

# 極値統計解析における2種類の外挿とその限界

北野利一 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る外挿

再現期間は、超過確率の逆数で与えられる。通常、年最大値分布の超過確率を用いる。ここでは、年最大値分布の母数  $\theta = \{\mu, \log \sigma, \xi\}$  を用いて、閾値  $y$  をこえる外力の生起率  $\lambda(y; \theta) = \exp\left\{-\frac{1}{\xi} \log\left(1 + \frac{y - \mu}{\sigma}\right)\right\}$  とし、再現期間  $R$  の確率外力  $y_R$  に対する  $\frac{1}{R} = \lambda(y_R; \theta)$  という関係を用いる。この時、母数が与えられれば、確率外力は  $y_R = \mu + \sigma \frac{e^{-\xi \log \lambda} - 1}{\xi} \Big|_{\lambda = \frac{1}{R}}$  と求まる。

対数スケールにおける再現期間の増大（生起率の低下）に伴い、観測された外力の範囲から抜けて、確率外力が延伸される。図-1のオランダの潮位データに対して、図-2は、再現期間の延伸に伴い、確率潮位が外挿される様子を描いている。図中の窓枠は、その外挿の限界を示すものであり、次式の経験度により与えられる（北野ら，2008）。

$$\frac{1}{K} = V\left\{\log \lambda(y_R; \theta)\right\} \quad (1)$$

一般に誤差を定量化するために、誤差分散を用いられることが多い。母数の誤差分散は、最尤推定においては情報行列  $I$  を介して求められる。  $V(\hat{\theta}) = I^{-1}$  (2) 式(1)と(2)を比較してわかるとおり、経験度と情報は、いずれも分散の逆数を表す量である点では同等であるが、情報行列は多元の量であるのに対して、経験度はスカラー量である。経験度の値は、確率外力  $y_R$  を推定するために用いたデータの実質の数を表している（それゆえ、 $K \geq 2$  を有効範囲とする理由が解釈可能）。対象となる確率外力に的を絞り、多元の量である情報行列を整理して表した量が経験度であると言える。経験度の具体的な算出にあたり、式(1)を展開して、  $K = (\nabla_{\theta} \log \lambda I^{-1} \nabla_{\theta} \log \lambda)^{-1}$  (3) という情報行列  $I$  を用いた演算が必要であり、情報行列のスカラー化であることが明確となる。ところで、確率外力の誤差分散  $V(y_R)$  もスカラー量であり、推定誤差の検討に用いられてきている実績から、経験度のような新しい指標を取って用いる必要を感じないかもしれない。しかしながら、次元量で表される確率外力  $y_R$  の誤差分散  $V(y_R)$  を直接的に扱うのは、表面的には分かり易いものかもしれないが、相対的な比較しか検討できず、外挿の限界に対する判断の指標には適さない。

## 2. 観測期間から将来への時間の延伸による外挿

非定常モデルを検討する場合には、 $\theta = x'\beta$ ;  $x = \begin{pmatrix} 1 \\ t \end{pmatrix}$ ;  $\beta = \begin{pmatrix} \mu_0 & \log \sigma_0 & \xi_0 \\ \beta_{\mu} & \beta_{\log \sigma} & \beta_{\xi} \end{pmatrix}$  により、母数  $\theta$  と時間変数  $t$  をリンクさせて、生起率を取り扱うことになる。非定常モデルの場合、期待情報行列  $I$  は、母数  $\theta_0 = \{\mu_0, \log \sigma_0, \xi_0\}$  の定常モデルに対する情報行列  $I_0$  の拡大行列  $I = \begin{pmatrix} I_0 & I_1 \\ I_1' & I_2 \end{pmatrix}$ ;  $I_1 = I_0 \mathbf{1}$ ;  $I_2 = \mathbf{1}' I_0 \mathbf{1}$ ;  $\mathbf{1}_{(\mu)} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ ,  $\mathbf{1}_{(\mu, \log \sigma)} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}, \dots$  である。その部分行列  $I_{j=1,2}$  は、

情報行列  $I_0$  から時間に依存する母数に対する成分だけを抜き出した行列である。このように得られる期待情報行列を式(3)に代入する際に、母数による微分  $\nabla_{\theta}$  は、ベック作用素による  $\text{vec} \beta'$  あることに注意し、シュアの補元やクロネッカー積などの行列演算のテクニックを利用すれば、非定常モデルの経験度を、一般的に、  $\frac{1}{K} = \frac{1}{K_0} + \frac{N}{N-1} \frac{MD^2(x)}{M}$  (4) と分解できる。  $N$  は観測記録の年数であり、 $K_0$  は、定常モデルに対する情報行列  $I_0$  から算出される経験度であり、 $M$  は、時間に依存する母数に対する成分だけを抜き出した部分行列  $I_2$  を用いて算出される経験度である。したがって、全ての母数が時間に依存するなら、 $M = K_0$  となり、位置母数  $\mu$  のみが時間に依存する時 ( $\beta_{\sigma} = \beta_{\xi} = 0$ )、ガンベルタイプ ( $\xi = 0$ ) となる年最大値上位  $r$  番目までの観測記録を使用できる場合には、 $M/N = r$  となる（ガンベル分布に従う年最大値の場合は、 $M/N = 1$ ）。

$MD^2(x)$  は、共変量  $x$  に対するマハラノビス平方距離であり、この場合、  $\frac{MD^2(x)}{N-1} = \frac{(t - \bar{t})^2}{\sum_i (t_i - \bar{t})^2}$ ;  $\bar{t} = \frac{\sum_i t_i}{N}$  となる。

以上で得られた非定常モデルの経験度において注目すべき点は、右辺のいずれの項にも、時間の変化に依存する度合いを表す係数が（陽には）含まれていないことである。したがって、形式的に、式(4)を右辺から左辺の向きに適用すれば、定常モデルの経験度  $K_0$  に、時間の経過による効果を合成することができる。このように定常モデルに対する経験度を拡張するものとして、耐久性  $K$  を導入できる。

図-3は、再現期間と実際の時間（観測期間からの経過時間）による2つの外挿のため、耐久性が低下する様子を3次元空間内の曲面で表したものである。

耐久性  $K$  が与えられれば、信頼区間は容易に得られる。対象とする確率外力  $y_R$  の生起率の推定値  $\hat{\lambda} (= \lambda(y_R, \hat{\theta}))$  が、耐久性の値を形状母数に与えたガンマ分布（ポアソン分布の自然共役分布）に従うと考え、さらに、所与の生起率  $1/R$  が平均となるようにガンマ分布の尺度母数を定める。信頼係数95%の場合には、超過確率と累積確率が0.025となるガンマ分布のパーセント点を生起率  $\hat{\lambda}$  が取りうる範囲の上下端  $\lambda_U$  および  $\lambda_L$  として与えることができる。よって、  $\lambda_L \leq \lambda(y_R, \hat{\theta}) \leq \lambda_U$  を満足するように、生起率関数に含まれる確率外力  $y_R$  に対して解けば、その信頼区間を求めることができる。

## 3. 応用例～オランダの潮位記録（験潮記録に基づく近年データ + 堆積物と古文書による曖昧さを伴う歴史データ）

歴史データが追加されると、図-4に比較するとおり、等高線の楕円が反時計向きに傾く。時間軸に沿った線対称の軸（緑点線）は、歴史データの観測時間側に引き寄せられる。信頼区間の変化は再現期間ごとには対称となるが、再現期間の増大とともに、その対称軸が近年データの観測期間内から歴史データのそれに移行している。近年データのみによる経験度  $K = 2$  の境界線を茶点線で表しており、例えば、20年確率潮位の推定は、延伸される将来区間が顕著に増える（区間a）。その一方で、1250年確率潮位は、線対称軸が歴史データの区間に大幅に引き込まれるため、将来への延伸量はわずかである（区間b）が、歴史データの観測期間が長い分、過去への延伸は顕著である（区間c）。また、図-4左上の水平線 ( $\log_{10} R \approx -0.18$ ) は、各年毎に  $K$  が最大値をとる再現期間であり、それ未満となる再現期間に対しては、たとえ、 $K \geq 2$  であっても興味の対象外である（図-2を参照）。そのような境界線も傾き、20年確率潮位の過去への延伸は制限される（区間d,e）。得られる信頼区間は、多くの観点で、時間軸に沿った対称性がくずれている。

参考文献：北野利一，極値理論の工学への応用 (9, 8, 7, 6), 統計数理研究所 共同研究リポート 274 (2012), pp.51-58; 261 (2011), pp.50-64; 246 (2010), pp.9-12; 224 (2009), pp.7-12.

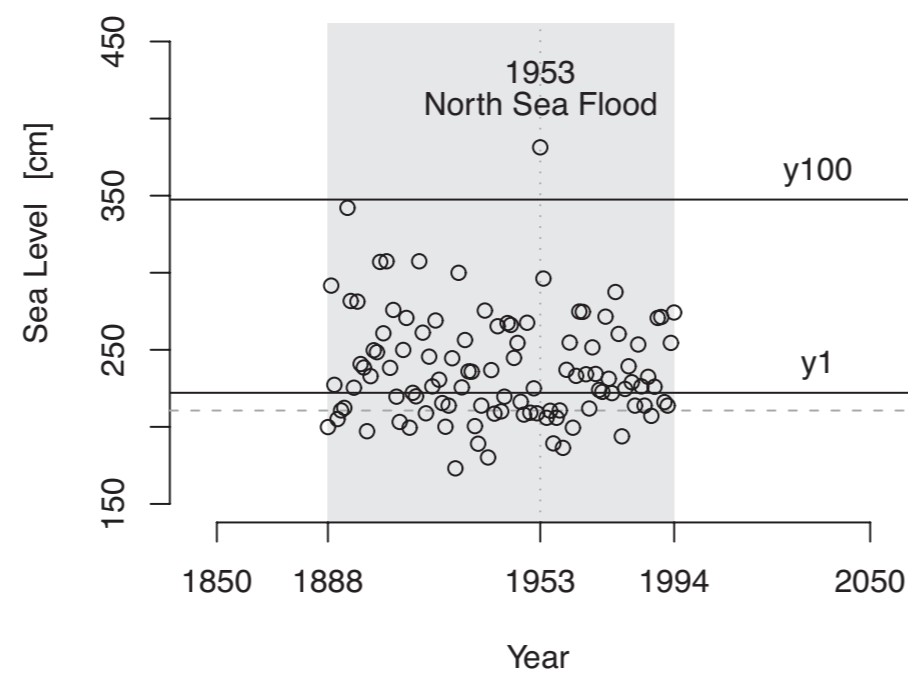


図-1 年最大潮位の時系列（左）と再現期間に対する潮位（右）

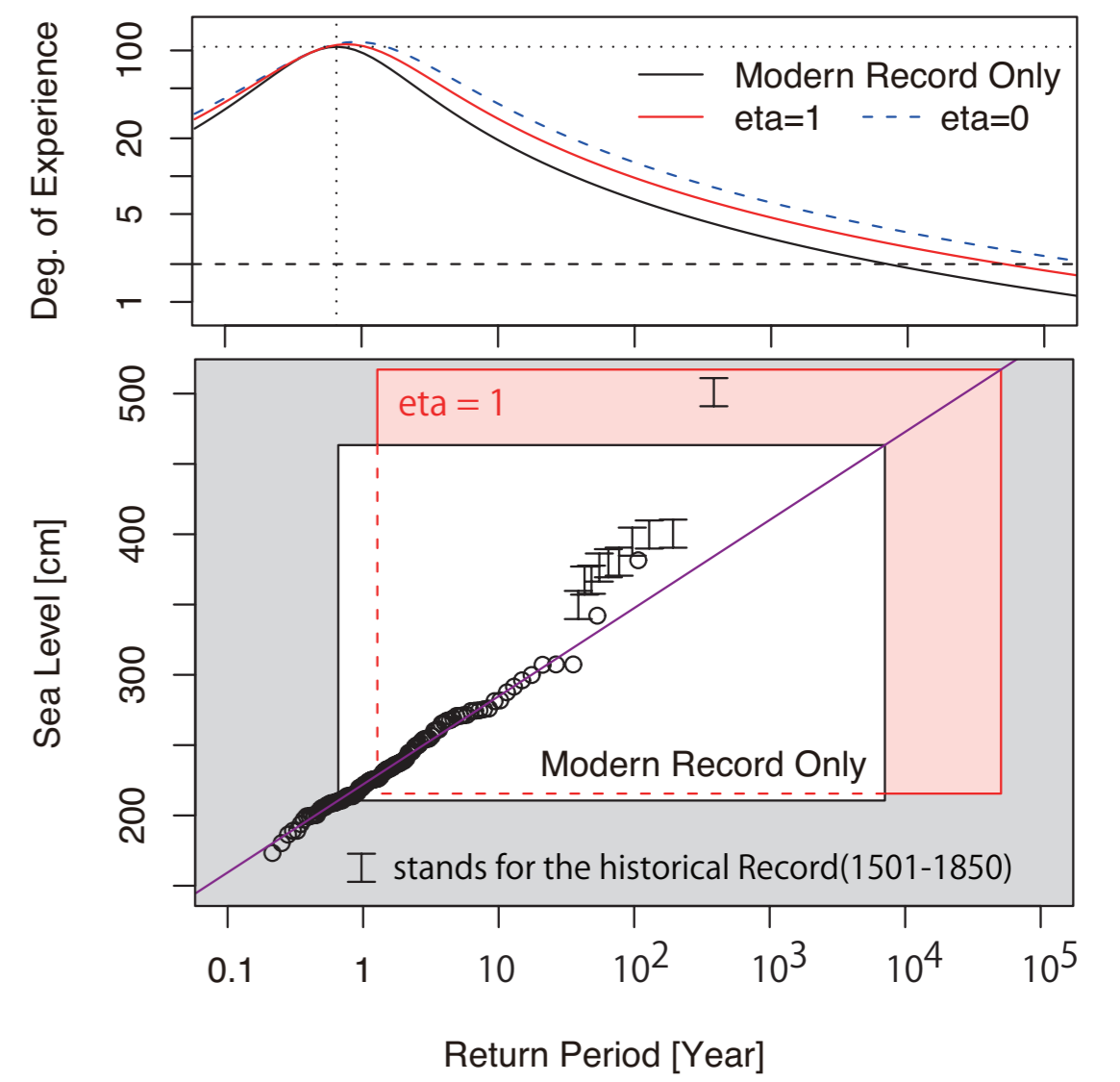


図-2 ガンベル確率紙における窓枠

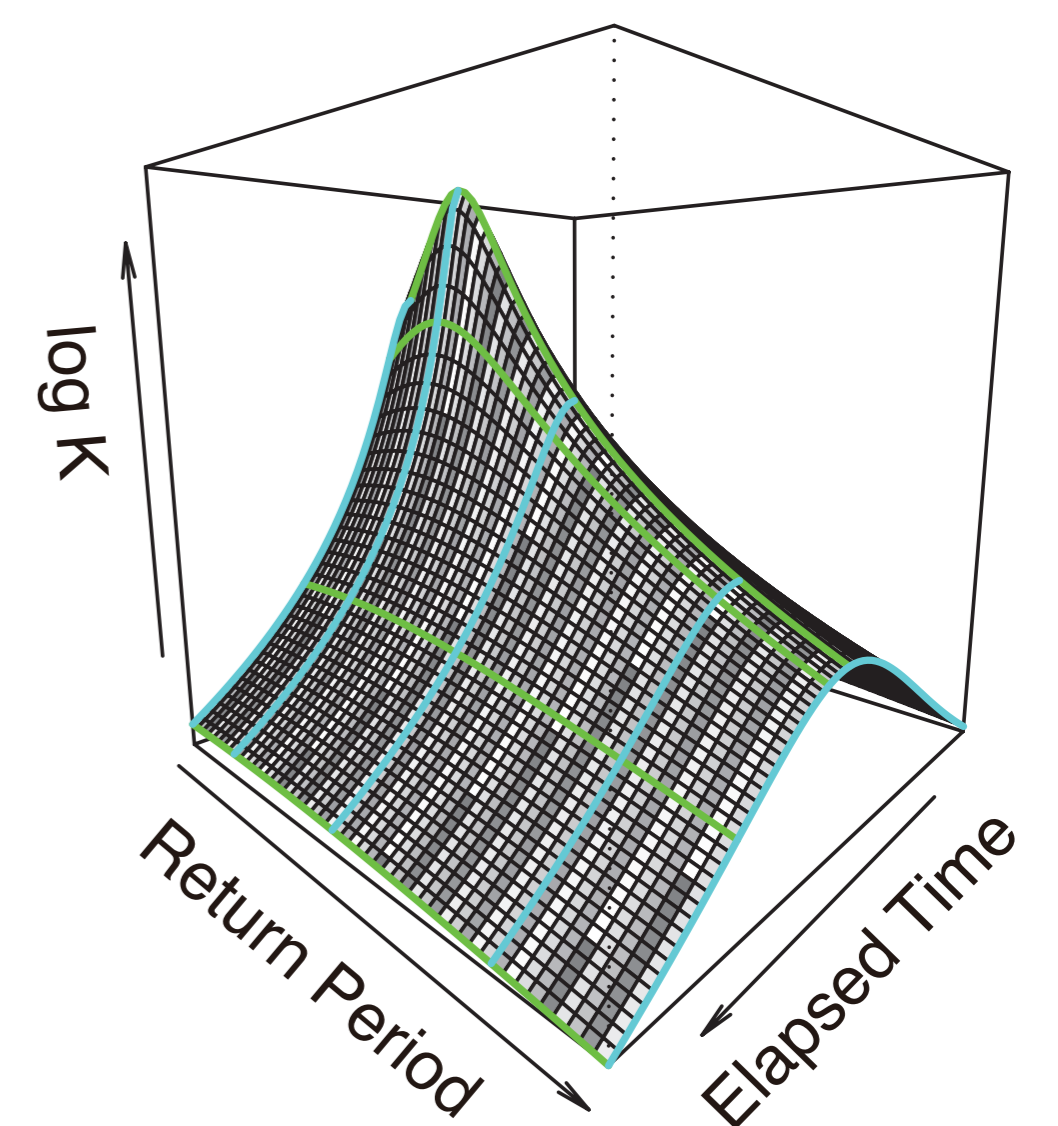


図-3 2種類の時間による外挿

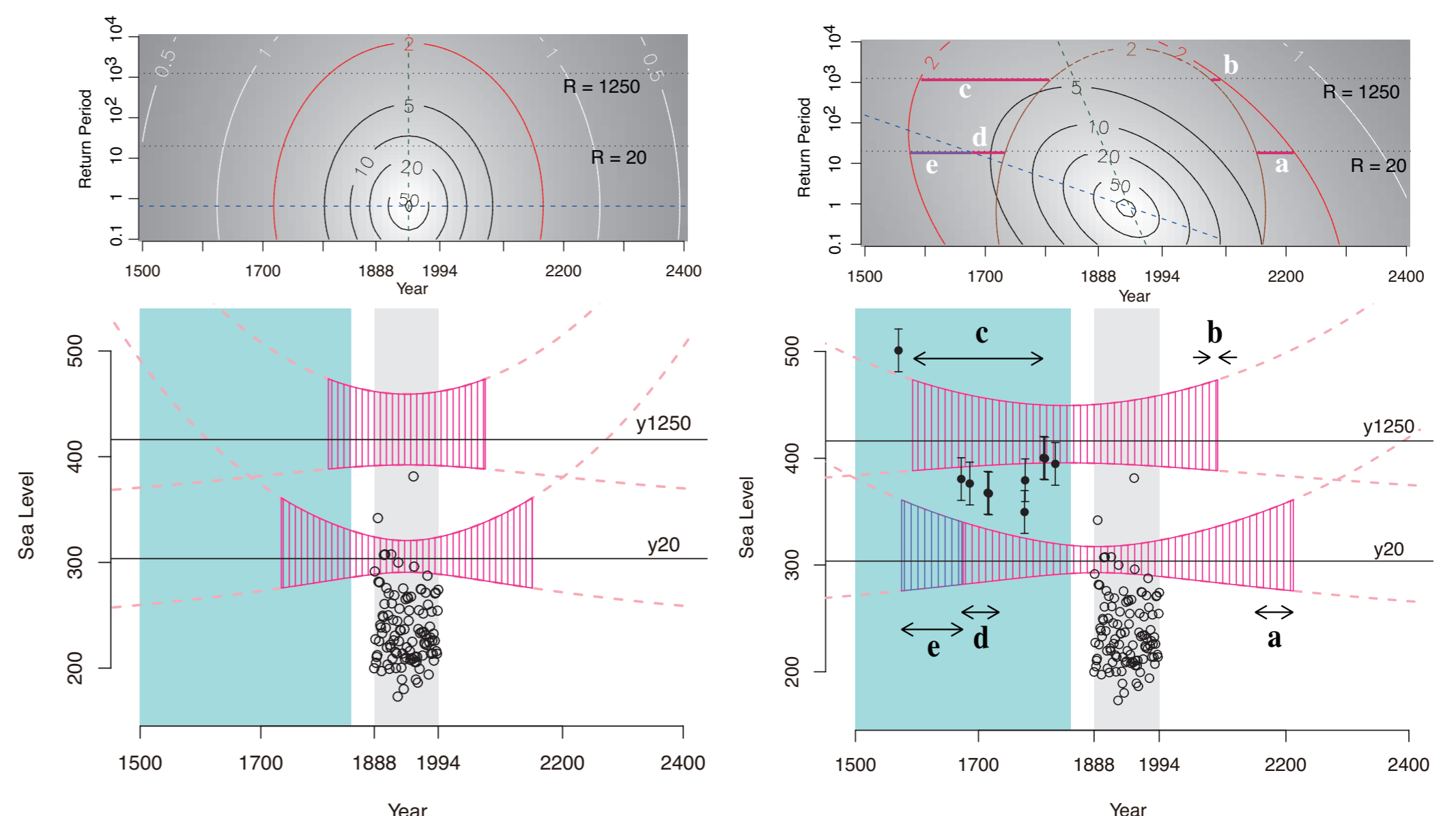


図-4 データの不均質・非対称性による推定誤差の回折効果の非対称化