

論 文

あけぼの衛星のデータベースを用いたプラズマ波動スペクトルの自動分類法

秋元 陽介^{†*} 後藤 由貴[†] 笠原 禎也^{†**} 佐藤 亨[†]

Automatic Classification of Electromagnetic Waves from Database Obtained by the Akebono Satellite

Yosuke AKIMOTO^{†*}, Yoshitaka GOTO[†], Yoshiya KASAHARA^{†**}, and Toru SATO[†]

あらまし 本論文では、科学衛星「あけぼの」で取得したプラズマ波動の電磁界強度データを用いて、他の衛星観測にも応用できる汎用的なプラズマ波動現象の種別分類アルゴリズムを提案する。まず、人が波動の種別分類を行う際に用いる様々な指標を特徴量として数値化し、求めた特徴量にクラスタ分析を適用することでサンプルデータから代表的な波動現象（代表分類）を作成するアルゴリズムを提案する。その際にクラスタ分析における計算時間の増大を抑制するため、取得されるデータの特性を生かして、分類精度を下げずに計算時間を短縮する手法を考案した。次に、情報量基準（AIC）を用いた評価関数により、客観的に波動の代表分類数を決定した。また、得られた代表分類に対して全観測データの当てはめを行うために、ベイズ識別方式を導入し、サンプルデータに用いなかった観測データについても波動現象の識別ができることを示した。最後に、どの代表分類にも属さない例外的な観測データの検出法を提案し、実際にこのアルゴリズムで検出できる例外データの事例を示した。これにより、従来の人手による解析法では抽出が難しかった特異な観測データを自動抽出できることが示された。

キーワード スペクトル認識, パターン認識, クラスタリング, プラズマ波動

1. ま え が き

地球は大きなダイポール型の磁場をもち、地球周辺プラズマは、太陽から噴き出す高速のプラズマ流（太陽風）により閉じ込められた地球磁気圏という領域を形成する。この宇宙プラズマは、地球磁場との相互作用により、大規模磁気嵐による人工衛星の故障など、人類に大きな影響を及ぼす。近年では、測位や地球環境モニタ用人工衛星、国際宇宙ステーションや宇宙太陽発電所計画など、様々な形での宇宙利用が進みつつあり、より正確な宇宙天気予報のために地球周辺プラズマ環境の研究の重要性が高まっている。

磁気圏内にはホイスラー、VLF ヒス、コーラスなど、種々のプラズマ波動現象が生起し、伝搬している。これらは磁気圏環境を反映する高感度な情報源とし

て、地球周辺プラズマ環境の調査に重要な役割を果たす。地球物理においては、太陽活動や下部電離層の影響など極めて多くのパラメータに支配される種々の物理過程が同時に混在するため、観測された各々の現象の明確な解釈が困難である。そこで、まず顕著な観測例を集めてその特徴を整理し、単純化したモデルを仮定し、より多くの例についてそのモデルの可否を検討する手法がとられる。プラズマ波動現象についても同様で、研究手法が本来発見的であり、観測される個々の波動について十分に機構が解明されていないのが現状である。我々のグループは、科学衛星「あけぼの」搭載の VLF 観測器 [1], [2] で取得された 20 kHz 以下のプラズマ波動のデータ解析を担当し、波動現象の解析を行っている。あけぼの衛星が 1989 年に地球近傍磁気圏に軌道投入されてから、現在までに 1.5 Tbyte に及ぶデータが蓄積されているが、後述のように取得データ一つひとつを、専門知識を有する人間が精査し解釈するにはデータ量が膨大なため限界があり、多くのデータが手つかずとなっているのが現実である。これに対し、近年、これらのデータを計算機上で統括的

[†] 京都大学大学院情報学研究所, 京都市
Graduate School of Informatics, Kyoto University, Kyoto-shi,
606-8501 Japan

* 現在, NTT アクセスサービスシステム研究所

** 現在, 金沢大学工学部情報システム工学科

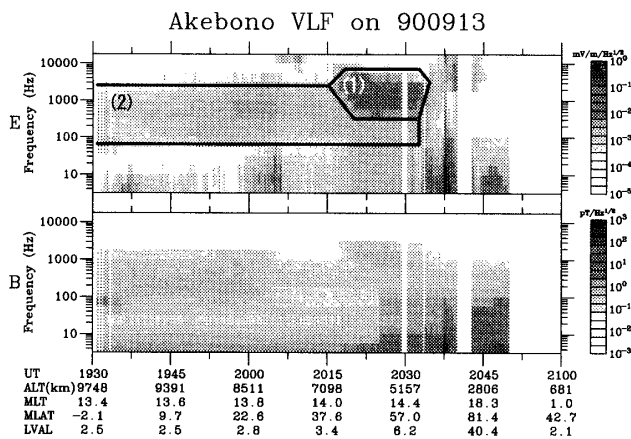


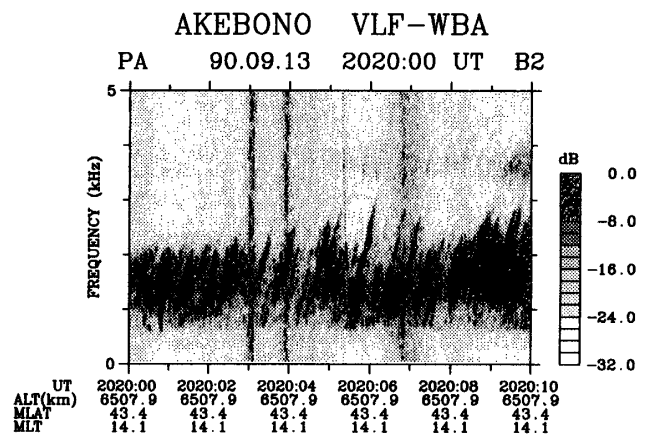
図 1 MCA によるプラズマ波動観測例

Fig. 1 An example of plasma waves observed by MCA.

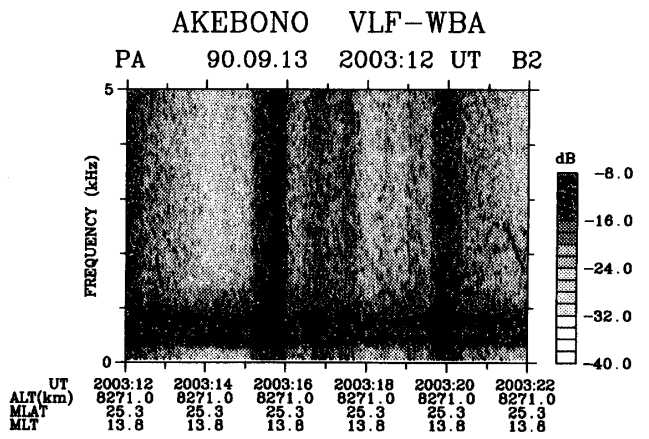
に取り扱う試みがなされており [3], 波動現象の機械分類はその重要な開発要素の一つである。

本研究では, VLF 観測器で得た電磁界のスペクトル画像データから, プラズマ波動現象の汎用的な種別分類アルゴリズムを開発する. あけぼの衛星で観測されたプラズマ波動の例を図 1 に示す. これは VLF 観測器のサブシステムの一つである多チャンネル ELF/VLF 強度測定器 (MCA) で観測された波動のスペクトルを, 横軸を世界標準時 (UT), 縦軸を波動の周波数で表しており, 各時刻, 周波数における波動の電磁界強度を濃淡で示している. MCA は時間分解能が 0.5 秒で, 3.16 Hz~17.8 kHz の周波数帯の対数的に等間隔な 16 周波数チャンネルで電磁界各 1 成分を観測する. 図中の (1) で示した領域 (20:15~20:33 UT, 1 kHz) に存在する波動はコーラス, (2) で示した領域 (19:30~20:17 UT, 500 Hz) に存在する波動はプラズマ圏ヒスと呼ばれる現象である. この他, 極低周波ノイズ (20:25~20:50 UT, 10 Hz) やオーロラヒス (20:40~20:50 UT, 3 kHz) と呼ばれる現象も見られる. 個々の波動現象の特徴をわかりやすく示すため, VLF 観測器のサブシステムの一つで高時間高周波数分解能観測を行う広帯域波形受信器 (WBA) で取得したスペクトル画像を図 2 に示す. 図 2(a), (b) はそれぞれ 20:20:00 UT, 20:03:12 UT から 10 秒間のスペクトルで, 図 2(a) の 1~2 kHz 近辺の波動がコーラス, 図 2(b) の 0.4~1 kHz 近辺の波動がプラズマ圏ヒスである. なお WBA の観測は波動観測全体の一部に限られるため, 以後の波動種別の分類は MCA のデータを用いて行う.

このようにプラズマ波動の種別は, 異なる物理過程



(a) chorus



(b) plasmaspheric hiss

図 2 高分解能観測器 (WBA) による波動観測例; 濃淡は, 観測波形が衛星から地上にアナログ伝送される際の搬送波強度を 0 dB として表示している. また, 図中の 0 kHz から 5 kHz まで縦に伸びた筋はデータの伝送時にのったノイズであり波動現象ではない

Fig. 2 An example of plasma waves observed by fine resolution receiver.

に起因するスペクトルの微細構造の違いに加え, 波動の電磁界比 (屈折率) などから現象論的に分類・命名されている. 波動現象は一般に, コーラスなど一つのカテゴリーに属する現象の中でも磁気圏の状態や観測条件によって周波数や継続時間・スペクトル形状などが多様に変化するため, 図 1 の例に示したスペクトル画像の巨視的な描像から分類するのは困難である. したがって人手で個々の波動現象の種別を同定するには, 微視的 (数秒~数十秒程度) な構造が判別できる程度までスペクトル図を拡大する必要がある. しかし 13 年にわたる長期観測データを人間の手ですべて精査することは事実上不可能である. 一方, 対象となる波動について十分に理解されていない状況下で現象論的な差異に基づいて機械的に分類する場合, 典型的なイベ

ントを教師データとして用いる方法では、事例が偏って抽出されてしまい、特異な観測例や、未知の現象を捜し出すプラズマ波動の研究本来の発見的な研究手法にそぐわない。そこで本研究では、スペクトル画像の微細構造の特徴から衛星で観測される様々な波動を自動分類すると同時に、従来のカテゴリーに含まれないような観測頻度の低い波動を発見するためのアルゴリズムの開発を目指す。

前述のように波動現象は主に微視的なスペクトル構造の違いから分類されることから、本論文では MCA で取得されたスペクトル画像にテクスチャ解析とパターン認識の考え方をもとに、教師データを必要としない分類アルゴリズムを提案する。テクスチャ解析の手法において、現在、Gabor フィルタやウェーブレット変換を用いた多重解像度解析等が主流であるが [4], [5], 波動現象の分類はそれに加えて電磁界比など多次元の特徴量が加味されるため、自然波動現象に特化した分類法を新たに考案する必要がある。外的な分類概念が与えられないパターン認識の手法として、自己組織化マップ (Self-Organizing Map, SOM) [6], 適応共鳴理論 (Adaptive Resonance Theory, ART) [7], [8], クラスタ分析 [9] などがある。SOM は入力に対する出力の類似関係を得る手法であるため、明確に分類概念間の区切りを設けることが困難である。ART は分類と学習を同時に行えるため認識対象の変化に柔軟なアルゴリズムであるが、学習により分類された事象に偏りが出ることがある。これに対してクラスタ分析はある法則で定義された類似度により、順序立てされた分類概念の構成を行うため SOM, ART に比べてそれぞれの分類概念がもつ特徴量の分布を評価しやすい。本研究では、認識対象の類似関係や適当に決められた分類ではなく、物理的に意味のある分類を得ることを目標としていること、また解析対象は自然現象であるためその確率密度分布がある程度予測可能なことから、クラスタ分析を用いた分類アルゴリズムを採用する。

2. プラズマ波動現象の特徴の定量化

本章では、波動現象をクラスタ分析を用いて分類するために波動の違いを表す指標となる特徴量を導出する。まず、各種波動のスペクトル画像の微細構造の違いを表す指標を導出する。図 3 に例として、コーラス、プラズマ圏ヒス、オーロラヒスの電界強度の時間変化 (スペクトル画像の横方向の変化) を 128 秒 (256 点) の FFT により周波数解析した結果を示す。ただし、

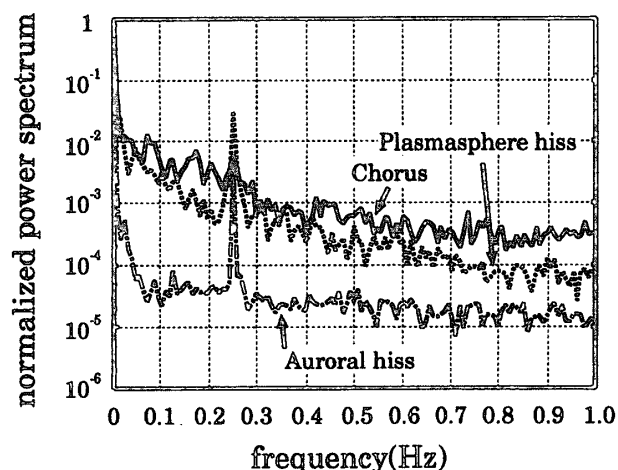


図 3 波動の電界強度の時間変動例
Fig. 3 Examples of time variation of electric field intensities.

データは図 1 で示した例を用いている。各グラフは、時間変動成分の大小を評価するために DC 成分で正規化している。図に示すように、各種波動現象を数秒～数十秒の時間範囲で見たときのスペクトル構造の違いは 1 Hz 以下の時間変動成分の違いに現れる。例えば、図中 0.25 Hz のピークは衛星のspin運動に起因したもので、この成分が相対的に強い波動は電界強度の時間変動がspin周期に比べ十分ゆっくりであることを示している。他の時間変動成分についても同様、各種波動のスペクトル構造の特徴を反映しており、これらを定量化することがすなわち波動を現象論的に分類することと等価となる。またスペクトルの微細構造以外に、波動の平均電界強度、電磁界比 (屈折率) も波動の種別分類の指標として有効と考えられる。

本研究では特徴量として (a) 平均電界強度, (b) 電界強度のspin成分を除く時間変動成分, (c) spin成分を DC 成分で正規化した値, (d) 電界強度の時間変動成分の分散, (e) 電磁界比を用いる。各特徴量を図 1 で示した例に関して数値化した結果を図 4 に示す。ただし (b) に関してはspinを除くすべての成分の平均値を表している。図では 128 秒間の平均電界強度が受信器のノイズレベル以下の部分は除去している。各特徴量は対数で定義されており、図中では濃淡によりその値が示されている。各特徴量を対数で定義したのは、観測波動の頻度分布が電磁界強度に対し対数的に分布しているという経験則に基づいている。特徴量の定性的な比較として、例えばコーラスとプラズマ圏ヒスでは表 1 のような違いがある。この様に特徴量として波動の違いそのものを表す指標を採用することに

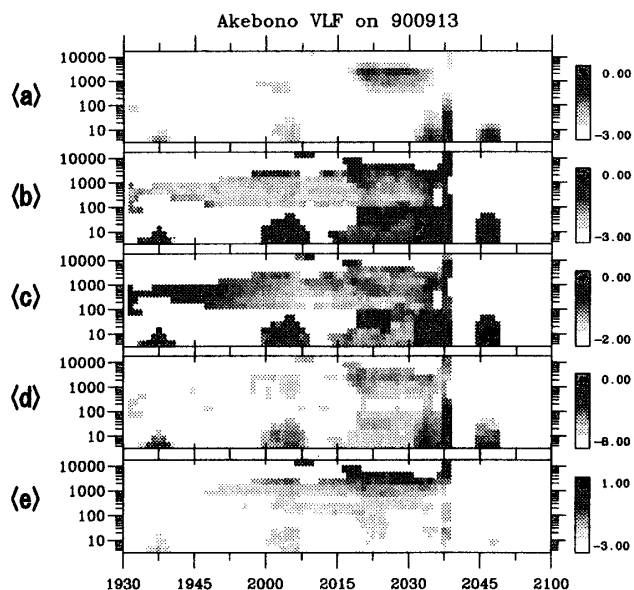


図4 波動現象の特徴量. (a) 平均電界強度, (b) 電界強度の時間変動成分, (c) スピン成分, (d) 電力分散, (e) 電磁界比

Fig. 4 Characteristic properties of plasma waves. (a) averaged of electric field intensity, (b) time variation of electric field intensity, (c) intensity of spin modulation, (d) variance of power spectrum, (e) ratio of electric field to magnetic field.

表1 代表的な波動現象の特徴量の比較
Table 1 Relative features of plasma waves.

	コーラス	プラズマ圏ヒス
DC成分	大	小
スピン成分ピーク	有	有
電界強度の時間変動	大	小
電力密度の分散	大	小
電磁界比	大	小

より, クラスタリング処理により類似性の高い波動を統合し, 波動の分類を行う.

3. サンプル抽出による代表クラスの決定法

3.1 ウォード法の改良

クラスタ分析には様々な類似度が利用されているが, 本論文ではノイズに対する耐性が強いウォード法を用いる [10]. ウォード法では, クラスタ p と q の類似度 ΔS_{pq} を

$$\Delta S_{pq} \equiv \frac{n_p n_q}{n_p + n_q} \sum_{i=1}^{i_{\max}} (\bar{x}_{ip} - \bar{x}_{iq})^2 \quad (1)$$

で定義する. ただし, n_p, n_q はそれぞれクラスタ p, q に所属するデータ数, i_{\max} は特徴量として用いるパラメータ数, $\bar{x}_{ip}, \bar{x}_{iq}$ はクラスタ p, q に属する i 番

目のパラメータの平均値を表す. クラスタ分析には前章で導入した数種類の特徴量を用いるが, これらのパラメータは必ずしも互いに独立とはいえない. そのため単純にクラスタ分析を適用すると, 相関の高いパラメータの影響が支配的になり, 正しい分類が行えない. そこで, 相関のある要素同士の影響を軽減するために, ウォード法の一般式である式 (1) に対して

$$w_i = 1 / \sum_{i'=1}^{i_{\max}} \rho_{ii'} \quad (2)$$

で表される重みを追加した. ここで $\rho_{ii'}$ はパラメータ i, i' の相関係数を表す. また要素の分散を考慮するために

$$\sigma_i^2 = \frac{1}{n-1} \sum_{j=1}^n (x_{ij} - \bar{x}_i)^2 \quad (3)$$

で表される正規化パラメータを導入した. n はすべてのクラスタを含めた全要素数, x_{ij} はすべてのクラスタを含めた要素 j , パラメータ i の観測値である. これらを用いて再定義したクラスタ p と q の類似度は

$$\Delta S_{pq} = \frac{n_p n_q}{n_p + n_q} \sum_{i=1}^{i_{\max}} \frac{(\bar{x}_{ip} - \bar{x}_{iq})^2}{\sigma_i^2} w_i \quad (4)$$

となる. これにより, 特徴量が互いに冗長な情報をもっていた場合, 適切な重み付けがなされる.

3.2 代表クラス作成アルゴリズム

本研究ではあらかじめ分割された最小単位のクラスタ (以下初期クラスタ) を式 (4) で定義した類似度を用いて順に統合することにより, クラスタ分析を行う. 本節ではクラスタ分析を用いて代表的なプラズマ波動現象を表す代表クラスの作成法について議論する. クラスタ分析では比較対象のクラスタ間の類似度を繰り返し計算するため, 全クラスタ間の類似度を総当たりで計算すると, クラスタ数の2乗に比例して計算時間が増大する. このため解析対象となるデータ数が多い場合, 計算上の工夫が必要となる. そこで本研究では解析データ数をできるだけ減らさずに計算時間を短縮し, かつ波動現象分類の精度を高めるアルゴリズムを開発した.

スペクトル画像データにクラスタ分析を適用するには, まず画像全体を初期クラスタに分割する必要がある. 本研究では時間・周波数領域を等間隔な格子に区切り, 初期クラスタを作成した. ただし時間方向の格

子幅は 1 分, 周波数方向の格子幅は 1 チャネル (図の縦幅の $1/16$) である。

科学衛星による観測では衛星運用の都合上, 1 日 24 時間すべてのデータが得られるわけではなく, 例えばあけぼの衛星のデータの場合 1~2 時間の観測パスと呼ばれる単位で 1 日当り 10 数個というように, データが断片的に取得される. そこで代表クラス作成のために以下の 2 段階の手順を踏むことにする.

(1) 観測パスごとのデータを用いてクラスタ分析を適用する. ただし観測パス内の異なる波動現象が同一現象として誤って統合されない程度に, クラスタ数を多めに処理を打ち切る.

(2) 各観測パスごとに作成したクラスタを, 複数パス合わせて再度クラスタ分析を適用する. これによって代表的な波動分類に対応するいくつかの代表クラスを作成する.

このようにクラスタ分析を 2 段階に分けることによって, 約 1 か月のサンプルデータを用いた代表クラスの分類処理が汎用ワークステーションで約 5 時間で実現可能になった.

3.3 各観測パスごとのクラスタ分析

第 1 段階のクラスタ分析では, 一つの波動現象は十分近傍のクラスタのみによって構成され, 同じ特性をもつ波動現象でも時間・周波数的に離れたものは別のクラスタ (波動現象) として取り扱うことにする. このとき, 初期クラスタは, 時間・周波数領域双方について, 対象とするプラズマ波動の広がりより十分小さいため, 時間・周波数的に十分近傍にあるクラスタのみを類似度計算の比較対象にしても問題は生じない. 本研究では少なくとも数分間持続するプラズマ波動の分類を対象とするため, 時間方向は 3 分間, 周波数方向は統合の自由度をもたせるために 3 チャネルの近傍領域を設けた.

提案アルゴリズムを実観測データに適用しクラスタが統合される様子を, 図 1 で示した観測例を用いて説明する. この図の前半 60 分, 周波数 16 チャネルに対してクラスタ分析を適用した結果を図 5 に示す. ただし特徴ベクトルは平均電界強度, 電界強度の時間変動 4 成分, スピン成分, 電力分散, 電磁界比の 8 パラメータから構成する. 図 5(a) は対象となるスペクトル画像の上に解析者が経験的に定めた波動現象の分類を示したものである. それに対し, 図 5(b), (c), (d) は提案アルゴリズムを用いて, それぞれ 11, 10, 9 個のクラスタを形成した結果である. ただし, 図 5(a) にお

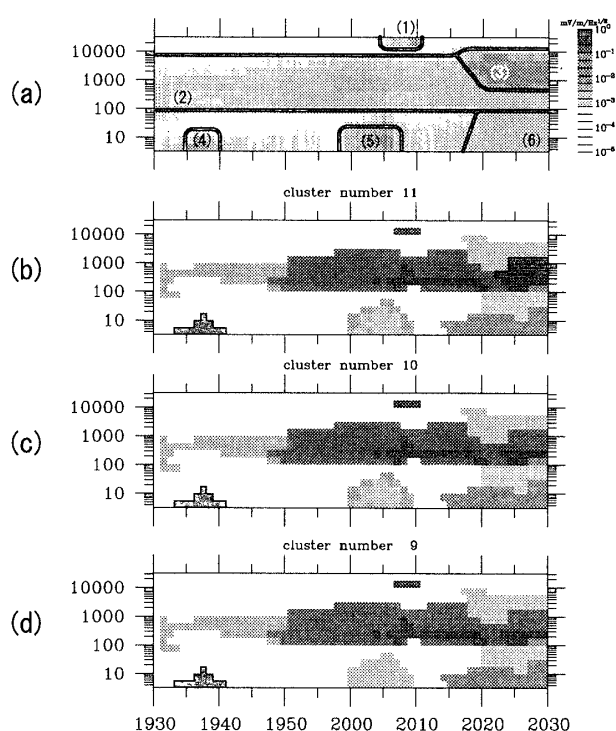


図 5 1990 年 9 月 13 日 (19:30~20:30 UT) の観測データの分類

Fig. 5 Classification of plasma waves observed on September 13, 1990.

いて番号の付けられていない領域は波動強度がノイズレベル以下であるため前処理において除去している.

クラスタ数が 11 個のとき, 同一の波動現象を複数クラスタに分割している箇所は存在するが, 異種の波動現象については統合することなく分類を構成している. この状態から更にクラスタを統合していくと, クラスタ数が 10 個のときに領域 (3) に属すべき 20:20~20:30 UT, 500 Hz~1.00 kHz 付近のクラスタが領域 (2) に属するクラスタに統合されている. 同種の波動を別クラスタとして分類してしまうケースは第 2 段階の代表クラス決定過程で再結合可能であるが, 逆に異種波動がいったん結合してしまうと後で分離できないため, 第 1 段階のクラスタ分析を終了させるクラスタ数の決定は重要である. この決定法は次節で議論する.

3.4 最適クラス数の決定

波動現象の代表クラスを作成する第 2 段階のクラスタ分析で用いるクラスタは, 各観測パスで行ったクラスタ分析の結果を初期値として用いる. すなわち特徴ベクトルの初期値は, 各クラスタがもつ各パラメータの平均値である. また, クラスタ分析の処理の終了に関して以下のような評価基準を定めることにより, 客

観的にクラス数を決定する。

まず、本研究での識別対象であるプラズマ波動現象は自然現象であることから、観測される波動の特徴ベクトルが各パラメータに対して多変量正規分布

$$p(\mathbf{x}|C_m) = \frac{1}{(\sqrt{2\pi})^Q \sqrt{|\Sigma_m|}} \cdot \exp\left\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_m)^T \Sigma_m^{-1}(\mathbf{x} - \boldsymbol{\mu}_m)\right\} \quad (5)$$

に従うと仮定する。ここで Q は特徴ベクトルの次元数、 $\boldsymbol{\mu}_m$ は平均値ベクトル、 Σ_m は共分散行列を表す。本研究でのクラス分析は、特徴空間においてクラスが密になっている代表クラスの中心付近でクラスが統合が頻繁に行われ、徐々に各代表クラスが形成される。そこで、クラス統合の各段階において理想的な多変量正規分布を仮定した場合のクラス自身のゆがみ度 [11] を計算し、AIC [12] に基づくモデルの評価により最適クラス数を決定する。

AIC はパラメトリックな統計モデルに対して

$$\text{AIC} = -2(\text{最大対数ゆがみ度}) + 2(\text{調整可能なパラメータ数}) \quad (6)$$

で定義され、一般にこの値が小さいほど良いモデルとされる。多変量正規分布を仮定した場合にクラス分析で得られる代表クラス C_m の対数ゆがみ度は、 C_m が観測される確率 $P(C_m)$ とクラス C_m に属する各要素の特徴ベクトル \mathbf{x} の条件付き確率密度関数 $p(\mathbf{x}; \boldsymbol{\theta}_m | C_m)$ を用いて

$$l_m(\boldsymbol{\theta}_m) = \sum_{i=1}^{N_m} \log\{P(\mathbf{x}_{im}; \boldsymbol{\theta}_m | C_m)P(C_m)\} \quad (7)$$

と表すことができる。ただし、 N_m は代表クラス C_m に属するサンプルデータ数、 \mathbf{x}_{im} は代表クラス C_m の i 番目のサンプルデータ、 $\boldsymbol{\theta}_m$ はクラス C_m のモデルを予測する際に調整可能なパラメータである。 $\boldsymbol{\theta} = \{\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_M\}$ とすると、すべてのクラスを考慮して得られる対数ゆがみ度は

$$l(\boldsymbol{\theta}) = \sum_{m=1}^M l_m(\boldsymbol{\theta}_m) \quad (8)$$

で与えられる。

一方、各クラスが多変量正規分布を満たすと仮定した場合、2乗残差から求められる最ゆがみ解は平均ベクトル $\boldsymbol{\mu}_m$ 、共分散行列 Σ_m で与えられるため、最大

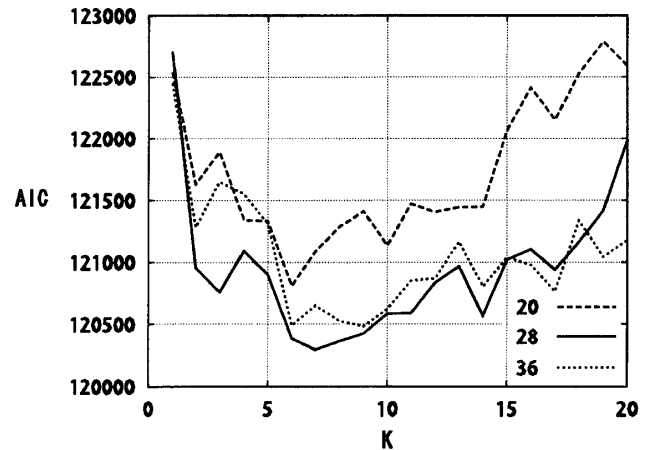


図 6 代表クラス数を決定する評価関数
Fig. 6 Evaluation function for decision of the representative class number.

ゆがみ度を与えるパラメータ数 F は特徴ベクトルの次元数 Q 、クラスの数 M を用いて

$$F = M \left(Q + \frac{Q(Q-1)}{2} \right) \quad (9)$$

と表される。ただし括弧内の第 1 項は平均ベクトルの次元に対応するものであり、第 2 項は各特徴量間の共分散行列に対応するものである。以上の値を用いて AIC を求めることにより、データをいくつの多変量正規分布を仮定したクラスに分類するのが最適か評価することが可能となる。

1 か月分 (1989 年 3 月) のデータセットに対して代表クラスの作成を行い、各クラス数において求めた AIC を図 6 に示す。ただし各観測パスにおけるクラス分析で作成するクラス数を 20、28、36 と変えたときの結果を示している。図より、各観測パスで作成するクラス数が小さすぎると AIC が大きくなるのがわかる。これは各観測パスでのクラス分析において、既に異なる現象同士を統合してしまったことが主な原因と考えられる。つまり 1 観測パスで作成するクラス数はある程度大きくする必要がある。本研究では AIC による評価で改善の見られなくなるクラス数 28 を各観測パスにおけるクラス分析の終了判定値とする。また図 6 より、最適な代表クラス数として $M = 7$ を採用する。すなわち、観測された波動を 7 種類に分類する。

4. 全観測データの判別

4.1 ベイズ識別方式による判別

本章では代表クラスの決定に用いたサンプルデータ

以外の観測データに対してベイズ識別方式を適用することにより、各観測データを代表クラスに識別する方式について議論する。ここでは、各クラスが特徴ベクトルに対して平均 μ_m 、共分散行列 Σ_m の多変量正規分布に従い、観測データが必ず代表クラスのどれかに属すると仮定した場合に最適な識別関数を求める。

ベイズ識別方式では、誤って判別した場合の損失を一定とした場合、観測データ \mathbf{x} が属するクラスを事後確率 $P(C_m|\mathbf{x})$ が最大となるクラス C_m に決定することにより最適な識別とする。識別値 $g_m(\mathbf{x})$ はベイズの公式及び式 (5) から計算される $P(C_m|\mathbf{x})$ の対数值から代表クラスに独立な項数の対数を差し引き、

$$g_m(\mathbf{x}) = \log P(C_m) - \frac{1}{2}(\mathbf{x} - \mu_m)^T \Sigma_m^{-1}(\mathbf{x} - \mu_m) - \frac{1}{2} \log |\Sigma_m| \quad (10)$$

と定義する。識別対象はこの値が最大値となるクラス C_m に分類する。

4.2 実観測データへの適用例

式 (10) で表される判別値を用いた識別方式をあげほの衛星の観測データに適用した結果を示す。まず代表クラスを決定する際に、各観測パスにクラスタ分析を適用したのと同様の手順で1観測パスのデータごとに28個のクラスタを作成する。こうして作成したクラスタごとに特徴ベクトルの平均値 $\bar{\mathbf{x}}$ を求め、各クラスタは $g_m(\bar{\mathbf{x}})$ が最大となるクラス C_m に分類する。ただし簡単のため $P(C_m)$ はクラスにかかわらず一定とした。

図7は、分類後の各クラスのデータ分布を電界強度の時間変動を表す特徴量に対して描いたものである。図より各クラスのデータ分布は単峰的ではあるものの完全な正規分布にはなっていないことがわかる。そこで、各クラスに属するデータが特徴空間において不適切に混ざり合っていないことを確認するために、 k -NN (nearest neighbor) 識別規則を用いた検証を行った。すなわち、分類後のデータ分布を教師データとして、個々のデータの各クラスに対する事後確率を求めることにより本来どのクラスに属するのが適切であるかを求めた。 $k = 25$ としたときの結果を表2に示す。表中の各セルの上段には、データの度数が k -NN 法により識別されたクラスに対して示されている。また各セルの下側には、 k -NN 法で識別されたクラスのデータの何%が提案法により分類されたクラスに属している

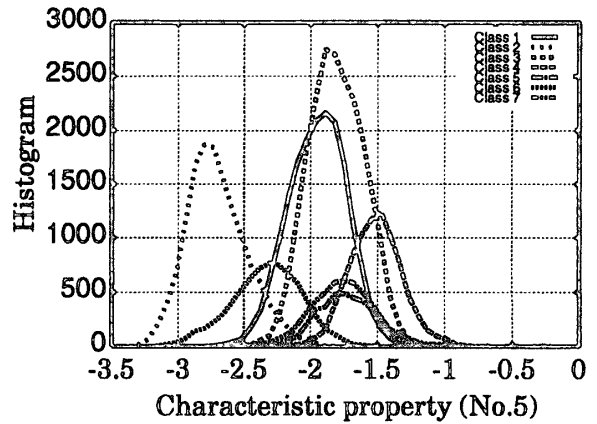


図7 電界強度の時間変動を表す特徴量に対する各クラスの分布
Fig. 7 Distributions of the classes for a characteristic property.

表2 分類されたデータ分布を元に事後確率最大となるクラスを再度 k -NN 法により識別した結果 ($k = 25$)
Table 2 Identification by the k -NN rule using the classified distributions.

	k-NN 法により識別されるクラス						
	1	2	3	4	5	6	7
Class1	14478 79%	408 3%	1382 6%	816 10%	125 4%	278 4%	181 4%
Class2	629 3%	13129 93%	7 0%	0 0%	0 0%	348 5%	0 0%
Class3	954 5%	2 0%	19414 84%	702 9%	242 7%	135 2%	258 6%
Class4	1268 7%	2 0%	876 4%	6153 77%	141 4%	0 0%	49 1%
Class5	115 0%	0 0%	606 3%	187 2%	2925 85%	0 0%	0 0%
Class6	363 2%	641 5%	228 1%	0 0%	0 0%	5582 81%	327 8%
Class7	465 3%	0 0%	652 3%	118 2%	0 0%	520 8%	3399 81%
	100%	100%	100%	100%	100%	100%	100%

かを示している。表2の結果より、クラス1, 3, 4とクラス6, 7においてそれぞれ10%に満たないデータの混じりがあるものの、およそ80%以上のデータが対角成分に集中しており、分類されたデータが特徴空間において混ざり合わずに分類できているのがわかる。 $k = 50, 100, 150$ においても数%の誤差で同様の結果が得られた。以上の結果から、正規分布を仮定した簡便な本識別法により、実用上問題ない精度で各クラスを独立な分布として分類できることが確認できた。

図1で示した例について分類を行った結果を図8に示す。ただし、この観測データは代表クラス作成のサンプルとして用いていない。上の図は電界スペクトルを示しており、下の図は提案アルゴリズムを適用して

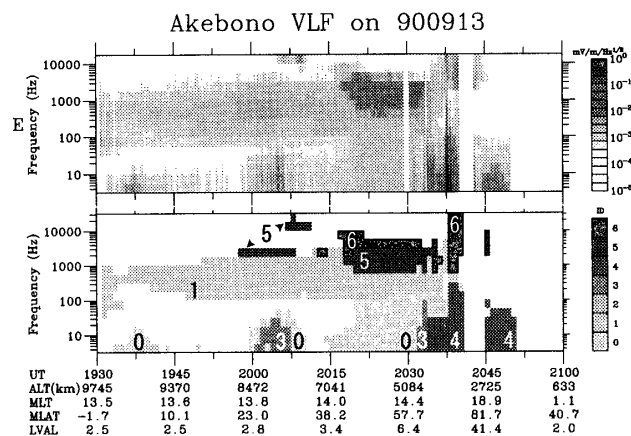


図 8 観測データの代表クラスへの当てはめ例

Fig. 8 An example of discrimination of plasma waves to representative classes.

得られた対応クラスの番号を濃淡と番号で示している。図 1 で示された波動は、オーロラヒスを除いて、経験ある研究者と同等、ほぼ完全に別々の波動として分類できているのがわかる。オーロラヒスに関しては、本来の領域に加えて、コーラスの上部も含めて分類されている。分類されたクラスの特徴量を調べると、プラズマ圏ヒスに対応するクラス 1 の波動は、電界強度の時間変動が七つのクラスの中で最も緩やかであった。また、コーラスに対応するクラス 5 は電界強度の時間変動が激しいという特徴をもち、スペクトル画像の微細構造に基づいた分類が成功したといえる。その他、極域低周波ノイズは平均電界強度の強い波動として分類されている。これらの結果から、本研究で導入した特徴量がプラズマ波動の分類に有意であることが確認できた。

本研究では、観測波動の分類数をデータ分布の正規性から導出した AIC に基づいて七つに設定することで、特徴空間におけるデータ分布を独立に分類することに成功し、実用上有意な波動の識別ができることを示した。しかし、実際にはデータの正規分布からのずれにより最適分類数が変化することが十分考えられる。最適分類数の変化により分類結果がどのように変わるかは、後述の発生頻度の低い波動現象の影響等も考慮した上で、今後検討していく必要がある。

5. 例外的な観測データの検出法

前章までは、クラスタ分析による代表クラスの生成法と、それを元にベイズ識別方式を用いて全観測データをこれらの代表クラスに当てはめる手法について議論してきた。一方、自然現象には代表クラスに含まれ

ない発生頻度の低い波動現象が存在するが、これらを見逃す処理方法では新たな現象の発見が望めない。そこで本章では代表クラスに属さない例外データを検出するアルゴリズムについて検討する。

5.1 例外データの検出法

作成された代表クラスに属さない例外的な観測データには、特徴空間において観測値が複数のクラスの間中に存在する場合と、観測値がどのクラスからも遠く離れたところに存在する場合が考えられる。本研究では代表クラスが形成する代表ベクトルの分布と分類対象であるデータの間の距離を定義し、この距離の大小で一般的な現象か発見的な現象かを判断する。

特徴空間における観測データとクラスとの距離を表す指標としてマハラノビスの距離 [10] がある。マハラノビスの距離は共分散行列の逆行列を用いることにより、パラメータ間の分散、相関を考慮した距離として次式で表される。

$$D = \sqrt{(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})} \quad (11)$$

マハラノビスの距離は式 (10) からクラス間の相対評価に関係する第 1 項と第 3 項を見逃し、第 2 項だけを採用した値に対応する。こうして定義した D に対してあるしきい値を設けることにより、観測データの一般性を識別する。

5.2 例外データの検出

マハラノビスの距離は式 (5) で表される多変量正規分布の指数部に対応するため、あるクラスに所属する観測データのマハラノビスの距離があるしきい値内に収まる確率は解析的に求めることができる。本研究では全体の 95% の観測データが収まるようにマハラノビスの距離を設定する。このマハラノビスの距離をしきい値として検出した例外データの観測例を図 9 に示す。上図は電界スペクトル、下図は例外データを検出した結果で、灰色で表した領域がいずれかの代表分類に識別された一般的なデータ、黒で表した領域が発見的なデータを表す。各データについて時間間隔 2 分間に拡大したものをそれぞれ図 10、図 11 に示す。図 11 に示した 10 Hz 周辺のスペクトルの構造は、一般的に知られている図 10 のものと明らかに異なっており、この波動は従来のカテゴリーには含まれない新たに発見された波動であることがわかる。従来の人手による解析では、経験ある研究者でも、図 9 上図だけで新しい波動を発見するのは難しく時間方向に拡大した図を用いる必要がある。このように提案法は、単に人手の

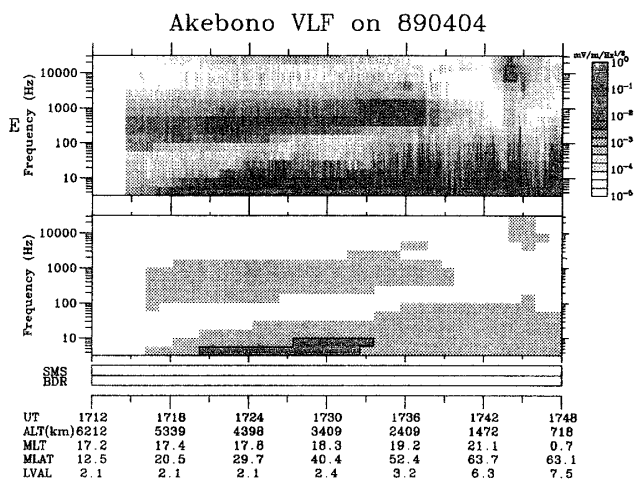


図9 マハラノビスの距離から検出された例外データ
Fig.9 Exceptional data detected by Mahalanobis distance.

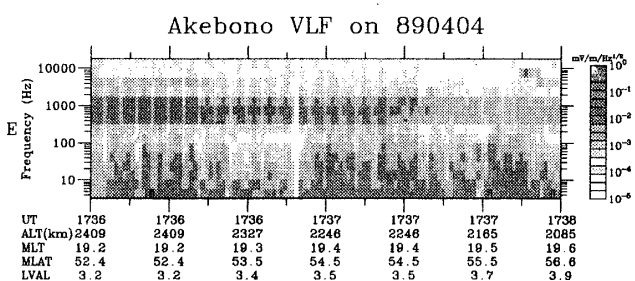


図10 一般的なデータとして識別された波動スペクトル
Fig.10 Spectrum of the normal data.

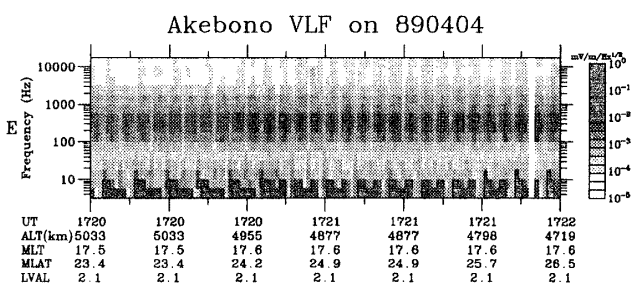


図11 発見的なデータとして識別された波動スペクトル
Fig.11 Spectrum of the exceptional data.

かかる作業を自動化しただけではなく、未知と思われる波動の検出に有効な画期的な方法といえる。

6. むすび

本論文ではあけぼの衛星のMCAで得られた電磁界強度データを用いて外的な分類要因を必要としない、様々な衛星観測に適用できる汎用的なプラズマ波動現象の種別分類アルゴリズムの検討を行った。

まず電界強度の時間変化を周波数解析するなど、人間が波動を分類する際に用いる指標を特徴量として数値化する手法を提案した。次にこの手法で求めたいく

つかのサンプルデータの特徴量にクラスタ分析を採用することで、頻繁に観測される代表クラスを作成するアルゴリズムを提案した。ただし、クラスタ分析は分類対象の数に対して計算時間が増大するため、MCAで得られるデータの特徴を生かして、分類精度を下げずに計算時間を短縮する工夫を施した。次に、クラスタ分析によって代表的な波動現象が適正に分類されているかを評価するために、AICを評価関数として導入して客観的に波動分類数を決定した。こうして得られた代表分類をもとに、全観測データの当てはめを行う方法と、どの代表クラスにも属さない例外的な観測データの検出法を提案した。その結果、提案アルゴリズムにより物理的に意味のある波動分類が行われると同時に、従来の人手による解析法では発見が困難な観測データが検出可能なことが示された。

本論文では、提案アルゴリズムによって代表的な波動分類のクラスの構成と識別、更には例外データの検出が可能であることを示したが、波動現象間の境界部分で誤判別が生じることも確認されており、この領域での分類精度を上げることが今後の課題である。提案アルゴリズムは、従来、専門家が波動の微視的なスペクトル構造などをもとに現象論的に波動の種別分類を行っていた手法を計算機上で再現したという意味で、本論文で用いたMCAのデータに限らず、他の波動観測器のデータに対しても汎用的に適用可能である。しかしその場合、それぞれの観測器の時間・周波数分解能やダイナミックレンジなどに応じて、分類に用いる特徴量をどの程度の時間・周波数間隔で与えるか、また観測器の特性に応じたより有効な特徴量の選択法など、数多くの調整の余地が残されている。これら最適な特徴量の選び方の指針も今後の検討課題といえる。

文 献

- [1] I. Kimura, K. Hashimoto, I. Nagano, T. Okada, M. Yamamoto, T. Yoshino, H. Matsumoto, M. Ejiri, and K. Hayashi, "VLF observations by the Akebono (EXOS-D) satellite," J. Geomagn. Geoelectr., vol.42, no.4, pp.459-478, 1990.
- [2] K. Hashimoto, I. Nagano, M. Yamamoto, T. Okada, I. Kimura, H. Matsumoto, and H. Oki, "EXOS-D (AKEBONO) very low frequency plasma wave instruments (VLF)," IEEE Trans. Geosci. Remote Sens., vol.35, no.2, pp.278-286, 1997.
- [3] 笠原禎也, 佐藤 亨, "科学衛星巨大データセットを用いた電磁波動現象の多次元発見的解析に関する研究," 人工知能学研資, SIG-FAI-A9901-35, pp.165-168, 1999.
- [4] E. Sorali and Z. Ling, "Texture segmentation using hierarchical wavelet decomposition," Pattern Recog-

論文／あけぼの衛星のデータベースを用いたプラズマ波動スペクトルの自動分類法

nit., vol.28, no.12, pp.1819-1824, 1995.

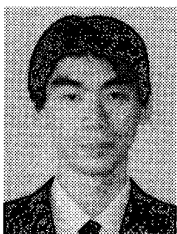
- [5] O. Pichler, A. Teuner, and B.J. Hosticka, "A comparison of texture feature extraction using adaptive Gabor filtering, pyramidal and tree structured wavelet transforms," *Pattern Recognit.*, vol.29, no.5, pp.733-742, 1996.
- [6] T. Kohonen, "The self-organizing map," *Proc. IEEE*, vol.78, pp.1464-1480, 1990.
- [7] S. Grossberg, "Adaptive pattern classification and universal recoding. I. Parallel development and coding of neural feature detectors," *Biol. Cybern.*, no.23, pp.121-134, 1976.
- [8] S. Grossberg, "Adaptive pattern classification and universal recoding. II. Feedback, expectation, olfaction, and illusions," *Biol. Cybern.*, no.23, pp.187-202, 1976.
- [9] R.C. Tryon and D.E. Bailey, *Cluster Analysis*, McGraw-Hill, New York, 1970.
- [10] 田中 豊, 脇本和昌, 多変量統計解析法, 現代数学社, 1983.
- [11] 中溝高好, 信号解析とシステム同定, コロナ社, 1988.
- [12] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Autom. Control*, vol.AC-19, no.6, pp.716-723, 1974.

(平成 14 年 3 月 29 日受付, 11 月 1 日再受付)



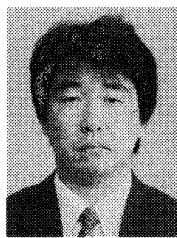
秋元 陽介 (正員)

平 12 京大・工・電気電子卒. 平 14 同大大学院情報・通信情報修士課程了. 在学中, 宇宙空間中のプラズマ波動の自動分類法の研究に従事. 現在, NTT アクセスサービスシステム研究所勤務.



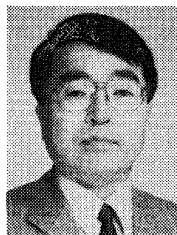
後藤 由貴 (正員)

平 10 京大・工・電気第二卒. 平 12 同大大学院情報・通信情報修士課程了. 現在, 同大学院博士後期課程在学中. 科学データベースからの知識発見, 宇宙空間中のプラズマ波動の伝搬の研究に従事. 平 14 国際電波科学連合若手研究者賞受賞.



笠原 禎也 (正員)

平元京大・工・電気第二卒. 平 3 同大大学院修士課程了. 現在, 金沢大・工・情報システム助教授. 科学データベースからの知識発見, 宇宙空間中のプラズマ波動の伝搬, 波動-粒子相互作用の研究, 衛星通信プロトコルの研究に従事. 工博. 地球電磁気・地球惑星圏学会, 米国地球物理学学会連合各会員.



佐藤 亨 (正員)

昭 51 京大・工・電気第二卒. 昭 53 同大大学院修士課程了. 昭 56 同博士課程研究指導認定退学. 現在京大・情報・通信情報教授. レーダによる大気, 降雨, スペースデブリの観測, 地下探査レーダの信号処理, 衛星通信プロトコルの研究に従事. 工博. 航空宇宙学会, IEEE, 地球電磁気・地球惑星圏学会各会員. 昭 61 地球電磁気・地球惑星圏学会田中館賞受賞.