

# 大規模データの発見的全探索： 大規模地球電流系構造の自動同定

統計数理研究所 樋 口 知 之

(受付 1999 年 4 月 13 日；改訂 1999 年 6 月 30 日)

## 要 旨

地球の磁気圏には、太陽風との相互作用の結果、大規模な電流系が構成されている。その規模、形態などは時間・空間的に変動の大きいものである。本研究では、一日 10 数回絶えず極軌道を周回している人工衛星のデータから、この電流系の様相を自動的にかつオンライン処理で特徴づける手法を提案した。一回人工衛星が極域を通過するたび、主として磁気経度方向成分一変量にのみ、大きい磁場変動が観測される。このスカラー空間系列は、観測された 3 次元空間系列に対して、主成分分析から得られる情報をもとに適切な座標変換を行うことにより導出される。さらにこのスカラー変量に、可変ノードスプライン関数を最小 2 乗法であてはめ、その節点の位置と値で、電流系の特徴付けを行った。この当てはめにより、5 年間のすべてのデータセットの大規模地球電流系構造の自動同定及びタイプ別への分類が実行され、地球物理学的見地からはさまざまな新しい知見が得られつつある。

キーワード：AIC, 可変ノードリニアスプライン, 主成分分析, 情報圧縮, 発見科学。

## 1. 研究の背景

### 1.1 理解から予測へ

移動体通信や GPS (Global Positioning System) 等に代表されるように、人間生活が個人レベルにおいても瞬時かつグローバルな情報利用に非常に依存した社会になりつつあることを踏まえると、それを可能にする人工衛星の保守管理はきわめて重要な課題である。人工衛星の保守の観点から、動的要因として考慮すべき主たる環境因子は、つきつめれば太陽から吹き出す超音速プラズマ流 (そのスピードは数百 [km/s]), いわゆる太陽風活動の時間変動である。商用の人工衛星は、静止軌道 (高度地球半径の 6.6 倍の赤道上空) または、低高度周回軌道 (周期 100 分, 高度約 800 km 程度) に位置しているため、太陽表面でおこった突発的な激しい活動が地球近傍の宇宙空間に到達するまでには時間差がある。したがって、地球から遙か遠く太陽風上流中に常駐している人工衛星 (Fig. 1 では、左端に示した人工衛星) での太陽風物理量の直接観測や、リモートセンシング技術にもとづく太陽表面の活動状況のモニターにより、非常時には前もって、これらの人工衛星に対しその搭載機器を保護するための司令を送ることが可能である。

太陽風は地球固有の内部磁場と複雑な相互作用の結果、Fig. 1 に模式的に示すように、地球磁気圏と呼ばれる磁場の領域を地球近傍の宇宙空間に形成する。これと同時に、地球の電離層と地球磁気圏の境界面をつなぐ巨大な電流系を生み出す。極域地方に特有の自然現象である

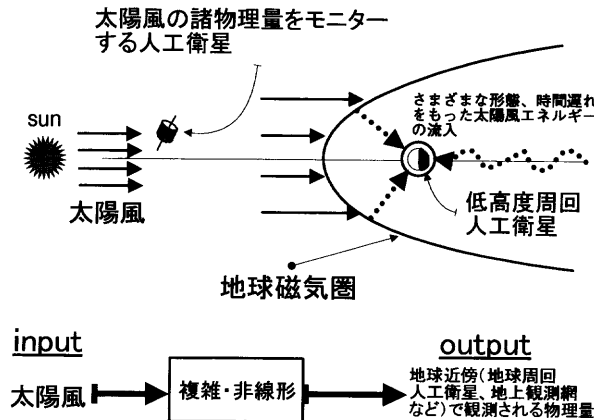


Fig. 1. 太陽風と地球磁気圏の関係の模式的図.

オーロラは、この大規模な電流系—— Large Scale Field Aligned Current systems, 以後 LSFAC と呼ぶ—— が関連した物理プロセスの一部分である。LSFAC は、磁気緯度の絶対値で  $70^{\circ}\sim 80^{\circ}$  の、その中心が夜側に系統的にずれたリング状に局在する。太陽風からのエネルギーの流入メカニズムは、我々日常生活で想起できるような単純なものではない(詳しくは、入門的教科書 Kivelson and Russell (1995) を参照されたい)。その理由として大きく次のような点があげられる。

- ・宇宙空間に存在する粒子どうしは、ほとんど衝突することがないほど宇宙空間は真空に近い。
- ・粒子は電荷をおびたバラバラの状態、いわゆるプラズマの状態にある。プラズマは、磁力線に沿った方向にはきわめて移動しやすいが、その垂直方向に移動しにくいという、粒子移動に関する異方性がいちじるしい。
- ・磁力線は自己組織的に変態(実際には変わるのは、磁力線のつくる流れの場)し、複雑なかたちで自分自身のもつ磁場エネルギーを開放する。

この結果、Fig. 1 に模式的に示すように、太陽風のエネルギーは、境界面前面から直接流入したり、磁気圏の後ろ(磁気圏尾部とよぶ)から時間遅れをもって流入したりと、そのエネルギー流入の物理プロセスはかなり複雑である。したがって、太陽風エネルギーの地球への流入物理メカニズムの理解は、太陽風活動が地球圏へ及ぼす影響の予測に必須な基礎研究テーマである。

本研究では、この物理プロセス上特に重要な役割をはたす LSFAC に注目する。宇宙空間では電流そのものを測定することは難しいが、電流によって生じる磁場変動の測定は簡単である。そのため、人工衛星で観測された磁場データを解析し大規模電流系の 3 次元構造を推定する研究がこの 20 年来かなり精力的になされてきた。その結果、太陽風エネルギーの地球磁気圏への侵入メカニズムの大筋は少なくとも定性的には解明されたと理解されている。しかしながら一歩進んだ、太陽風活動が地球圏へ及ぼす影響の定量的な予測には、太陽風が input, 低高度人工衛星で(あるいは地上で)観測される物理量を output とする信号処理的センスで粗視化した理解において(Fig. 1 下部分を参照)、その両者をつなぐブラックボックスの数理モデルが必要である。

## 1.2 偶然的発見プロセスからの離陸

上述したように定量的な予測には、太陽風と地球磁気圏の複雑かつダイナミックな相互作用を、精緻にかつ柔軟に記述できる数理モデルの開発が求められる。モデル開発研究の大きな流れは、現象を理想化した物理モデルにもとづくシミュレーションを利用した演繹的なアプローチである。これとは逆の次のような帰納的なアプローチも有力であろう。さまざまな人工衛星によって蓄積されてきた超大量のデータからのすべての情報を有機的に統合することによって、データベースと非線形関数を組み合わせたような機能性をもつ数理モデルの構築をめざすものである。

後者のプロジェクトの端緒として、人工衛星によって観測された約5年分の磁場データをすべて解析し、データからLSFACの構造を自動的に抽出・分類する作業に取り組んでいる。Fig. 2に示したように、LSFACは、同緯度に沿ったシート状——カレントシートと呼ばれる——のものが数層重なった構造をもつことが知られている (Iijima and Potemra (1976))。1.1節で説明したプラズマ粒子の移動のしやすさに関する異方性のため、基本的に電流は地球固有の磁力線方向 (図でMagnetic field と表したベクトル) に沿って流れるので、電流の向きは地球に流れ込む (Downward) かそれとも流れ出す (Upward) かの2方向である。Fig. 2には、低緯度側から (Up, Down, Up) の3層のカレントシートが模式的に示してある。従来の地球電磁気学の知見においては、カレントシートの数はただだか3層であったが、4層を示唆する磁場変動パターンがたまたま発見された (Ohtani et al. (1995))。その論文には4個の4層電流系の発見が報告され、その解釈のために新しい相互作用のモデルが仮説として提案されている。そもそも本研究の発端は、この4例しか見つかっていない新しいタイプの電流系は、実はさまざまな偶然的物理条件が重なり合った結果の希な現象なのか、それとも今まで見過ごされてきた実はごく普通に存在するものなのかを明らかにすることにあつた。そのため、低高度の極周り人工衛星の今までに蓄積された超大量の磁場データを丹念に調べあげ、新しいタイプの電流系を発見するプロジェクトに取り組み始めた。

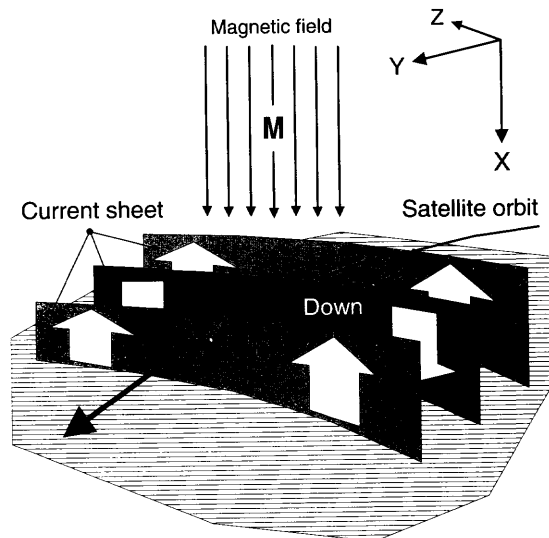


Fig. 2. LSFAC 構造の模式図。3層のカレントシートを人工衛星が横断する例を示している。

LSFACは同じタイプ(例えばシート数が同じ)の電流系であっても、その空間的位置や電流値は、時々刻々と変化する太陽風の影響や日照条件(極域地方の太陽となす角度)の違いなどさまざまな要因でかなり変動する。このため従来この作業は、解析に慣れた熟練した研究者の目による作業にほぼ依存していた(樋口(1995))。よって、LSFACの空間的位置や電流値の変動性を容認しつつ、タイプ別に分類する柔軟な手法の開発が必須である。このような手法を開発し、新しいタイプの電流系の観測例を巨大なデータセットから組織的かつ自動的に探索することが本研究の目的である。詳しい手続きの説明はHiguchi and Ohtani(1998)にあるため、本論文では、採用された統計モデルと物理モデルとの関係及び、モデル選択がLSFACの自動同定においてどのような役割を果たしているのかについて重きをおいて解説する。

## 2. 物理モデルとしての piecewise linear function

### 2.1 低高度周回極周り人工衛星データ

解析に使用した人工衛星データは、DMSP-F7と呼ばれるアメリカの人工衛星によって1983年12月から1988年の1月までの期間取得された、サンプリング時間が1秒の磁場3成分データである。この磁場成分には、もともと観測された磁場データから地球固有のモデル磁場の値(Fig. 2で $M$ と示したもの)が予め差し引かれた、LSFACに関連する磁場変動のみが含まれる。DMSP-F7の軌道は、低高度周回(地球1周回時間が約100分程度、平均高度835[km])で、極周り、つまり地球儀での縦周りである。前述したようにLSFACは、磁気緯度 $70^{\circ}\sim 80^{\circ}$ 上に同磁気緯度に沿ったシート状の構造をもつので、この人工衛星の軌道は、Fig. 2に示すようにLSFACを垂直に横断するものとなり、LSFACの空間構造を観測するのに適している。

Fig. 3(a)に北半球、(b)に南半球の各極域で、DMSP-F7がカバーする領域を示す。解析したデータセットに含まれるすべての衛星軌道を実際にプロットすると、この示された領域を隙間なく稠密に埋める。地球の磁気圏の物理特性は、太陽の方向との相対的關係でほぼ決まっているので、地方時が経度方向の座標として意味をもつ。そのためFig. 3では、座標は地理的座標系でなく、PACE coordinates というモデルと計算法を用いて地理的座標系を射影・変換した、地磁気上の緯度、地方時に相当する座標系(Magnetic Latitude, Magnetic Local Time) — 普通(MLat, MLT)と略して記載される — で表示してある。人工衛星の軌道は地理学的に制御

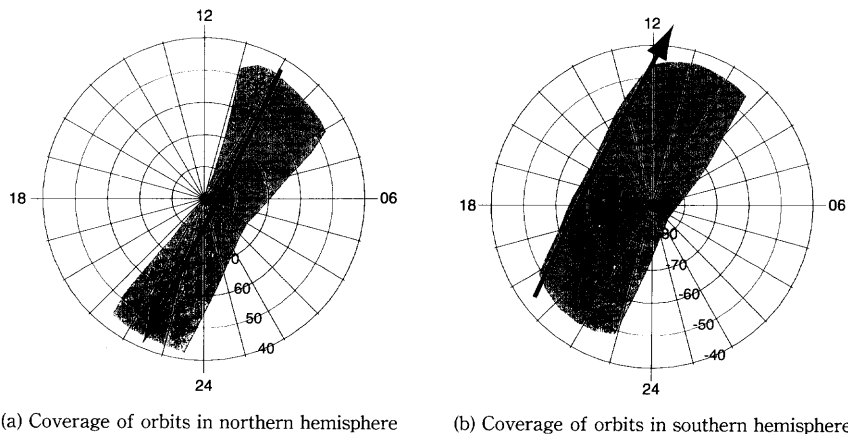


Fig. 3. (MLT, Mlat)の座標系における、全期間のDMSP-F7の軌道がカバーする領域。(a)北半球、(b)南半球。南半球の矢印は、skimming orbitの例を示している。

されている一方、カレントシートの形状は地磁氣的に定まるので、軌道はいつもカレントシートを垂直に近いかたちで横断するとは限らない。時には Fig. 3(b) に示した軌道のように、軌道がほぼカレントシートに沿うようになる場合、いわゆる *skimming orbit* もあることには注意する必要がある。

本研究では、連続して得られる観測データを人工衛星の軌道情報を利用して、まず半球毎のデータに分ける。上述したように LSFAC は極域にリング状にあるので、半球毎に 2 回この領域を通過することになる。分割した半球毎のデータを、一つのデータセットが一つの LSFAC に関連した磁場変化を含むよう、人工衛星軌道の磁気緯度を参照してさらに 2 分割する。荒っぽくいえば、1 周回のデータを、北半球・昼間側、北半球・夜側、南半球・夜側、南半球・昼間側の 4 つに分割することになる。明らかに LSFAC に関係ない区間は、衛星の軌道情報を元にあらかじめ切り捨てておく。

このようにして作成された一つのデータセットには、約 600~800 点の 3 次元磁場データが含まれる。本研究では、7 万以上の個数のデータセットを取り扱う。今、あるデータセットを

$$(2.1) \quad \mathcal{B} = [\mathbf{B}_1, \mathbf{B}_2, \dots, \mathbf{B}_N]$$

と記す。 $\mathbf{B}_n = [B_{x,n}, B_{y,n}, B_{z,n}]'$  は、 $n$  番目の磁場 3 成分のベクトル値。また  $'$  は転置をあらわす。また、測定単位は [nT] である。 $\mathbf{B}_n$  のレンジは、あるデータセットでは 20 [nT] 程度である一方、1,000 [nT] になるくらい大きいものもある。 $B_x$  は地球の中心方向、 $B_y$  は  $B_x$  軸に垂直な平面に衛星の速度ベクトルを射影した方向、 $B_z$  は右手系で座標系を作るように定義されている。後述のために、この衛星に固定された座標系を衛星座標系と呼ぶことにする (Fig. 2 右上部分を参照)。前述したように、電流は地球の固有磁場に沿って、つまりダイポール磁場の磁力線に沿って流れるため、LSFAC に関連した磁場変動は  $YZ$  平面内にほぼ含まれる。さらに Fig. 3(a) に太矢印で示したように、DMSP-F7 の軌道は第 0 近似として子午面内 (真昼と真夜中をとる面) に近いので、 $B_z$  成分がカレントシートに沿う方向になる。

データセット中のベクトル数  $N$  は各データセット毎にかわる。時間でいえば、約 10 分から 13 分のデータを一つのデータセットとして取り扱う。この 10~13 分の間にも太陽風の状態は大きく変化することもあるかもしれないが、このプロジェクトでは、一つのデータセット内の磁場変動はほぼ LSFAC の空間構造をみているものだという大前提にたつ。つまり  $\mathcal{B}$  を空間系列として取り扱う。これらの磁場データ以外に実際のデータには、各時刻での衛星の軌道情報があたえられている。

## 2.2 物理モデルと統計モデルの対照

カレントシートに関連した磁場変動は、カレントシートが経度方向に十分幅広く展開しているという物理モデルの仮定のもとに、シートに沿った方向——これ以後  $A$  軸方向と呼ぶ——の磁場成分、 $B_A$  成分にのみあらわれる。このことを、Fig. 2 で示した 3 層のカレントシートを、人工衛星が断続的に横切った場合を例にとって具体的に説明する。いま、Fig. 4 に示すようにカレントシートは、 $Z$  軸から  $\phi$  だけずれた方向 (図右上部分を参照) に平行に広がっているものとする。人工衛星は、図の上から下に向けカレントシートを横断し、理想的状況のもとでは、その軌道上の各点において図中にベクトルで示したような磁場変動 (図では Magnetic perturbation) を観測する。LSFAC に関連したベクトル (図では太いベクトルで示してある) の主たる成分は、 $A$  軸方向となるように描いてある。いま、この時の  $B_A$  成分を軌道にそってプロットすると、Fig. 5(a) に点線で示すような、その平均的構造が区分的に直線 (図では太実線) となるような空間系列となる。この時、直線が折れ曲がる点、いわゆる節点 (図では黒太点) の位置は、電流の向きがかわるところ、つまりシートの境界面になる。また、各線分の縦方向の gap はカ

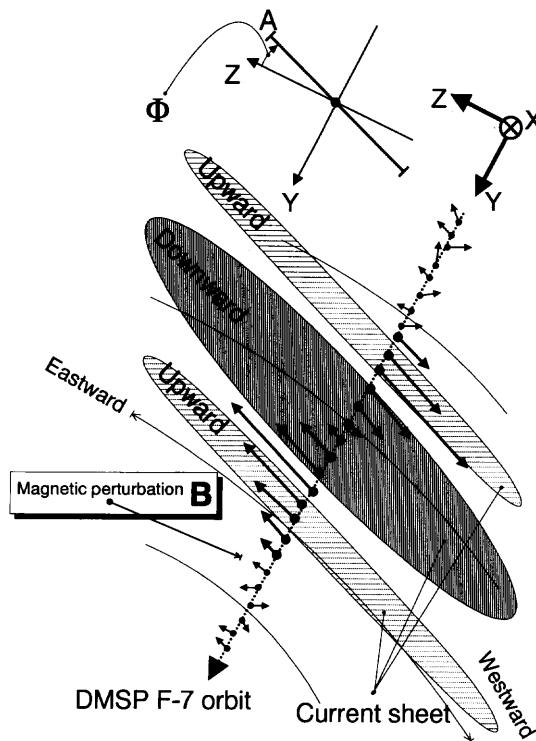


Fig. 4. Fig. 2 で示した LSFAC を横断するとき観測される磁場変動を模式的にあらわした図.

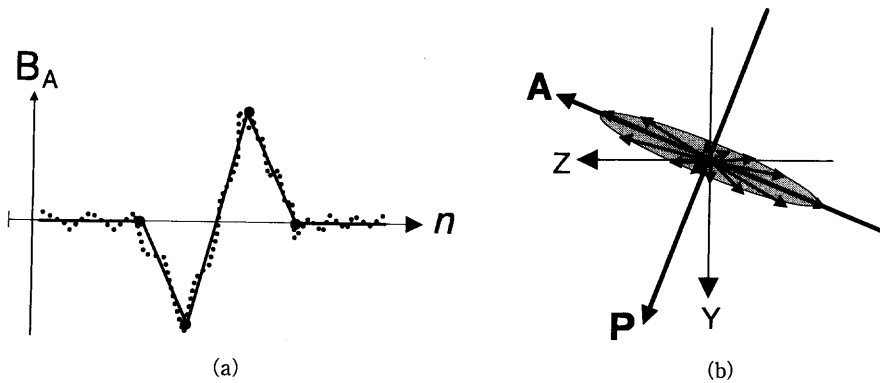


Fig. 5. (a) Fig. 4 で示したベクトルの、カレントシートに沿う方向 (A 軸方向) 成分 (点線) と、その観測値にあてはめる関数 (実線). (b) Fig. 4 に示したベクトル系列のうち、LSFAC に関連したものの、YZ 平面へのプロット.

レントシートの強さに対応する。

観測されたスカラー空間 (あるいは時) 系列を  $y_n$  ( $n = 1, \dots, N$ ) と記すことにする。本研究では、一つのデータセット中の  $B_A$  成分が  $y_n$  に相当する。上述の区分的直線は、piecewise linear function, あるいは polyline ととも簡単に呼ばれる、可変節点リニアスプライン関数 (Williams (1978), Hiragi et al. (1985), Friedman and Silverman (1989)) で表現できる。いま、節点の位

置 (node position) を  $t_i$  ( $i = 1, \dots, M$ ), 値 (node value) を  $g_i$  ( $i = 0, 1, \dots, M, M+1$ ) で表すことにする.  $g_0$  と  $g_{M+1}$  は端 ( $n = 1$  及び  $n = N$ ) での node value である. 場所  $n$  での polyline の値を  $f(n|\mathbf{T}_M, \mathbf{G}_M)$  と記すことにする. ただし,  $\mathbf{T}_M = [t_1, \dots, t_M]'$  および  $\mathbf{G}_M = [g_0, g_1, \dots, g_M, g_{M+1}]'$  である. 本研究では,  $\mathbf{T}_M$  及び  $\mathbf{G}_M$  を, 観測量と polyline の残差の二乗誤差

$$(2.2) \quad \text{RSS}(\mathbf{T}_M, \mathbf{G}_M|M) = \sum_{n=1}^N [y_n - f(n|\mathbf{T}_M, \mathbf{G}_M)]^2$$

を最小にするように定める. ここで注意すべき点は, 通常のスプライン回帰と違い, node position の位置および, さらには節点数  $M$  もデータセットに依存することである.

この  $f(n|\mathbf{T}_M, \mathbf{G}_M)$  は,  $\mathbf{T}_M$  が所与のもとで,  $M+2$  個の 1 次 B-スプラインの基底関数  $F_m(n|\mathbf{T}_M)$  をもちいて, 次式のように書き換えることができる.

$$(2.3) \quad f(n|\mathbf{T}_M, \mathbf{H}_M) = \sum_{m=0}^{M+1} h_m F_m(n) = \mathbf{F}_n'(\mathbf{T}_M) \cdot \mathbf{H}_M,$$

$$\mathbf{H}_M = [h_0, h_1, \dots, h_M, h_{M+1}]',$$

$$\mathbf{F}_n(\mathbf{T}_M) = [F_0(n|\mathbf{T}_M), F_1(n|\mathbf{T}_M), \dots, F_M(n|\mathbf{T}_M), F_{M+1}(n|\mathbf{T}_M)].$$

$F_m(n|\mathbf{T}_M)$  は,  $n$  軸上の線分  $[t_{m-1}, t_{m+1}]$  を底辺,  $n = t_m$  線上の正の部分に頂点をもつ三角形である. 具体的な数式は, 例えば Hiragi et al. (1985) 等を参照していただきたい.  $\mathbf{F}_n(\mathbf{T}_M)$  は既知のベクトルであるから,  $\mathbf{T}_M$  が与えられれば, (2.2) は以下のように  $\mathbf{H}_M$  に関して 2 次形式となる.

$$(2.4) \quad \text{RSS}(\mathbf{T}_M, \mathbf{H}_M|M) = \sum_{n=1}^N [y_n - \mathbf{F}_n'(\mathbf{T}_M) \cdot \mathbf{H}_M]^2.$$

この  $\mathbf{T}_M$  が与えられたもとの最小 RSS を,

$$(2.5) \quad \text{RSS}(\mathbf{T}_M) = \text{RSS}(\mathbf{T}_M, \mathbf{H}_M^* = \mathbf{H}_M(\mathbf{T}_M)|M)$$

とあらわすことにする. 残念ながら (2.5) は  $\mathbf{T}_M$  に関しては非線形関数であるため, その最適化は難しい. 本研究で採用した最適化手法は 4 章にて説明する.

本研究の一つの目標は, 最終的に  $(\mathbf{T}_M, \mathbf{G}_M)$  により LSFAC による磁場変動を代表させるような, つまりある意味でのデータの情報圧縮である. このような polyline の最小二乗法による当てはめは, 節点数が少ない (1 から 2) モデルの場合には変化点を探索する統計的手法として, またやや節点数の多い (3 から 5 程度) 時には, 加法的観測ノイズの影響が大きい少数データの平滑化として行われるケースが多い. 変化点問題に関しては例えば Kashiwagi (1991) を参照していただきたい. また後者は, Friedman and Silverman (1989) に詳しい解説がある. これらは, 観測量の特性を明示的にはとりこんでいない, 解析対象の属性を捨象したある意味での統計モデルの利用としてとらえられる. 一方, 本研究における polyline の当てはめは, 観測対象の背後にある物理モデルから導出される観測量のモデルが, polyline となる事実に基づいている. 従って少なくとも本研究においては, LSFAC の構造に対する物理モデルが実際の LSFAC をある程度近似できる限りにおいて, 高次オーダーのスプライン関数の当てはめなどは考える必要がないことを強調しておく. これと同じ理由で, 非常に多数 — 具体的にはデータの個数と同程度, あるいはそれ以上 — の基底を用意するスプライン関数の当てはめ (石黒・荒畑 (1982), 最近の詳しい解説は Eilers and Marx (1996)) は考えない. このアプローチは, Penalized-Spline から P-スプライン, あるいはノンパラメトリック・スプラインと呼ばれている.

### 2.3 モデル選択の役割

以上の表現法および計算法は、次のような統計モデル（データ  $y_n$  発生確率的構造）

$$(2.6) \quad p(y_n | n; \mathbf{T}_M, \mathbf{H}_M) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{(y_n - f(n | \mathbf{T}_M, \mathbf{H}_M))^2}{2\sigma^2} \right\},$$

を考えたもとの、 $\mathbf{T}_M$  と  $\mathbf{H}_M$  の推定に最尤法を用いたことと同等である。 $f(n|\cdot)$  は、 $2M+2$  個のパラメータを含むため、この統計モデルのパラメータ数は  $\sigma^2$  を加えた、 $2M+3$  個であることに注意する。

$M$  が所与のとき (2.4) を最小にする  $(\mathbf{T}_M, \mathbf{H}_M)$  を  $(\mathbf{T}_M^*, \mathbf{H}_M^*)$  と記すことにする。電流系のタイプを決定する節点数  $M$  の選択は、赤池情報量規準 (AIC) (Akaike (1973, 1974)) によって客観的に行う。節点数が  $M$  の場合 AIC は以下のように定義される (Hiragi et al. (1985))。

$$(2.7) \quad \begin{aligned} \text{AIC}(M) &= N \log \frac{\text{RSS}(\mathbf{T}_M^*, \mathbf{H}_M^* | M)}{N} + 2(2M+3) + C \\ &= N \log \frac{\text{RSS}(\mathbf{T}_M^*, \mathbf{H}_M^* | M)}{N} + 4M + C' \end{aligned}$$

$C$  及び  $C'$  は、 $M$  に依存しない定数である。AIC を用いたモデル選択の場合、相対値しか意味をもたないのでこの値を具体的に定める必要はない。以後、AIC 最小化によって選択された最適な節点数を  $M^*$  とあらわすことにする。節点数  $M$  の選択は、電流系のタイプを決定することに相当する。

ここで注意すべき点は、(2.6) で与えられるモデルのパラメータ数の取り扱いである。 $\mathbf{T}_M$  を変化させることは、モデルのデータへの当てはまりの良さと関係してくるので、これは対数尤度で測っていると考え、したがってバイアス補正項であるパラメータ数は、 $\mathbf{H}_M$  の  $M+2$  個に  $\sigma^2$  を加えた、 $M+3$  でいいという考え方もできる。つまり、(2.7) のようにバイアス補正項を  $4M$  ではなく、 $2M$  でモデル選択をするわけである。両者のどちらが適切であるかは、真のモデル  $f(n|\cdot)$  の形にも実は関係してくるため (北川 他 (1994))、モデル選択にどの基準をもちいるべきかは注意深く行う必要がある。例えば、BIC, CV, GCV などのほかの選択基準の併用が、理論的観点からは望ましい (井元・小西 (1999))。ただ、本研究では次のような2つの理由により、モデル選択基準として (2.7) を採用した。一つには、最終的なカレントシート数の推定には、さまざまな地球物理的な条件——例えば、となりあう区間で、もし傾きが同じ符号であるならば、連続した一つの区間とみなす。あるいは、あるレベル以下の大きさの構造は、カレントシートとは見なさない等々——を加え、頑健な構造を抽出する手続きがさらに必要である。よって、最初から少なめの節点数を推定しておくほうが、実質的なデータ解析の上では適当である。もう一つの理由は、明瞭な構造をもつカレントシートのみを最終的な地球物理学的研究対象として拾い上げるので、そのような場合は、RSS( $\cdot|M$ )がある  $M$  で急激に減少し、それ以上のノード数では極めてゆっくりと減少するといった変化を示すため、モデル選択においては  $2M$  と  $4M$  の差はほとんどでてこないといった、解析結果にもとづくものである。

## 3. 座標変換

### 3.1 主成分分析の適用

2.1 節で触れたが、第0近似として  $B_z$  成分に主として、カレントシートに関連した磁場変動が現れる。しかしながら、衛星の軌道は必ずしも磁気緯度線を垂直に横断するわけでないので、衛星座標系の  $B_z$  成分が常に磁気緯度線に沿うことは保証されていない。実際 Fig. 3(b) に太矢



印で示した skimming orbit の場合、部分的区間において、 $B_z$  成分は磁気緯度線にほとんど垂直なものとなる。この事実を踏まえると、各観測点での同 MLT (いわゆる磁気経度) 方向と磁気緯度方向を新しい  $B_r$ ,  $B_z$  とするような、いわば局所磁気経度・緯度座標系に座標系を変換したほうが適切であるように思われる。ところが実際の LSFAC は多様な構造を示すために、第 0 近似としてはカレントシートは磁気緯度線に沿っているが、同 MLT (つまり磁気経度線) 方向よりは緯度線方向を向いているといった程度にも、その緯度線からのずれは大きいものである。よって Fig. 4 に模式的に示したように、データセット毎にその観測時のカレントシートに沿う方向に座標系を回転することが望ましい。問題は、データからどのようにしてカレントシートに沿う方向、つまり  $A$  軸方向を算出するかである。

今、Fig. 4 中のベクトルのうち、LSFAC の領域にある部分 (図で太ベクトルで示したもの) だけを拾い出し、YZ 平面にプロットしてみる。その結果は、Fig. 5(b) にあるようなものになることが予想される。本研究では、 $(B_r, B_z)$  の 2 成分空間系列に主成分分析を適用した後、最大固有値 (つまり大きいほうの) 成分を、LSFAC に関連した磁場変動のカレントシートに沿う成分とする便宜的な手法を採用した。主成分分析にもとづき、最大主成分方向の磁場成分を、前述の  $B_A$ 、またもう一方を  $B_P$  とする。 $P$  は poleward 方向を意味する。ここで  $B_A$  の正の方向は、 $B_z$  の正の方向からのずれの小さいものをとる。

大きいほうの固有値を小さいほうの固有値で割った値の平方根を  $\alpha$  と記して、実際の LSFAC の構造をカレントシートでどの程度近似できているのかをあらわす目安とした。理想的物理モデルにおいては  $\alpha$  は無限大となるが、実際観測されるデータにおいては、LSFAC に関連した磁場変動は他成分にも染み出してくるため、現実には  $\alpha \approx 1$  のような、明らかにカレントシートの近似モデルとしての妥当性が疑わしい場合もある。この染み出しの原因として大きなものは、

- i) シート構造の経度方向への広がり有限である
- ii) 地球固有モデル磁場と実際の地球固有磁場の値との系統的なずれがある

の二つがあげられる (Higuchi and Ohtani (1988)).

### 3.2 初期準備的 LSFAC 区間の同定法

主成分分析をデータセットにそのまま適用するには次のような問題がある。カレントシートは上述したように極域に局在しているので、LSFAC に関連した磁場変動も、ほぼカレントシートの近傍にのみ観測される。LSFAC が存在しない領域では、観測される磁場擾乱 (Fig. 4 では細いベクトルで示したものは非常に小さい。従って、本研究の狙いに即した意味ある座標変換を行うためには、限定された区間に主成分分析を適用する必要がある。このため、あらかじめ LSFAC の領域 (区間) を大まかにでも同定する必要がある。この初期準備的 LSFAC 区間の同定の手続きは、詳しくは Higuchi and Ohtani (1998) にあるので、ここでは北半球昼間側の観測を例に取りながら、その手続きの流れの簡単な説明にとどめる。

3.1 節に説明したように、極域での多くの軌道においては、 $B_z$  成分は経度方向 (緯度線に沿う方向) の磁場を観測している。従ってまず  $B_z$  成分の磁場変動に注目する。課題は  $B_z$  の空間系列を、赤道側 (Equatorward), LSFAC 内 (inside of LSFAC), 及び極域側 (Polarward) の 3 区間 — 以下にこれらの区間をおおの  $I_E^Z$ ,  $I_L^Z$ , および  $I_P^Z$  と各々記す — に分割することである。分割点は、赤道側から  $L_E^Z$  および  $L_P^Z$  とあらわす。ここで添え字の  $Z$  は、初期準備的 LSFAC 区間の同定の手続きが  $B_z$  成分をもとに行われたことを明示的に示すためである。

$I_E^Z$  や  $I_P^Z$  においては、3.1 節で言及した ii) の原因により、その系統的なずれが直線的トレンドとしてあらわれることがしばしば観測される。また直線の変動からのずれ — 残差系列 — は、

i.i.d. でその分散は比較的小さいと仮定できる。今仮に  $L_z^E$  と  $L_z^P$  を所与の量とする。ここでこの残差系列の大きさを見積もるために以下のような量を計算する：

$$(3.1) \quad U_z^E = \frac{1}{L_z^E - 1} \sum_{n=1}^{L_z^E - 1} [B_{z,n} - (a^E n + b^E)]^2$$

$$(3.2) \quad U_z^P = \frac{1}{N - L_z^P} \sum_{n=L_z^P}^N [B_{z,n} - (a^P n + b^P)]^2,$$

ここで  $a^E$  と  $b^E$  ( $a^P$  と  $b^P$ ) は、 $\mathcal{I}_z^E$  ( $\mathcal{I}_z^P$ ) の区間における最小 2 乗当てはめによって定まる定数である。以降の解析のために、 $\mathcal{I}_z^E$  外での磁場擾乱の大きさの代表値として、(3.1) と (3.2) の二つの量の、観測点数の重みつき平均値

$$(3.3) \quad U_z^{E,P} = \frac{(L_z^E - 1) U_z^E + (N - L_z^P) U_z^P}{L_z^E - 1 + N - L_z^P}$$

を計算しておく。

$\mathcal{I}_z^E$  では一方、もし LSFAC が存在すれば、 $\sqrt{U_z^{E,P}}$  と比較して少なくとも数倍以上の大きさの構造をもった磁場変化が観測される。いいかえれば、区分的線分で近似できるような変化のその線分間の縦方向の gap は、もっとも小さいものでも  $\sqrt{U_z^{E,P}}$  よりも数倍以上の大きさをもつことが期待できる。逆にいえば、数倍以上の大きさをもつような変化のみを LSFAC に関連した磁場変動として注目し解析することになる。問題はこの構造をもった磁場変化の大きさを、何で見積もるかである。

厳格に計算を行うとすれば、2.2 節で言及した polyline の当てはめを行い、隣接する節点間の node value の差の絶対値を求めるべきであろうが、ここで求められているのは、座標回転を行うために主成分分析を適用する区間のおおまかな推定である。本研究では計算量の軽減のために、以下のような逐次的計算の必要の無い簡便な量で代用することとした。まず LSFAC に関連する磁場変動を浮き彫りにするため、133 点の半値幅をもつ線形平滑化フィルタ（いわゆるローパスフィルタ）を  $B_{z,n}$  に適用し、 $\mathcal{I}_z^E$  区間で、その平滑化された系列、 $\tilde{B}_{z,n}$  ( $n = 1, \dots, N$ ) の一階差分の分散を計算した。その量を具体的に式であらわせば

$$(3.4) \quad \tilde{U}_z^E = \frac{1}{L_z^P - L_z^E} \sum_{n=L_z^E+1}^{L_z^E} [(\tilde{B}_{z,n} - \tilde{B}_{z,n-1}) - b^F]^2$$

となる。ここで、 $b^F = 1/(L_z^P - L_z^E) \cdot \sum_{n=L_z^E+1}^{L_z^E} (\tilde{B}_{z,n} - \tilde{B}_{z,n-1})$  である。133 点という中途半端な値は、実際に目によって定めた  $|L_z^P - L_z^E|$  の 300 データセットの平均値である。平滑化の具体的手続きは Higuchi (1991) を参照されたい。 $\tilde{B}_{z,n}$  の分散でなく、その一階差分の分散を計算したのは次のような理由からである。 $B_{z,n}$  で  $L_z^E$  や  $L_z^P$  が明瞭に同定（視認）できる場合、ローパスフィルタの適用により、 $\tilde{B}_{z,n}$  においてはそれらが不明瞭になってしまう。初期準備的 LSFAC 区間の同定ではあるが、計算量にあまり負荷がかからない範囲内でなるべく精度良く同定しておきたい。よって、一階差分に注目するという、画像処理におけるエッジをきわだたせる処理と同じ思想でもって、エッジをぼやかしてしまうローパスフィルタの影響を軽減することとした。ただ、何ら高階差分成分の採用を妨げる理由はないことだけは述べておく。

最適な  $L_z^E$  及び  $L_z^P$  は、

$$(3.5) \quad Q(L_z^E, L_z^P) = \frac{\tilde{U}_z^E}{U_z^{E,P}}$$

を最大化することで求まる。N 個のデータ点に対し、上記の量を計算すると比較的計算量が多

くなるので、精度を損なわない程度、具体的に言えば地球物理的に存在が疑わしいような非常に薄いカレントシートの幅以上、具体的には10点毎に間引いた点に対してのみ、 $Q(L_z^E, L_z^I)$ を計算した。この10点は、平均的衛星の速度、高度から計算すると、ほぼ $0.6^\circ$ の緯度幅に対応する。

章を終わる前に、この章で説明したことを実際の手続きの流れに従ってまとめておく。

**Z.1**  $Q(L_z^E, L_z^I)$ を最大にする $(L_z^E, L_z^I)$ の探索。

**Z.2**  $(B_{Y,n}, B_{Z,n})$  ( $n = L_z^E, \dots, L_z^I$ ) に対し主成分分析を適用。座標変換行列の計算。

**Z.3** 座標変換をすべての点に対して行い、 $B_{A,n}$  ( $n = 1, \dots, N$ ) を求める。

#### 4. 可変節点位置の最適化

この章では、 $T_M$  に関して非線形関数である (2.5) の最小化問題を取り扱う。Hiragi et al. (1985) では、 $T_M$  そのものを変数とした準ニュートン法を採用することで解決している。しかしながらこのアプローチを直接適用すると、解の更新時に節点が交叉してしまい、結果として最適解の探索が破綻してしまう。従って実装時には、解の更新時において各節点位置  $t_i$  の移動に制限を設けた巧妙な制約を課す必要がある。この問題を避けるために簡単な解決法として：

- ・隣接する節点が接近してきたなら、お互いに反発するような擬似ポテンシャルを最適化関数にとりこむ
- ・最適化変数  $T_M$  を多変数ロジット変換により区間幅に変換する

が想起される。前者はどのようなポテンシャル関数を導入すべきかを定めるのが難しいことと、最適化に非常に時間がかかることで適切でないと判断した。また後者は、一つの区間幅の変更がすべての節点の位置に波及するので、準ニュートン法的逐次解法には馴染まないため採用しなかった。本研究では、まず逐次的処理でない既存の統計的手法を組み合わせた方法で可能な限り良い初期値を求め、それからの微小な改良を準ニュートン法でおこなうという、泥臭い方法を採用した。

##### 4.1 初期値の探索

適切な初期値の設定は、次のような手続きで求めた節点候補に対して、残差が最も小さくなる節点の組み合わせを探索することで達成した。

###### 4.1.1 節点候補の拾い上げ

座標変換の際 (3.2 節で説明) に利用した、初期準備的 LSFAC 区間の同定の手続きを再度  $B_A$  に対して行い、LSFAC の区間を大まかに同定する。求められた区分点を  $L_A^E$  及び  $L_A^I$  とあらわす。添え字の  $A$  は、これらの区分点が  $B_A$  成分をもとに求められたことを示すためである。次に、定まった3区間—— $\mathcal{I}_A^E$ ,  $\mathcal{I}_A^I$  及び  $\mathcal{I}_A^P$ ——各々において、空間系列の自己相関関数を計算する。自己相関関数から得られる特徴的空間距離をもとに、各々の区間でローパスフィルタを設計し平滑化を行う。特徴的空間距離は、自己相関関数が負の値をとる最短ラグの4倍、もし負の値をとらない場合は極小値をとる最短ラグの2倍の値で定義した。これらの特徴的空間距離からのローパスフィルタの設計は Higuchi (1991) を参照していただきたい。Fig. 6(a) に、1986年8月26日11周回目の北半球昼間側の  $B_A$  成分を示す。このデータセットにおいては、 $\alpha = 6.3$  また  $\Phi = 9.0^\circ$  である。二つの長い縦線は、 $L_A^E$  と  $L_A^I$  の位置に対応する。図中で示された太曲線は、平滑化された空間系列である。

各区間において、平滑化された空間系列の極大値、極小値のすべてを節点候補として拾い上

げる。そのような節点候補を

$$(4.1) \quad T^E = [t_1^E, \dots, t_{j^E}^E]$$

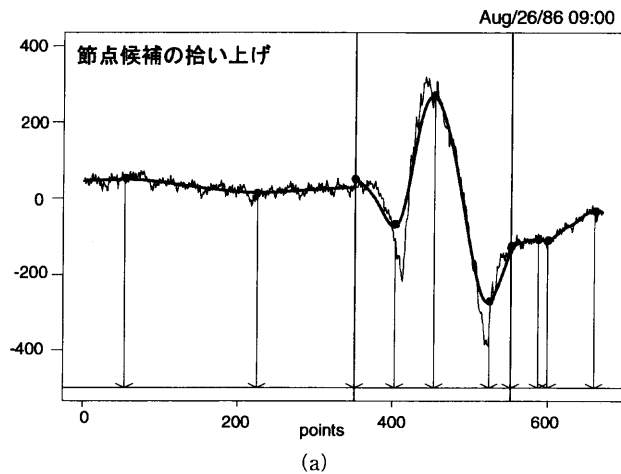
$$(4.2) \quad T^F = [t_1^F, \dots, t_{j^F}^F]$$

$$(4.3) \quad T^P = [t_1^P, \dots, t_{j^P}^P]$$

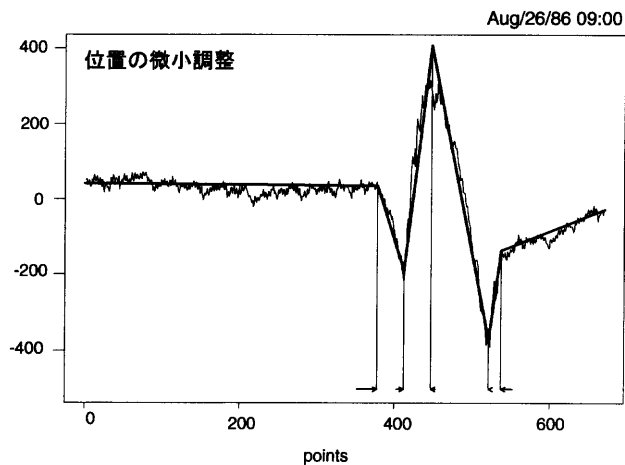
とあらわす。ここで  $J^E$ ,  $J^F$ , 及び  $J^P$  は、各区間での極大点と極小点の個数の和である。Fig. 6に示した例では、 $J^E = 2$ ,  $J^F = 3$ , そして  $J^P = 3$  である。これらに  $L_A^E$  と  $L_A^F$  の二つも節点候補に取り込む。よって後述のため、すべての節点候補をまとめ

$$(4.4) \quad T_J = [t_1^E, \dots, t_{j^E}^E, L_A^E, t_1^F, \dots, t_{j^F}^F, L_A^F, t_1^P, \dots, t_{j^P}^P]$$

のように表記しておく。ここで  $J = J^E + J^F + J^P + 2$  である。Fig. 6 の場合は  $J = 10$  となる。



(a)



(b)

Fig. 6. (a)  $B_A$ 成分。太曲線は平滑化されたものを示す。黒太点は極大、極小点および  $L_A^E$  と  $L_A^F$ 。(b)最終的にあてはめられた polyline.

#### 4.1.2 組み合わせ的最適化

今、節点数 $M$ 個の polyline のあてはめを考える。まず先ほどの候補  $T_j$  から $M$ 個を選び出し、節点位置を与えるベクトル —  $T_M(k)$  ( $k = 1, \dots, K_M$ ) — を構成する。ここで  $K_M$  は  $K_M = J C_M$  である。与えられた  $T_M(k)$  に対し  $RSS(T_M(k))$  を求め、その中で最小のものをあたえるベクトルを  $T_M^*$  と表す。一つのデータセットにおいて、節点候補数 $J$ は多目に見積もっても約20個、普通は10程度である。またあてはめる polyline の節点数 $M$ は、現在共通の理解を得ているカレントシート数は最大4であるから、従来の知見では最大  $M = 5$  までを考えれば十分であることを踏まえて、さらに複雑な LSFAC の構造の存在発見の可能性を探る意味で最大7まで調べることにした。必要とされる残差計算の回数  $K_M$  は比較的大きな数になるが、それでも十分計算できる範囲である。

#### 4.2 解の微少調整

求まった  $T_M^*$  を初期値に、 $t_i$  そのものを最適化変数、 $RSS$  を目的関数とした準ニュートン法を適用する。この際、値の更新時において節点の交叉の問題を避けるために、 $t_i$  の移動に制限を設けている。この手続きにより改良された節点の位置を  $T_M^{**}$  と表すことにする。モデル選択はこの  $T_M^{**}$  でもって計算された AIC の値、つまり

$$(4.5) \quad AIC(M) = N \log \left( \frac{RSS(T_M^{**})}{N} \right) + 4J + \text{constant}$$

の比較でもって行う。Fig. 6(b)に、最終的な polyline のあてはめを示す。このデータの場合最適な節点個数は5となった。最下の横ベクトルは、各節点がどの程度微調整をうけたのか、つまり  $T_M^{**} - T_M^*$  の各成分を表している。図中において中央の W 部分が大規模電流系に関連した磁場変動である。

#### 4.3 アルゴリズム

2.2 節及びこの章で説明した手続きをまとめると以下のようなになる。

##### A.1 $M = 0, \dots, 7$ に対して

- $T_M^{**}$  を求める
  - 節点候補  $T_j$  の構成
    - a) 初期準備的 LSFAC 区間の同定法の適用  $\implies L_A^E$  と  $L_A^P$  を定める
    - b) 各区間において
      - i) 自己相関関数の計算
      - ii) 特徴的空間距離の算出
      - iii) ローパスフィルタの設計
      - iv) 平滑化の適用
    - c) 平滑化曲線からの極大・極小点の拾い上げ
  - $k = 1, \dots, K_M$  に対し  $RSS(T_M(k))$  の計算
  - $RSS(T_M(k))$  を最小にする  $T_M(k)$ ,  $T_M^*$  の探索
  - 準ニュートン法により  $T_M^*$  を微小移動、 $T_M^{**}$  を求める
- AIC( $M$ ) を求める

##### A.2 AIC( $M$ ) を最小にする $M$ を $M^*$ とする

## 5. ま と め

このような polyline の当てはめにより、5年間のすべてのデータセットの LSFAC の自動同定及びタイプ別への分類が実行された。地球物理学的見地からはさまざまな新しい知見が得られ、現在多数の論文としてまとめつつある。ここでは一つの例に絞り、実際にどのような発見があったのかを報告するとともに、従来の LSFAC の理解の範囲内において新しいタイプとして先に報告された (Ohtani et al. (1995)) 4層のシート構造をもつデータセットは、わずか4例であったが、両半球で計953例にもおよぶ4層の LSFAC を見つけることができた。また、Ohtani et al. (1995) で報告された例はすべて北半球で観測されたものであったが、南半球でも数多く観測されていることがわかった。また、W型と電流の向きが逆のM型もこのプロジェクトで初めて発見された。これらの結果、この4例しか見つかっていない新しいタイプの電流系は、さまざまな偶然の物理条件が重なり合った結果の希な現象ではなく、今まで見過ごされてきた実はごく普通に存在するものであることを初めて明らかにすることができた。

実学的観点からは、次のような新しい方向性が有望であることを初めて示すことができた。Fig. 1にもあるように、太陽風のデータは地球の遙か太陽風上流中に常駐している別の人工衛星によって常時モニターされている。このデータを用いて、太陽風と本研究での解析結果である LSFAC の発生パターン間の、その両者間の入出力関係を柔軟に表現できる非線形関数の探索により、宇宙天気予報の実現へ一歩踏み出せることが期待できる。

本研究では DMSP-F7 という一つの人工衛星によって観測されたデータセットのみを取り扱った。しかしながらここで行ってきた作業結果の意味するところは、読み込み部分の変更のみにより、蓄積されてきた膨大な他の人工衛星データからの発見的な全探索が初めて約束されることである。その結果もたらされる地球科学、特に電流系に関する知見は、はかりしれないものになるであろう。

## 謝 辞

査読者の有益なコメントに感謝いたします。また、統計数理研究所・田邊教授には、節点位置の非線形最適化技法について広範な経験談をいただきました。

本研究で使用された DMSP-F7 の磁場データは、F. J. Rich 博士によって提供された。本文の一部分は、The Johns Hopkins University, Applied Physics Laboratory (JHU/APL) の大谷晋一博士との共同研究の成果をもとにしている。また、JHU/APL での研究は、NASA 及び NSF の補助を受けている。

この研究は、平成 10-11 年度科学研究費奨励研究 A 「超大規模時空間データセットの全探索のためのモデルの開発とその実装」、また一部は平成 10~12 年度科学研究費 特定領域研究 (A) 「発見科学：巨大学術情報からの知識発見に関する基礎研究 (研究代表者 有川 節夫)」の援助を受けている。

## 参 考 文 献

- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle, *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, Foundations and Basic Theory* (eds. S. Kots and N.L. Johnson), 610-624, Springer, New York, (1992)).
- Akaike, H. (1974). A new look at the statistical model identification, *IEEE Trans. Automat. Control*, **AC-19**, 716-723.
- Eilers, P. H. C. and Marx, B. (1996). Flexible smoothing with *B*-splines and penalties, *Statist. Sci.*,

- 11 (2), 89-121.
- Friedman, J. H. and Silverman, B. W. (1989). Flexible parsimonious smoothing and additive modeling, *Technometrics*, **31** (1), 3-29.
- Higuchi, T. (1991). Frequency domain characteristics of linear operator to decompose a time series into multi-components, *Ann. Inst. Statist. Math.*, **43**, 469-492.
- 樋口知之 (1995). 人工衛星時系列データの処理, 『時系列解析の実際II』(赤池弘次, 北川源四郎 編), 133-149, 朝倉書店, 東京.
- Higuchi, T. and Ohtani, S. (1998). Automatic identification of large-scale field-aligned current structure, Research Memo., No. 668, The Institute of Statistical Mathematics, Tokyo. (<http://www.ism.ac.jp/~higuchi/AIFACpaper.html>)
- Hiragi, Y., Urakawa, H. and Tanabe, K. (1985). Statistical procedure for deconvoluting experimental data, *Journal of Applied Physics*, **58** (1), 5-11.
- Iijima, T. and Potemra, T.A. (1976). Field-aligned currents in the dayside cusp observed by Triad, *Journal of Geophysical Research*, **81**, 5971-5979.
- 井元清哉, 小西貞則 (1999). B-スプラインによる非線形回帰モデルと情報量規準, *統計数理*, **47**, 359-373.
- 石黒真木夫, 荒畑恵美子 (1982). ベイズ型スプライン回帰, *統計研彙報*, **30** (1), 29-36.
- Kashiwagi, N. (1991). Bayesian detection of structural changes, *Ann. Inst. Statist. Math.*, **43**, 77-93.
- 北川源四郎, 石黒真木夫, 坂元慶行 (1994). 情報量規準 AIC と EIC, 電子情報通信技術研究報告 (信学技報), **92**(503), 49-62.
- Kivelson, M.G. and Russell, C.T. (1995). *Introduction to Space Physics*, Cambridge University Press, Cambridge.
- Ohtani, S., Potemra, T. A., Newell, P. T., Zanetti, L. J., Iijima, T., Watanabe, M., Blomberg, L. G., Elphinstone, R. D., Murphree, J. S., Yamauchi, M. and Woch, J. G. (1995). Four large-scale field-aligned current systems in the dayside high-latitude region, *Journal of Geophysical Research*, **100-A1**, 137-153.
- Williams, C. M. (1978). An efficient algorithm for the piecewise linear approximation of planar curves, *Computer Graphics and Image Processing*, **8**, 286-293.

Automatic Identification of the Large Scale Field  
Aligned Current Systems as an Example of  
Knowledge Discovery from the Large Database

Tomoyuki Higuchi

(The Institute of Statistical Mathematics)

Characteristics of large-scale field-aligned currents (LSFAC) observed above the Earth's ionosphere are highly variable, and we have been depending on visual examination to identify LSFAC systems. The objective of this paper is to report a new procedure that we developed to automatically identify the spatial structure of LSFAC from satellite magnetic field measurements. Depending on the number of LSFAC sheets crossed by a satellite and also on the intensity and flow direction (upward/downward) of each LSFAC, a plot of the east-west magnetic component can have any shape. The required task is to automatically fit line segments to the plot. The procedure is based on the concept of the first-order  $B$ -spline (polyline) fitting with variable node positions. The number of node points, which determines the number of FAC sheets, is one of the fitting parameters and is optimized for each orbit so that the Akaike Information Criterion (AIC) is minimized.