

地震予測の評価法について

庄 建倉¹・尾形 良彦²

(受付 2015 年 1 月 1 日；改訂 4 月 6 日；採択 4 月 8 日)

要 旨

地震予測可能性の研究には予測結果を客観的に評価できることが必要不可欠である。滅多に起こらない大地震と小さな地震ではその発生確率が極めて違い、予測が当たった場合の評価も大きく違う。地震活動度による地域差も評価に関わる。先ず各地域に適合した経験的な地震活動による基準予測を確立する必要がある。確率予測の成績を測るものとして情報利得が合理的である。新しい予測モデルが提案されれば、基準のモデルの予測と比較して、予測能力がどの程度向上しているか否かの評価ができる。赤池情報量規準 AIC や ABIC は提案モデルの将来の予測の良さを予め現在のデータのみで推定する評価スコアとして有用である。予測のアルゴリズムや予測の経験が発展途上であるため、確率の数値予測を出すことが難しい場合がある。したがって、その多くは警告型の予測(二値予測)である。本稿では更に警告型地震予測を評価するためのグラフ法やギャンブルスコア法を解説する。これにも地震の大きさや活動度の経験的な基準確率(相場)の設定が不可欠である。相場に基づいた公平な賭けのもと、警告型予測の成功または失敗の結果の得失スコアを比較する評価法である。経験的基準確率としては、ゲーテンベルグ・リヒター則(指数分布)を地震の大きさの出現頻度とする。時間・空間に一樣な地震発生モデル(定常ポアソン過程)が考えられている。しかし、より現実的に、地震活動度に合わせた空間非一樣なポアソン過程や地震の連鎖過程などを基準モデルに設定すると、現状の警告型予測が基準予測より良い評価を得るのは難しくなる。

キーワード：地震確率予測，基準予測，警告型地震予測，情報利得，ギャンブルスコア評価，赤池情報量規準。

1. はじめに

2009 年 4 月のイタリア中部ラクイラ地震(M6.3)は地震予知や地震情報伝達のあり方で世界の地震関係者に衝撃を与えた。その前の数か月来の群発地震で、独自に地震予知情報を出す人も現れた。しかし、ごくたまに大地震の予知が当たったという話もあるが、失敗したケースは山ほどあり、ラクイラのように、権威筋の予測失敗を除けば、ニュースにならない。このように通常の状態と違う異常現象に遭遇したり観測したりしたときの地震予測の公開は責任関係機関においては極めて慎重だが、他方で独自の観測により地震予知情報を出す個人や団体が世界中に増えてきている。それゆえ一連の予測が公表されればそれらの予測結果の成績を評価できるインフラを整備する必要がある。

地震発生の仕組みの研究は進んでいるが、確度の高い予知は現状では難しい。これは地殻内

¹ 統計数理研究所：〒 190-8562 東京都立川市緑町 10-3

² 統計数理研究所 名誉教授：〒 190-8562 東京都立川市緑町 10-3

部の断層やストレスが直接的に見えないうえ、その仕組みが複雑で地域的に多様だからである。しかし地震の発生は全くの無秩序ではなく、統計的な経験法則にもとづいて確率的な予測はある程度可能である。このことについては本特集号の尾形(2015)を参照されたい。

地震予知の特効薬探しではなく組織的に着実な予測研究を進展させるべきだという機運が高まってきた(e.g., Jordan, 2006)。そこで地震予測可能性を探る国際的共同研究である Collaboratory for the Study of Earthquake Predictability(CSEP, 2014)が主要地震国の研究機関で連携して進められている。これは一方で地震活動の統計的モデルの開発を促し、他方では予測の性能を客観的に評価することを当面の目標とする。それは、地震活動、地殻変動や電磁気変動などの様々な観測異常による各種の地震予測法の有意性と予測性能を評価できる科学的なインフラ(共通基盤)を整備することでもある。

実際これまで少なからず各種異常現象に基づいた大地震予知手法が提案されているが、それらの有効性をめぐっての論争は絶えず、評価の定まったものは無いといわれている(Jordan et al., 2011)。したがって客観的に予測能力を評価する基盤が必要であり、これが無かったら論争は不毛なものに終始する。CSEP プロジェクトは先ず、標準の確率予測を与えるために、世界の各地域に適合した基準の地震活動モデルの成立と、それらの改訂を進めようとするのである。その際、予測の成績を測るものとして情報利得(information gain, 本稿 4 節参照)が合理的なものと考えられている。もし有用な知見が組み込まれた新予測モデルが出てくれば、基準モデルと比較して、予測力が向上したか否かの評価ができる。この様に地震予測モデルは進化すべきものである。

本稿では、地震統計学的観点に立って、各種モデルの予測結果からそれぞれの性能を評価する方法について紹介する。なお、本稿に引用されていない関係文献も網羅した参考解説論文として Zechar (2010)を挙げておく。

2. 警告型予測の評価ダイアグラム

警告型予測が対象とする地震は大地震である。時折メディアを通じて公衆の注目を引く地震予測が散見される。これらの予測は多くの場合、確率ではなく、Yes/No 型の二値予測である。その様な警告型予測については予測結果について不偏で公正な評価方法が必要である。事後に予測性能の評価ができるためには、事前に、1)発生領域、2)期間、および3)マグニチュード範囲を、曖昧さを排して定量的に定める必要がある。さもないと評価に議論の余地が生ずる(e.g., Kagan, 1996; Zechar and Zhuang, 2010)。警告型地震予測は、これらの範囲を明示していなければ客観的に評価できない。大地震は起きるか否かの予測として「警告予測」または「反予測(安心予測)」を時間・空間・マグニチュードの範囲を附して公表する。これら以外は「無予測(分からない)」となる。警告のレベルを与える予測もあろうが、警告のレベルは確率のクラスに対応していると考えられるので、これは 4 節以降に述べる確率予測に含めることができる。反予測の例は社会的インパクトが無いので稀であるが、先に述べたラクイラ地震の場合や、今村恒明の関東地震予測に対する大森房吉の反予測などは有名である。また実際には、何等の警告予測無しに大地震が起きる場合が多いのであるが、このことについての警告型予測の実績(予知率; 宇津, 1977)は本稿では考えていない。あくまで予測されたものについての評価(的中率; 宇津, 1977)のみを考える。平均的に活動度の高い時期と領域に対して、警告期間と警告領域を大きくとれば予測成功の率は高くなる。そこで、警告を宣言した期間と領域の面積と活動度を掛けた地震確率を横軸にとって、縦軸にそのような予測の失敗率をプロットした評価ダイアグラムを作る。評価ダイアグラムを比較するために縦軸と横軸のスケールは規格化する。これは Molchan ダイアグラム(Molchan, 1991; Molchan and Kagan, 1992)と呼ばれる。そして、同じ警告の設定

状況のもと、標準的な地震活動に基づいてランダムな大地震をシミュレーションで発生させたダイアグラムを幾つも作れば、警告実績が有意であるか、単にまぐれ当たりだったかを可視化できるというものである。

例えば、ロシアグループの M8 アルゴリズム (Keilis-Borok and Malinovskaya, 1964; Kosobokov, 1997) による地震活動の異常現象に基づく大地震の予測が電子メールで定期的に世界の地震学者に送り付けられている。或る Molchan ダイアグラムによれば、成績は有意であると言われているが、警告に対する的中率は標準的な地震発生確率の高々数倍といわれている (Jordan et al., 2011)。しかも、この M8 アルゴリズムを検証したダイアグラムの横軸は、当該地域内の地震活動度の違いを無視して、警告領域の面積や警告時間幅に比例した量にとり、有意性を調べる標準的な地震活動のシミュレーションは一様分布に基づいている。実際の通常地震活動度に比べた成績を評価するには、原論文 (Molchan, 1991) にあるように、少なくとも警告領域内の地震活動度に合わせた、非一様な空間分布で大地震をシミュレーションすべきである。

Wang et al. (2013) は南カルフォルニア、日本 (図 1(a)), ニューゼaland における GPS の観測点間の基線距離の日別変化の時系列を解析している。10 日間のウィンドウ内での基線の伸縮率の最大変動幅に関する異常を定義し、これらの異常をもって周辺部での中規模地震を一定期間警告し、実際にその期間に地震が発生したか否かを Molchan ダイアグラムを使って示している。異常が発現した直後から一定期間に亘って地震の警告予測をするのだが、警告の期間を長く取れば取るほど、その様な予測が当たる確率は高くなるのが道理である。したがって、横軸に警告期間の長さを取り、縦軸に予測が外れた場合の割合を取った Molchan ダイアグラムに

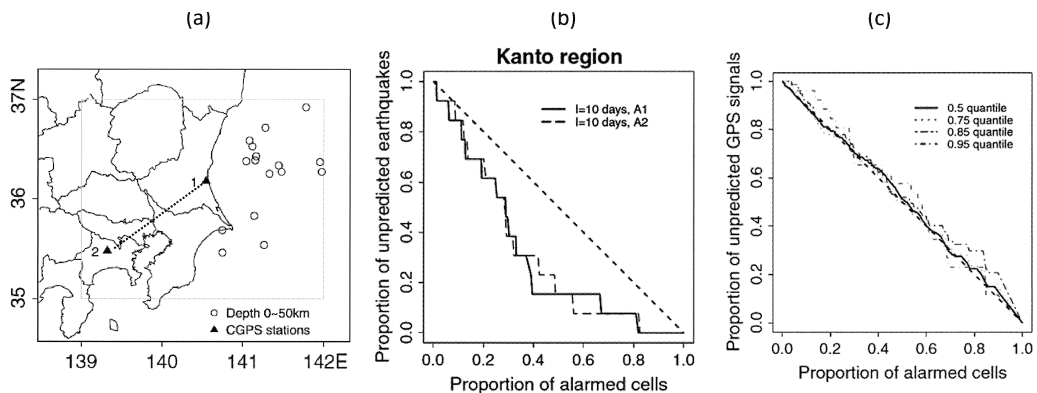


図 1. (a)GPS 観測点、鉾田-厚木間の基線と 1997 年から 2010 年までの $M \geq 5.5$ の地震の震央。(b)GPS 変動異常による地震に対する警告型予測期間の長さとして予測失敗率に関する Molchan ダイアグラム。(c)地震の発生から GPS 基線距離の時系列変動異常 (異常値レベルは図中で示している) を予測に関する Molchan ダイアグラム。

Fig. 1. (a) Earthquakes (16 in total) from the JMA catalog, with magnitudes ≥ 5.5 , during the period from 1 January 1997 to 31 December 2010. The gray segment is baseline distance between GPS stations 1 (Hokota) and 2 (Atsugi), daily time series of which is used to define anomalous variability (Wang et al., 2013). (b) Results from using the Molchan diagram to examine the alarm-level type predictions made based on GPS anomalies. (c) Results from using the Molchan diagram to examine the predictions of anomalous variability (with the different levels of variability) of the time series after the earthquakes against the relative length of the warning period (Wang et al., 2013).

なる。伸縮率異常と中規模地震の発生に因果関係が無い場合、予測の当否が一様にランダムとなり、ダイアグラムでは逆対角線上に並ぶが、因果性が見られれば図 1(b) のように下に凸になり、その漸近的な傾きから確率利得 (Aki, 1981) が推定され、この場合はほぼ 2 となる。さらに、地震が基線距離の時系列の異常伸縮変化を引き起こしたか否かを Molchan ダイアグラムで調べたのが図 1(c) であり、逆向きの因果関係は見られない。

また Wang and Bebbington (2013) は同様な異常性の有無と発生した地震の有無の件数を分割表にして有意性を示唆している。予測量 (確率など) と予測結果をカテゴリに分割した分割表のカイ 2 乗検定や対数尤度による有意性を測るやり方は良く使われる。

測地計測の時系列には、GPS の測定誤差だけでなく、周辺地域の中小地震や遠地の大規模地震の地震時やその前後のすべり変動が混ざり合っている。以上の結果は、それらと異常事象を分離判別する統計モデルや解析法を開発する余地があることを示唆している。

3. ギャンブルスコアによる評価

警告型予測に対して、それが予測範囲に実際に起きたか否かの一連の結果が整っている場合、以下に述べるギャンブルスコアによる評価法が適用可能である。ギャンブルスコアについて数学的な定義は Zhuang (2010) で述べられているが、ここでは話を分かり易くするために競馬を例にとって説明する。簡単のために胴元 (主催者) の取り分は無いことにする。競馬の実際に近い評価方式については Zechar and Zhuang (2014) を参照されたい。

競馬の常連客や予想屋は、下馬評や経験に基づく競争馬の状態や実績などから、レースの予測をする。彼らの成績を評価する原理は簡単である。彼らが予測する勝ち馬にある金額を賭けたとして、多くのレースでの損得額の合計を総掛金で割った量の多少で比べるのである。レース前にある馬が勝つ確率 p_0 とする。言うまでもなく、レース前の確率の高い本命馬と低い穴馬では予測の当たり外れに対する評価が大きく違う。これはオッズ $p_0/(1-p_0)$ に反比例して決まる配当金が大いに違うことから分かる。オッズは観客に買われた馬券の数によって計算される各馬の競走力の相場である。

同様に、滅多に起こらない大地震と頻繁に起こる小さな地震ではその発生予測が当たった場合の評価が大きく異なる。しかも、それは地域によっても違う。したがって公平な評価のためには、先ず地震発生の予測領域 (時間幅, 地域, マグニチュード幅) での基準確率 p_0 をどのように見積もるかが大事である。また反予測 (安心予測) をすることも考えられるが、この場合に対する基準確率は $1-p_0$ となる。評価では予測ごとに予測者が「評価スコア」 (reputation score) 1 単位を賭けたと考える。予測が当たれば配当率 $(1-p_0)/p_0$ 単位の評価スコアを受けるが、外れれば予測ごとに 1 単位の評価スコアを失うこととする。

評価の一例として RTP (Reverse Tracing of Precursors) アルゴリズム (Shebalin et al., 2004) を取り上げる。このアルゴリズムは世界各地の地震活動の RTP 異常パターンを基に決めた領域内での最長 9 カ月間の期間の大～中規模地震の予知を目指した警告型予測の一種である。Zechar and Zhuang (2010) は、Shebalin らが実施した警告予測の全記録と実際に警告通りの地震が発生したか否かを厳密に検討し、地震活動とマグニチュード分布に基づいた簡単な基準モデル (予測領域面積と時間幅に比例したポアソン型発生率と G-R 則に基づく) から計算した警告範囲の各対象地震の基準確率を推定し、RTP アルゴリズムの予測の結果をギャンブルスコアで評価した。予測者が成功したとしている或るケースに予測内容の曖昧さがあったので、厳格な評価と予測者の言い分を取り入れた寛大な評価の両方を試みた。結果は後者の寛大な評価の場合の RTP 予測なら 29 事例中 6 事例の中で、評価スコアの総計は +84.42 単位、予測一件あたりのスコア (ギャンブルスコア) は 2.9 単位である。これは有意である。しかし厳格な評価の予測結果

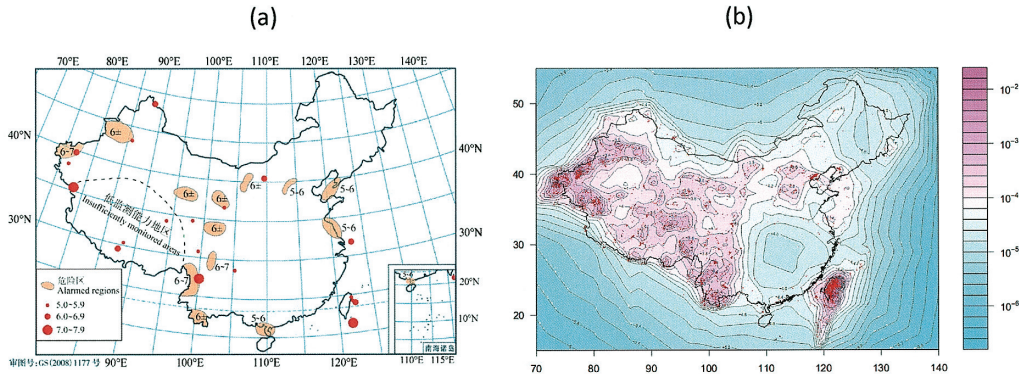


図 2. (a) 中国地震局年間地震予測 1996 年版. 警報地域は肌色で塗られていて、傍に予測マグニチュードの範囲が 0.5 誤差幅で書かれている. 同年に実際に起きたマグニチュード 5 以上の地震が赤丸で表示されている. (b) 空間非一様ポアソンの強度関数 $\lambda_0(x, y, M)$ で推定された中国各地の地震発生率.

Fig. 2. (a) Annual Earthquake Predictions in 1996 made by China Earthquake Administration. The alarmed regions are marked in yellow and the earthquakes with magnitudes of 5 and above are represented by red dots and disks. The numbers on the alarmed regions are the magnitude ranges (± 0.5) of the expected future earthquakes. (b) The inhomogeneous Poisson rate $\lambda_0(x, y, M)$ for the seismicity in China.

は、26 事例中 3 事例の中で評価スコアの総計は -4.15 単位、ギャンブルスコアは -0.16 単位である。これは、かなりの確度で、RTP アルゴリズムによる予測は基準確率より悪いことを意味している。結局これらの評価例の意味するところは、予測について、発生領域、期間、マグニチュード範囲の事前の厳密な設定が必要であることに帰する。

一般的に地震活動が高い地域では低い地域より予測が当たり易い。また仮に基準確率通りの予測をすれば、そのあたりはずれは平均 0 の評価スコアとなる。それゆえ適切な相場の地震活動を基準にして予測結果を評価することが大事である。多くの場合、簡便性から、一様な発生率の下で面積や期間に比例した発生確率を基準モデル(定常ポアソン過程)とすることがとられ、日本周辺は世界の約 2 割が発生する地域なので、ロシアグループの M8 アルゴリズムは日本周辺での予測を出すことが多い。しかし例えば日本周辺の常時地震活動度(尾形, 2015 の図 3(a)参照)を基準確率と考えると、実際その後 15 年間の大地震の大勢が常時地震活動度の極めて高いところで起きている。日本およびその周辺についての警告予測は、これに基づいて評価をする方が妥当である。

このような地震活動度の地域性のモデルを適用した中国における事例 (Zhuang and Jiang, 2012a, 2012b) を紹介する。中国地震局は毎年、大地震予測について機関内部で公式に記録して、一定の年数が経てば予測内容の情報を公表している (Center for Analysis and Prediction of the State Seismological Bureau (中国地震局), 1990–2003)。図 2(a) は 1996 年の地方の各省の地震局が出したものを纏めた大地震の予測である。この際重要なことは、予測結果を評価するときに基準確率 p_0 を中国各地の地震活動度に基づいて算出する必要がある。基準確率は当該予測地域の地震活動度と地震のマグニチュード頻度分布のゲーテンベルグ・リヒター則から計算される(尾形, 2015 の 4 節参照)。実際、予測地域内の発生確率を過去の地震データを使って推定した空間非一様ポアソン過程の地震活動度 $\lambda(x, y)$ (図 2(b)) と G-R 則とから基準の空間・マグニチュー

下の強度関数 $\lambda_0(x, y, M) = \lambda(x, y) \cdot 10^{-b(M-M_c)}$ を得る. 震源域 A で一年間にマグニチュード M_0 以上の地震が発生する基準の確率は

$$p_0(A) = \iint_A \int_{M_0}^{\infty} \lambda_0(x, y, M) dx dy dM$$

によって計算される. さらに, 基準モデルの強度関数 $\lambda_0(x, y, M)$ を使って中国全土の大地震の確率シミュレーションを繰り返し行い, 得られた予測の成績のギャンブルスコアが如何に偶然的に得られ難いものであるか否かの評価ができる. これは予測が基準の統計モデルでは得難い何らかの有用な予測情報に基づいた結果であるかも知れないことを意味し, その理由を検討するに値する.

総じて, 14 年間で各年の評価は大きく異なり, 毎年 10 件程度の警告予測があるが, そのうち予測通りに大地震が起きなかったケースは全体の予測件数の 8 割以上で, 予測失敗の件数自体は少なくない. しかし予測が当ればオッズ倍の評価スコアの配当を得る. 実際は 133 件の予測に対して総計 299.83 単位の評価スコア, 予測一回当たり平均約 2.25 倍 (299.83/133) の評価スコアを得ているということになる. これらの評価スコアの不確定性 (uncertainty) についての検討については Zhuang and Jiang (2012b) を参照されたい.

4. 確率予測の評価規準としての対数尤度の意味

異常事象データと地震データが十分豊富になってくると, 異常事象の出現と実際の地震発生 of データから, 異常事象の出現による確率予測のモデル作成とその評価が可能になる (宇津, 1979; Aki, 1981; 尾形, 2015). その様な予測評価を促進する目的で, 確率予測の実績を試すための基盤を整える国際プロジェクト CSEP (Collaboratory for the Study of Earthquake Predictability, <http://www.cseptesting.org/>; 例えば Jordan, 2006) が 2006 年から進行中である. これは世界の各地域に適合した地震活動モデルの成立を進めようとするものである. CSEP プロジェクトは元来カリフォルニアでの 5 年間の予測を競った RELM プロジェクト (Regional Likelihood Models, 2007) を引き継いだものであり, 日本周辺の予測に関しては CSEP 日本実験センターが担当している (楠城 他, 2008; 平田 他, 2011; Nanjo et al., 2011).

例えば CSEP 日本の 3 カ月予測には 50 近い確率予測アルゴリズムの応募がある. これらでは 0.1 度刻みの空間メッシュ (図 3) にマグニチュード値を合わせた 3 次元のビン (bin) に 3 カ月間で M_5 以上の地震が起こる確率を当てはめてもらう. かくして, すべてのビンの予測確率が, 実際起こった地震に対して, どうだったかということの評価検証するのである. そのような確率予測の成績を測る規準として, 対数尤度の比較に基づく尤度比統計量 (情報利得) など様々な検定が採用されている (Schorlemmer et al., 2010; Zechar et al., 2013). 以下では, 評価基準の主概念である対数尤度がなぜ確率予測の評価として合理的であるのか, 初歩的な例を挙げて説明する.

応募された確率予測の評価の原理は以下のようなものである. 先ず, 空間・マグニチュードのビンを図 3 の表のように $i = 1, 2, \dots, N$ と並べたときに, i 番目のビンの予測確率が p_i であるとする. そしてそれぞれのビン内で実際に地震が m_i 個起きたとする. それらの正規化した頻度 (相対頻度) は $\nu_i = m_i/N$ である. これを比べるために, 予測確率 $\{p_i, i = 1, \dots, N\}$ が実際に起こった地震の相対頻度にどれほど近いかを評価する. Boltzmann (1878) はこのような確率予測の評価 P を次のように計算している.

$$P = \frac{n!}{m_1! m_2! \dots m_n!} p_1^{m_1} p_2^{m_2} \dots p_n^{m_n} \approx e^{N \sum_i \nu_i \ln \frac{p_i}{\nu_i}} = e^{N \sum_i \nu_i \ln p_i - N \sum_i \nu_i \ln \nu_i}$$

この多項分布の階乗をスターリングの式で近似したときに, 上記の指数項の第一項が対数尤度に

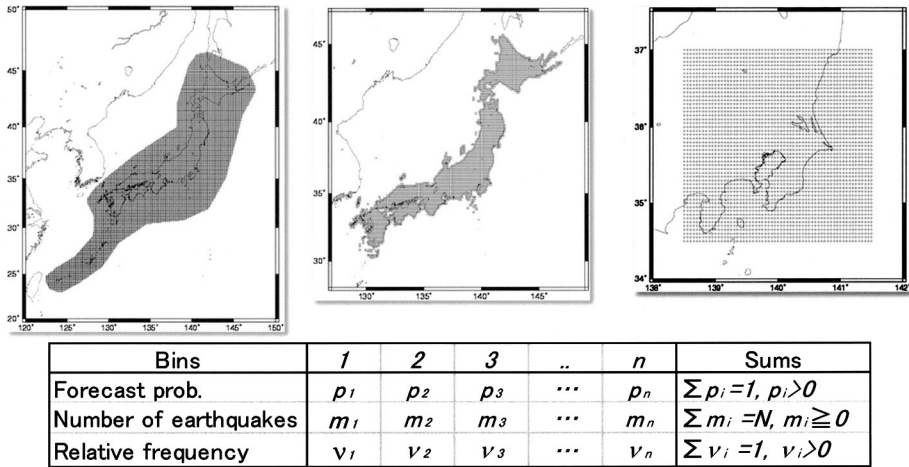


図 3. CSEP の日本版実験センターの 3 カ月間ごとの確率予測(全日本, 内陸, 関東)に関する取り決め. 空間 0.1 度平方刻みのピクセル上にマグニチュード 0.1 毎に定義されたビンに予測確率を与える. 空間・マグニチュードのビンをも $i = 1, 2, \dots, N$ と並べたときに, ビン i の確率 p_i が CSEP の予測確率であるとする. そして実際に起こった地震がそれぞれのビンに対して m_i 個起きたとする. それらの相対頻度は $v_i = m_i/N$ である.

Fig.3. Grids used in the 3-month forecasting protocols of the CSEP Japan Testing Center for issuing earthquake probabilities, which are the discrete bins sized $(0.1 \times 0.1 \text{ deg}^2)$ in space and 0.1 in magnitude. The outcomes of the bins are shown in Table 1, including the forecast probabilities, outcomes of observations and their relative frequencies.

なっている. ビンを極限的に小さくし連続形の分布にした場合を含めて, これはよく知られているボルツマンの相対エントロピーと呼ばれるものである. 相対的エントロピーを予測の評価に使う例は Ogata (1995), Kagan and Jackson (1995), Ogata et al. (1996), Vere-Jones (1999), Harte and Vere-Jones (2005), Schorlemmer et al. (2010), Zechar et al. (2010), Nanjo et al. (2012), Ogata et al. (2013) などがある.

上記 P そのものは尤度ということになり, これが予測の出現率を表す量である. ただし, 中規模地震の発生率を考えると, 上記で定められたビンの体積サイズが小さく, 殆どの m_i は 0 個で, 残りには地震の活動度に見合って 1 個が見られ, そして 2 個以上は珍しい. 因みに「点過程」(尾形, 2015 参照)と呼ばれる確率過程はこう云った確率論の特徴を有する.

未来のデータ y を予測するとき, 予測者は現在までのデータ X を使って確率予測モデル $P_k(y|X)$ を構成し, CSEP でとり決められたピクセルに確率値を埋めるものとして CSEP 実験センターに預けられる. 実験センターは確率予測値を各ピクセルにおけるポアソン分布の平均値を与えたものと見做し未来の実現データ Y (ピクセル内の地震数) を使って各モデル k の対数尤度 $\log P_k(Y|X)$ を比べる (Schorlemmer et al., 2010; Zechar et al., 2010, 2013). たとえば表 1 は CSEP 実験センターに提出された或る 3 カ月間の予測を対数尤度の大きい順に並べたものである. 10 番目の予測は日本周辺全領域で地震が等確率で起こる, すなわち空間一様に起こるといふモデルである. それを基準にして比べると対数尤度値が表 1 のように違ってきている. 対数尤度値を尤度値に直すと, それは一番良かったモデルに比べて他のモデルが相対的にどのく

表 1. CSEP 日本の 3 カ月予測評価の一例の対数尤度, 尤度, 相対尤度および正規化尤度. 10 番目のモデルは標準 G-R 分布で日本域全体一様の予測.

Table 1. An example of log-likelihood and likelihood scores of 3-month forecasts for $M \geq 5$ earthquakes in the Japan region. The scores are listed in the descending order. Model #10 assumes spatially uniform distributed seismicity over the whole Japan region with a standard G-R magnitude frequency.

	対数尤度	尤度	相対尤度	正規化尤度
Models	$\Delta \ln L$	Likelihood	Likelihood0	density
1	325.82	3.1756E+141	1	0.521558
2	325.39	2.0658E+141	0.6505091	0.339278
3	324.29	6.8763E+140	0.2165357	0.112936
4	322.83	1.5969E+140	0.0502874	0.026228
5	282.07	3.1728E+122	9.991E-20	5.21E-20
6	268.16	2.8867E+116	9.09E-26	4.74E-26
7	247.61	3.4329E+107	1.081E-34	5.64E-35
8	252.67	5.4099E+109	1.704E-32	8.89E-33
9	229.10	3.1395E+99	9.886E-43	5.16E-43
10	0.00	1	1.85E-110	9.6E-111
sum		6.0887E+141	1.9173322	1

らの確率の比で実現性があるかを見ていることになる. この例で言うと 4 番目くらいまでが確率的に有意であり, 4 番目以下はかなり小さな確率になってしまつて予測の実現性が相当悪いということの意味する.

この他に時間経過とともに地震確率がどう変化するかを見るために時間軸も入れた日別予測確率を実際起こった地震を通して検証する, 日ごとの予測(日別予測)もあり, それらの予測実験が世界各地(カリフォルニア, 日本, イタリア, ニュージーランド)で進行中である. 例えば, 東北沖地震前後に限った期間での様々な時空間 ETAS 型の予測についての結果が, CSEP で採用されている評価法の問題点と合わせて, 議論されている(Nanjo et al., 2012; Ogata et al., 2013).

かなり多様な統計的及び物理的モデルによる予測の評価に関して, 1992 年の M7.3 カリフォルニアのランダース地震の余震活動の時空間の予測性能を遡及的に比べている論文(Woessner et al., 2011)がある. これは Epidemic-type Aftershock Sequence(ETAS)モデルと Short-Term Earthquake Probability(STEP)モデルと Coulomb-Rate-and-State(CRS)モデル, それぞれのモデルの予測性能を比べたものである. STEP モデルは原理的に大きい地震を起点にして大森・宇津の減衰式を重ね合わせたものによる予測であり, CRS モデルは原理的にクーロン破壊ストレス(Coulomb Failure Stress)の変化とすべり速度と状態に依存する摩擦法則(Rate and state dependent friction law)に基づく物理モデル(Dieterich, 1994)による予測である. これらのモデルの対数尤度値が図や表として並べてあつてランクも付けられているけれども, 大事なものは, 部分的なデータにおける予測モデルの尤度の良し悪しから導かれる, 予測モデルの診断内容である. 時空間予測の評価から示唆されるモデルの短所はモデル改善の参考になる. たとえば, この論文の空間分布の検証の対数尤度(S-test; Zechar, 2010)の日別のスコア列を見ると, どのモデルも本震後最初のうちは大変に成績が悪い. 何故かという, ランダース地震の本震が引き起こした(トリガーした)初期の余震群の空間分布の形状が, どのモデルでも, うまく捉えられていないからである.

CSEP の評価は、上記の各種テスト評価量についての不確定性、有意性について、大量のシミュレーションで比較検討している。これらについては Schorlemmer et al. (2010), Zechar (2010), Zechar et al. (2010, 2013), Rhoades et al. (2011) を参照されたい。また Vere-Jones の応力解放モデル, ETAS モデル, 各種更新過程といった確率モデルを仮定して定常ポアソン過程に対する情報量利得の理論的な評価が研究されているが、これらについては Vere-Jones (1999), Imoto (2004), Kagan (2007) などを参照されたい。

ただし、以上のような CSEP で試されている評価は大中小の地震総体の予測に対する評価になっている。もし大地震のみの確率予測の評価について論じるとなると、主にマグニチュード頻度分布モデルに基づいて予測することになる。しかし前出の ETAS, STEP, CRS などの時間依存性のモデルは、マグニチュード頻度分布についてはゲーテンベルグ・リヒター則によるので、これは独立の同じ指数分布である。2 節と 3 節で議論した警告型予測のように、過去の地震異常履歴に依存したり、何らかの前兆的な事象かも知れない異常データに基づいて変化するようなマグニチュードの予測分布は考えられてない。これらの予測こそが究極的には一般社会から期待される確率予測に求められているものであろう。

CSEP の評価のためのプロトコルには則っていないけれど、一連の地震が予め大きな地震の前震と疑われる確率(前震確率)を予測することは、そのような試みの一つと考えられる。前震確率を予測するモデルの概略は本特集号の尾形論文で解説されているが、気象庁震源カタログの 1993 年までのデータから推定されたモデルを使い、1994 年から 2009 年までの前震確率を予測した。これによると、地震の群 i の前震確率 p_i は、Ogata and Katsura (2012) にあるように 0~40% の値をとり、これは実際の前震の出現頻度と良く合っていた。さらに予測結果は対数尤度比に基づき次のように評価検証できる。第一に、確率変数 η_i は地震群 i が結果的に前震型であれば 1、さもなければ 0 という値をとり、 p_i は予測確率である。前震の予測確率列 $\{p_i; i = 1, 2, \dots\}$ を学習期間での平均前震確率 ($p_0 = 6.8\%$) による予測と比べた。ここで、 $\{\eta_i\}$ の独立性を仮定して、各予測の情報量利得(相対エントロピー)の累積和(対数尤度比スコア)

$$\ln \frac{L_1}{L_0} = \sum_i \left\{ \eta_i \ln \frac{p_i}{p_0} + (1 - \eta_i) \ln \frac{(1 - p_i)}{1 - p_0} \right\}$$

が予測列全体の評価となる。累積相対エントロピーの値は 15.2 となり予測は前震の経験的平均値の確率(6.8%)で予測した場合より大きく、尤度で比べると 4×10^6 倍も予測実現の性能が優れている。

さらに予測の確率を分割(categorize)して各確率区間の頻度と前震群の出現と非出現の頻度の分割表(クロス表)の従属性を調べることができる(表 2 参照)。クロス表の独立性モデルに対する非独立性モデルの AIC の差は -21.47 であった。これも極めて有意である。なお時空間 ETAS モデルに前震確率を加味したマグニチュード予測を併せれば CSEP のプロトコルに則るので、このような時空間マグニチュードの予測の評価が可能である(Ogata and Katsura, 2014)。

5. 確率予測評価量の推定量としての AIC と ABIC

前節では現在までのデータから将来を予測し、時間経過とともに検証するデータが得られた場合の対数尤度による評価法を述べた。すなわち、確率モデル族 $P(y|\theta)$ で現在までのデータ \mathbf{X} を使って最尤推定値や最大事後分布推定値などの最適なパラメータ値 $\hat{\theta}(\mathbf{X})$ を求め、それを代入した予測(プラグイン予測)の確率 $P(y|\hat{\theta}(\mathbf{X}))$ にその後時間が経過して得られたデータ \mathbf{Y} を代入した対数尤度の値 $\log P(\mathbf{Y}|\hat{\theta}(\mathbf{X}))$ によって予測性能を評価するのであった。

しかし、例えば大地震は発生頻度が非常に低く、相当長い期間待たないと検証のデータが得られない稀有な事象ほど通常の検証期間ではなかなか起きない。このため或るモデルで大地震の発

表 2. クラスターが複数の地震を含むとき、それらが前震であるという予測確率に対する実際の結果の分割表. 表の各区間における実際の前震の比率は、予測確率と良く対応している. 889 回の予測実験の結果、いつも日本全国均一の 7.9% の確率で予測した場合に比べて AIC の改善が -21.47 の予測であり、ほぼ $\exp(21.47/2)$ 倍優れた予測パフォーマンスであることを示している.

Table 2. Contingency table of observational outcomes for the clusters with multiple earthquake members. The table shows that the ratios of outcomes in each interval are consistent with the corresponding forecast probabilities. The AIC improvement is -21.47 for 889 forecasts, which indicates that actual frequency of the foreshock-type clusters is substantially dependent on the forecast probabilities.

Forecast	0–2.5 per cent	2.5–5 per cent	5–10 per cent	10–15 per cent	15 per cent	All
Foreshocks	4	10	30	12	14	70
Others	179	211	263	115	51	819
All types	183	221	293	127	65	889
Ratio (per cent)	2.2	4.5	10.2	9.4	21.5	7.9

生がどの程度良く予測できるかを評価することは難しい。ではどうするか。例えば、未来の実現データ \mathbf{Y} の代わりに現在までのデータ \mathbf{X} を代入した評価量 $\log P(\mathbf{X}|\hat{\theta}(\mathbf{X}))$ を使うことが考えられる。しかし、これは予測モデルの調節と検証の評価を同じデータ \mathbf{X} で行うため、 $\log P(\mathbf{Y}|\hat{\theta}(\mathbf{X}))$ の推定値としてはバイアス(偏り)が出る。そこで、現在のデータ \mathbf{X} と未来のデータ \mathbf{Y} が同じ確率法則に従うと仮定すればバイアスは予測モデルのパラメータ数 $\dim \hat{\theta}(\mathbf{X})$ にほぼ等しいので、対数尤度評価スコア $\log P(\mathbf{Y}|\hat{\theta}(\mathbf{X}))$ の代わりに $-AIC/2 = \log P(\mathbf{X}|\hat{\theta}(\mathbf{X})) - \dim \hat{\theta}(\mathbf{X})$ が推定値として使える。因みに AIC は赤池情報量規準 (Akaike, 1973, 1977, 1985) で、モデルの予測の悪さの評価量を推定しているものである。これを前節の表 2 で議論した尤度と同じ意味で用いて、どのくらい予測の実現確率が違うかを、モデルの尤度で見なければ、 $\exp(-AIC/2)$ を比べることになる (Akaike, 1978a)。

長いこと待たないと検証のデータが得られない場合の典型的な例は大地震の長期予測の評価である。たとえばブラウン運動到達過程 (Brownian Passage Time, BPT; Matthews et al., 2002) と呼ばれる更新過程モデルを使って大地震の発生間隔を予測する。パラメータは 2 つあり、 μ 値は再来平均間隔である。 α 値は間隔のバラツキである。その物理的な意味は、隣近辺の地震発生の影響を受けて活断層地震の発生がどの程度早まったり遅まったりしているかという指標に対応するものである。活断層地震の予測の評価には長期間(たとえば数百年以上)のデータの出現を待たなければならない。したがって、これまでのデータと最尤法 (Ogata, 2002; Nomura et al., 2011) を使った BPT モデルの AIC 値が予測の評価の手掛かりになる。

日本の内陸活断層地震の BPT モデルの α 値が 0.24 のままで良いかということが問題になっている (地震調査委員会, 2010, 2.6.1 節参照)。この値の違いが予測確率にどう反映するかについては本特集号の野村論文を参照されたい。野村・尾形 (2014) は履歴として 2 つ以上の大地震が確認されている 61 の活断層に対して共通の α 値を仮定して最尤推定値を求めると $\hat{\alpha} = 0.44$ となる。これを地震調査委員会の推定値 $\alpha_0 = 0.24$ の場合と比べると AIC が 25.6 改善された。これによれば 61 断層の総体の予測として、 α 値が 0.44 の場合は 0.24 の場合に比べて $\exp(-\Delta AIC/2) \approx 1.2 \times 10^{11}$ 倍程度の子測力の向上が期待できる。

活断層地震や海溝型歴史地震などの地震発生間隔を、BPT モデル以外にも、対数正規分布などで与えた更新過程モデル (Utsu, 1984, など) や応力解放点過程モデル (Vere-Jones, 1978; Imoto,

2001, など)などの予測力も同様に評価できる。

さて、単一の活断層の場合などデータ量に比して尤度のパラメータ θ の数(たとえ2つでも)が多過ぎて、プラグイン予測が実用的でないとき、 θ の事前分布を仮定して事後確率分布の平均化予測(ベイズ予測)

$$p(y|\mathbf{X}) \propto \int p(y|\theta)\pi(\theta|\mathbf{X}, \hat{\tau})d\theta$$

を使う方法がある(Akaike, 1978b; Ogata, 2002)。ここで $\pi(\theta|\mathbf{X}, \hat{\tau})$ は最適な超パラメータ値の事後確率分布であり、赤池ベイズ情報量規準

$$\text{ABIC} = -2 \int \pi(\theta|\mathbf{X}, \hat{\tau})d\theta + 2 \dim(\tau)$$

は上記の予測モデルの悪さを与える。データの少ない活断層地震のプラグイン予測に対するベイズ予測の優位性を示す評価比較に関するシミュレーション実験については Nomura et al. (2011) を参照されたい。

本特集号の尾形論文にあるように、一般に点過程は突発的な確率的事象を抽象化した数学的モデルであるが、なかでも中心的概念である「条件付き強度関数」は事象発生の切迫度(確率の時間微分)の予測と直結している。たとえば ETAS モデルは元来、通常の地震活動を表現するために、余震の減衰の改良大森関数を或る重み付きで重ねあわせたものとして創出されたものであるが、余震活動そのものを、純粋に減衰する場合から群発型の複雑な経過を辿る場合まで、定量的に良く表現、予測できる。また入出力型の点過程モデルによって、異なる地域の地震活動の因果関係や相互作用の検証、季節性や応力場の変化など第3の因子による地震発生率変化の探索などができる。他地域の大地震からの応力急増加による誘発効果(尾形, 2014, 2015)など様々な地球物理学的情報によってモデルの拡張、複雑化を余儀なくされるが、それらが予測として意味があるかの判定も AIC や ABIC で定量的に検証できる。

この様に目的に応じて点過程の条件付き強度関数による自由自在なモデル化が可能である。従来の予測モデルに比べた新提案モデルの有意性と予測性能は AIC で定量的に比較、検証できる。

6. まとめと補遺

数値 0~100% を出す確率予測、その算出が難しい場合の警告型予測、それぞれの予測結果についての評価法を解説した。確率予測は対数尤度値の差(情報利得)または尤度比の値による大小の評価ができる。警告型予測は予測に「評価1単位」を賭けてギャンブルスコアに基づく評価単位の得失和による評価スコアの大小で評価できる。警告レベルを与えている予測は、適切な確率に翻訳して確率予測と見做すことができる。それぞれのスコアを比較する際に必要不可欠なのは、予測される対象の地震の公正な基準確率を見積ることである。確率予測の評価では対数尤度(尤度)値を比較する際の基準モデルによる予測確率であり、警告型予測の評価ではオッズ比を算定する基準確率である。したがって、基準の確率をどのように設定するかが重要である。本稿では地震活動の地域的非常性がより妥当な大地震の基準の発生確率を与えると強調してきたが、さらに時間依存モデルなかつ地震の連鎖性を考慮したモデルが基準確率算定に必要であると考えられる(e.g., Kagan, 1991; Hardebeck et al., 2008)。

究極的に目指すべき地震予測は警告型でなく確率予測である。冒頭に述べたラクイラ地震の場合のように、異常事象で社会的関心が高まった場合に、ある程度定量的な情報を整え、確率予測が出来るように用意しておくことが肝要であろう。

予測能力をより良くするモデルを構成する最適パラメータ(最尤推定値など)を求めるため

に、現在までのデータの制約のもと、対数尤度を最大化したり、最適予測モデルを選択するために AIC や ABIC を最小化したりするけれども、その時の AIC 値や ABIC 値から計算される $\exp(-AIC/2)$ や $\exp(-ABIC/2)$ がモデルの予測力に比例するものになっている。なお赤池による統計的予測に関する貢献は多岐多数にわたっており、詳しくは赤池論文選集 (Parzen et al., 1998) を参照されたい。

とくに基準地震活動モデルの確率予測は、天気予報の様に、少なくとも準リアルタイムで常時計算されている必要がある。注目されている地域のみならず、なるべく地震発生可能な全体領域をカバーするようにする。このことが予測結果を評価する際のデータを蓄積し、予測を改善するために有用である。確率予測は平均的確率や基準確率予測からの改善幅(確率利得; Aki, 1981) が大きいほど予測の情報が効率的に働いていることを示す。その意味で独立な複数観測項目(尾形, 2015 参照)による確率予測が有効であり、これは各地震について対数尤度比の値(情報利得)と合わせて評価される。警告型地震予測は、「警告予測」のみならず「反予測」も常時公表できるのが望ましい。

謝 辞

査読者のコメントは本稿の改善にとって大変有益であった。また本特集の松浦充宏担当編集委員からは多くの適切な表現上の助言をいただいた。ここに記して感謝する。

参 考 文 献

- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle, *Proceedings of 2nd International Symposium on Information Theory* (eds. B. N. Petrov and F. Csaki), 267–281, Akademiai Kiado, Budapest (Reproduced in *Breakthroughs in Statistics*, Vol. I (eds. S. Kotz and N. L. Johnson), Springer-Verlag, New York, 610–624 (1992)).
- Akaike, H. (1977). On entropy maximization principle, *Applications of Statistics* (ed. P. R. Krishnaiah), North-Holland Publishing Company, 27–41, Amsterdam.
- Akaike, H. (1978a). On the likelihood of a time series model, *The Statistician*, **27**, 217–235.
- Akaike, H. (1978b). A new look at the Bayes Procedure, *Biometrika*, **65**, 53–59.
- Akaike, H. (1985). Prediction and entropy, *A Celebration of Statistics*, The ISI Centenary Volume (eds. A. C. Atkinson and S. E. Fienberg), 1–24, Springer-Verlag, New York.
- Aki, K. (1981). A probabilistic synthesis of precursory phenomena, *Earthquake Prediction* (eds. D. W. Simpson and P. G. Richards), Maurice Ewing Series, 4, 566–574, American Geophysical Union, Washington, D.C.
- Boltzmann, L. (1878). Weitere Bemerkungen uber einige Probleme der mechanischen Warmetheorie, *Wiener Berichte*, **78**, 7–46.
- Center for Analysis and Prediction of the State Seismological Bureau (1990–2003). *Annual Summary Report on National Earthquake Tendency, 1990, ..., 2003*.
- CSEP (2014). *Collaboratory for the Study of Earthquake Predictability*, <http://www.cseptesting.org/>.
- Dieterich, J. H. (1994). A constitutive law for rate of earthquake production and its application to earthquake clustering, *Journal of Geophysical Research*, **99**(B2), 2601–2618.
- Hardebeck, J. L., Felzer, K. R. and Michael, A. J. (2008). Improved tests reveal that the accelerating moment release hypothesis is statistically insignificant, *Journal of Geophysical Research*, **113**, B08310, doi:10.1029/2007JB005410.
- Harte, D. and Vere-Jones, D. (2005). The entropy score and its uses in earthquake forecasting, *Pure and Applied Geophysics*, **162**, 1229–1253.

- 平田直, 鶴岡弘, 楠城一嘉 (2011). CSEP 日本の取り組みについて, 地震予知連絡会会報, **66** (12-2), 433–436.
- Imoto, M. (2001). Application of the stress release model to the Nankai earthquake sequence, southwest Japan, *Tectonophysics*, **338**, 287–296.
- Imoto, M. (2004). Probability gains expected for renewal process models, *Earth Planets Space*, **56**, 563–571.
- Jordan, T. H. (2006). Earthquake predictability, brick by brick, *Seismological Research Letters*, **77**(1), 3–6.
- Jordan, T. H., Chen, Y.-T., Gasparini, P., Madariaga, R., Main, I., Marzocchi, W., Papadopoulos, G., Sobolev, G., Yamaoka, K. and Zschau, J. (2011). Operational earthquake forecasting: State of knowledge and guidelines for implementation, Final Report of the International Commission on Earthquake Forecasting for Civil Protection, *Annals of Geophysics*, **54**(4), 315–391, doi:10.4401/ag-5350, <http://www.annalsofgeophysics.eu/index.php/annals/article/view/5350/5371>.
- Kagan, Y. Y. (1991). Likelihood analysis of earthquake catalogues, *Geophysical Journal International*, **106**, 135–148.
- Kagan, Y. Y. (1996). VAN earthquake prediction — An attempt at statistical evaluation, *Geophysical Research Letters*, **23**, 1315–1318.
- Kagan, Y. Y. (2007). On earthquake predictability measurement: Information gain and error diagram, *Pure and Applied Geophysics*, **164**, 1947–1962.
- Kagan, Y. Y. and Jackson, D. (1995). New seismic gap hypothesis — 5 years after, *Journal of Geophysical Research*, **100**, 3943–3959.
- Keilis-Borok, V. I. and Malinovskaya, L. N. (1964). One regularity in the occurrence of strong earthquakes, *Journal Geophysical Research*, **69**(14), 3019–3024, doi:10.1029/JZ069i014p03019.
- Kossobokov, V. G. (1997). User manual for M8, *Algorithms for Earthquake Statistics and Prediction* (eds. J. H. Healy, V. I. Keilis-Borok and W. H. K. Lee), **6**, IASPEI Software Library, Seismological Society of America, Albany, California.
- Matthews, M. V., Ellsworth, W. L. and Reasenber, P. A. (2002). A Brownian model for recurrent earthquakes, *Bulletin of the Seismological Society of America*, **92**, 2233–2250.
- Molchan, G. M. (1991). Strategies in strong earthquake prediction, *Physics of the Earth and Planetary Interiors*, **61**(1.2), 84–98.
- Molchan, G. M. and Kagan, Y. Y. (1992). Earthquake prediction and its optimization, *Journal Geophysical Research*, **97**, 4823–4838, doi:10.1029/91JB03095.
- 楠城一嘉, 鶴岡弘, 遠田晋次, 平田直 (2008). 地震活動の評価に基づく地震発生予測: 世界と日本の動向, 日本地震学会ニュースレター, **20**, http://www.zisin.jp/modules/pico/index.php?content_id=1029.
- Nanjo, K. Z., Tsuruoka, H., Hirata, N. and Jordan, T. H. (2011). Overview of the first earthquake forecast testing experiment in Japan, *Earth Planets Space*, **63**(3), 159–169, doi:10.5047/eps.2010.10.003.
- Nanjo, K. Z., Tsuruoka, H., Yokoi, S., Ogata, Y., Falcone, G., Hirata, N., Ishigaki, Y., Jordan, T. H., Kasahara, K., Obara, K., Schorlemmer, D., Shiomi, K. and Zhuang, J. (2012). Predictability study on the aftershock sequence following the 2011 Tohoku-Oki, Japan, earthquake: First results, *Geophysical Journal International*, **191**, 653–658.
- 野村俊一, 尾形良彦 (2014). 地震の再来間隔分布にかかる空間的変動と長期確率予測への影響, 日本地球惑星科学連合 2014 年大会, パシフィコ横浜, 横浜市, 2014 年 4 月 29 日.
- Nomura, S., Ogata, Y., Komaki, F. and Toda, S. (2011). Bayesian forecasting of the recurrent earthquakes and its predictive performance for a small sample size, *Journal of Geophysical Research*, **116**, B04315, doi:10.1029/2010JB007917.
- Ogata, Y. (1995). Evaluation of probability forecasts of events; invited discussion as a commentary on

- “Forecasting Earthquakes and Earthquake Risk” by Prof. D. Vere-Jones, *International Journal of Forecasting*, **11**, 539–541.
- Ogata, Y. (2002). Slip-size-dependent renewal processes and Bayesian inferences for uncertainties, *Journal of Geophysical Research*, **107**, 1–14, doi:10.1029/2001JB000668.
- 尾形良彦 (2014). 常時地震活動・余震・誘発地震の予測能力と評価, 地震予知連絡会会報, **93**(12-7), 415–426, http://cais.gsi.go.jp/YOCHIREN/report/kaihou93/12_07.pdf.
- 尾形良彦 (2015). 地震の確率予測の研究—その展望, 統計数理特集「地震予測と統計モデル」, 統計数理, **63**(1), 3–27.
- Ogata, Y. and Katsura, K. (2012). Prospective foreshock forecast experiment during the last 17 years, *Geophysical Journal International*, **191**, 1237–1244.
- Ogata, Y. and Katsura, K. (2014). Comparing foreshock characteristics and foreshock forecasting in observed and simulated earthquake catalogs, *Journal of Geophysical Research*, **119**, Issue 11, 8457–8477, doi:10.1002/2014JB011250.
- Ogata, Y., Utsu, T. and Katsura, K. (1996). Statistical discrimination of foreshocks from other earthquake clusters, *Geophysical Journal International*, **127**, 17–30.
- Ogata, Y., Katsura, K., Falcone, G., Nanjo, K. Z. and Zhuang, J. (2013). Comprehensive and topical evaluations of earthquake forecasts in terms of number, time, space, and magnitude, *Bulletin of the Seismological Society of America*, **103**, 1692–1708.
- Parzen, E., Tanabe, K. and Kitagawa, G. (eds.) (1998). *Selected Papers of Hirotugu Akaike*, Springer-Verlag, Tokyo.
- Regional Likelihood Models (RELM) (2007). *Seismological Research Letters*, **78**(1), Special Issue, <http://www.cseptesting.org/documents/relm>.
- Rhoades, D. A., Schorlemmer, D., Gerstenberger, M. C., Christophersen, A., Zechar, J. D. and Imoto, M. (2011). Efficient testing of earthquake forecasting models, *Acta Geophysica*, **59**, 728–747, doi:10.2478/s11600-011-0013-5.
- Schorlemmer, D., Zechar, J., Werner, M., Field, E., Jackson, D., Jordan, T. and RELM Working Group (2010). First results of the regional earthquake likelihood models experiment, *Pure and Applied Geophysics*, **167**(8-9), 859–876, doi:10.1007/s00024-010-0081-5.
- Shebalin, P. N., Keilis-Borok, V. I., Zaliapin, I., Uyeda, S., Nagao, T. and Tsybin, N. (2004). Advance short-term prediction of the large Tokachi-oki earthquake, September 25, 2003, $M = 8.1$: A case history, *Earth planet, Space*, **56**, 715–724.
- 宇津徳治 (1977). 地震予知の適中率と予知率, 地震 II, **30**, 179–185.
- 宇津徳治 (1979). 地震予知の適中率の計算(伊豆大島近海地震を例として), 地震予知連絡会報, **21**(7-4), 164–166.
- Utsu, T. (1984). Estimation of parameters for recurrence models of earthquakes, *Bulletin of the Earthquake Research Institute, University of Tokyo*, **59**, 53–66.
- Vere-Jones, D. (1978). Earthquake prediction: A statistician’s view, *Journal of Physics of the Earth*, **26**, 129–146.
- Vere-Jones, D. (1999). Probabilities and information gain for earthquake forecasting, *Computational Seismology*, **30**, 248–263.
- Wang T. and Bebbington, M. (2013). Identifying anomalous signals in GPS data using HMMs: An increased likelihood of earthquakes?, *Computational Statistics and Data Analysis*, **58**, 27–44.
- Wang, T., Zhuang, J., Kato, T. and Bebbington, M. (2013). Assessing the potential improvement in short-term earthquake forecasts from incorporation of GPS data, *Geophysical Research Letters*, **40**, 2631–2635, doi:10.1002/grl.50554.
- Woessner, J., Hainzl, S., Marzocchi, W., Werner, M. J., Lombardi, A. M., Catalli, F., Enescu, B., Cocco, M., Gerstenberger, M. C. and Wiemer, S. (2011). A retrospective compara-

- tive forecast test on the 1992 Landers sequence, *Journal of Geophysical Research*, B05305, doi:10.1029/2010JB007846.
- Zechar, J. D. (2010). Evaluating earthquake predictions and earthquake forecasts: A guide for students and new researchers, *Community Online Resource for Statistical Seismicity Analysis*, doi:10.5078/corssa-77337879, Available at <http://www.corssa.org>, Article Theme VI, <http://www.corssa.org/articles/themevi/zechar/zechar.pdf>.
- Zechar, J. D. and Zhuang, J. (2010). Risk and return: Evaluating reverse tracing of precursors earthquake predictions, *Geophysical Journal International*, **182**, 1319–1326, doi:10.1111/j.1365-246X.2010.04666.x.
- Zechar, J. D. and Zhuang, J. (2014). A parimutuel gambling perspective to compare probabilistic seismicity forecasts, *Geophysical Journal International*, **199**(1), 60–68, doi:10.1093/gji/ggu137.
- Zechar, J., Gerstenberger, M. and Rhoades, D. (2010). Likelihood-based tests for evaluating space-rate-magnitude earthquake forecasts, *Bulletin of the Seismological Society of America*, **100**, 1184–1195.
- Zechar, J. D., Schorlemmer, D., Werner, M. J., Gerstenberger, M. C., Rhoades, D. A. and Jordan, T. H. (2013). Regional earthquake likelihood models I: First-order results, *Bulletin of the Seismological Society of America*, **103**(2A), 787–798, doi:10.1785/0120120186.
- Zhuang, J. (2010). Gambling scores for earthquake predictions and forecasts, *Geophysical Journal International*, **181**, 382–390, doi:10.1111/j.1365-246X.2010.04496.x.
- Zhuang, J.-C. and Jiang, C.-S. (2012a). Evaluation of the prediction performance of the Annual Consultation Meeting on Earthquake Tendency by using the gambling score (in Chinese with English abstract), *Chinese Journal of Geophysics*, **55**, 1695–1709, doi:10.6038/j.issn.0001-5733.2012.05.026.
- Zhuang, J. and Jiang, C. (2012b). Scoring annual earthquake predictions in China, *Tectonophysics*, **524–525**, 155–164, doi:10.1016/j.tecto.2011.12.033.
- 地震調査委員会 (2010). 活断層の長期評価手法(暫定版)報告書の公表について, http://www.jishin.go.jp/main/choukihyoka/katsu_hyokashuho/honpen.pdf.

Evaluation Methods of Earthquake Forecasts

Jiancang Zhuang¹ and Yosihiko Ogata²

¹The Institute of Statistical Mathematics

²The Institute of Statistical Mathematics, Professor Emeritus

Objective evaluation of forecasting performance is essential in research on earthquake predictability. Since the occurrence probabilities of large and small earthquakes are completely different, the score for a successful prediction of a rarely occurring large earthquake should be significantly different from that of small earthquakes. Similar reasons should be applied to predictions in non-active and active seismic regions. First of all, it is necessary to build empirical models for seismicity in different regions, which can be used as references for forecasting future seismicity. The significance of earthquake forecasts can be evaluated by using the log likelihood ratio of the performance to the reference, or the information gain. The Akaike information criterion (AIC) is useful to estimate the information gain and to determine whether the proposed model will have better predictive performance than the reference model using currently available data. Due to the underdevelopment of forecasting algorithms and the lack of prediction experience, it is often the case that predictions are not given in the format of probabilities, but as earthquake warnings (binary predictions). This article also explains how to use a gambling score to evaluate such binary predictions. This method also needs a reference model. Each time the prediction succeeds or fails the predictor is rewarded or penalized by using a fair gambling rule according to the reference model. As the reference model, the uniform distribution (homogeneous Poisson process) for the occurrence times and locations of earthquakes has been used in addition to the Gutenberg-Richter law (exponential distribution) for earthquake magnitudes. But, when a more reasonable nonhomogeneous Poisson process is used as the reference model, the warning-type predictions that are currently available rarely have better scores.