東海豪雨級の大雨は，統計的に推定可能か？それとも予測可能なのか？

Extremes should be treated to be estimable, or predictable?

北野利一　リスク解析戦略研究センター　客員准教授／名古屋工業大学　社会工学専攻・高度防災工学センター　准教授

1．再現期間の延伸により外挿して求める確率外力では説明しにくい極値

図 -1 は，いわゆる**ガンベル確率紙**に，名古屋における日降雨量の年最大値（1901 年から 2008 年までの 108 年間）をプロットしたものである．横軸に**再現期間** R をとり，極値分布の母数 $\boldsymbol{\theta}_n = \{\mu_n, \sigma_n, \xi\}$ を用いて表される**生起率関数** $\lambda(y, \boldsymbol{\theta}_n)$ が，単位時間（この場合，$n = 1$ 年）あたりに閾値 y を越える平均生起数 $\lambda_1 = \lambda(y, \boldsymbol{\theta}_1)$ が，再現期間の逆数に等しくなることから，次式の関係が満足する．

$$\lambda_1 = \lambda(y, \boldsymbol{\theta}_1) = \frac{1}{R}, \qquad for \qquad \lambda(y, \boldsymbol{\theta}_n) = \exp\left\{-\frac{1}{\xi}\log\left(1 + \xi\,\frac{y-\mu_1}{\sigma_1}\right)\right\} \quad (\xi \neq 0), \qquad or \qquad = \exp\left(-\frac{y-\mu_1}{\sigma_1}\right) \quad (\xi = 0) \tag{1}$$

$\xi = 0$ の場合，上式を解いて，縦軸の外力（ここでは，日降雨量）y は，再現期間の対数 $\log R$ と線形の関係で表される．また，$N = 108$ 個の順序統計量の実現値となるデータに対して，その超過確率の割り付けを考えることにより，順序統計量 $y_{(i)}$（i は降順に数えた序数）に対する再現期間 R_i が次式に示すものとなるように，図 -1 に描かれている．

$$R_i = 1 \Big/ \left\{-\log\left(1 - \frac{i}{N+1}\right)\right\} \tag{2}$$

図 -1 で突出した値 (428.0 mm) をとっているのは，東海豪雨 (2000 年 9 月 11 日) における日降雨量である．本来的に確率変数である**順序統計量**に対して，その実現値であるデータに超過確率を割り付けるという考えには正解はなく，便宜的にならざるを得ない．しかし，たかだか 100 年程度の観測期間に対して，その最大値の超過確率に再現期間が数万年となる超過確率を割り付けるのも無理な話であり，図に示されるとおり，外力と再現期間の対数の線形関係から大きく外れることになる．また，たかだか 100 年程度に観測期間が限られているにも関わらず，外力と再現期間の対数の線形関係を，データの疎となる範囲にどこまでも延伸できるわけがない．その外挿の限界（図中の白い領域の右端および上端）を示すために，次式で算定される**経験度** K を用いて，「２度あることは３度ある」という諺を基準に，$K = 2$ を外挿が可能な限界としている (Kitano, 2013 を参照)．

$$\frac{1}{K} = V\left(\log\lambda(y, \hat{\boldsymbol{\theta}}_1)\right) = \left.\frac{\nabla'\lambda\ \mathcal{I}^{-1}\ \nabla\lambda}{\lambda^2}\right|_{\lambda = \lambda(y, \hat{\boldsymbol{\theta}}_1)}, \qquad \nabla' = \nabla'_{\boldsymbol{\theta}_1} = \left(\frac{\partial}{\partial\mu_1}, \frac{\partial}{\partial\sigma_1}, \frac{\partial}{\partial\xi}\right) \quad (\xi \neq 0), \qquad or \qquad = \left(\frac{\partial}{\partial\mu_1}, \frac{\partial}{\partial\sigma_1}\right) \quad (\xi = 0) \tag{3}$$

なお，$\mathcal{I}$ は **Fisher 情報行列**である．東海豪雨の日降雨量に対しては，図 -1 に示す灰色の領域（$K < 2$）にあり，推定結果を保留するべきであることを意味している．

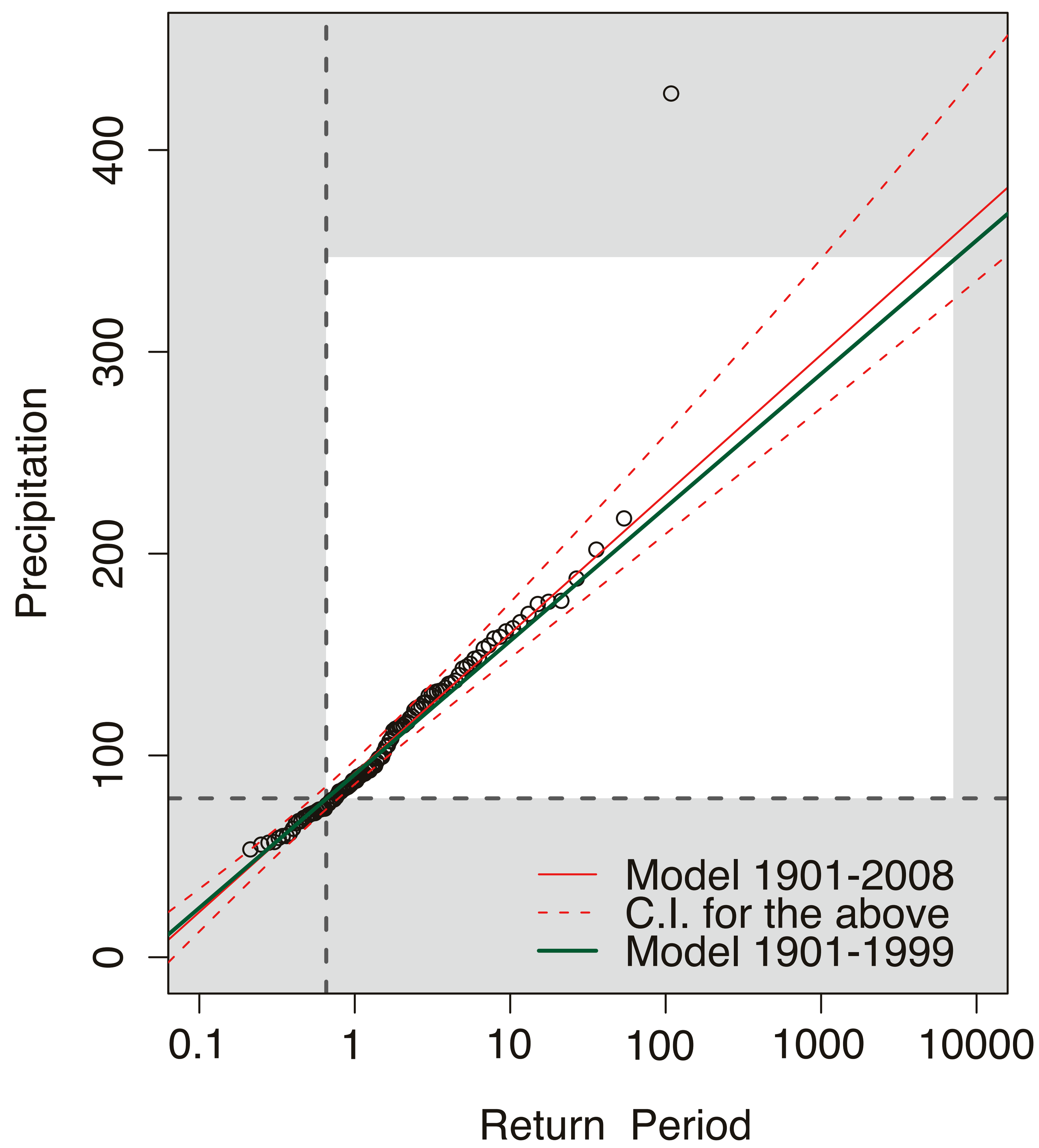

図 -1　ガンベル確率紙における推定の限界を示す窓枠（文献 1 より転載）

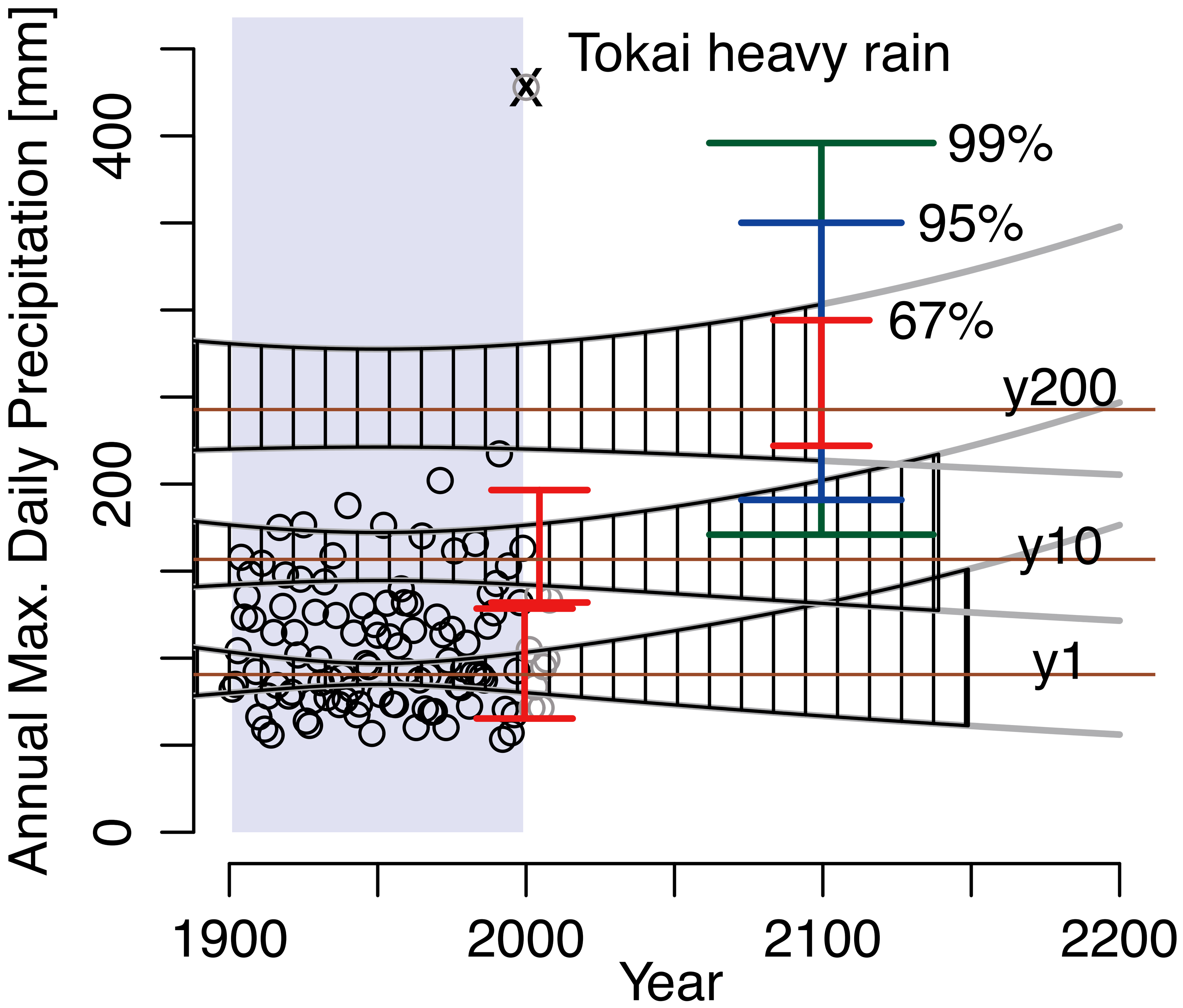

図 -2　経過時間に伴う推定の延伸の限界と予測区間（文献 2 より転載）

2．観測期間から将来への時間の延伸により外挿して求める確率外力の区間推定でも説明しにくい極値

推定における再現期間の扱いは，確率外力の超過確率として　現実に未来に横たわる実時間を対象にしていない．そこで，時間変数 t を導入すれば（t_j は連続的に観測された極値 y_j の観測年およびその平均 $\bar{t} = \sum_j t_j / N$），時間の経過を考慮に入れた経験度（これを定常モデルに対する**耐久性**，もしくは，観測データに対する**鮮度**とよぶ）は，

$$\frac{1}{K} = \frac{1}{K_0} + \frac{(t - \bar{t})^2}{\sum_j (t_j - \bar{t})^2} \tag{4}$$

で与えられる（Kitano, 2013）．K_0 は，１．に示されている式 (3) で得られる経過時間を考慮しない経験度である．$t = \bar{t}$ において，経過時間に伴う経験度の低下（鮮度の低下）が見られないことになる．図 -2 における縦縞の帯は，式 (4) の経験度により求められる信頼区間であり，時間の経過と伴に幅が広くなっていることがわかる．このような広がりを，エネルギーが密の領域から疎の領域に " 漏れる " という回折現象にたとえて，推定誤差の**回折効果**とよぶ（Kitano, 2013）．なお，$K = 2$ で，その延伸を打ち切っている．

3．将来に生起するであろう極値の予測分布

再現期間を与えて定まる確率外力（あるいは，より一般的に言えば，**再現レベル**とよばれる）は，母数の真値が既知であれば，1 つの値を示す定数である．母数が未知であるから，その推定として，信頼区間を求めて区間推定を行っている．これに対して，将来に生起するであろう極値は，本質的に確率変数である．これがとりうる範囲を示すために，予測分布が必要となる．例えば，母分布が正規分布であれば，標本平均を標準化した統計量がティ分布に従うように，予測分布もティ分布で表される．残念ながら，極値の予測分布を厳密に誘導できるとは考えにくい．しかし，さいわいなことに，経験度を導入することにより，**指数分布族ではない極値分布**の特性と，指数分布族であるポアソン分布の特性を近似的に結びつけることができる．このようにして，将来 R 年間に生起するであろう**極値の予測分布**は次式で与えられ，$R = 200$ として予測区間を図 -2 の中に示した．

$$F_R^*(y) = \left\{1 + \frac{\lambda_R(y)}{K}\right\}^{-(K+1)} \tag{5}$$

参考文献：1) 北野利一・高橋倫也・田中茂信 (2011): : 確率降雨量の統計解析におけるモデル適合性の限界，土木学会論文集 B1（水工学），67(4), pp.I_277-I_282.
2) 北野利一・高橋倫也・田中茂信 (2013): 降水量の極値の予測区間 ～ 確率降水量の信頼区間を誤解していませんか？，土木学会論文集 B1（水工学），69(4), pp.I_271-I_276.
3) Kitano, T. (2013): Properties of the log-transformed occurrence rate, The ISM Cooperative Research Report 299, pp.61-66.

（極値理論による応用全般に関して，共同研究を歓迎する）