

教師なし HMM に基づいて ELF 帯磁界の波形画像から 異常検出する手法への背景信号除去の導入を含めた再検討

毛利 元昭（愛知大学経営学部）

安川 博（愛知県立大学情報科学部）

内匠 逸（名古屋工業大学情報工学科）

要旨

地殻活動に伴う電磁界異常が報告されている。大地震の前兆としての電磁界異常の検出を目指し、筆者らが所属する研究グループでは極超長波（Extremely Low Frequency, ELF）帯磁界を全国で観測してきた。そのデータを可視化した画像から異常検出する手法として、隠れマルコフモデル（Hidden Markov Model, HMM）を利用したものが過去に提案された。しかし、HMM へ入力する特徴量を正規化することの是非や、HMM のベースである状態遷移モデルで重要な入力順序についてなど、未確認の点も多かった。また、ELF 帯磁界観測データが有する日変化や年変化の影響について論じられていなかった。そこで、それら未確認の点について調査するとともに、背景信号除去を導入するなど手法の再検討を行った。コンピュータシミュレーション実験より、主に偽陰性（見逃し）を減らす場合は、背景信号除去あり、特徴量の正規化なし、画像の下から順という形に手法を変更することが効果的であることが示唆された。また実際の ELF 帯磁界データを用いた実験を通して、手法の改善とともに、背景信号除去の有用性を確認した。

キーワード： 環境電磁界、地殻活動、異常検出、隠れマルコフモデル、信号推定

1. はじめに

日本では大地震が度々発生し、その度に甚大な被害が生じている。地震の発生予測を正確に行うことができれば、被害の減少や早期の復興に繋げることができる。ここで、伝統的に行われている地震発生予測は、活断層に対するトレンチ調査や地震発生周期、GPS での地表歪み評価に基づいたもの¹⁾であるが、その予測

として出される情報は、今後数年から数十年という期間における発生確率と被害規模であり、調査対象は特定の地震のみである。より時間解像度が高く、対象を限定しない地震発生予測を行うには、異なるアプローチが必要である。

地殻活動に伴う電磁界異常が報告されており²⁾、地震や火山活動の先駆けとされたものもあった。その検出を目的として、大地電位・電流を観測・解析する研

究や、電磁波の伝搬状況を観測・解析している研究もなされている^{3, 4, 5)}。筆者らが所属する研究グループでは、極超長波 (Extremely Low Frequency, ELF) 帯 223 [Hz] の電磁波について、東西方向・南北方向それぞれの磁束密度として全国各地で継続的に観測し、その絶対値をさらにダウンサンプリングしたもの（以下、観測信号）を時系列波形のグラフとして画像化し、ウェブ上で公開してきた^{6, 7, 8, 9, 10)}。公開画像の例を図 1 に示す。縦軸は磁界の強度 $[pT/\sqrt{Hz}]$ を、横軸は時間を表している。波形は解像度が低くアンチエイリアシングもなされていない赤線で表されている。これらの画像は、ウェブサーバによって要求に応じて gnuplot で作成され、GIF 形式で表示・ダウンロードできる⁶⁾。

観測信号は、磁気圏・電離圏や雷放電、電波塔や鉄道など様々な自然現象・人間活動に由来する電磁波放射が重畳した状態で得られる。また、時間帯や季節などによる傾向（日変化や年変化）も有している。例えば、図 1 には夜間に高く昼間に低いという日変化が、観測地点に依らず見られる。図 2 に示す、観測信号の度数分布をマップ表示（黒いほど値が大きい）した例からは、夏季に高く冬季に低いという年変化が見られる。このような日変化や年変化は、地殻活動とは関係のない背景的な電磁波放射によるものと考えられている。また、いずれにも突発的

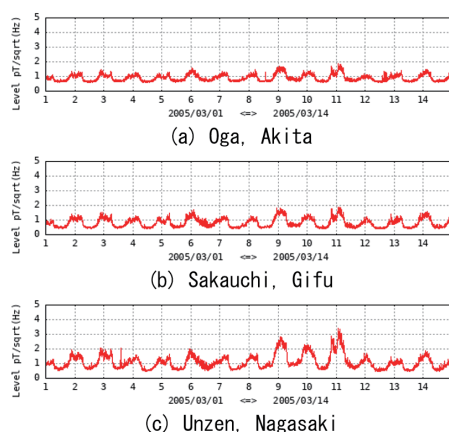


図 1: 公開用に画像化した観測信号の例
(14 日間, 3 地点・南北方向)

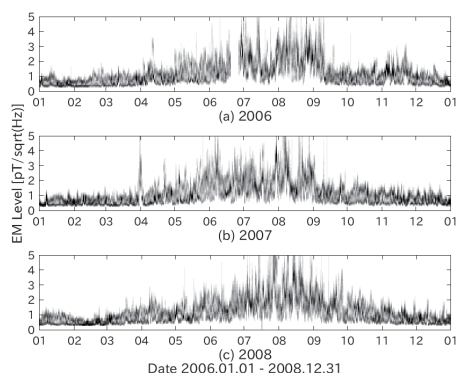


図 2: 観測信号の度数分布をマップ表示した例（3 年間, 1 地点・宮城県栗原市・南北方向）

に高い値を持つ場合が見られる。確度の高いデータ解析のためには、地殻活動や地震との関連している可能性のある異常信号を選別する必要がある。当初は、公開していた画像を人の目で確認・評価し、平常時のデータと異常を含むデータを判別・分類していた。それは膨大な作業で

あり、また作業者の経験や知識に大きく左右されるものでもあったため、自動化が求められていた。

これまで、筆者らの研究グループでは、様々な異常検出手法を試みてきた。例えば、過去年度における平均的な観測波形を算出し、それとのズレを評価する平年値法がある¹⁰⁾。この手法はシンプルで扱いやすかったものの、同一観測地点における過去複数年に渡る膨大なデータが必要となり、新規に設置した観測地点では利用できない、あるいは検出精度が低下するものであった。また、平常時データに線形予測分析を適用して平均的な線形予測係数を算出しておき、それを用いて対象データを線形予測した際の誤差を評価する手法も提案された¹¹⁾。この手法は異常の検出精度が高かったものの、その結果と地殻活動との関連性が決して高いものではなかった。他にも、多層ニューラルネットを利用した信号の弁別手法も提案された¹²⁾。しかし、そのモデルの構築には地震との高い関連が疑われるデータを多数必要とするものであり、その数が十分に得られているとはいえない状況もあり、十分な結果が得られるものではなかった。

これら先行研究の一環として、伊藤および浦田らによって、ウェブ上に公開されているグラフ画像から隠れマルコフモデル (Hidden Markov Model, HMM) を利用して異常検出を行う手法が提案され

た^{13, 14, 15)}。この手法に対して、筆者らは多くの疑問とともに、再検討の必要性を見出した¹⁶⁾。例えば、HMM は状態遷移を伴うことから、シンボルの入力順序についての検討は重要であるが、それがなされていなかった。また、日変化や年変化が異常検出に及ぼす影響について十分に論じられていなかった。

そこで、本論文では、伊藤および浦田らによって提案された手法 (以下、従来手法) について再検討を行いつつ、その改善を図る。本論文は全 6 章からなる。第 2 章では、従来手法の概要を述べる。第 3 章では、背景信号の除去手法の導入を含めた、従来手法に対する再検討について述べる。第 4 章では、異常検出手法をコンピュータシミュレーション実験によって検証する。第 5 章では、実際の ELF 帯磁界データに対して手法を適用し、結果を比較する。最後に、第 6 章にてまとめと今後の課題について述べる。

2. 教師なし HMM ベースに基づいて ELF 帯磁界の波形画像から異常検出する手法 (従来手法)

伊藤らによって提案された、教師なし HMM に基づいて ELF 帯磁界の波形画像から異常検出する手法^{13, 14, 15)}は、パターン認識技術がベースになっている。図 3 および以下にその流れを示す。

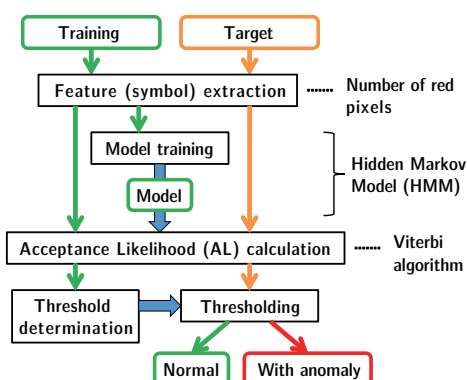


図 3: 異常検出の流れ

2.1. 特徴量の抽出

まず、公開されている ELF 帯磁界の波形画像の中から典型的なものを多数選び、それらの特徴量を抽出する。このとき、HMM におけるトレーニングデータや入力として扱うため、抽出される特徴量は離散的なシンボル列となるもの考える必要がある。従来手法では、解像度が低くアンチエイリアシングもなされていない波形画像のピクセルを利用して簡易的に度数分布を算出し、それをシンボル列としていた。14 日間分のデータを 0 から 5 [pT] の範囲でグラフ化したものが最適であるとされた。図 2 から分かるように、夏季においてもほとんどの観測信号がその範囲に収まるためである。加えて、浦田らの報告¹⁵⁾によれば、14 日間分が描画された画像から 2 日間分の波形画像を切り出し、それを 1 日ごとにスライドさせつつ特徴量を抽出する形が良い

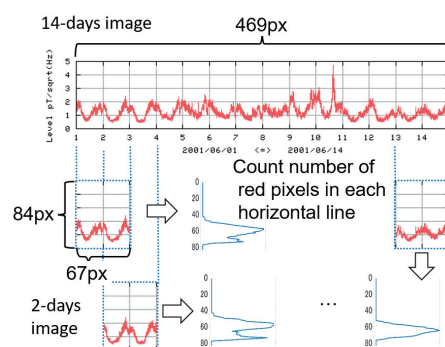


図 4: 波形画像からの特徴量抽出の流れ (14 日間画像から 2 日間分を切り出し)

とされた。これらに基づき、従来手法における波形画像データから特徴量を抽出する流れを図 4 および以下の箇条書きに示す。

1. 14 日間分の波形データを描画する (このときの、グラフの描画領域のサイズは横が 469 [px], 縦が 84 [px])
2. 1 の画像から 2 日間分 (横が 67 [px], 縦が 84 [px]) を切り出す
3. 2 の画像にて、それぞれの水平走査線上にある赤いピクセルの数を数えて特徴量の元 (シンボル列 s) とする item 抜き出し位置を 1 日間分ずらしながら、1 から 3 を繰り返す

従来手法では、切り出し幅の検討のためか、シンボル列 $s = (s_1, s_2, \dots, s_{84})$ に対し以下の式による正規化が適用された s' が、実際の特徴量として用いられていた。

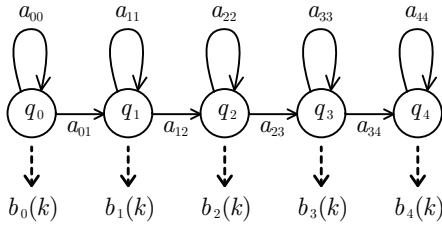


図 5: left-to-right 型 HMM の状態遷移図
(状態数 5 の例)

$$s' = \left\lfloor \frac{31 \times s}{\max_i \{s_i\}} \right\rfloor + 1 \quad (1)$$

ここで、床関数により、 s' は $1 \leq s'_i \leq 31$ となる整数の値しか持たない離散的なシンボル列（要素数 84）となる。つまりこれは HMM への入力として利用可能である。また、シンボル列の並びは、画像で言えば水平走査線の上から順となっていた。

2.2. 教師なし HMM のトレーニング

次に、典型的なデータから抽出された特徴量を用いて HMM を教師なしでトレーニングし、典型的なモデルを得る。HMM とは確率的な状態遷移と確率的なシンボルの出力を備えたオートマトンのモデルである。まず、有限個の状態 q_i が連鎖した系列があり、その中を確率的に (q_i から q_j へ確率 a_{ij} で) 遷移していく。遷移のたびに、遷移後の状態 q_j が持つ確率分布 $b_j(k)$ に基づいてシンボル k を 1 つ出力する。このように、遷移を

繰り返しながらシンボル列が出力されていく系列を表したものである。従来手法では、図 5 に示すような、単純な left-to-right 型の HMM を利用していた。HMM のパラメータは、まず状態数を任意に設定した後、状態遷移確率 a_{ij} および出力確率分布 $b_j(k)$ について、Baum-Welch アルゴリズムを用いてトレーニング用のデータから推定する。Baum-Welch アルゴリズムは、与えられたシンボル列を用いて、それぞれの状態について前向き確率と後ろ向き確率をそれぞれ求めながら、EM アルゴリズムによってパラメータを推定する手法である。 N 個の状態 $\mathbf{q} = (q_1, q_2, \dots, q_N)$ を仮定し、与えられたシンボル列を $\mathbf{s} = (s_1, s_2, \dots, s_T)$ とした際の、パラメータ推定の過程を以下に簡潔に記す。

● 前向き確率の算出

時刻 t において状態 q_i にあり、その時刻までに出力されるシンボル列が $s_{1,t} = (s_1, s_2, \dots, s_t)$ となる確率 $\alpha_t(i)$ は、次の式で算出する。

$$\alpha_t(i) = \sum_{j=1}^N \alpha_{t-1}(j) a_{ij} b_j(s_t) \quad (2)$$

● 後ろ向き確率の算出

時刻 t において状態 q_i にいるとして、時刻 $t+1$ から最後までに出力されるシンボル列が $s_{t+1,T} =$

$(s_{t+1}, s_{t+2}, \dots, s_T)$ となる確率 $\beta_t(i)$ は、次の式で算出する。

$$\beta_t(i) = \sum_{j=1}^N \beta_{t+1}(i) a_{ij} b_j(s_{t+1}) \quad (3)$$

● パラメータの更新

状態遷移確率 a_{ij} および出力確率分布 $b_j(k)$ は、Kronecker のデルタを用いて次の式で更新する。

$$a_{ij} \leftarrow \frac{\sum_{t=1}^T \alpha_t(i) \beta_{t+1}(j) a_{ij} b_j(s_{t+1})}{\sum_{t=1}^T \alpha_t(i) \beta_t(i)} \quad (4)$$

$$b_j(k) \leftarrow \frac{\sum_{t=1}^T \delta_{s_t, k} \alpha_t(j) \beta_{t+1}(j)}{\sum_{t=1}^T \alpha_t(i) \beta_t(i)} \quad (5)$$

HMM のパラメータ推定に先立って、その状態数を設定しておく必要がある。浦田らの報告¹⁵⁾では、実験結果から今回の目的に最適な状態数は2とされていた。

2.3. 対数受理尤度の算出

与えられたシンボル列が、トレーニング済みのモデルに従うものであるかどうかの度合いを算出することが、異常の程度を評価することとなる。モデルとの合致度合いが高ければ平常時のデータだとみなすことができ、低ければ異常を

含むデータだと考えられる。HMM では Viterbi アルゴリズムを用いて尤度の形でモデルとの合致度合いを算出できるため、その値の対数を用いる。Viterbi アルゴリズムでは、時刻 t において状態 q_i にいることの尤度 $f(i, t)$ を、動的計画法を用いて次の式で計算する。

$$f(i, t) = \max_j \{f(i, t-1) a_{ji} b_j(s_t)\} \quad (6)$$

これにより、最も可能性の高い状態遷移系列上での生起確率を通して、受理尤度が近似される。従来手法では、こうして得られた尤度関数にて、時刻 T において状態 q_N にいることの対数尤度 $\log \{f(N, T)\}$ を、モデルに対するシンボル列の合致度合いとみなしていた。

2.4. 閾値による異常検出

最後に、評価対象のデータから得られたシンボル列について、トレーニング済みの HMM における対数受理尤度を計算し、それがある値（閾値）を下回った場合に異常として検出する。ここで、典型的な平常時のデータから得られたシンボル列については、異常として検出されない程度に設定する必要がある。

従来手法では、HMM のトレーニングに用いていない複数の平常時データについて対数受理尤度を算出し、それらの最小値を異常検出の閾値として設定していた。

3. 従来手法への疑問と再検討

3.1. 背景信号の影響について

第1章で述べたように、ELF帯の観測信号は様々な要因によるものが重畳した状態で得られる。筆者らが着目している地殻活動だけでなく、雷放電、太陽活動、人間活動などである。ここで、典型的な平常時の観測信号の時系列波形に見られる日変化および年変化（図1および図2は、日本から遠方にて巨大なエネルギーを伴って放射された電磁波によるものだと考えられる。観測地点に依らず、同時に同程度の大きさで観測されるこの信号を、筆者らは背景信号と呼んでいる。従来手法では、この背景信号を平常時データの特徴として扱っていた。しかし、小さな異常は背景信号に埋もれ、大きな異常に対しても、その性質を調査する際にノイズとなる可能性が高い。この背景信号を含む場合と取り除いた場合で、異常検出の結果を比較することが必要だと言える。

ELF帯の背景信号の除去手法は、以前に筆者らが提案した⁹⁾。その手法の流れを、図6および以下に記す。

1. 複数の地点における観測信号を多変量データと見なし、ブラインド信号分離（BSS）アルゴリズム¹⁷⁾を利用して源信号を推定する。

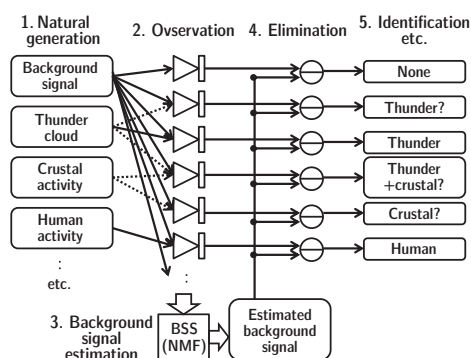


図 6: 背景信号の除去の流れ

2. 推定された源信号の中から背景信号によるものを特定する。現行では、観測信号への寄与の合計が最大のものを選ぶ。
3. 各観測信号への寄与を乗算した背景信号を、各観測信号から減算することで、観測点固有の信号を得る。

以後、この背景信号除去を適用した信号を局所信号と呼ぶ。

局所信号の例を図7および図8に示す。これらのグラフは、それぞれ図1および図2に対応するものであり、1日ごとに算出した局所信号から作成した。BSSアルゴリズムには筆者らが過去に提案したQL1-NMF⁹⁾を用いた。図7に着目すると、日変化がほぼ取り除かれていることが分かる。図8に着目すると、度数分布の範囲が狭まっており、夏季以外では年変化が抑制されていることが分かる。夏季には雷雲が多く発生し、それらに起因する電磁波はその近隣の地域のみで観測

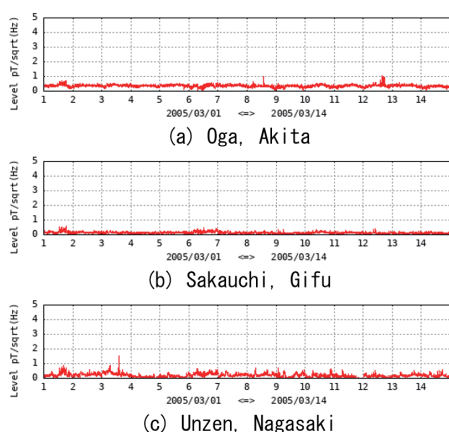


図 7: 局所信号の例 (図 1 に対応)

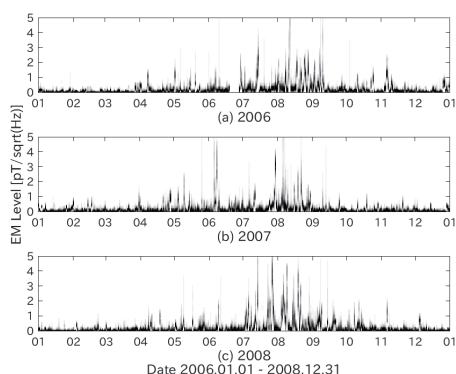


図 8: 局所信号の度数分布をマップ表示した例 (図 2 に対応)

できることから、それらが取り除かれず局所信号となった可能性が考えられる。

3.2. 特徴量の抽出について

局所信号から異常検出を行う際の注意点としては、特徴量の抽出を 1 日ごとのデータで行えるように、手法を変更する必要があることである。過去の経験か

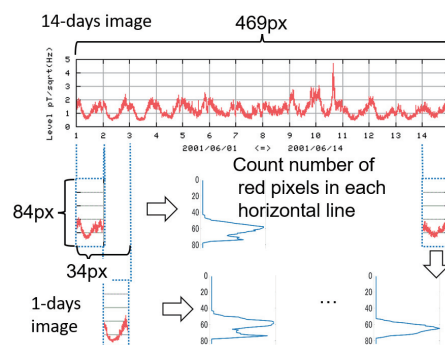


図 9: 変更後の特徴量抽出の流れ (14 日間画像から 1 日間分を切り出し)

ら、背景信号除去に置いて、観測信号に含まれる源信号の数を制限しつつ、分離に必要なデータ量を賄うに適した単位が、1 日ごとであったためである。また、そのように推定された局所信号は、BSS が持つバイアスの不定性から、日付が変わる部分で不連続の状態となる。2 日間分を切り出した画像から度数分布を求めると、性質の異なるものを混ざってから特徴を見ることとなり、目的と合致しない。そこで筆者らは、波形画像からの特徴量抽出の流れを、図 9 に示すように 14 日間画像から 1 日間分を切り出す形に変更する。もちろん、HMM のトレーニングも改めて行う必要がある。従来手法では、特徴量をシンボル列とする際に、0 から 31 の整数列になるよう式 (1) で正規化を行っていた。しかし、物理的な理由や統計的な理由が明確とはいえない処理であり、個々のデータへの依存性が強く、データに含まれる情報を歪める恐れがある。

ここで、 s の各要素の最大値は、2 日間分を切り出す場合で 67、1 日間分を切り出す場合で 34 となるが、実際の値では 32 以上となる場合が少なかった。正規化前の s をそのままシンボル列とした場合の検討も必要だと言える。

シンボル列については、もう 1 点、検討が必要な事項がある。従来手法では、単純に画素の位置情報に基づいて、上から順の並びとしていた。しかし、HMM は状態遷移を伴い、対数受理尤度の算出に動的計画法を用いるものであるため、シンボルをどのような順序で与えるかは重要である。そして、その順序には、物理的あるいは統計的な意味合いなど、何らかの理由付けができるものが望ましい。今回の場合、シンボル列が物理量（磁界強度）の度数分布に準ずるものであり、分布の下限である 0 [pT] が下側であることから、下から順の並びについても検討が必要だと言える。

3.3. HMM の拡張と状態数について

本論文では従来手法より拡張した HMM を用いる形に手法を変更する。状態数 5 の例を図 10 に示す。拡張前の HMM（図 5 と比較して、右側への遷移にスキップを許容した柔軟なモデルになっている。これは拡張前の HMM を包含するものであるため、本論文では拡張の有無について詳細な検討を行わない。

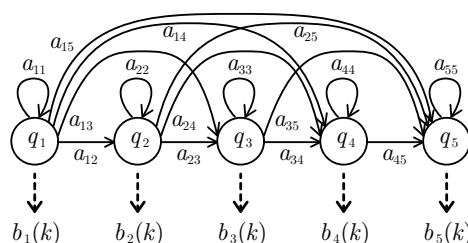


図 10: 拡張 left-to-right 型 HMM の状態遷移図（状態数 5 の例）

状態数について、浦田らの報告¹⁵⁾では 2 が最適とされていた。しかし、本論文ではデータおよびモデルの変更を伴うため、今回の目的に最適な状態数は改めて検討する必要がある。

3.4. 異常検出の閾値について

従来手法では、HMM のトレーニングに用いなかった複数の平常時データについて対数受理尤度を算出し、それらの最小値を異常検出の閾値として設定していた。しかし、トレーニング用データを削減することは、モデルの妥当性を失わせる。また、少ないサンプルから閾値を決定することは、結果の信頼性を失わせる。そこで筆者らは、トレーニングに用いたデータ全てを用いて閾値を設定する形をとる。閾値の具体的な候補は、従来手法に準拠すれば最低値であり、異常検出の慣例であれば、平均と標準偏差からなる $\mu - 2\sigma$ である。

3.5. 再検討事項のまとめ

本章で述べた、再検討が必要な事項について、以下にまとめる。なお、以降は各箇条書きの末尾の形で略記する場合がある。

(A) 取り扱う画像データ

- (1) 14 日間分の観測信号の波形画像から 2 日間分を切り出し (RAW-2d)
- (2) 14 日間分の局所信号の波形画像から 1 日間分を切り出し (BSE-1d)

(B) 特徴量の正規化

- (1) 行う (s')
- (2) 行わない (s)

(C) シンボル列の順序

- (1) 画像の上から順 (H2L)
- (2) 画像の下から順 (L2H)

(D) 拡張した HMM の状態数

(E) 異常の検出閾値

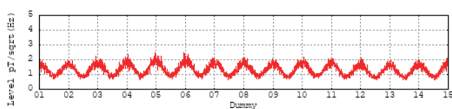
- (1) 全トレーニング用データから算出した対数受理尤度の最小値
- (2) 全トレーニング用データから算出した対数受理尤度の $\mu - 2\sigma$

4. コンピュータシミュレーション実験

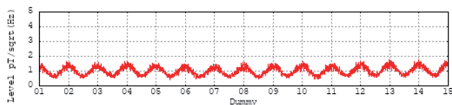
4.1. 実験条件

まず、典型的な平常時の ELF 帯磁界データを模した、14 日間分の信号を 80 個生成した。そのうちの 40 個をトレーニング用データとして扱い、残りの 40 個を異常検出の対象用データとして扱った。異常検出の対象用データのうちの 20 個には、それぞれ異常信号を加えたものも用意した。異常信号は、継続期間について約 1 日間、約 2 日間、約 3 日間の 3 通り、大きさについて小、中、大の 3 通りを組み合わせた 9 通りをデータごとに用意した。それらを画像化した例を図 11 に示す。それぞれ、縦軸は振幅、横軸は日数に相当する。加算した異常信号は、図のようにバイアスが継続期間において山なりに変化する形とし、ほとんどの日に含まれるように異常信号どうしの時間間隔を空けてある。これらのうち、0.06[pT] 以上の異常信号が 45 分間分以上含まれる日を、異常日として設定した。これは、振幅も時間も画像化した際に 1[px] 分に相当する最低値に基づいている。

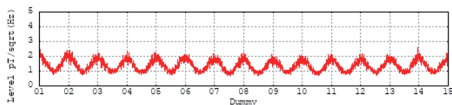
続いて、生成された信号について、1 日間ごとに背景信号除去を適用した状況を模して局所信号を生成した。それらの例を図 12 に示す。トレーニング用データと異常検出の対象用データの振り分け、および異常日の設定は、背景信号除去前の



(1) RAW 01-0-0

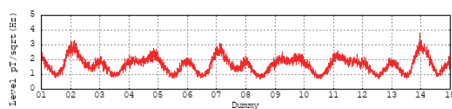


(2) RAW 21-0-0

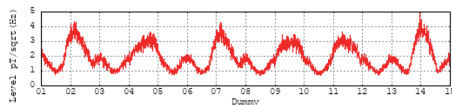


(3) RAW 61-0-0

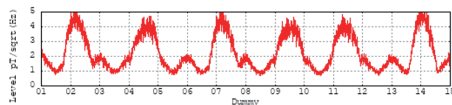
(a) 生成した平常時の信号 (3 個分)



(1) RAW 61-1-1

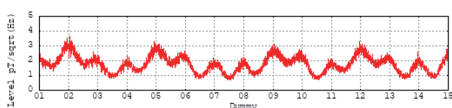


(2) RAW 61-1-2

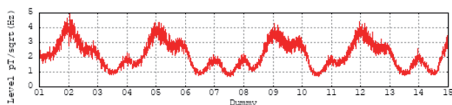


(3) RAW 61-1-3

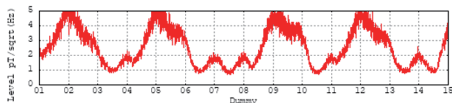
(b) 約 1 日間の異常信号を加算
(上から小中大)



(1) RAW 61-2-1

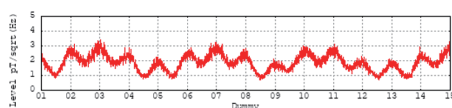


(2) RAW 61-2-2

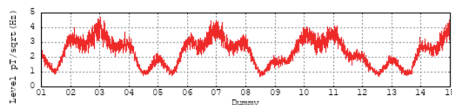


(3) RAW 61-2-3

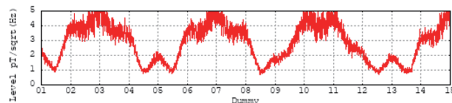
(c) 約 2 日間の異常信号を加算
(上から小中大)



(1) RAW 61-3-1



(2) RAW 61-3-2



(3) RAW 61-3-3

(d) 約 3 日間の異常信号を加算
(上から小中大)

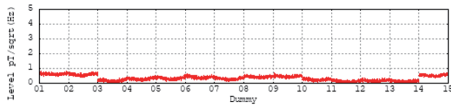
図 11: 生成した 14 日間分の信号の例

ものと対応する形とした。

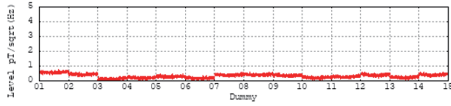
4.2. 偽陽性（誤検出）についての検証

こうして作成された画像データを用いて、HMM のトレーニングや異常検出実験を行った結果を比較する。表 1 から表

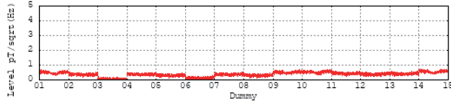
4 に、偽陽性（誤検出）の割合を示す。これらの表は、利用した画像データと閾値がそれぞれ異なっている。「RAW-2d」が付記されているものが 14 日間分の生成信号を画像化し 2 日間分を抜き出した場合、「BSE-1d」が付記されているものが 14 日間分の局所信号を画像化し 1 日間分



(1) BSE 01-0-0

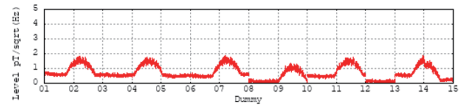


(2) BSE 21-0-0

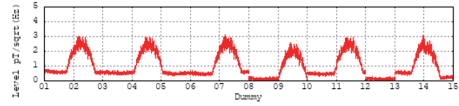


(3) BSE 61-0-0

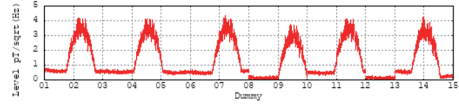
(a) 生成した平常時の信号 (3 個分)



(1) BSE 61-1-1

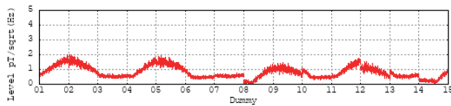


(2) BSE 61-1-2

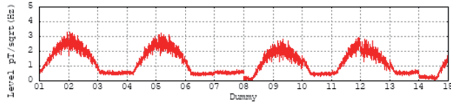


(3) BSE 61-1-3

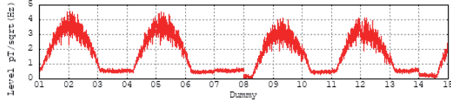
(b) 約 1 日間の異常信号を加算
(上から小中大)



(1) BSE 61-2-1

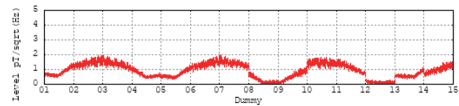


(2) BSE 61-2-2

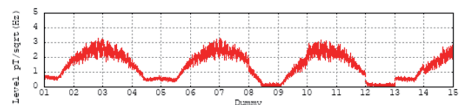


(3) BSE 61-2-3

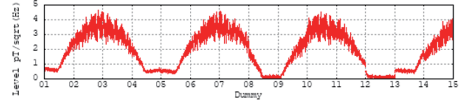
(c) 約 2 日間の異常信号を加算
(上から小中大)



(1) BSE 61-3-1



(2) BSE 61-3-2



(3) BSE 61-3-3

(d) 約 3 日間の異常信号を加算
(上から小中大)

図 12: 生成した局所信号の例 (図 11に対応)

を抜き出した場合である。また、閾値を最小値とした場合と $\mu - 2\sigma$ とした場合でも分けている。表中のシンボル欄は、正規化を行ったものを「s」、行っていないものを「s」でまとめ、順序欄はシンボル列の順序が上から順のものを「H2L」、下から順のものを「L2H」として分けてい

る。状態数は、HMM をトレーニングさせる際に設定した値である。

各表から分かる傾向として、閾値は最小値の方が、誤検出率が低いことが挙げられる。これは、検出の条件がより厳しいためである。また、順序は H2L の方が、多くの場合において誤検出率が低い

表 1: 偽陽性（誤検出）の割合
（データ：RAW-2d, 閾値：最小値）

シンボル	順序	状態数に対する誤り率 [%]						
		2	3	4	5	6	7	8
s'	H2L	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	L2H	1.2	0.8	0.4	1.2	1.2	0.4	0.0
s	H2L	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	L2H	0.0	0.0	0.0	0.0	0.0	0.0	0.0

表 2: 偽陽性（誤検出）の割合
（データ：BSE-1d, 閾値：最小値）

シンボル	順序	状態数に対する誤り率 [%]						
		2	3	4	5	6	7	8
s'	H2L	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	L2H	0.0	0.0	0.0	0.0	0.0	0.2	0.0
s	H2L	0.8	0.8	0.8	0.8	0.8	0.8	0.8
	L2H	0.9	0.8	0.9	0.8	0.9	0.8	0.9

表 3: 偽陽性（誤検出）の割合
（データ：RAW-2d, 閾値： $\mu - 2\sigma$ ）

シンボル	順序	状態数に対する誤り率 [%]						
		2	3	4	5	6	7	8
s'	H2L	2.3	2.3	2.3	1.9	1.9	1.9	1.9
	L2H	2.3	2.3	2.3	2.3	2.3	2.3	1.9
s	H2L	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	L2H	1.5	1.5	1.5	1.2	1.2	1.5	0.8

表 4: 偽陽性（誤検出）の割合
（データ：BSE-1d, 閾値： $\mu - 2\sigma$ ）

シンボル	順序	状態数に対する誤り率 [%]						
		2	3	4	5	6	7	8
s'	H2L	0.9	0.9	0.9	0.9	0.9	0.9	0.9
	L2H	0.9	1.9	1.9	2.5	2.1	3.2	3.6
s	H2L	2.3	2.3	2.3	2.3	2.3	2.3	3.0
	L2H	3.0	2.3	2.5	2.8	3.6	2.5	3.4

ことが分かる．その原因は定かではないが，H2L の場合はシンボル列の冒頭で1（赤いピクセルの数が0）が連続しやすく，その後緩やかにシンボルの値が増加するため，left-to-right 型のモデルで追従しやすかった可能性がある．

他に分かる傾向として，RAW-2d の場合は s' の方が，BSE-1d の場合は s の方が，誤検出率が低いことが挙げられる．ここで，RAW-2d の場合は正規化によって実質的なシンボルの種類が減ることも多く，逆に BSE-1d の場合は正規化によって実質的なシンボルの種類が増えることが多い．つまり，シンボルにはある程度の制限があった方が良いことを示唆している可能性がある．

一方，状態数については，いずれも明確な傾向が見られなかった．本来，状態数が多い方が複雑なモデルに対応できるが，計算コストが高くなるだけでなく，過学習に陥りやすくなるという欠点もある．従来手法の通り，状態数は2で十分な可能性がある．

4.3. 偽陰性についての検証

さて，誤検出が生じること自体は，地殻活動など危険を検出するシステムとして深刻な問題ではない．トレードオフとなる偽陰性（見逃し）が多数生じる方が深刻な問題となる．偽陰性（見逃し）の割合を表 5から表 8に示す．表の並びや

見方は表1から表4と同様である。

各表から分かる傾向として、閾値は $\mu - 2\sigma$ の方が、見逃し率が低いことが挙げられる。これは、検出の条件がより緩いためである。また、RAW-2d と BSE-1d を比較すると、いずれのケースにおいても BSE-1d の方が、見逃し率が低くなっている。つまり、切り出し幅を1日まで短くしても、背景信号除去が見逃しを減らしていると言える。

シンボルについては、ほとんどのケースにおいて s の方が、見逃し率が低くなっている。これは、前述した「シンボルの種類には制限があった方が良い可能性がある」とは異なる状況であるが、見逃し率の観点からは、正規化が情報を歪めている可能性を示唆している。また、シンボル列の順序については、ほとんどのケースにおいて L2H の方が、見逃し率が低くなっている。前述の誤検出に関しての結果を含め、L2H の方が検出の感度を上げる傾向があることを示唆している。

状態数については、いずれも明確な傾向が見られなかったが、比較的パフォーマンスが良いのが4~6であった。状態数が1だけ異なることが深刻な差を生むとは考えにくいため、この実験において十分な状態数は5だと結論付ける。

これらの結果を受け、異常検出手法を見直すと、状態数5のHMMにBSE-1dから得られたシンボル列 s を L2H で与え、 $\mu - 2\sigma$ を閾値とする形が、最良の修

表 5: 偽陰性（見逃し）の割合
(データ：RAW-2d, 閾値：最小値)

シンボル	順序	状態数に対する誤り率 [%]						
		2	3	4	5	6	7	8
s'	H2L	9.9	9.9	9.9	11.8	11.9	11.9	11.9
	L2H	8.4	8.3	8.3	7.9	7.9	8.0	10.6
s	H2L	8.6	8.6	8.6	8.6	11.2	11.2	11.2
	L2H	7.3	7.0	7.0	6.2	6.2	6.5	7.8

表 6: 偽陰性（見逃し）の割合
(データ：BSE-1d, 閾値：最小値)

シンボル	順序	状態数に対する誤り率 [%]						
		2	3	4	5	6	7	8
s'	H2L	5.4	5.4	5.4	5.4	5.4	5.4	5.4
	L2H	5.2	5.3	5.2	5.1	5.2	5.1	5.4
s	H2L	4.0	4.0	4.0	4.0	4.0	4.0	4.0
	L2H	3.9	5.0	3.8	3.9	4.1	4.0	3.9

表 7: 偽陰性（見逃し）の割合
(データ：RAW-2d, 閾値： $\mu - 2\sigma$)

シンボル	順序	状態数に対する誤り率 [%]						
		2	3	4	5	6	7	8
s'	H2L	6.8	6.8	6.8	8.3	8.2	8.2	8.4
	L2H	6.6	6.2	6.2	6.2	6.1	5.9	7.1
s	H2L	5.3	5.2	5.2	5.2	6.3	6.3	6.2
	L2H	4.6	4.4	4.3	3.9	3.9	4.1	4.4

表 8: 偽陰性（見逃し）の割合
(データ：BSE-1d, 閾値： $\mu - 2\sigma$)

シンボル	順序	状態数に対する誤り率 [%]						
		2	3	4	5	6	7	8
s'	H2L	4.1	4.1	4.1	4.1	4.1	4.1	4.1
	L2H	4.1	3.4	3.3	3.3	3.1	3.0	3.0
s	H2L	2.8	2.8	2.8	2.8	2.8	2.8	2.8
	L2H	2.8	2.7	2.9	2.8	2.5	2.5	2.6

正候補だと言える。ただし、実際の ELF 帯磁界データでも同様である保証はないため、実際の信号に対しては、いくつかの候補について検討が必要だと言える。

5. ELF 帯磁界データへ適用した例

前章で述べたように、コンピュータシミュレーション実験でのパフォーマンスが、実際の信号でも同様となる保証はない。そこで、下記に示す候補について比較を行う。

- Type-A: RAW-2d, s' , H2L, 状態数 2
- Type-B: RAW-2d, s' , H2L, 状態数 5
- Type-C: RAW-2d, s , L2H, 状態数 5
- Type-D: BSE-1d, s , L2H, 状態数 5

Type-A と Type-B の違いは状態数、Type-B と Type-C の違いはシンボルとその順序、Type-C と Type-D の違いは背景信号除去となっている。

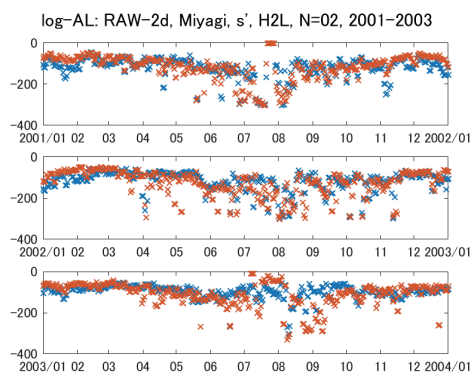
実際の ELF 帯磁界データから算出した対数受理尤度のグラフを図 13 に示す。宮城県栗原市の 2001 年から 2003 年までのデータに、それぞれ (a) Type-A から (d) Type-D の条件の下で求めた。各グラフの横軸は時間であり、目盛が月、上中下段が年の推移である。また、青色のマークが東西方向、朱色のマークが南北方向のデータのものである。対数受理尤度であるため、値が小さい方がモデルとの合

致度が低い、つまり異常性が高いことを表している。

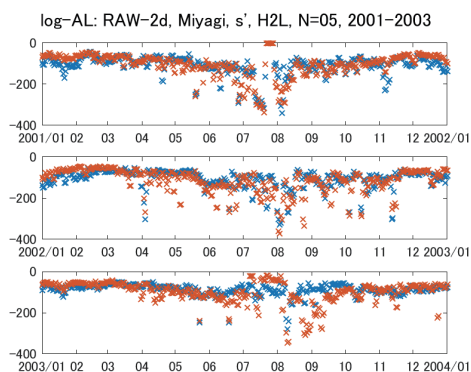
グラフより、Type-A から Type-C はほぼ同様となっていることが分かる。6 月から 9 月の値が小さく、閾値にもよるが、異常として検出されやすい状態である。一方、Type-D については、6 月から 9 月も含めて全体的に値が 0 付近に近づいている。とはいえ、小さな値を維持している日もあるため、中間的な値が減ってグループが分かれた状態と言える。Type-C との比較から、これは背景信号除去による効果だと言える。

Type-A と Type-D について、宮城県栗原市の 2001 年 7 月 31 日から 8 月 26 日までの東西方向のデータを図 14 に示す。上から (1) 観測信号、(2) 局所信号、(3) 対数受理尤度である。(3) について、青色の「×」マークと一点鎖線は Type-A によるもので、一点鎖線は閾値（上側が最小値、下側が $\mu - 2\sigma$ ）である。朱色の「+」マークと破線は Type-D によるものである。この期間中の 2001 年 8 月 14 日に青森県東方沖で M6.4 の地震が発生しており、検出された異常と関連している可能性がある。

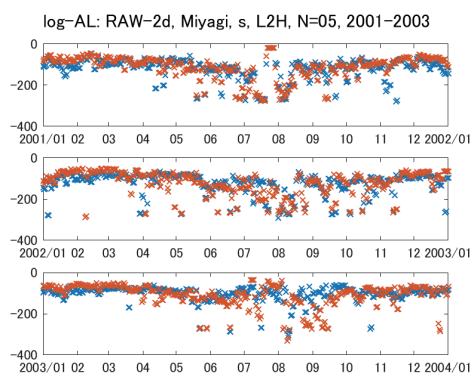
対数受理尤度が閾値以下となる日の多くは、Type-A と Type-D で共通している。しかし例えば、8 月 6 日は Type-A でのみ異常と判定されている。8 月 6 日の観測信号には異常が含まれているようには見えない。これは、Type-A が 2 日



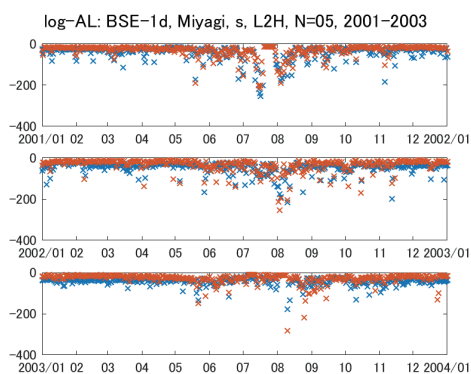
(a) Type-A



(b) Type-B

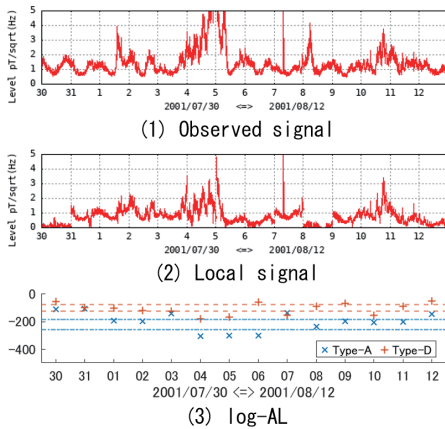


(c) Type-C

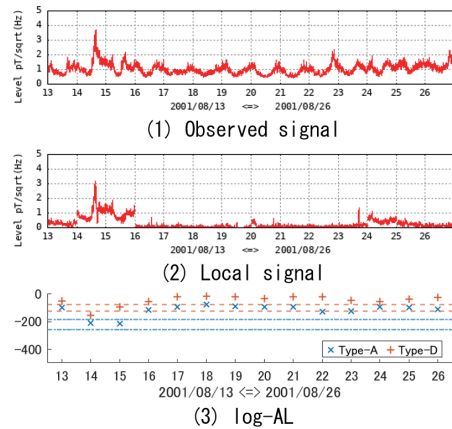


(d) Type-D

図 13: ELF 帯磁界データの対数受理尤度の例（宮城県栗原市，2001 年～2003 年）
青色が東西方向，朱色が南北方向の値

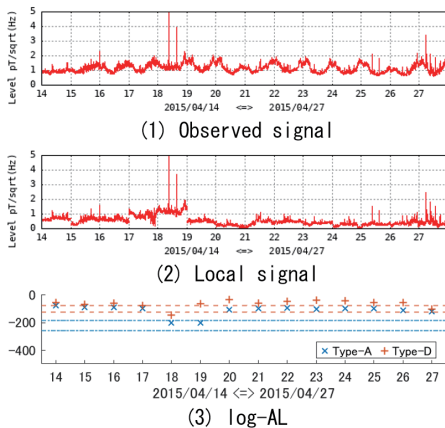


(a) 2001 年 7 月 30 日～8 月 12 日

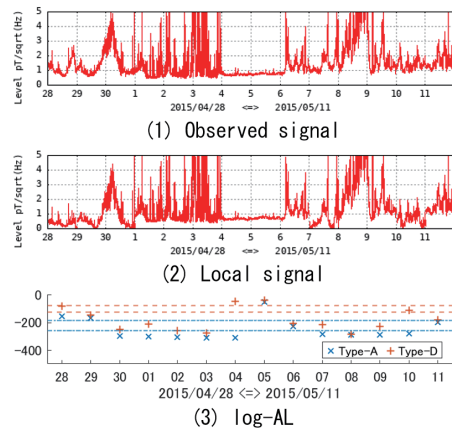


(b) 2001 年 8 月 13 日～8 月 26 日

図 14: 地震前後のデータと異常評価 (宮城県栗原市・東西方向)
2001 年 8 月 14 日に青森県東方沖で M6.4 の地震が発生



(a) 2015 年 4 月 14 日～4 月 27 日



(b) 2015 年 4 月 28 日～5 月 11 日

図 15: 地震前後のデータと異常評価 (神奈川県湯河原町・南北方向)
2015 年 5 月 6 日に箱根山噴火警報が発令

間の切り出し画像から特徴量を抽出しているため、前日に含まれている異常に反応したものだと考えられる。また、8 月 10 日と 14 日は Type-D でのみ異常と判

定されている。両日も昼間の観測信号に大きな上昇がみられるため、異常として検出される方が望ましい。背景信号除去後の特徴量を学習時および評価時に用

いたことで、異常を見逃しにくくなったと言える。

2015 年の 8 月 6 日に箱根山噴火警報が発令された。これに関連している可能性のある異常が、神奈川県湯河原町の南北方向のデータに表れていた。それらを図 15 に示す。図の見方は図 14 と同様である。

対数受理尤度が閾値以下となる日のうち、Type-A と Type-D で異なっているものに着目する。例えば、4 月 18 日は Type-D でのみ異常と判定されている。同日には短時間の外れ値が 2 回観測されており、これを捉えたものと考えられる。地殻活動に関係のある異常は長時間であると考えられているため、これは誤検出だと言える。一方、7 月 29 日や 8 月 6 日も Type-D でのみ異常と判定されているが、こちらは長時間であることから、異常として検出される方が望ましい。

他に検出される方が望ましかった異常としては、4 月 27 日のものがある。Type-D でも、閾値を $\mu - 2\sigma$ とした場合には検出されない。閾値に関するさらなる調査・検討の必要性を示唆するとともに、異常検出手法の単独運用の限界を示唆していると言える。

6. まとめ

本論文では、伊藤および浦田らによって提案された、教師なし HMM に基づい

て ELF 帯磁界の波形画像から異常検出する手法について再検討を行いつつ、その改善を図った。具体的には、背景信号除去の有無とそれに付随する切り出し区間の変更、特徴量の正規化の有無、シンボル列の順序、HMM の状態数、異常の検出閾値について再検討を行った。コンピュータシミュレーション実験より、主に偽陰性（見逃し）を減らす場合は、背景信号除去あり、特徴量の正規化なし、シンボル列は画像の下から順、状態数は 5 という形に手法を変更することが効果的であることが示唆された。また実際の ELF 帯磁界データを用いた実験を通して、手法の改善とともに、背景信号除去の有用性を確認した。

今後の課題としては、閾値に関するさらなる調査が必要と考えられる。その際、誤検出と見逃しはトレードオフの関係であるため、運用上の望ましさを含めて検討する必要がある。一方で、教師なし HMM に基づいて ELF 帯磁界の波形画像から異常検出する手法の限界が示唆されていること、および、近年は ELF 帯磁界データが得られなくなっていることから、これまでと異なる公開データを利用した異常検出手法についての研究を模索する必要もある。

謝辞

本稿を執筆するにあたり、本学国際コミュニケーション学部の梅垣敦紀教授に作成して頂いた pL^AT_EX のクラスファイル 18, 19) を利用し、またその際に様々なご助力も頂いた。同氏に厚く感謝を申し上げる。

参考文献

- 1) C. Hashimoto, A. Noda, T. Sagiya and M. Matsu'ura: "Interplate seismogenic zones along the Kuril-Japan trench inferred from GPS data inversion," *Nature Geoscience*, vol. 2, pp. 141-144, 2009.
- 2) M. B. Gokhberg, V. A. Morgunov, T. Yoshino and I. Tomizawa, "Experimental measurements of EM emissions possibly related to earthquakes in Japan," *Journal of Geophysical Research*, vol. 87, pp. 7824-7829, 1982.
- 3) M. Hayakawa and Y. Fujinawa, "EM phenomena related to earthquake prediction," *Terra Scientific (TERA-PUB)*, 1994.
- 4) K. Maeda and N. Tokimasa, "Decametric radiation at the time of the Hyogo-ken Nanbu earthquake near Kobe in 1995," *Journal of Geophysical Research Letters*, vol. 23, pp. 2433-2436, 1996.
- 5) E. Petraki, D. Nikolopoulos, C. Nomicos, J. Stonham, D. Cantzos, P. Yannakopoulos and S. Kottou, "Electromagnetic Pre-earthquake Precursors: Mechanisms, Data and Models-A Review," *Journal of Earth Science & Climatic Change*, vol. 6, issue 1, 205, 2015.
- 6) 毛利 元昭, 内匠 逸: "環境電磁波観測に関する研究", <https://halo.aichi-u.ac.jp/~mouri/EQ/>
- 7) I. Takumi, M. Nagai, S. Kato, M. Hata and H. Yasukawa: "Estimation of EM Source Location From 223 Hz EM Field Power Data at Multiple Spots", *Proc. of the International Geoscience and Remote Sensing Symposium*, pp.3191-3193, 2001.
- 8) H. Yasukawa, S. Niwa, M. Hata and I. Takumi: "A Simple Signal Detection Method for Great Earthquake Prediction by Observation of Seismic Electromagnetic Wave Radiations", *XI European Signal Processing Conference*, Volume III, pp.419-422, 2002.

- 9) M. Mouri, I. Takumi and H. Yasukawa: “Development of Reliable and Stable QL1-NMF Algorithm for Analyzing Environmental ELF Magnetic Signals”, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 11, no. 6, pp. 1821-1831, 2018.
- 10) S. Niwa, H. Yasukawa, M. Hata and I. Takumi: “A Signal Detection on Precursor of Earthquake Using Normal Value for ELF Electromagnetic Wave Observation”, Proc. of The International Symposium on Information Theory and Its Applications, pp.863-866, 2002.
- 11) M. Mouri, A. Funase, I. Takumi, A. Cichocki, H. Yasukawa, M. Hata: “Effectiveness of Global Signal Elimination from Environmental Electromagnetic Signals for Earthquake Prediction,” 2008 International Symposium on Information Theory and Its Applications, pp. 1128-1133, 2008.
- 12) A. Itai, H. Yasukawa, M. Hata and I. Takumi: “Anomalous Signal Detection in ELF Band Electromagnetic Wave using Multi-layer Neural Network with Wavelet Decomposition,” IEEJ Trans. FM, Vol.129, No.12, pp.875-883, 2009.
- 13) Y. Ito, A. Itai, H. Yasukawa, I. Takumi and M. Hata: “HMM Based Anomalous Signal Detection From ELF Electromagnetic Wave Signals,” Proc. of 2010 European Signal Processing Conference, pp.1861-1864, 2010.
- 14) Y. Ito, A. Itai, H. Yasukawa, I. Takumi and M. Hata: “Performance of anomalous signal detection with HMM approach in electromagnetic wave observation using moving window,” 2011 IEEE International Geoscience and Remote Sensing Symposium, pp.4074-4077, 2011.
- 15) S. Urata, H. Yasukawa, A. Itai, I. Takumi: “A study on clustering for anomalous signal detections from electromagnetic wave data,” 2012 IEEE International Geoscience and Remote Sensing Symposium, pp.6083-6086, 2012.
- 16) M. Mouri, A. Itai, I. Takumi, H. Yasukawa: “A STUDY ON FEATURES EXTRACTION FOR HMM BASED ANOMALOUS SIGNAL DETECTION FROM WAVEFORM IM-

AGES OF ELF MAGNETIC SIGNALS,” Proceedings 2019 IEEE International Geoscience and Remote Sensing Symposium, pp. 9439-9442, 2019.

- 17) A. Cichocki, S. Amari: Adaptive Blind Signal and Image Processing: “Learning Algorithms and Applications,” John Wiley & Sons Ltd., 2002.
- 18) 梅垣 敦紀: “愛知大学情報メディアセンター紀要「COM」への $\text{\LaTeX}$ 投稿の試み”, 愛知大学情報メディアセンター紀要「COM」, Vol. 33, No. 1, pp.31-36, 2024.
- 19) 梅垣 敦紀: “愛知大学情報メディアセンター紀要「COM」への $\text{\LaTeX}$ での投稿”, 愛知大学情報メディアセンター紀要「COM」, in printing.

**Title: A Revisional Study with
Implementing Background Signal
Elimination on Unsupervised Hidden Markov Model based Anomalous Detection from ELF Magnetic Waveform Images**

**Author: Motoaki MOURI, Hiroshi
YASUKAWA, Ichi TAKUMI**