

# 海底地殻変動観測の統計的な精度評価手法<sup>†</sup>

石川直史\*

Statistical method for accuracy evaluation of seafloor geodetic observation<sup>†</sup>

Tadashi ISHIKAWA\*

## Abstract

The GPS-Acoustic (GPS-A) seafloor positioning is a composite technique which combines GPS positioning with acoustic ranging. Due to the complexity of the technique, there are many inestimable errors from each component of measurement. Therefore, it is difficult to evaluate an accuracy using an error propagation formula. To evaluate an accuracy of the technique, we perform a statistical evaluation of reproducibility of measurements through campaign observations.

Here we summarize a statistical method for evaluating the accuracy of GPS-A seafloor positioning. We also investigate an observation time dependence of accuracy using this method.

## 1 はじめに

海上保安庁海洋情報部で実施している GPS—音響測距結合方式による海底地殻変動観測では、GPS 測位により求められた測量船の位置と音響測距によって求められた測量船と海底局の間の距離を組み合わせることで、海底に設置した海底局の位置を求めている。

本観測は、GPS 測位と音響測距という 2 つの測位技術を結合させるため、様々な観測データに由来する誤差の影響を受ける。誤差の要因は、大気や海水の環境条件に起因するものから、観測機器のハードウェアに起因するものなど多様であり、それらの誤差全てを定量的に評価することが困難となっている。そのため、個々の誤差に起因する測定の不確かさを足し上げることで精度を評価することはせずに、複数回の観測における測位

結果の再現性を評価することで、精度を評価している。

本稿では、現在行っている精度評価手法について、必要となる統計学の基礎事項とともにまとめた。また、その手法の応用例として、観測データ数と測位精度の関係についての検討を行った結果についても述べる。

## 2 海底地殻変動観測の測位結果の不確かさ

### 2.1 観測の誤差

本観測の誤差要因を大きく分けると、「GPS 測位に起因する誤差」、「音響測距に起因する誤差」、「船体の動揺計測に起因する誤差」の 3 つになる。このうち動揺計測の誤差については、使用している装置の公称精度が本観測で求められる精度を上回っているため、動揺計測に起因する誤差

<sup>†</sup> Received September 18, 2015; Accepted October 19, 2015

\* 技術・国際課 Technology Planning and International Affairs Division

の影響は無いと考えて差し支えない。

GPS 測位に関しては、海上で時々刻々で変化する測量船の真の位置を知ることはできないため、測量船の推定位置の真値からのずれを評価することができない。藤田・矢吹 (2003) は、測量船の上下の位置については、一義的には海面に拘束されているという条件をもつことに注目し、測位解 3 成分のうちの上成分に対して潮汐とジオイド高の補正を行うことで、測位結果の安定性の評価を行った。この手法は、測位解の全体の傾向の良し悪しを判断する手法として優れているものの、時々刻々で推定される位置の不確かさを定量的に評価することは難しい。

音響測距における最大の誤差要因は海中の音速度である。海中音速度は時間と空間（水平方向と鉛直方向の両方）で複雑に変化しているため、観測にかかる時間・空間の全体にわたって真値を把握することは事実上不可能である。そのため、海中音速度に起因する誤差の影響についても定量的に評価することが難しい。

音響測距におけるもう一つの主要な誤差要因は、音波の往復走時を求めるときの波形の読み取り誤差である。この誤差は、波形解析の性質から、全体としては概ね 1 波長程度の幅でばらつく偶発的な性質の誤差であるとみなせる。

## 2.2 インバージョン解析における解の分散

最終的には、得られた観測データを結合して、藤田・他 (2004) によるベイジアンインバージョン法によって、海底局の位置推定を行う。

ベイジアンインバージョン法では、正規分布する誤差  $e$  を持つ観測データ  $y$  とモデルパラメータ  $x$  の間に、 $y = Ax + e$  の観測方程式が成り立つときにモデルパラメータ  $x$  とその分散を推定するものである。

我々の観測では、音波の往復走時を観測値  $y$  として、海底局位置  $x$  を推定する。また、観測値の誤差  $e$  の標準偏差として、音波の読み取り誤差に起因する 1 波長分の時間 (0.1 ms) を与えている。

しかしながら、この方法では、往復走時の読み取り誤差以外の様々な誤差要因の影響、特に、最大の誤差要因である海中音速度の影響が考慮されていない。また、上述のようにそれら誤差の定量的な評価ができない以上、その影響を観測方程式に反映することはできていない。そのため、インバージョン解析の結果として出力されるモデルパラメータの分散は、必ずしも測位結果の不確かさの良い指標とはならない。

## 2.3 測位結果のばらつきの評価

以上のように、誤差伝播則等を用いて個々の誤差要因を積み上げることで最終結果の不確かさを定量的に評価することが難しいのが現状である。

そこで、繰り返し測位の再現性評価として、以下で述べるような観測データをいくつかのサブセットに分割して、そのばらつきを評価することを行ってきた (例えば、藤田・佐藤, 2004; 佐藤・他, 2009; 石川・佐藤, 2012)。

本観測では、各種の偶発的な誤差の影響を低減するために、ある程度の量の音響測距データが必要になる。そのため、1 回の観測では、所定の精度を得るため、経験的に定められた数の音響測距データ (1 観測点について約 5200 ショット) を取得することとしている。

また、音響測距データの空間的配置 (海底局に対する測量船の位置) が精度に大きな影響を与えるため、水平方向の誤差が互いにキャンセルされるように海底局に対して対称となるような位置で観測を行う。そのような、海底局の配置に対して対称となる測量船の測線の組を 1 セットとして、観測の最小単位としている。

この時の測線配置の広がりには水深によって決まり、水深が深いほど測線が広範囲となる。音響測距観測は一定時間間隔で行うため、測線配置が広いほど 1 セットで取得できる音響測距データ数が多くなる。例えば、水深約 2900 m の観測点では 1 セットあたり約 1300 ショットの音響測距データが取得できるのに対して、水深約 1200 m の観測点では約 650 ショットとなる。そのため、1 回

の観測に必要な音響測距データを取得するために要するセット数は、観測点毎に異なる。

なお、音響測距は一定の時間間隔ごとに音波を発射して行うため、音響測距データ数は観測時間に比例する。取得すべき音響測距データの総数は各観測点において同数になるようにしているため、必要とされる観測時間も各観測点で同一となる。現在は、1観測点あたり約1日の時間をかけて必要なデータ数を確保している。

最終的には、各観測点において定められたセット数全てのデータを使用して推定した位置をその観測回の最終成果としている。このときに、全セット分のデータから推定された解（全セット解）をその観測回における最適解と解釈すると、複数のサブセットに分けた解（サブセット解）の全セット解に対するばらつきがその観測回における精度評価の指標となりうる。

ただし、全セット解の位置はサブセット解の平均位置として求めているわけではないため、サブセット解のばらつきは必ずしも全セット解の良い精度評価指標とはならない。例えば、データ数の少ないサブセット解では存在した誤差が、データ数を増やした全セット解では解消されることがありうる。その場合は、サブセット解のばらつきが大きくても、全セット解の精度が低いとは一概には言えなくなる。

またこの評価手法の問題は、基準としている全セット解自体が誤差を含んでいることにある。そもそも、この全セット解の持つ誤差こそが本観測における測定誤差と解釈されるべきものである。

そこで以下では、複数回の観測結果から線形回帰で推定された回帰直線を最適解と解釈し、回帰直線に対する測位解のばらつきを評価することで、測位精度の評価を行う。

### 3 海底地殻変動観測の測位精度の評価手法

実際の観測データを例に、現在行っている測位精度の評価手法について紹介する。なお、必要となる統計学の基本的事項は、Appendix にまとめた。

ここでは、Table 1 に示した土佐沖1の観測点で2012年から2015年にかけて行った9回の観測で得られたデータを使用する。表の数値は、最初の観測（2012年1月27日）の測位解に対する位置を示している。

データを線形回帰したときの傾きとその分散、および東西成分と南北成分の間の共分散は、(A.5-7)、(A.10) 式から求めることができる。

線形回帰による推定結果を Table 2 に示す。今の場合、データ数は9個なので、線形回帰した時の自由度は7となる。回帰直線の傾き、すなわち変位速度の推定の95%信頼区間は、Table A.1 で示した自由度7の  $t$  分布の両側5%点である2.365を、傾きの分散の平方根に乗じて、東西、南北について、それぞれ  $-3.8 \pm 1.8$  cm/y,  $3.6 \pm 1.2$  cm/y と表される。

時系列プロットを Fig.1 に示す。赤線が推定された回帰直線で、青線が(A.8)式で表される95%信頼区間の双曲線である。

通常線形回帰による分析では、推定された傾きとその推定の幅から、線形回帰の妥当性の検定を行う。単純には、推定された傾きよりも推定の幅が大きければ、2変数の間に有意な線形関係は無いと考えるのが普通である。しかしながら、我々の推定している海底の観測点の変位速度は、相対的な値であり、変位の固定点をどの場所に設定するかで絶対値は変化する。従って、推定された傾きの大きさとその信頼区間の幅を比較することは、意味を持たないということに注意が必要で

Table 1. An example of result of seafloor geodetic observation obtained at TOS1 site.

表1. 土佐沖1における海底地殻変動の結果の例。

年 [year]	2012.074	2012.444	2012.786	2013.068	2013.521	2014.038	2014.668	2014.951	2015.414
東方向 [cm]	0.00	-3.67	-8.25	-5.58	-8.64	-5.30	-10.13	-12.94	-16.84
北方向 [cm]	0.00	-0.33	4.04	0.30	4.63	3.72	8.29	9.85	11.86

Table 2. Result of regression analysis for data in Table 1.

表 2. Table 1 のデータを回帰分析した結果.

	傾きの推定値 [cm/y]	標本分散 [(cm/y) <sup>2</sup> ]	傾きの分散 [(cm/y) <sup>2</sup> ]	傾きの共分散 [(cm/y) <sup>2</sup> ]	自由度	95%信頼水準の係数	傾きの 95% 信頼区間 [cm/y]
東方向	-3.8	2.5	0.55	-0.25	7	2.3646	1.8
北方向	3.6	1.7	0.25				1.2

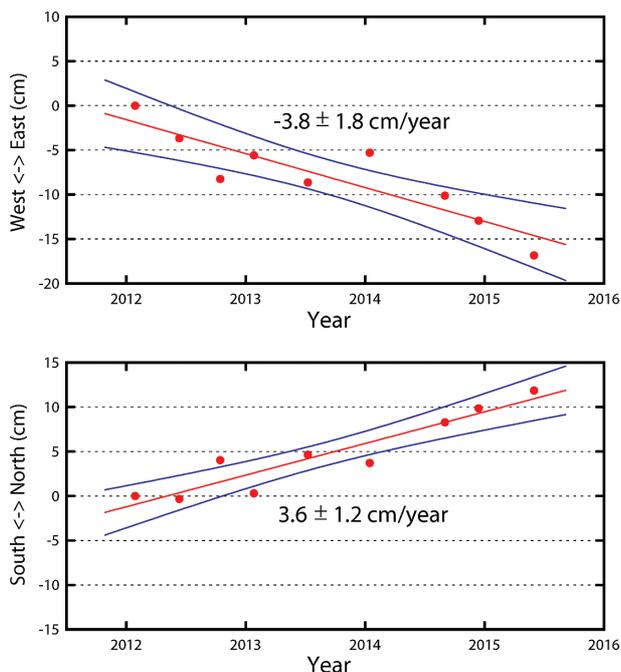


Fig. 1. An example of time series. The linear trends and their 95% two-sided confidence intervals are shown with red and blue lines, respectively.

図 1. 時系列プロットの例. 赤線は回帰直線, 青線はその 95% 信頼区間.

ある.

次に, 推定された変位速度を地図上の 2 次元ベクトルとして表示するときの信頼区間を考える. 2 次元の信頼区間は, (A.4) 式で表される楕円と

なる. ここで, 分散共分散行列は Table 2 で示した傾きの分散・共分散を用いる. (A.4) 式は, 座標軸の回転で分散共分散行列を対角化することで, 以下の標準形式に変形できる.

$$\frac{x'^2}{z^2 \lambda_a} + \frac{y'^2}{z^2 \lambda_b} = 1$$

ここで,  $\lambda_a, \lambda_b$  は分散共分散行列の固有値である. Table A.2 から自由度 7 のときの係数  $z = 3.079$  を固有値の平方根に乗じることで, 95% 信頼区間を表す楕円の長半径・短半径が求まる. また, 長径の向く方向は対応する固有ベクトルの方向となる. 得られた結果を Table 3 に, 地図上に表示したベクトル図を Fig.2 に示す.

#### 4 応用例：観測時間と測位精度の関係についての検討

精度評価手法の応用例として, 観測時間 (音響測距データ数) と測位解のばらつきについて検討を試みた例について述べる.

##### 4.1 目的

観測点の変位速度の推定精度となる回帰直線の推定精度の向上のためには, (A.6) 式からわかるように, 1 回 1 回の観測における測位精度 (標本分散) の向上に加え, 観測回数を増やさなければ

Table 3. Two-dimensional confidence interval of result in Table 2.

表 3. Table 2 の結果の 2 次元信頼区間.

分散共分散行列 [(cm/y) <sup>2</sup> ]	分散共分散行列の固有値 [(cm/y) <sup>2</sup> ]	95%信頼水準の係数	楕円の長径・短径 [cm]	長径が東西軸となす角度 (東向きが 0 度で反時計回りが正) [°]
$\begin{pmatrix} 0.55 & -0.25 \\ -0.25 & 0.25 \end{pmatrix}$	0.69 0.11	3.079	5.1 2.1	-29

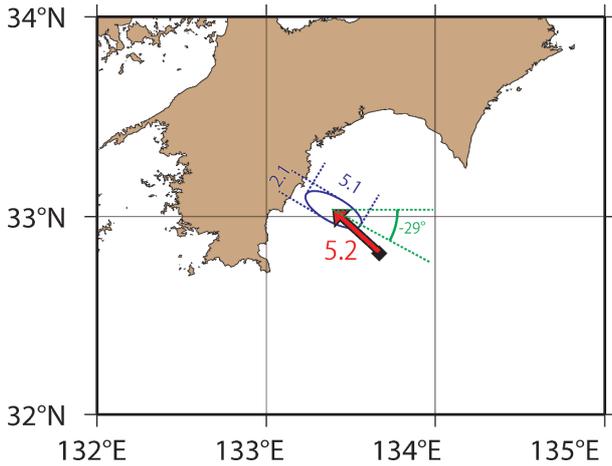


Fig. 2. An example of two-dimensional velocity vector and its 95% confidence interval on map.

図 2. 地図上への 2 次元速度ベクトルとその 95% 信頼区間の図示例.

ばならない。微小な地殻変動を有意に捉えるために要する 1 cm/年 (95% 信頼区間) という精度を達成するには、現状では年 2-3 回の観測を 4-5 年程度続ける必要がある。

ここで、4-5 年間のデータを線形回帰するということは、その間の変位速度が一定であることを仮定している。しかしながら、スロースリップイベントや巨大地震後の余効変動などにみられるように、地殻の変動は必ずしも一定速度で長期間安定しているわけではない。従って、地殻の変動状態をより正確に把握するためには、より短期間のデータで観測点の変位を高精度に推定する必要がある。

測位精度に関しては、本観測の開始以来、ソフト・ハードの両面において改善が進められてきた (佐藤・藤田, 2012)。現在では、繰り返しの測位

精度は 2-3 cm (標準偏差) 程度となっている。

観測回数に関しては、測量船の運航日数からくる制限があるため、全体の観測日数を変えずに観測回数を増加させるには、1 回の観測にかかる観測時間を削って、その分を観測回数の増加に回す必要がある。

観測時間を減らすことは、一般的に測位精度の低下を招くが、一方で観測回数の増加は回帰直線の推定精度を向上させるというトレードオフの関係にある。両者の最適な関係の検討を行うためには、観測時間を減らした場合の精度の低下度合いの検証が必要となる。

#### 4.2 手法

観測時間短縮の効果を見るため、過去に取得した観測データを用いて、音響測距データ数を 100% 使用した結果から、75%、50%、25% と減らしていったときの測位解の変化を見る。2.3 で述べたように、本観測におけるデータの最小単位は 1 セットであるため、以下では、データ数の単位としてセット数を基準にして考える。

ここでは、Table 4 に示した南海トラフ沿いの 9 ヶ所の観測点において 2012 年以降に取得したデータを使う。解析は、現在ルーティン解析で用いている松本・他 (2008) の複数エポック一括解析の手法を用いた。なお、既に述べたとおり水深の違いによって、各観測点で 100% のセット数が異なるが、精度に影響する音響測距のデータ数は各観測点で同一となっている。

セット数の減らし方は、例えば全体で 4 セット [ABCD] ある場合、75% は [ABC]、50% は

Table 4. Data used in this study.

表 4. 解析に使用したデータ.

観測点名	東海沖 3	熊野灘 2	室戸沖 1	土佐沖 1	土佐沖 2	足摺沖 1	足摺沖 2	日向灘 1	日向灘 2	
観測回数	6	7	7	7	8	7	6	8	8	
セット数	100%	8	5	7	9	5	8	4	6	5
	75%	6	3	5	6	3	6	3	4	3
	50%	4	2	3	4	2	4	2	3	2
	25%	2	1	1	2	1	2	1	1	1

[AB], 25%は[A], というように後ろのセットから順次減らしていく方法をとった. 本観測では, 1セットを最小セットとしているため, それ以上の分割はできない. そのため, セット数を分割するときに割り切れない場合は, 小数点を以下切り捨てた.

ここでは, 100%全てのセットを使って推定した回帰直線をもっとも確からしい値として, その直線の周りの測位解のばらつきについて評価した.

ばらつきの指標として (A.7) 式の標本分散を用いるが, セット数を減らしたときの結果については, そのデータから推定した回帰直線ではなく, 100%セットのデータを用いて推定した, つまり外部から与えられた直線を用いるため, データの自由度を示す分母は  $n$  となる. 100%セット解については, 回帰直線の推定に自由度が2使われているので, 分母は (A.7) 式のとおり  $n-2$  となるが, ここでの目的はセット数の違いによるばらつきの変化の検証であるため, 100%セット解の場合も, 回帰直線は外部から与えられたものと解釈し, 分母を  $n$  として計算した. 同様に東西・南北成分の共分散の計算の時の分母も  $n$  とした.

### 4.3 結果

本観測は, 海水温等の環境要因に大きく影響を受けるため, 観測地点ごとの固有の誤差があるものと考えられるが, 個々の観測点におけるデータ数が少ないため, 観測地点ごとの詳細な特徴を有意に検出できるまでに至っていない. ここでは, 本観測手法の平均的な傾向を把握するために, 全ての観測点のデータをまとめて評価した.

得られた測位解のばらつきの分布を平面で図示したものを Fig. 3 に示す. 図のプロットは全セット解のデータから推定した回帰直線からの差を示しており, 中心が回帰直線上の点となる. 楕円は, (A.4) 式において, 上述のように自由度を  $n$  として計算した分散共分散行列を用い,  $z=1$  とした ( $1\sigma$  の楕円).

Table 5. Relationship between the number of data and the standard deviation.

表5. セット数 (データ数) と標準偏差の関係.

セット数	100%	75%	50%	25%
$\sigma_d$ [cm]	1.6	1.7	1.9	2.4

ここで, 2次元のばらつきを評価するため, 方向によらない標本標準偏差として次式で定義される  $\sigma_d$  を導入する.

$$\sigma_d = \sqrt{\frac{\sigma_{ew}^2 + \sigma_{ns}^2}{2}}$$

ここで,  $\sigma_{ew}^2$ ,  $\sigma_{ns}^2$  はそれぞれ, 東西・南北方向の標本分散である. セット数と  $\sigma_d$  の関係を Table 5 に示す.

一般に, 標本標準偏差の推定の幅は大きく, 観測データ数が10個程度の場合, 平均的に25%程度の不確かさがある. 今の場合, 観測データ数は64個であるが, この場合でも10%程度の不確かさがある (例えば, 田中・高津 (2014) の第3章における議論を参照). そのことを考慮すると, 100%セットの結果と75%セットの結果では, 標本標準偏差に有意な差はないといえる.

### 4.4 測位精度の考察

統計処理上, セット数の多寡が影響するのは, 主に偶発的な誤差要因に対してである. 偶発的な誤差要因として考えられるのは, 音波の往復走時の読み取り誤差のほか, 大気や海水の短周期の擾乱がGPS測位や音響測距に与える誤差である.

今回の結果に見られる, セット数の増加とともに  $\sigma_d$  が減少していく様子は, これら偶発的な誤差の影響が統計的に抑えられていることを表しているが, その一方で減少率は頭打ちとなり, 75%セットと100%セットでは, 有意な違いは現れていない. このことは, 偶発的な誤差の影響は75%セットにおいて十分に抑えられていることを示唆している.

一方, セット数の増加では対処できない誤差要因としては, 観測時間を超えるような長周期の大気や海水の擾乱による系統的な誤差が考えられる.

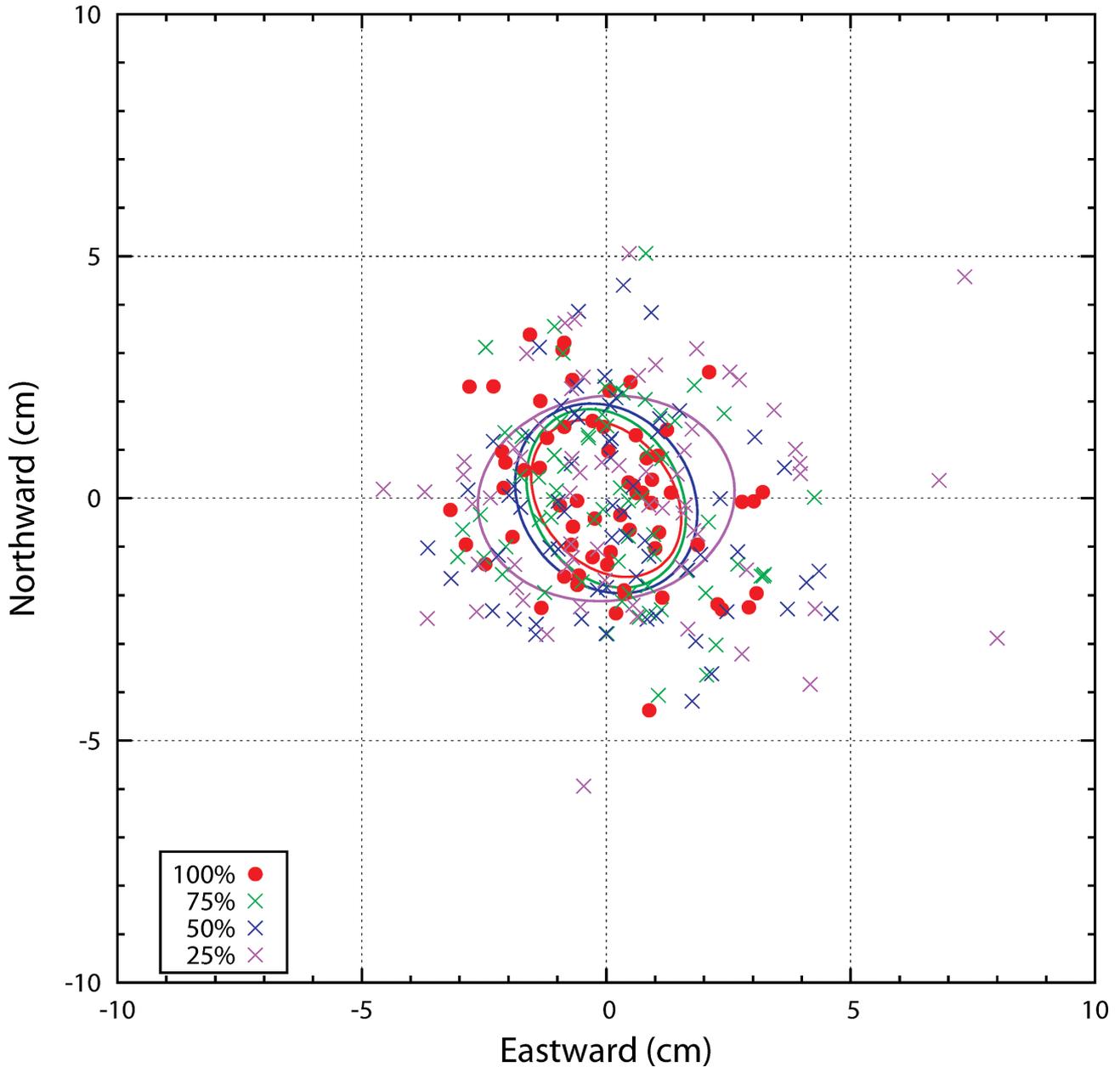


Fig. 3. Scatter plot of difference between estimated positions and linear trend line.

図 3. 回帰直線の周りの測位解の分布。原点が回帰直線上の点を表す。

特に海水温の空間的不均質は、本観測における最大の誤差要因となっている。これは、主として海流や冷水塊・暖水塊の影響などが考えられるが、これらの影響は場合によっては数日から数週間継続するため、現在の約1日の観測時間を多少増やしたところで、有効な対処法とはなりえない。

こうした長周期の誤差に対しては、観測手法、解析手法の両面でのイノベーションが必要にな

り、今後のさらなる研究開発が必要とされる。

#### 4.5 変位速度の精度の考察

次に回帰直線の傾き、すなわち観測点の変位速度の推定の不確かさについて考える。

(A.9) 式から信頼区間は、データのばらつきの標準偏差（測位精度）、データ数（観測回数）、データの幅（観測期間）に依存する。そこで、それぞれの値を変化させたときの回帰直線の95%

Table 6. 95% confidence interval of linear regression line.

表 6. 回帰直線の傾きの推定の 95% 信頼区間.

標準 偏差 [cm]	回帰直線の傾きの推定の 95%の信頼区間 [cm/year]												年間の観測回数
	3 回				5 回				6 回				
	2年	3年	4年	5年	2年	3年	4年	5年	2年	3年	4年	5年	
1.50	1.5	0.9	0.6	0.5	1.3	0.8	0.5	0.4	1.2	0.7	0.5	0.3	
1.75	1.8	1.0	0.7	0.5	1.5	0.9	0.6	0.4	1.4	0.8	0.6	0.4	
2.00	1.9	1.2	0.8	0.6	1.7	1.0	0.7	0.5	1.6	0.9	0.6	0.4	
2.50	2.4	1.5	1.0	0.8	2.1	1.3	0.8	0.6	2.0	1.1	0.8	0.6	

信頼区間を Table 6 に示す. なお, ここでは簡便のため, 観測を等間隔で行うものとして計算している. 我々の観測では, 傾きの推定の幅として 1 cm/y を達成することをひとつの目安としているが, これは例えば, データのばらつきの標準偏差が 1.75 cm の場合, 年 3 回の観測を 4 年続けることで達成できる.

観測時間と観測回数はトレードオフの関係にある. 仮に, 観測時間を短縮することで, 標準偏差が悪化したとしても, 短縮した時間分を観測回数の増加に回すことで, 傾きの推定精度が保つことができる. 例えば, Table 5 の結果から, 現状 1.5 cm 程度の標準偏差であるものが, セット数を半分にすることで標準偏差が 2.0 cm 程度まで悪化したとしても, 年間の観測回数を現行の 3 回から, 5-6 回に増加させることで, 同程度の精度が確保できることがわかる.

精度が同程度であれば, 観測回数が多い方がメリットが大きい. なぜならば観測回数の増加は, 海水温の空間不均質等の現在の手法では対処できない系統的な誤差に対して, 時期を変えた観測を増やすことで, 時系列全体としてその影響を低減できるからである. さらに, 観測データが時間的に密になることで, 現在は捉えきれない非線形な変動を検出できる可能性も高くなる.

#### 4.6 今後に向けて

今回の例では, セット数を 75% に減らしても有意な測位精度の低下はみられないため, 測位精度の低下というデメリットを受けずに, 観測回数を増やすことができることがわかった.

この結果を踏まえ, セット数を減らし観測回数を増やす方向で観測計画を立てることで, より地震学的・測地的に有効な観測成果が得られると期待される.

ただし, 実際の観測計画では, 全体の観測日数のうち測量船の回航にかかる時間の比率が大きいいため, 単純に観測時間を半分にしたとしても, 単純に観測回数を倍にできるわけではない.

非線形な変動の検出など, 本観測の今後のさらなる進展のためには, 必要とされる推定の精度を考慮し, 観測時間・観測回数を適切にバランスさせた観測計画を立てることが必要となってくる.

#### Appendix 統計処理の基礎的事項

ここでは, 本稿で行う統計処理に必要な関係式についてまとめておく. 本項目をまとめるにあたっては, 田中・高津 (2014) と東京大学教養学部統計学教室編 (1991) を参考とした.

### A.1 地殻変動がない場合の測位解のばらつき

最初に変動の無い地点における測位を考える。この場合、求めたい真値は変動の無いある一点の座標値となる。

真値は、通常は知ることのできない値であるので、複数回の測定から得られた測位解からもっとも確からしい値の推定を行う。ここでは、推定の母集団として平均値  $\mu$  の周りに分散  $\sigma^2$  で正規分布する無限個の観測データの集団を想定する。

測定において、正規分布する母集団のばらつきとは別の要因による誤差が生じていなければ、母集団の平均値が求めたい真値の推定値となる。ここで別の要因とは、例えば測定機器自体が持っている系統的な誤差や、測定時の環境要因による系統的な誤差などである。前者は、機器の校正によって予め取り除かれるべき誤差であり、後者は、観測手法や解析手法を工夫することでできるかぎり軽減しておくべき誤差である。以下で行う統計的手法による測定の不確かさの評価は、正規分布する偶発的な誤差を対象としているため、このような系統的な誤差は予め取り除かれている（もしくは十分軽減されている）ことを前提としている。ただし、海底地殻変動観測では、系統的な誤差を適切に分離することができていないため、全ての誤差を偶発的なものとして扱っている。

実際に行われる複数回の測定は、母集団から有限個の標本を抜き出すことに相当する。無限回の測定は行えないため、母集団の平均値（母平均）や分散（母分散）を直接測定することはできない。そこで、測定で得られた標本集団から、母集団の統計量（母数）を推定する。

母集団から  $n$  個の標本を取り出したときの標本集団の標本平均  $\bar{x}$  と標本分散  $s^2$  は次式で与えられる。

$$\bar{x} = \frac{\sum x_i}{n}, s^2(x) = \frac{\sum (x_i - \bar{x})^2}{n-1} \quad (\text{A.1})$$

以下、とくに断りの無い限り、和記号  $\Sigma$  は  $i$  について 1 から  $n$  までの和を取ることにし、和

の範囲を省略して書く。標本分散は、標本平均に対する標本のばらつきを表しており、標本平均の計算で自由度が 1 使われているので、分母の自由度は  $n-1$  となる。

それぞれの期待値は、

$$E(\bar{x}) = \mu, E(s^2) = \sigma^2$$

となり、標本平均・標本分散は、平均的には母平均・母分散に一致する推定量となる。このように「偏り」なく、母数の推定ができる推定量は不偏推定量と呼ばれる。

母集団がばらつきを持つため、そこから  $n$  個の標本を取り出すという行為を何度か繰り返した場合、標本平均  $\bar{x}$  もその都度変動する。標本平均の母分散は、 $\sigma^2/n$  で与えられ、確率分布は、平均値が  $\mu$ 、分散が  $\sigma^2/n$  の正規分布となる。これを標準化して、平均値が 0、分散が 1 の標準正規分布に従う形にすると

$$-z < \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} < z$$

となり、 $z$  で指定される範囲に含まれる確率を正規分布表から求めることができる。この式を変形すると、

$$\bar{x} - z \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z \frac{\sigma}{\sqrt{n}} \quad (\text{A.2})$$

となり、測定で得られた標本平均とその分散から、母平均が存在する範囲が示される。例えば、 $z=1.96$  とすると、母平均は 95% の確率でこの範囲に存在することになる。

ただし、ここでは  $\sigma^2$  は母分散であるため、通常は知ることができない量である。そこで、母分散の推定値として標本分散  $s^2$  を使うと、

$$\frac{\bar{x} - \mu}{s(x)/\sqrt{n}}$$

で定義される量は、自由度  $f$  の  $t$  分布に従うことが知られている。ここで、自由度は標本分散を計算するときの自由度であり、今の場合  $f=n-1$  である。(A.2) 式と同様の変形を行うと、

$$\bar{x} - z \frac{s(x)}{\sqrt{n}} < \mu < \bar{x} + z \frac{s(x)}{\sqrt{n}} \tag{A.3}$$

となり，自由度  $n-1$  の  $t$  分布  $t_\alpha(n-1)$  から信頼水準  $\alpha$  で母平均の区間推定ができる．例えば  $n=10$  の場合，95% の信頼区間は， $t$  分布の両側 5 パーセント点， $z = t_{0.025}(9) = 2.262$  から得られる． $t$  分布は，正規分布よりも幅の広い分布であり，自由度が大きくなるに従い分布の幅が狭くなる．自由度が無限大の極限で正規分布と一致する．正規分布と  $t$  分布の両側 5 パーセント点の値を Table A.1 に示す．

ここで，(A.2) 式や (A.3) 式から，測定回数  $n$  を増やすことによって，推定の幅をいくらで小さくすることができることがわかる．これは，測定の不確かさを示す母分散の大きさにかかわらず，非常に狭い範囲で母平均の推定ができてしまうことを意味している．ただし，(A.2) 式や (A.3) 式が示す範囲は，あくまで母平均の推定の幅であり，測定の誤差，すなわち真値からのずれを示しているわけではないことに注意が必要である．通常は，母数と真値がなるべく一致するように，測定に存在する系統的なずれなどは注意深く取り除いた上でデータ処理を行うが，こうしたずれは完全に取り除くことは不可能であるため，真値と母数は完全に一致することはない．そのため，必要以上に測定回数を増やして母数推定の幅を狭めても，そこに真値が含まれるかどうかは不明である．よって，測定においては，母集団がどのように定義されているか，想定される不確かさはどの程度かなどを吟味し，必要とされる精度，測定回数を把握しておく必要がある．

次に，変数が 2 つある場合，つまり 2 次元の場合を考える．変数が複数ある場合は，母集団のば

らつきを示す統計量として，それぞれの変数の分散に加え，変数間の共分散も考慮する必要がある．2 次元の座標軸を  $(x, y)$  とすると，標本共分散は次式で表され，

$$s_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

その期待値は母共分散  $\sigma_{xy}$  と一致する．

母平均がある確率で存在する範囲は，次式で示される 2 次元平面上の楕円となる．

$$(\bar{X} - \bar{\mu})^T \Sigma^{-1} (\bar{X} - \bar{\mu}) = z^2$$

$$\bar{X} = \begin{pmatrix} \bar{x} \\ \bar{y} \end{pmatrix}, \bar{\mu} = \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \Sigma = \begin{pmatrix} \text{Var}(\bar{x}) & \text{Cov}(\bar{x}, \bar{y}) \\ \text{Cov}(\bar{x}, \bar{y}) & \text{Var}(\bar{y}) \end{pmatrix} \tag{A.4}$$

$\Sigma$  は標本平均の分散共分散行列で， $z$  は母平均が存在する範囲の確率を定める係数である．(A.4) 式は， $x$  と  $y$  の 2 次式式であるため，座標系の適当な回転によって  $xy$  項を含まない標準系に変形できる．この座標変換は， $\Sigma$  を対角化することに対応しており，楕円の長径と短径は  $\Sigma$  の 2 つの固有値で表され，それぞれの方向は対応する固有ベクトルの指す方向となる．

母分散・母共分散が既知の場合，確率分布は 2 次元の正規分布となり， $z$  は自由度 2 の  $\chi^2$  分布  $\chi^2(2)$  から計算される．例えば，95% の信頼区間は， $\chi^2$  分布の上側 5 パーセント点から求められ， $z = \sqrt{\chi^2_{0.05}(2)} = 2.448$  となる．

母分散・母共分散が未知の時は，標本分散・標本共分散を推定値として使用する．この場合の確率分布は，2 次元の  $t$  分布となり， $z$  は自由度  $(2, n-1)$  の  $F$  分布から， $z^2 = 2F(2, n-1)$  として求まる．ここで，各自由度における  $z$  の上側 5 パーセント点の値を Table A.2 に示す．

最後に，これまでに出てきた 4 つの確率分布の関係について Table A.3 にまとめる．

Table A.1. One-dimensional two-sided 5% significance level.

表 A.1. 1 次元の両側 5% 点.

		t 分布: $z = t_{0.025}(n)$											正規分布
自由度: $n$	1	2	3	4	5	6	7	8	9	10	20	100	$\infty$
$z$	12.706	4.303	3.182	2.776	2.571	2.447	2.365	2.306	2.262	2.228	2.086	1.984	1.960

Table A. 2. Two-dimensional upper 5% significance level.

表 A. 2 2次元の上側 5%点.

	F 分布: $z = \sqrt{2F_{0.05}(2, n)}$												$\chi^2$ 分布
自由度: $n$	1	2	3	4	5	6	7	8	9	10	20	100	$\infty$
$z$	19.975	6.164	4.371	3.727	3.402	3.207	3.079	2.986	2.918	2.685	2.643	2.485	2.448

Table A. 3. Relationship among probability distribution of sample.

表 A. 3. 標本集団が従う確率分布の間の関係.

	母分散既知	母分散未知 (データの自由度: $f$ )
1次元	正規分布	自由度 $f$ の $t$ 分布 $z = t_\alpha(f)$
$n$ 次元	自由度 $n$ の $\chi^2$ 分布 $z^2 = \chi_\alpha^2(n)$	自由度 $(n, f)$ の $F$ 分布 $z^2 = nF_\alpha(n, f)$

### A. 2 地殻変動がある場合の測位解のばらつき

次に地殻変動がある場合の測位を考える。この場合、観測点は1点にとどまらず時間と共に移動していくので、複数回の観測で得られた測位解の平均値を求めても真値の良い推定とはならない。

そこで、変動が線形であると仮定し、測位解から最小二乗法で求めた線形回帰直線を真値の推定値とすることを考える。

ここでの母集団は、ある直線の周りに母分散  $\sigma^2$  で正規分布する無限個の測位解の集団を想定し、有限回の測定で得られた標本集団から母集団の直線の推定を行う。

時刻変数  $t$  と位置変数  $x$  の間に直線関係があるとき、

$$x = \alpha + \beta t$$

の関係式が成立する。この時の切片  $\alpha$ 、傾き  $\beta$  の真の値は母数である。ここで、離散的なデータの組を考えると  $i$  番目のデータに対して、

$$x_i = \alpha + \beta t_i + \epsilon_i$$

が成立する。ここで、 $\epsilon_i$  は誤差項で、その平均値は0、分散は  $i$  によらず  $\sigma^2$  であるとする。

母回帰係数の  $\alpha$  と  $\beta$  とは、直接知ることができないので、標本集団から推定する。最小二乗法によって推定された標本回帰係数は以下の式で表

される。

$$\begin{aligned} \hat{\alpha} &= \bar{x} - \hat{\beta} \bar{t} \\ \hat{\beta} &= \frac{\sum (t_i - \bar{t})(x_i - \bar{x})}{\sum (t_i - \bar{t})^2} \end{aligned} \tag{A. 5}$$

ここで、ハットは推定値であることを示し、バーは (A. 1) 式で計算される標本平均である。求めた標本回帰係数の期待値は、

$$E(\hat{\alpha}) = \alpha, E(\hat{\beta}) = \beta$$

となり、母回帰係数の不偏推定量となっていることがわかる。それぞれの分散は、

$$\sigma^2(\hat{\alpha}) = \sigma^2 \frac{\sum t_i^2}{n \sum (t_i - \bar{t})^2}, \sigma^2(\hat{\beta}) = \sigma^2 \frac{1}{\sum (t_i - \bar{t})^2} \tag{A. 6}$$

で与えられる。ここで、 $\sigma^2$  は知ることのできない母分散であるので、その推定値として次の標本分散を使う。

$$\hat{\sigma}^2 = \frac{\sum \{x_i - (\hat{\alpha} + \hat{\beta} t_i)\}^2}{n - 2} \tag{A. 7}$$

標本分散は推定された直線のまわりのデータのばらつきの指標であり、直線の推定で自由度が2使われているので、分母の自由度は  $n - 2$  となる。

次に、推定した回帰直線がどの程度正しいかを考える。今の場合、推定された線形モデルは、

$$x_0 = \hat{\alpha} + \hat{\beta} t_0$$

と表される。線形モデルの出力値である  $x_0$  の分散は、

$$\sigma^2(x_0) = \left( \frac{1}{n} + \frac{(t_0 - \bar{t})^2}{\sum (t_i - \bar{t})^2} \right) \sigma^2$$

となる。これまでと同様に知ることのできない母分散を標本分散で置き換えて推定値とすると、

$$\begin{aligned}\hat{\sigma}^2(x_0) &= \left( \frac{1}{n} + \frac{(t_0 - \bar{t})^2}{\sum (t_i - \bar{t})^2} \right) \hat{\sigma}^2 \\ &= \hat{\sigma}^2(\hat{\beta}) \left\{ t_0^2 - 2\bar{t}t_0 + \frac{\sum (t_i^2)}{n} \right\}\end{aligned}$$

となり、これから  $x_0$  の信頼区間は、 $t$  分布を用いて以下のように表される。

$$\begin{aligned}\hat{\alpha} + \hat{\beta}t_0 - t_\alpha(n-2)\hat{\sigma}(x_0) < x_0 < \hat{\alpha} + \hat{\beta}t_0 \\ + t_\alpha(n-2)\hat{\sigma}(x_0)\end{aligned}\tag{A.8}$$

観測点の変位速度は回帰直線の傾きで表されるので、母回帰係数  $\beta$  の区間推定について考える。  $\beta$  の標本分散を標準化した

$$\frac{\hat{\beta} - \beta}{\sqrt{\hat{\sigma}^2 / \sum (t_i - \bar{t})^2}}$$

は自由度  $n-2$  の  $t$  分布に従うので、次の式から  $t$  分布を用いて信頼水準  $\alpha$  で母数  $\beta$  の区間推定が行える。

$$\begin{aligned}\hat{\beta} - t_\alpha(n-2)\sqrt{\frac{\hat{\sigma}^2}{\sum (t_i - \bar{t})^2}} < \beta < \\ \hat{\beta} + t_\alpha(n-2)\sqrt{\frac{\hat{\sigma}^2}{\sum (t_i - \bar{t})^2}}\end{aligned}\tag{A.9}$$

次に、2次元平面における地殻変動を考える。この場合、 $(x, y)$  平面において時間  $t$  とともに直線的に移動する点を想定し、 $x$  と  $y$  のそれぞれについて回帰直線を推定する。

$$\begin{aligned}x &= \alpha_x + \beta_x t \\ y &= \alpha_y + \beta_y t\end{aligned}$$

各成分についての推定は1次元のときと同様であるが、2次元の場合さらに共分散を考える必要がある。傾きの標本共分散は次式で与えられる。

$$\begin{aligned}\text{Cov}(\hat{\beta}_x, \hat{\beta}_y) &= \\ \frac{\sum \{x_i - (\hat{\alpha}_x + \hat{\beta}_x t_i)\} \{y_i - (\hat{\alpha}_y + \hat{\beta}_y t_i)\}}{n-2} & \frac{1}{\sum (t_i - \bar{t})^2}\end{aligned}\tag{A.10}$$

回帰係数がある確率で存在する範囲を表す楕円は、(A.4)式における分散共分散行列の行列要素を回帰係数の分散と共分散で置き換えたもので与えられる。

## 文 献

- 藤田雅之・矢吹哲一郎 (2003) 海底地殻変動観測における K-GPS 解析結果の評価手法について、海洋情報部技報, 21, 62-66.
- 藤田雅之・佐藤まりこ・矢吹哲一郎 (2004) 海底地殻変動観測における局位置解析ソフトウェアの開発、海洋情報部技報, 22, 50-56.
- 藤田雅之・佐藤まりこ (2004) 海底地殻変動観測における局位置決定再現性の評価、海洋情報部研究報告, 40, 85-92.
- 石川直史・佐藤まりこ (2012) 海底地殻変動観測における重心推定法の評価、海洋情報部研究報告, 48, 74-84.
- 松本良浩・藤田雅之・石川直史 (2008) 海底地殻変動観測における複数エポック一括局位置解析手法の導入、海洋情報部技報, 26, 16-22.
- 佐藤まりこ・浅倉宜矢・齋藤宏彰 (2009) 船底トランスデューサによる航走海底地殻変動観測の評価 (速報)、海洋情報部技報, 27, 56-65.
- 佐藤まりこ・藤田雅之 (2012) GPS/音響測距結合方式による海底地殻変動観測技術の進展—この10年の歩み—、海洋情報部研究報告, 48, 26-40.
- 田中秀幸・高津章子 (2014) 分析・測定データの統計処理—分析化学データの扱い方—、朝倉書店、東京.
- 東京大学教養学部統計学教室 (1991) 統計学入門、基礎統計学 I、東京大学出版会、東京.

## 要 旨

GPS-A 音響測距結合方式による海底地殻変動観測は、複数の技術を結合させた測位技術であるため、個別の誤差要因を定量的に評価することが難しい。そのため、本観測の測位精度の評価として、繰り返し観測の再現性評価を行っている。ここでは、現在行っている評価手法について、必要となる統計学の基礎事項とともにまとめた。

また、本評価手法の応用例として、観測データ数と測位精度との関係についての検討を行った。