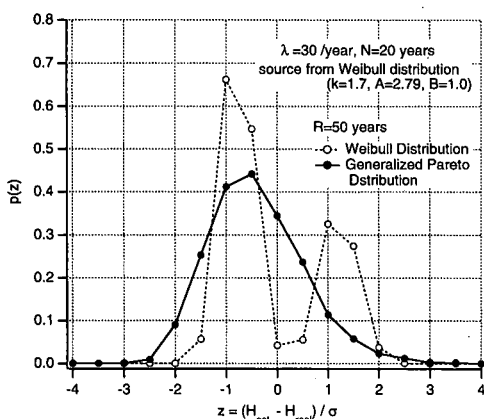
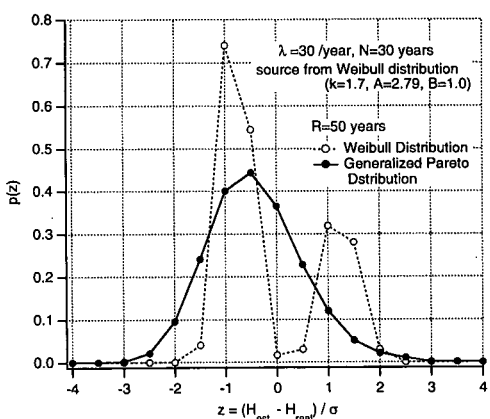
(a) $N=10, k=1.7$ (b) $N=20, k=2.0$ (c) $N=20, k=1.7$ (d) $N=30, k=1.7$

図-1 Weibull 分布を母集団分布とする場合の 50 年確率波高に関する誤差分布

らすべての分布形を近似する分布形として Gumbel 分布 ($A=0.79, B=-0.67$) を示した。この Gumbel 分布は以下に示す誤差分布も近似できる。

図-3 は、母集団の分布関数が Gumbel 分布である標本資料に対して、50 年確率波高に関する真値と推定値の誤差分布を示したものである。Weibull 分布をあてはめて確率波高を推定した場合の誤差分布は、図-2 と同様な 2 山型分布になるが、GPD を用いると誤差分布はほぼ同じになる。このケースにおいても、100 年確率波高に対する誤差分布もまとめて図示すると、図-4 が得られる。この図においても、Gumbel 分布 ($A=0.79, B=-0.67$) がこれらの誤差分布形をほぼ近似できる。

図-5 はデータ数が少ない標本を用いて、これまでと同様に誤差分布を示したものである。(a) 図は母集団が Weibull 分布の場合、(b) 図は Gumbel 分布の場合である。Weibull 分布の形状母数は 2.0 としたので、誤差分布は 1 山型分布になっている。GPD を用いると、これまで

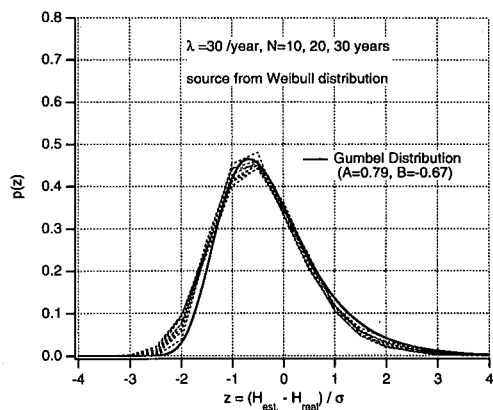


図-2 Weibull 分布母集団からの標本資料に GPD をあてはめたときの誤差分布

と同様の分布形が得られている。50 年および 100 年確率波高に関する誤差をまとめた結果が、図-6 である。この

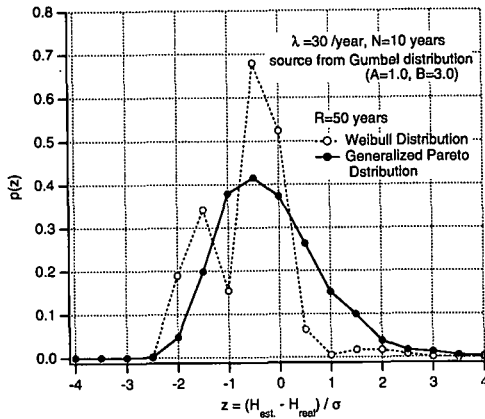
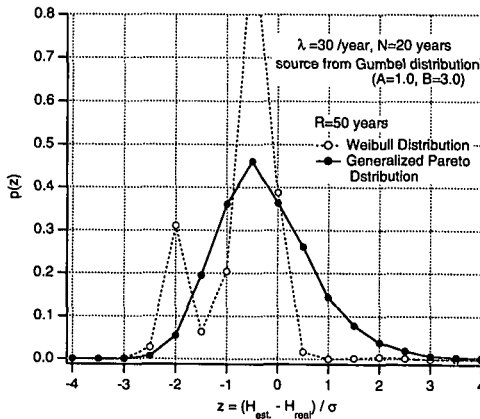
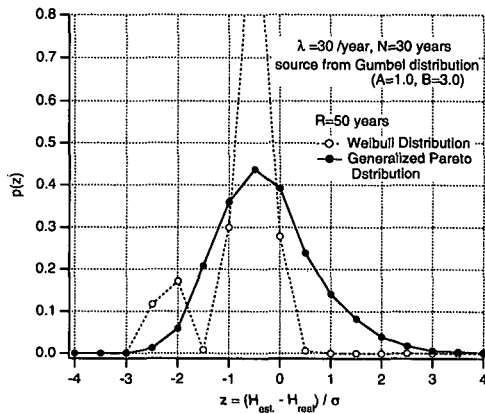
(a) $N=10$ (b) $N=20$ (c) $N=30$

図-3 Gumbel 分布を母集団分布とする場合の 50 年確率波高に関する誤差分布

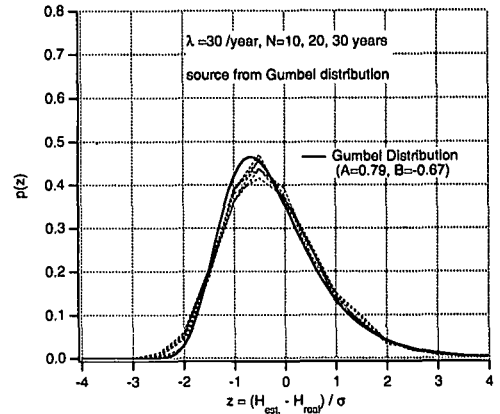
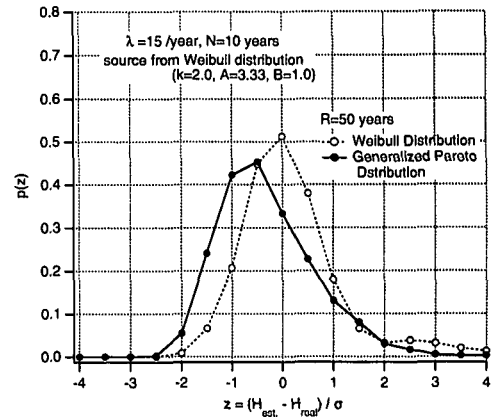
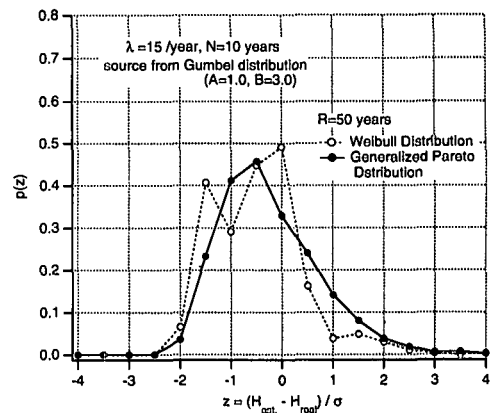


図-4 Gumbel 分布母集団からの観測資料に GPD をあてはめたときの誤差分布



(a) 母集団分布は Weibull 分布



(b) 母集団分布は Gumbel 分布

図-5 データ数が少ない場合の 50 年確率波高に関する誤差分布

図においても、図-2 および図-4 に示した Gumbel 分布 ($A=0.79$, $B=-0.67$) を示してある。

今回行ったシミュレーション条件下においては、閾値データに GPD を適用すると、確率波高の真値と推定値

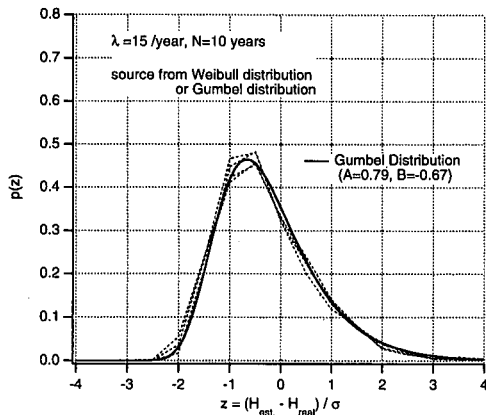


図-6 データ数が小さい場合の確率波高に関する誤差分布

(a)

(b)

(c)

(d)

(e)

(f)

(g)

(h)

(i)

(j)

(k)

(l)

(m)

(n)

(o)

(p)

(q)

(r)

(s)

表-1 確率波高に関する真値と推定値

 (a) 母集団は Weibull Distribution ($k=1.7$, $A=2.79$, $B=1.0$)

 $R=50$ year, Return Value=9.993 m

N	WD	GPD
10	10.021(0.917)	9.544(1.343)
20	9.920(0.821)	9.578(1.031)
30	9.892(0.781)	9.630(0.861)

 $R=100$ year, Return Value=10.485 m

N	WD	GPD
10	10.523(1.038)	9.913(1.664)
20	10.406(0.935)	9.941(1.281)
30	10.374(0.891)	9.997(1.078)

 (b) 母集団は Gumbel Distribution ($A=1.0$, $B=3.0$)

 $R=50$ year, Return Value=10.313 m

N	WD	GPD
10	9.560(1.181)	10.002(1.642)
20	9.610(0.974)	10.047(1.224)
30	9.669(0.856)	10.065(1.000)

 $R=100$ year, Return Value=11.006 m

N	WD	GPD
10	10.055(1.402)	10.638(2.117)
20	10.111(1.176)	10.673(1.590)
30	10.179(1.045)	10.683(1.301)

参 考 文 献

- 合田良夷(1991): 港湾構造物の耐波設計一波浪工学への序説一, 鹿島出版会, 333 p.
 田中茂信・宝 馨 (1999): 毎年資料と非毎年資料による確率水文量の評価, 水工学論文集, 第 43 巻, pp. 145-150.
 宝 馨・高埴琢馬 (1988): 水文頻度解析における確率分布モデルの評価基準, 土木学会論文集, 第 393 号/II-9, pp. 151-160.
 山口正隆・大木泰憲・中村雄二 (1997): シミュレーションデータに基づく波高の極値統計解析モデルの精度の検討, 自然災害科学, Vol. 16, No. 1, pp. 61-73.
 Davison, A. C. and R. L. Smith (1990): Models for exceedances over high thresholds (with discussion), Jour. Roy. Statist. Soc., Vol. 52, pp. 393-442.
 Jenkinson, A. F. (1955): The frequency distribution of the maximum or minimum of meteorological elements, Quart. Jour. Roy. Meteorol. Soc., Vol. 81, pp. 158-171.
 Pickands, J. (1975): Statistical inference using extreme order statistics, Ann. Statist., Vol. 3, pp. 119-131.
 Stedinger, J. R., R. M. Vogel and E. Foufoula-Georgiou (1993): Frequency analysis of extreme events, Chap. 18, Handbook of Hydrology, McGraw-Hill, New York, pp. 18.1-18.66.

(a)

(b)

(c)

(d)

(e)

(f)

(g)

(h)

(i)

(j)

(k)

(l)

(m)

(n)

(o)

(p)

(q)

(r)

(s)

(t)

(u)

(v)

(w)

(x)

(y)

(z)

(aa)

(ab)

(ac)

(ad)

(ae)

(af)

(ag)

(ah)

(ai)

(aj)

(ak)

(al)

(am)

(an)

(ao)

(ap)

(aq)

(ar)

(as)

(at)

(au)

(av)

(aw)

(ax)

(ay)

(az)

(ba)

(bb)

(bc)

(bd)

(be)

(bf)

(bg)

(bh)

(bi)

(bj)

(bk)

(bl)

(bm)

(bn)

(bo)

(bp)

(bq)

(br)

(bs)

(bt)

(bu)

(bv)

(bw)

(bx)

(by)

(bz)

(ca)

(cb)

(cc)

(cd)

(ce)

(cf)

(cg)

(ch)

(ci)

(cj)

(ck)

(cl)

(cm)

(cn)

(co)

(cp)

(cq)

(cr)

(cs)

(ct)

(cu)

(cv)

(cw)

(cx)

(cy)

(cz)

(da)

(db)

(dc)

(dd)

(de)

(df)

(dg)

(dh)

(di)

(dj)

(dk)

(dl)

(dm)

(dn)

(do)

(dp)

(dq)

(dr)

(ds)

(dt)

(du)

(dv)

(dw)

(dx)

(dy)

(dz)

(ea)

(eb)

(ec)

(ed)

(ee)

(ef)

(eg)

(eh)

(ei)

(ej)

(ek)

(el)

(em)

(en)

(eo)

(ep)

(eq)

(er)

(es)

(et)

(eu)

(ev)

(ew)

(ex)

(ey)

(ez)

(fa)

(fb)

(fc)

(fd)

(fe)

(ff)

(fg)

(fh)

(fi)

(fj)

(fk)

(fl)

(fm)

(fn)

(fo)

(fp)

(fq)

(fr)

(fs)

(ft)

(fu)

(fv)

(fw)

(fx)

(fy)

(fz)

(ga)

(gb)

(gc)

(gd)

(ge)

(gf)

(gg)

(gh)

(gi)

(gj)

(gk)

(gl)

(gm)

(gn)

(go)

(gp)

(gq)

(gr)

(gs)

(gt)

(gu)

(gv)

(gw)

(gx)

(gy)

(gz)

(ha)

(hb)

(hc)

(hd)

(he)

(hf)

(hg)

(hh)

(hi)

(hj)

(hk)

(hl)

(hm)

(hn)

(ho)

(hp)

(hq)

(hr)

(hs)

(ht)

(hu)

(hv)

(hw)

(hx)

(hy)

(hz)

(ia)

(ib)

(ic)

(id)

(ie)

(if)

(ig)

(ih)

(ii)

(ij)

(ik)

(il)

(im)

(in)