地磁気観測所テクニカルレポート　第１巻第１号　1-9頁　平成15年３月
Technical Report of the Kakioka Magnetic Observatory Vol.1, No.1, pp.1-9, March 2003

地磁気絶対観測における基線値の異常値判定基準の定量化に向けて

伊藤信和（鹿屋出張所）・藤井郁子（調査課）

2003年１月31日受付，2003年３月３日改訂，2003年３月７日受理

要　　旨

地磁気絶対観測において観測基線値に含まれる異常値を客観的に検出するため，ロバスト統計学を適用して，異常値の判定基準を定量化することを試みた．使用したデータは，鹿屋出張所で2000年２月25日～12月31日に観測されたH，D，Z成分の観測基線値データで，H・Z成分は58回分464個，D成分は72回分576個である．各回の観測の標本数は８個程度と少数でありロバスト統計を適用するのは難しかったが，1回の観測ごとに観測基線値から中央値を除いた残差を解析期間中にわたって集めた標本集団は，少数の標本を除いて正規分布的であることがわかった．そこで，約１年分の残差を使って異常値検出を行った．

約１年分の残差のQuantile-Quantileプロット（QQプロット）を作成し，標本分布が異常値のため正規分布的でなくなる部分を調査した．QQプロットの連続性が途切れるところをしきい値と設定したところ，H，Z成分で中央値からの差の大きさが中央絶対偏差の７倍，D成分で５倍以上を異常とみなすことになり，異常と判定された標本数はH成分が８個，D成分が13個，Z成分が12個であった．異常値を除いて計算した残差の平均と標準偏差を使って，異常値と平均との差を標準偏差で規格化してみると，本報告で用いた異常値判定は標準偏差の３倍と同程度かやや厳しい基準に基づいていることがわかった．

ロバスト推定の考え方を導入することにより，客観的に異常値を検出でき，また１回ごとでは埋もれていた異常値の検出も可能になることが示唆された．本報告で扱った2000年の観測基線値データに関しては，中央絶対偏差の５～７倍が異常値判定のしきい値の目安になると考えられるが，さらに精度の良い見積もりのためにはデータ数を増やしてしきい値を確定する必要がある．

１．はじめに

地磁気観測所では，ある時刻の地磁気ベクトルを観測するために，ベクトルの変化量を連続的に測る変化観測と，基線値を得ることを目的とした絶対観測を行っている．絶対観測はおよそ週１度の頻度で行われ，プロトン磁力計及び角度測定器を用いて地磁気ベクトル３成分の値を測定し，変化観測で得られた同時刻の地磁気と比較して観測基線値を導出している．

１回の絶対観測では，３成分ごとに原則として８個の観測基線値を平均することによって，その回の基線値を求めている．しかし，８個の観測基線値の中に何らかの原因で異常な値が含まれていると，真の基線値を推定するのに障害となる．そのため異常値を検出し取り除く操作（オミット）や再観測が取り入れられているが，異常値の検出に定量的な基準がないため観測者によるオミットの判定基準のばらつきが指摘されてきた．

本報告では，異常値が含まれるデータのためのロバスト推定（Huber，1981）の考え方を実際の観測基線値データに適用し，異常値のオミット基準を定量化することを試みた．

２．ロバスト推定

観測値の集合に異質な性質をもった異常値が含まれているとき，異常値の影響を除いた観測値分布の性質を調べるロバスト統計学（Huber，1981）が有用となる．本報告で適用した手法を，異常値が含まれる観測値分布のパラメーター，異常値の検出法，異常値の影響の除去法にわけて説明する．

２．１　異常値が含まれる観測値分布のパラメーター

ある物理量Uを推定するため，n回測定をくり返

し観測値（標本）の集合 $\{x_i\}$，i＝1～nを得たとする．観測エラーは平均0，標準偏差Eの正規分布に従うランダムエラーだけとすると，物理量Uと測定のばらつきの尺度Eの最も良い推定量は，それぞれ，標本平均μと標本標準偏差σで与えられる．ただし，

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i,\quad \sigma^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\mu)^2.$$

しかし，観測エラーにEの分布に従わない異常値が含まれている場合，μとσは異常値も演算の中に取り入れて計算するので，異常値の振幅によってはUやEからかけ離れた値になる可能性がある．

異常値の影響を受けにくい統計的な推定量をロバスト推定量という．上述のように，μとσはロバストな推定量ではない．一方，中央値と中央絶対偏差（Median Absolute Deviation：MAD）は異常値の影響を受けにくく，それぞれUとEのロバストな推定量であることが知られている．ここで，中央値Mとは，n個の標本 x_i を大きい順に並べたときのn/2番目の値であり，中央絶対偏差は x_i とMの残差の絶対値 $|x_i - M|$ を0を除いて大きさ順に並べたときの中央値と定義される．

MとMADがロバスト推定量であることを，例を使って示す．下に示した標本集合は鹿屋出張所において2000年4月19日の絶対観測で求められたH成分の観測基線値（単位：nT）である．

{33.33，34.61，34.54，34.62，34.41，34.68，34.79，34.59}

この8個の標本の平均値 μ_1，標準偏差 σ_1，中央値 M_1，中央絶対偏差 MAD_1はそれぞれ，34.45，0.46，34.60，0.07nTである．次に，1個目の値33.33は本報告の解析の結果，異常であるとみなされた値であるので（第4章参照），この値を除いて7個の平均値 μ_2，標準偏差 σ_2，中央値 M_2，中央絶対偏差 MAD_2 を求めると，それぞれ，34.61，0.12，34.61，0.07 nTとなる．

このとき平均値の変化率は0.46%であるが，中央値の変化率は0.03%である．この例より，中央値は平均値ほど異常値の影響を受けないことがわかる．平均値は0.16変化したが，異常値を除去した後の標準偏差は0.12であることから，この例における平均値の変化は標本のばらつき具合に対して有意な量である．標準偏差と中央絶対偏差も同様で，σ_1と σ_2には約4倍の差があり標本のばらつき具合がまったく異なる分布の印象を与えるが，MAD_1と MAD_2を比べると差がなく，異常値による影響が小さいことがわかる．

次にこの標本集合において中央値が真の基線値を代表しているかどうかを考える．異常値を除去した標本集合が本来の分布に従ったものとすると，中央値は異常値除去後の平均値に近い値になることから，中央値はこのような異常値を含む標本集合の真の基線値の推定量としての役割を果たすと考えられる．

2．2　異常値検出法

μやσを用いて標本に含まれる異常値を検出する方法がある．ある標本集団 x_i が正規分布に従う場合，x_i がμ±3σから外れる確率は0.3%しかないことから，

$$|x_i - \mu| > 3\sigma \qquad (1)$$

を満たす x_i を異常値とする方法（3σエディット・ルール）や，そのしきい値が標本数によって変化するトンプソンの棄却検定などである．しかし，前節で示したようにμとσは異常値のため標本の分布を正しく示さないことがあるので，これらの方法はしばしば破たんすることが指摘されている（Pearson, 2002）．

μとσのかわりにロバストなMとMADを使えば，異常値が含まれる標本集団でも標本の分布を客観的・定量的に見積もることができる．3σエディット・ルールを構成するμとσをそれぞれMとMADに置き換たものは，ハンペルの判定法と呼ばれる（Pearson, 2002）．本報告も同様の手法で解析を行っているが，しきい値を3MADに固定せず，

$$|x_i - M| > kMAD \qquad (2)$$

を満たす x_i を異常値とした．kはしきい値を決定するパラメーターである．

2．3　異常値の影響の除去法

2．2節の方法で標本の中に異常値が検出されたとき，異常値の影響を除いて標本集合の統計的な性質を求める方法には，大きくわけて2種類ある．一つは，異常値を完全に取り除くことで標本 x_i がUの本来の分布に従うようにするhard rejection法である．現行のオミットはこの方式である．二つ目は，異常値に重みをつけることで標本分布を本来の分布に近付けるsoft rejection法である．

本報告では異常値の除去法について特に検討しなかった．

3．観測基線値の性質

図1に，本報告で扱う鹿屋出張所の2000年2月25

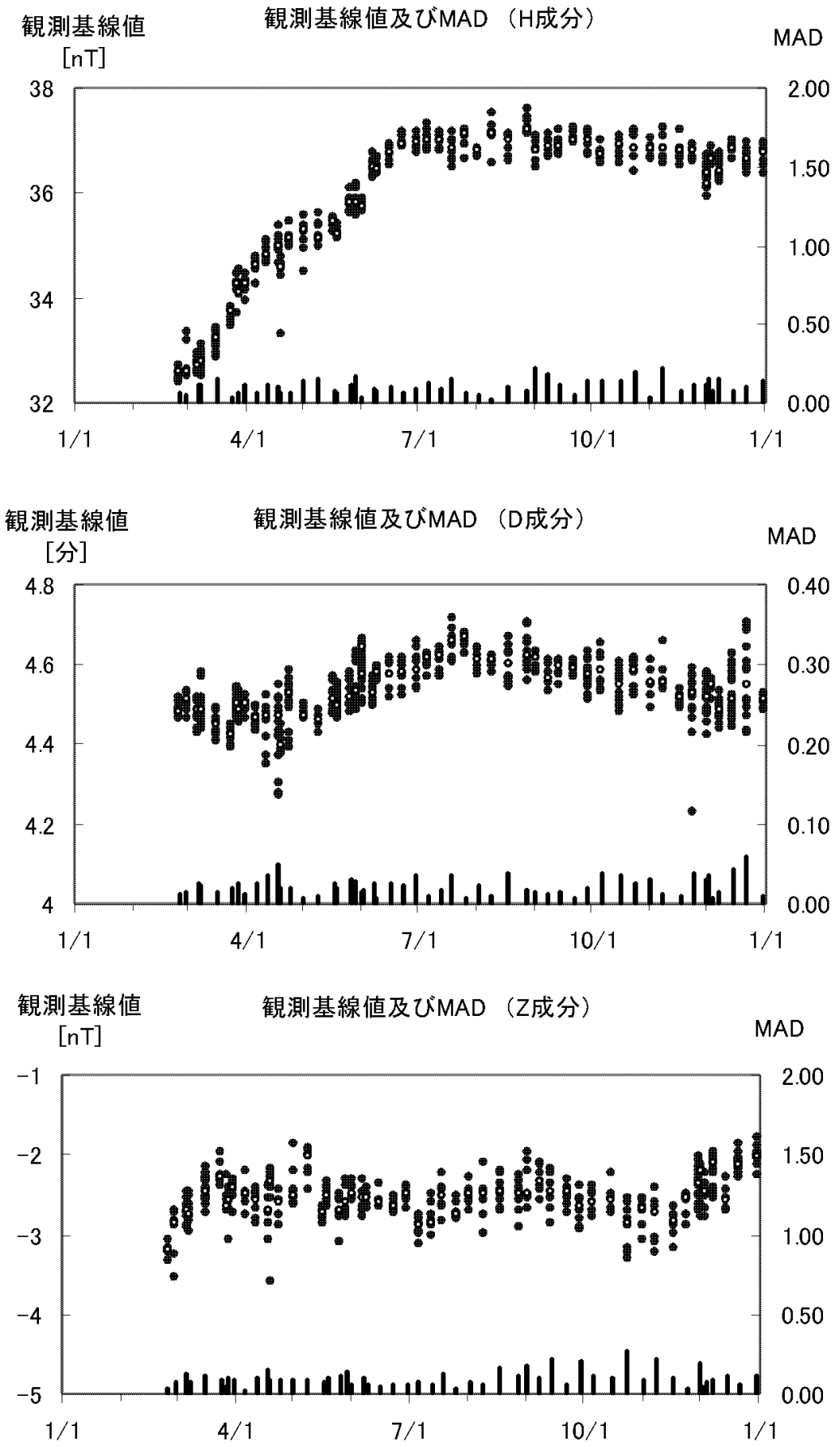

図１　2000年２月25日～12月31日に鹿屋出張所で観測されたＨ（上），Ｄ（中），Ｚ（下）成分の観測基線値，中央値及びMADの時間変化．グラフ上段には観測基線値を黒丸，中央値を白丸で，また下段にはMADを棒グラフで示す．

日～12月31日の観測基線値，それらの中央値，およびMADを示す．ここでは異常値と思われるデータが存在することがわかっているので，ロバストな推定量である中央値とMADを用いる．地磁気は大きさと方向を持つベクトル量であり，当観測所では水平成分（H），鉛直成分（Z），偏角（D）の３成分で表している．観測基線値の単位は，H・Z成分はnT，D成分は分である．

絶対観測はおおむね１週間ごとに行われ，原則として８個の観測基線値が得られるまで測定をくり返すが，所々で再観測が行われているため，その頻度や一回の観測での観測基線値の個数は一定ではない．図１の期間に絶対観測はH・Z成分で58回，D成分で72回行われ，得られた観測基線値の総数は，H・Z成分は464個，D成分は576個である．D成分の回数，個数が多いのは，再観測が多かったためである．

図１を見ると，中央値は永年変化や季節変化を反映して時間とともに変動しているが，一回ごとの観測基線値の集団は中央値のまわりのおおよそ同じような範囲に分布している．また，時折集団から外れたデータがあることがわかる．

一回の絶対観測の目的は，n個の観測基線値x_i，i＝1,...nから真の基線値Uを推定することである（x_iはH，Dあるいは Zを指す）．前述のように，現在の方式では基線値の推定量μはn個の観測基線値の平均によって求められている．n個のうちに他の観測基線値と大きく異なるデータが得られた場合は異常値とみなし，異常値を除いた平均によって基線値を得る．

異常値の検出基準の定量化のため，まず，観測基線値がどのような統計分布に従っているのかを調べた．図２はある平均的な一日の観測基線値が正規分布に従う仮定のもとに，確率密度関数$F(x_i)$

$$F(x_i)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right) \qquad (3)$$

を示したものである．この日の観測基線値は標本数が８個しかないため，ばらつきの大きい分布なのか，それともばらつきの小さい分布に異常値が含まれているのか，はっきりしない．

オミットのようなロバスト操作は，x_iが正規分布に従うことを仮定し，分布に合わないデータを異常と見なして除去することにより，残ったデータを正規分布と整合的にする操作と考えられる．しかし，図２のようにx_iがどのような分布に従うかはっきりしない場合は，正規分布を仮定した異常値検出では精度が下がることが予測される．例えば，図２の分布の場合，M＝34.25，MAD＝0.12なので，最も小さい観測基線値標本はMからMADの2.5倍離れていることになるが，より正規分布に近い標本集団でのMADの値0.07nT（第４章参照）を使うと4.3倍となる．仮に，式２のkを４とおくと，MADの値をどちらにするかによって，同じ標本でも異常かどうかの判断が分かれることになる．

図１から，観測ごとの中央値は時間変化するが，各回の中央値とn個の観測基線値の差は時間によら

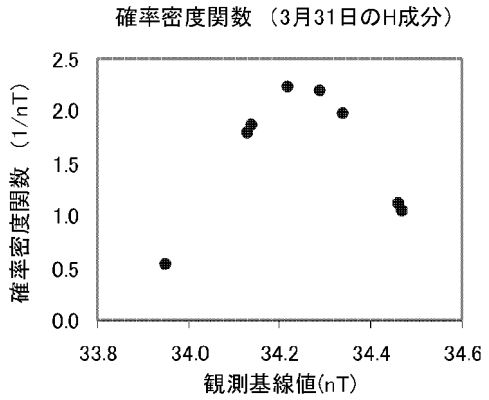

図２　３月31日に観測されたＨ成分の観測基線値の確率密度関数．

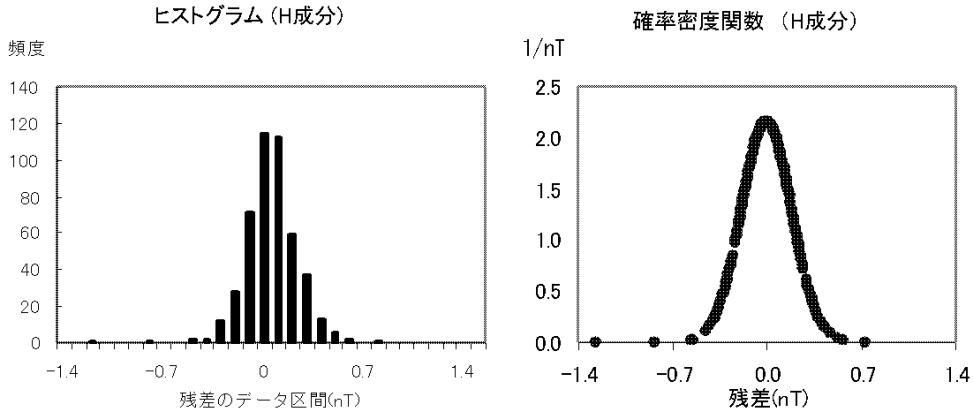

図３　2000年２月25日～12月31日に観測されたＨ成分の観測基線値のヒストグラム（左）と確率密度関数（右）．

ず同程度のばらつきをしていることが見て取れる．そこで，j回目の観測基線値をx_i^j，個数をn_j，中央値をx_M^jとしたとき，残差$\Delta x_i^j = x_i^j - x_M^j$をすべてのiとjについて計算し，一つの標本集合としてみたときの確率分布（ヒストグラム），及び確率密度関数を調べた．図３にH成分の場合を示す．ヒストグラムは0.1nTの間隔で個数を計上している．

図３の標本数は464個あるので，図２に比べて分布が鮮明になった．０を中心とした釣鐘型の分布に少数の振幅の大きいサンプルが重ねあわさった形をしており，確率分布が式３で定義される確率密度関数でおおむね表現できていることから，おおよそ正規分布的であることがわかる．これは，観測エラーの見積もりΔx_i^jが，少数の異常な場合を除いて，観測者の違いや季節の変化によらず均質である，言い換えれば，同じ分布に従うランダム変数であることを示唆している．

図３から，残差Δx_i^jの一年分の集合を使うことで標本数の不足を補い，残差分布のロバストなパラメータ推定を行う方法が考えられる．

４．解析

この章では，前章で説明した観測基線値データ中に含まれる異常値を検出するため，第２章のロバスト推定法を観測基線値データに適用する．前章で示されたように，一回ごとの観測基線値の集合では，標本数が少ないため分布のパラメーターの推定精度が下がり式２の異常値検出法がうまくいかない可能性があるが，残差Δx_i^jの一年分を使えば，標本数の不足を補って実効的な異常値検出ができると期待される．そこで，一年分の残差Δx_i^jを使ってMADを求め，式２を変形して，

$$\frac{|\Delta x_i^j|}{MAD} > k \qquad (4)$$

を満たす残差中の異常値を検出することを試みる．このとき，Δx_i^jはすでに中央値が０になるよう平行移動された値であるから，左辺上部の中央値との差の演算は省略できる．

まず，MADの値を計算する．一年分のΔx_i^jのMADは，H，D，Z成分でそれぞれ，0.07nT，0.022分，0.06nTであった．

次に，式２でしきい値を決めるパラメーターkの値を決定する．このときに，異常値やそのしきい値のおおよその値を視覚的に捕らえるための簡便な手段としてQuantile-Quantileプロット（QQプロット）がある．QQプロットとは，MADで規格化した残差$\Delta x_i^j/MAD$を小さい順に並べ，横軸に順番，縦軸に規格化した残差の値をプロットしたものである（図４）．残差が正規分布に従うとき，QQプロットの$\Delta x_i^j/MAD$は直線上に分布することが知られている（藤井・シュルツ，1999）．異常値は，直線から大きく外れた値としてプロットされる．

図４において，３成分とも図の中央部は直線的で，順番の最も小さい部分と大きい部分で曲線的，あるいは断続的になっている．明らかに異常と思われるものだけを選びだすような保守的な検出を目指

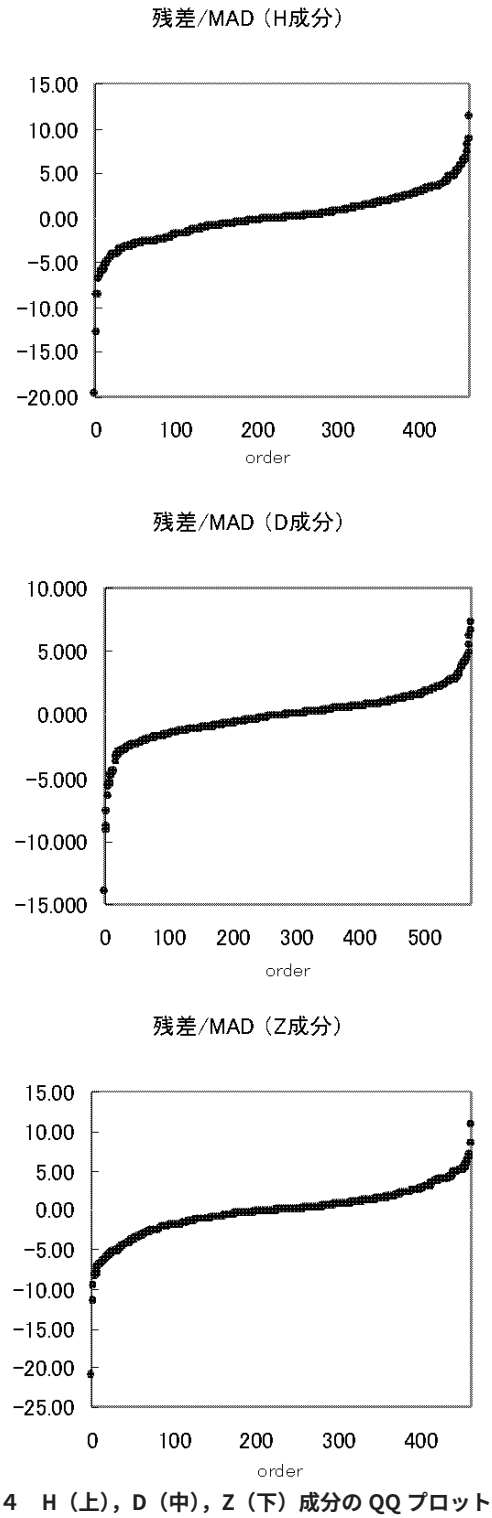

図４　H（上），D（中），Z（下）成分のQQプロット．

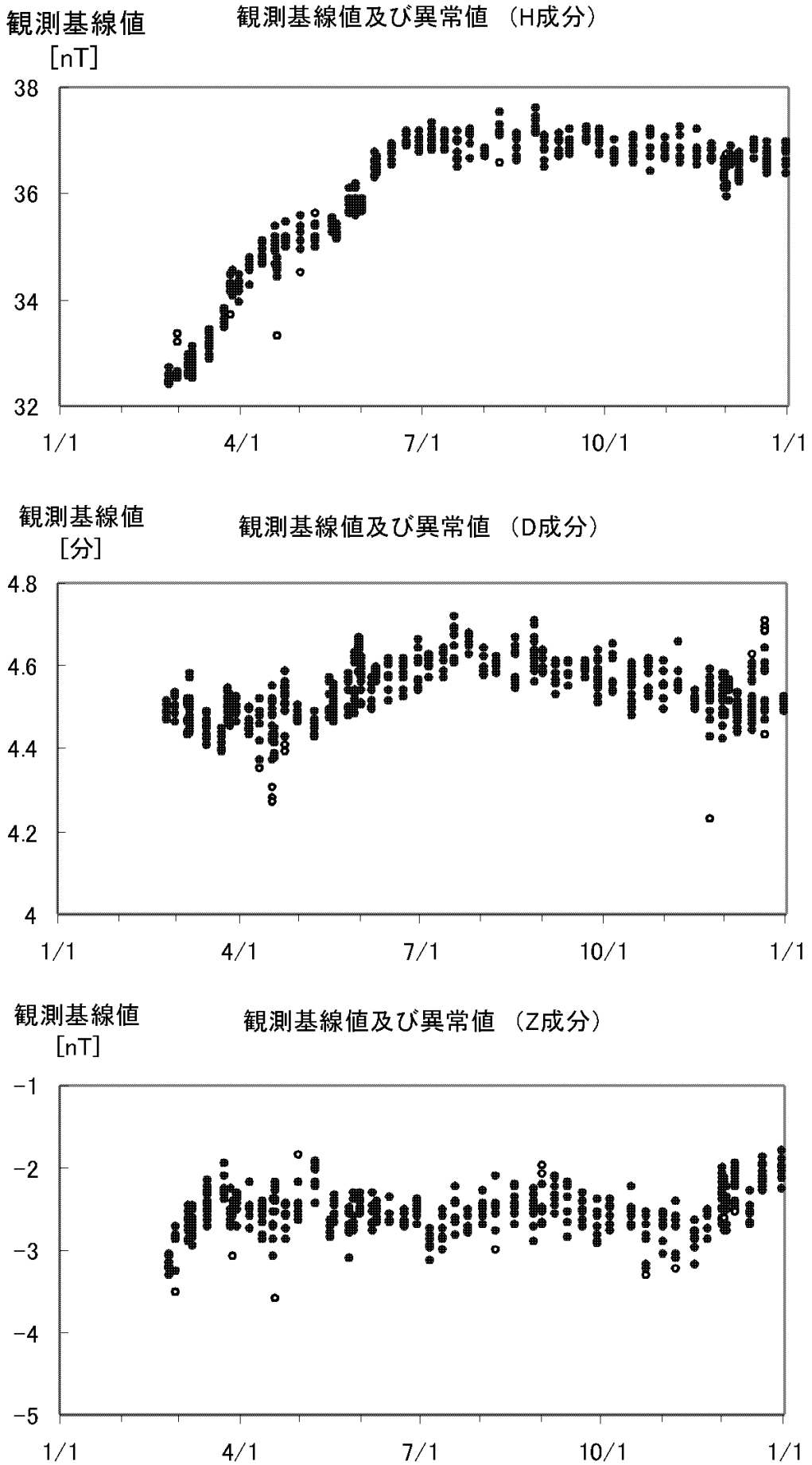

図5 H（上），D（中），Z（下）成分の観測基線値と異常値．観測基線値を黒丸，検出された異常値を白丸で示す．

すとすると，QQ プロットの連続的に並んだ標本からずれているものが候補になる．H・Z 成分は残差/MAD の大きさがおよそ７，D 成分はおよそ５で連続的でなくなっている．そこで H・Z 成分は k＝７，D 成分は k＝５をしきい値に設定した．これは，H・D・Z 成分の残差振幅が，それぞれ0.49nT，0.11分，0.42nT を越えるものを異常と判定するのと同値である．

このようにして異常値の検出を行うと，H 成分は８個，D 成分は13個，Z 成分は12個の計33個の残差が異常値と判定された．図５に，観測基線値と異常値を示す．明らかに異常と思われる観測基線値データだけでなく，その日の分布状況では判断するのが難しいデータも検出されていることが分かる．

k の値を小さくするほどしきい値が小さくなり，異常と判定される値が増える．ここで用いたしきい値がどのような意味を持つか知るために，３成分それぞれの残差から検出された異常値を除いて j 回目の平均値 μ^j と全体の標準偏差 σ を求め $x_i^j - \mu^j$ を調べてみた．33個のうち，H 成分７個，D 成分13個，Z 成分８個の計28個の異常値が ±3σ の範囲を超えた．これにより，本報告で採用した異常値判定は，3σ エディットルールと同程度かやや厳しい基準に基づくものであることが示唆される．

５．考察

５．１　実際のオミット状況との比較

第４章での異常値検出結果と実際に行われたオミットとを比較して，本報告での異常値検出法と現行の方法との差を考察する．

本報告で検出された33個の異常な観測基線値の実際の作業における取り扱い状況は，次の通りであった．再観測をせずにオミットされたのは H 成分１個（図４中の－19.54）と Z 成分１個（図４中の－21.00），再観測の対象になりオミットされたのは11個，オミットされなかったのは20個，の計33個である．

なお，本報告で異常と判定されなかった観測基線値の中で，D 成分の観測基線値データ（図４中の1.209）が単独でオミットされているが，理由は不明である．

オミットされなかった場合を見てみると，その観測日には，しきい値に近い観測基線値データが複数存在している場合が多かった．そのため，一回の絶対観測の中では判定できず，残っていたものと考えられる．

以上により，本報告の異常値判定法を用いると，判定の自動化・客観化に加えて，（１）再観測をせずに異常値判定ができる，（２）１回ごとの絶対観測では埋もれていた異常値の検出ができる，という２つの利点が得られる可能性が浮かび上がる．

５．２　しきい値について

本報告では QQ プロット図での連続性を考慮して，しきい値を H・Z 成分は k＝７，D 成分は k＝５と設定したが，しきい値決定により客観性を持たせるためにはさらに検討が必要である．

対象としているデータの性質にもよるが，藤井・シュルツ（1999）は，しきい値は経験的に３～12が採用されると述べている．一方，Xu（1989），徐（1998）は，そのしきい値は確率的には適切な意味を持たず任意に選ばれることから，ケース毎のロバスト推定の適用には限度があると主張している．しかし，第４章で述べたように，しきい値に対して異常値の影響を除いた残差の分布からおおよその意味付けを与えることも可能である．

本報告で扱った2000年の観測基線値データに関しては，しきい値 k＝５～７が一つの目安になると考えられる．オミット基準として実用化するためには，他の年のデータも調べて k の値を確定する必要がある．

５．３　標本数について

前節までに，ロバスト統計の適用により異常値を自動的に検出できることを示した．異常値を現行のオミットの対象とすれば，より真の値に近い基線値を求めることができると考えられるが，その際に問題となるのは標本数の減少であろう．第３章で見たように，１回ごとの観測基線値の数は正規分布を適用するには少ない状態にある．さらに標本数が減少すれば，最終的に得られる平均の基線値には大きなエラーバーがつくことになってしまう．オミット基準の策定に加えて，観測に必要な精度を確保するには最低何個の観測基線値が必要であるかを見積もり，オミットによりそれを下回った場合には再観測をするなどの規則の検討も必要であろう．

参考として，絶対観測で要求される精度を95%の確率で達成するのに必要なサンプル数 n を計算してみた．大きさ n の標本で標準偏差が E と分かっているとき，真の値 U と推定値 μ との最大誤差 U−μ は95%の確率で $1.96E/\sqrt{n}$ より小さくなる．H 成分の基線値の許容できる最大誤差を0.10nT，E はわからないので異常値を除いた標本標準偏差 σ＝0.16nT で代用すると，95%信頼限界の条件を満たす n は，

$$1.96 \times 0.16/\sqrt{n} \leq 0.10 \qquad (5)$$
$$\therefore n \geq 9.83$$

となる．従って，絶対観測のH成分が正規分布に従うとして，観測に要求される精度を95%の確率で満たすには，少なめに見積もって10個のサンプルが必要になる．精度を下げればnの値はもう少し小さくてもよい．

6．まとめ

鹿屋出張所のH，D，Z成分の観測基線値1年分を使って，Huber（1981）のロバスト統計の考え方で異常値を検出することを試みた．

異常値検出には，標本集団から中央値とMADを求め，各標本と中央値の差をMADで規格化することで，各標本が分布の中心からどのくらい離れているかを客観的に評価する．しかし，本報告で扱った絶対観測では，1回分の観測の標本数が一般に8個程度と少なく，中央値とMADの計算精度が悪くなるので異常値検出の精度があがらない恐れがあった．そこで，観測基線値と各回の中央値との差を1年分集めたものを標本集団として，異常値計算を行った．

1年分の残差はH・Z成分は464個，D成分は576個の標本で，0を中心としてほぼ正規分布し分布の端に少数の標本があった．H，D，Z成分のそれぞれのMADは0.07nT，0.022分，0.06nTであった．QQプロットにより，分布の連続性が途切れるところを異常値判定のしきい値と設定したところ，H，Z成分で中央値からの差がMADの7倍，D成分で5倍以上を異常とみなすこととなった．

このようにして検出された異常値は，H，D，Z成分でそれぞれ，8個，13個，12個あった．異常値を除いてから計算した残差の平均と標準偏差を使って，異常値と平均との差を標準偏差で規格化してみると，3と同程度かやや小さいしきい値を用いたのと同等であった．従来の異常値判定と比較すると，13個はどちらでも異常と判定されていた．今回新たに異常とされた20個について調べてみると，1回の絶対観測中に複数のしきい値に近い観測基線値が存在している場合が多かった．

このように1年分の観測基線値データを使って基準を設定して異常値の判定をしてみると，1日ごとでは埋もれていた異常値の検出も可能になり，ロバスト推定がオミット作業の効率化に役立つ可能性があることが示唆される．今後の課題としてもっと多くの事例を解析して，観測基線値に対するこの異常値検出法の客観性をより高めていくことが求められる．

参考文献

藤井郁子，シュルツ，A.，地球磁場観測ネットワークデータの解析手法について（その1），CA研究会論文集，97–104，1999.

Huber, P.J., Robust Statistics, John Wiley, New York, 308pp, 1981.

Pearson, R.，非線形フィルターでデータを洗浄する，EDN Japan，2002.

徐　培亮，ロバスト推定の信頼区間と定数の重要性，地球惑星科学関連学会合同大会予稿集，93，1998.

Xu, P.L., Statistical Criteria for Robust Methods, ITC Journal, No.1, 37–40, 1989.

On automatic determination of criteria to detect outliers for the absolute measurement of the geomagnetic field

by

Nobukazu Ito[1], and Ikuko Fujii[2]

[1]**Kanoya Magnetic Observatory**
[2]**Kakioka Magnetic Observatory**

Received 31 January 2003; received in revised form 3 March 2003; accepted 7 March 2003

Abstract

We propose a procedure for automatic determination of criteria to distinguish spurious baseline values in the absolute measurement of the geomagnetic field based on a robust statistical method. The data used in this study are baseline values of the horizontal force (H), declination (D), and vertical component (Z) observed at Kanoya Magnetic Observatory from February 25, 2000 to December 31, 2000. The absolute measurement sessions were conducted 58, 58, and 72 times for H, Z, and D, respectively, in the period of interest. Eight baseline values were obtained in each session, giving a total of 464 baseline values each for H and Z, and 576 for D. The data set of each session has too few samples to be analyzed with a robust procedure, so the whole data set comprising 464 or 576 values was used. After the median was subtracted from the data of each session, a set of the residuals for all sessions was analyzed statistically. As a result, we found that the residuals of all sessions follow a Gaussian distribution, except for a small portion of the data set.

Quantile-Quantile (QQ) plots of the residuals normalized by the median absolute deviation (MAD) were made for each component and were used to estimate the outliers. We assumed that the outliers are normalized residuals larger than 7 for H and Z, and 5 for D, at which discontinuities were seen in the QQ plots for each component. As a result, 8, 12, and 13 outliers were detected for H, Z, and D, respectively. These criteria are slightly more robust than those estimated as three times the standard deviation that are computed from the data set without the outliers.

The results of the robust procedure used in this study indicate that the outliers can be automatically detected and that the procedure could detect outliers hidden in each session. For the baseline values in the year 2000, 5 and 7 times the MAD are good estimates of the criteria to detect the outliers. Further investigation on the robust criteria by analyzing data for other time periods will be required to improve the accuracy of the procedure.