



ISSN 1349-1113
JAXA-RR-16-007

宇宙航空研究開発機構研究開発報告

JAXA Research and Development Report

宇宙科学情報解析論文誌 第六号

2017年3月

宇宙航空研究開発機構

Japan Aerospace Exploration Agency

目 次

データ処理、データ解析

1. 軌道計画への品質工学の応用可能性
池内正之, 角有司 1
2. 音声グラフ表示ソフトウェア **splot** を用いたスペクトル認識の可能性
宇野 伸一郎, 外谷 渉, 三浦 昭 11
3. Web 型データ解析ツール **ERGWAT** の開発と今後の発展計画
梅村 宜生, 瀬川 朋紀, 宮下 幸長, 桂華 邦裕, 三好 由純, 堀 智昭,
小路 真史, 田中 良昌, 関 華奈子, 篠原 育 25
4. SVM による深発月震分類のための有効な特徴量の分析
加藤 広大, 菊池 栞, 山田 竜平, 山本 幸生, 廣田 雅春,
横山 昌平, 石川 博 39
5. 深発月震源分類に適した機械学習法の研究
菊池 栞, 加藤 広大, 山田 竜平, 山本 幸生, 廣田 雅春,
横山 昌平, 石川 博 51
6. れいめい衛星搭載磁力計 **GAS** のデータベース作成・整備：沿磁力線電流の推定のために
益岡 葵, 高田 拓, 平原 聖文, 浅村 和史 63
7. Applications of data science techniques to disentanglement of X-ray
spectral variation of super-massive black holes1
S. Pike, K. Ebisawa, S. Ikeda, M. Morii, M. Mizumoto, and E. Kusunoki ... 73

アーカイブ開発

8. 宇宙科学資料データベースの開発
川上 修司, 本田 秀之, 小野 縁, 長木 明成, 斎藤 宏,
平井 寿美子, 周東 三和子, 竹島 敏明 89
9. 地球観測衛星データ処理における **JAXA** スパコン活用の効果検証
齋藤 紀男, 上田 陽子, 中西 功, 仁尾 友美, 小西 利幸,
斎藤 進, 田中 誠, 井口 茂, 井上 淳一, 鳥居 雅也 99
10. 火星大気データを地球大気解析ツールで解析できるようにするためのデータ整備
野口克行, 林寛生 109

データ利用拡大、アウトリーチ

11. WebGLによる「はやぶさ2」「あかつき」のリアルタイム軌道可視化
柏井 勇魚，宮崎 剛 117
12. 多様な環境においてデジタル立体地球儀を実現するためのダジック・アースの開発
齊藤 昭則，津川 卓也，市川 浩樹，島田 卓也 131
13. 科学データのドームを用いた映像化による研究成果の公表
田部一志 145
14. 全天周映像作成のための，簡易ドームマスター生成手法について
三浦 昭..... 155
15. 宇宙データを使った世界同時開催ハッカソン「International Space Apps Challenge」
の日本開催
湯村 翼 167

軌道計画への品質工学の応用可能性

○池内正之^{*1}, 角有司^{*1}

Feasibility Study of Quality Engineering for Trajectory Planning

Masayuki Ikeuchi, Yuji Kado

Abstract

This paper applies a new quality engineering tool based on Taguchi method to trajectory planning of non-cooperative approach for active debris removal, and evaluates the results of 1296 (L36) or 2500 (L50) cases by changing parameters, which comply with 7 inputs of products, 8 inputs of operations, and 10 outputs for evaluations. The results show the feasibility of applying this quality engineering tool and some suggestions to ADR trajectory planning.

Key Words: Quality Engineering, Trajectory Planning, JIANT, SVA, Non-cooperative Approach, Active Debris Removal, ADR

1. 目的および背景

本稿の目的は、軌道計画に品質工学を応用して、宇宙情報解析の有効性やプロセスを検討することである。軌道計画に品質工学を導入することにより、概念検討段階の軌道計画とシステムのロバスト性を可視化することが期待できる。

品質工学手法は、目標達成のために例えば製品の目標パラメータのバラつきを小さく設計するなど、これまでは詳細設計・製造設計に応用されてきた。

一方、プロジェクト決定前段階に、概念検討として軌道計画を立案し、ミッション実現性を示す必要がある。適切な軌道計画により、宇宙機設計に関するリスク低減とフロントローディングに貢献することが期待できる。筆者らは、概念検討分野の課題に、品質工学手法を応用することを方針として、昨年度、この手法¹⁾をデブリ近傍接近軌道計画²⁾に応用し³⁾、要求設計のロバスト性などの評価を試みた。

本稿は、昨年度得られたシミュレーション結果を基に、新しい品質工学の特長について検討を深め、概念検討に適用して得られた最適設計解および感度情報の有効性やプロセスについて検討する。

2. 品質工学と宇宙可視化解析のツール統合

ツールは、品質工学ツール(JIANT: JAXA Integrator for ANalysis Tools)¹⁾と宇宙可視化解析ツール(SVA

Spacecraft Visualization and Analysis tool)⁶⁾を結合したものをを用いている。

JAXA 安全・信頼性推進部では、宇宙機や一般製品の設計者が概念設計段階から品質工学(タグチメソッド)を利用しやすくするためのツールとそのプロセスの研究を進めている。JIANT は、インタフェースを適合させれば基本的に他のツールと結合でき、設計パラメータがシステムに与える感度やロバスト性を可視化できるので、軌道計画前段のマネージメント支援機能として応用することを考えた。

SVA は、宇宙機や天体の軌道、姿勢を模擬して、可視化するツールである。物体と物理量を組合せた多数のシステム変数とユーザの定義する変数を軌道設計に応じて定義できる特長がある。また、瞬時に表示可能な標準画面を備えており、高精度な軌道力学系を模擬できる。

これら独立した 2 つのツールを、あえて組合せて概念検討段階から利用することについて、当初は、直感的に有効性や費用対効果の点で疑問もあった。

しかし、昨年度実施した最適解探索の結果、図 2²⁾に示す様に、軌道フェーズ区切りの見方を変えると計画立案を改善させるとのヒントや、S/N 比を用いた航法誤差と軌道の関係が明らかになり、統合ツールが軌道計画に有効となる理由があがってきた。

品質工学と軌道計画の統合は昨年度始まり、最近では、ツールの利用法やプロセスの検討に移っている。

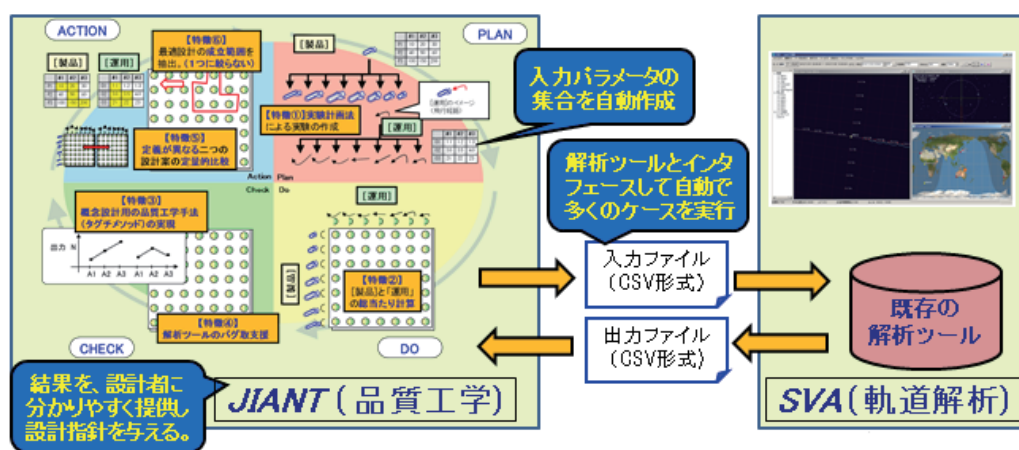


図1 JIANT(品質工学ツール)と SVA(軌道解析ツール)の統合

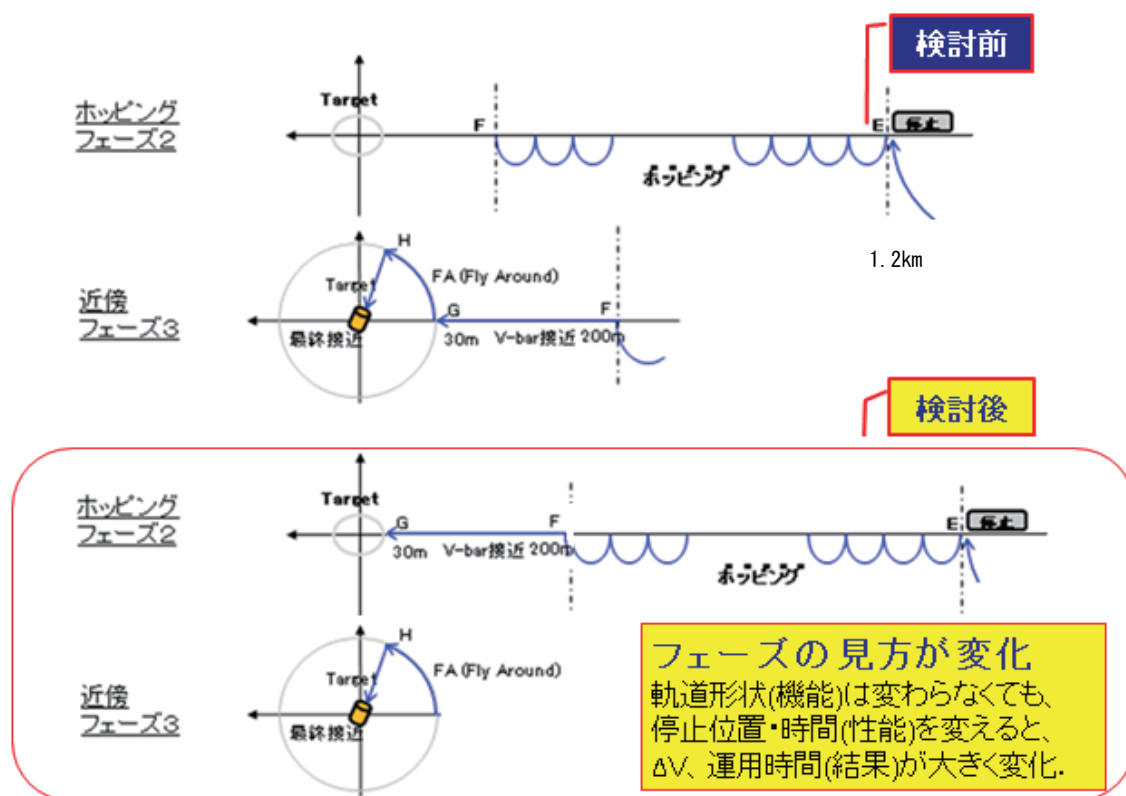


図2 シミュレーション実施前後での軌道のフェーズ区切りの見方の変化

3. 非協力接近シミュレーションの変化要因と出力

シミュレーション対象は、JAXA 研究開発部門等で実施している「導電性テザーを用いたデブリ除去の研究」における、「デブリ(非協力対象)に対する接近シミュレーション」とし、相対接近軌道に品質工学を応用した。

昨今、軌道上の衛星にスペースデブリが衝突するリスクが改めて認識され、積極的にデブリ(非協力対象)を排除(ADR: Active Debris Removal)する一方法として、デブリ除去衛星がデブリに接近して導電性テザーを取付けデオービット(除去)させる研究が進んでいる⁴⁾。本稿では、図2上段(検討前)のフェーズ2(ホッピング)およびフェーズ3(V-bar 接近から対象の捕獲まで)をシミュレーションの範囲とした。

デブリ近傍は、運用が複雑になる傾向があり関心がもたれる領域である。条件として以下を設定した。

- ・図 2 の軌道の基本形状は与えられたものとする。
- ・軌道設計には安全条件を反映している。何らかの原因で軌道制御が実施できない場合、除去衛星は対象に衝突しない軌道とする。
- ・捕獲までの主な接近用センサはカメラとする⁵⁾。
- ・次ステップに移行する際に運用者の判断を要するマイルストーンでは、除去衛星が送信する対象の画像を地上でモニタできるまで待機する。

変化要因入力となるパラメータを表 1 に示す。

JIANT と SVA のインタフェース条件に従い、製品と運用に整理して定義した (表 1 左側)。製品と運用は入替え可能で、シミュレーション結果を見てから変更することも可能である。

表 1 右側に、入力パラメータのふり幅を定義した。

始めは 3 因子の入力パラメータを指定し、後に 5 要因でシミュレーションを実行した¹⁾。最大因子数は、現在 11 まで設定することができる。

表 1 下側の出力は軌道に付する評価関数である。出力は軌道制御量 (ΔV)、運用時間 (T)、測距精度指標 (L/δ : [相対距離]/[測距精度]) などである。

図 3 に、「製品」P (Product)と、「運用」O (Operation)として定義した入力パラメータと相対接近軌道との対応を示す。

表 1 と図 3 は、JIANT+SVA 統合ツールを利用する際のインタフェース情報であり、シミュレーションのシステム設計に相当する。これらの図表は、品質工学と軌道解析の専門家、および他分野の関係者が相互確認できる情報となる。

4. 品質工学手法ツールの実行

JIANT は概念検討段階から利用できる機能を備え実験計画法に基づき、表 1 に示した「製品」と「運用」の全パラメータについて SVA にシミュレーションの実行を指示し、その出力から最適解の探索や解集合を視覚化することができる。

図 4 に、JIANT+SVA によるシミュレーション実行状況を示す。「製品」と「運用」について、総当たり計算する。ここで、正方形の点列の左側は「製品」、右側は「運用」のパラメータ変化に対応している。途中でエラーが発生した場合は、結果を正常終了時と異なる色で表示して、まずは全ケースを実行して正常終了とエラー終了をそれぞれ調査できる仕様とした。

表 1 軌道シミュレーションの入出力と製品／運用の定義

No.	入出力の 区別	製品 運用	パラメータ名	解析ツールから引当 可能なパラメータ	ノミナル値 (3因子)	単位
【製品】パラメータ 7個	1 入力	P	測位精度 (GPSなど)	URE	10, 20, 50	m
	2 入力	P	カメラ視野角	sensor.FOV	6, 10, 20	deg
	3 入力	P	カメラ測角精度	sensorError	0.02, 0.035, 0.05	deg
	4 入力	P	V-bar接近の制御時間間隔	vBarCtrlInt	10, 30, 60	s
	5 入力	P	フライアウトレ待機時の制御時間間隔	stopCtrlInt	10, 30, 60	s
	6 入力	P	フライアアウト中の制御時間間隔	faCtrlInt	10, 30, 60	s
	7 入力	P	捕獲位置移動中の制御時間間隔	catchCtrlInt	5, 10, 30	s
【運用】パラメータ 8個	1 入力	O	V-bar接近開始距離	vBarStartPos	200, 300, 400	m
	2 入力	O	V-bar接近速度	vBarVel	0.5, 0.6, 0.7	m
	3 入力	O	V-bar接近終了距離	faStartPos	30, 40, 50	m
	4 入力	O	フライアアウト角度(飛行角度)	faAngle	60, 70, 80	deg
	5 入力	O	1回のホッピングの長さ	hopLength	100, 120, 140	m
	6 入力	O	ホッピング高さ(長さ)の変化率	hopLengthRate	0.8, 0.9, 1.0	
	7 入力	O	フライアアウト時の角速度	faVel	0.2, 0.25, 0.3	deg/s
	8 入力	O	捕獲位置への移動速度	catchVel	0.2, 0.4, 0.8	m/s
出力パラメータ 10個	1 出力		総 ΔV 量	dvTotal		m/s
	2 出力		個々の ΔV 量 (X)	dvTotalX		m/s
	3 出力		個々の ΔV 量 (Y)	dvTotalY		m/s
	4 出力		個々の ΔV 量 (Z)	dvTotalZ		m/s
	5 出力		総時間	oprTime		s
	6 出力		待機時間	totalStbyTime		s
	7 出力		待機 ΔV 量	dvStbyTotal		m/s
	8 出力		L/δ の空間積分平均値	LD		
	9 出力		可視時間	totalLinkTime		s
	10 出力		観測可視時間	totalObsTime		s

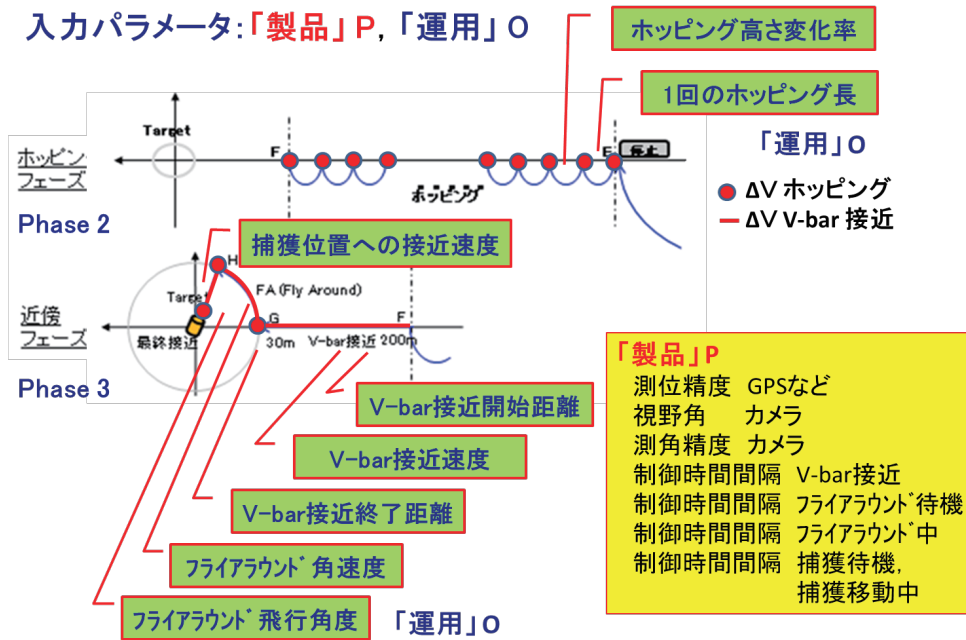


図 3 入力パラメータ:「製品」P,「運用」Oと軌道の対応

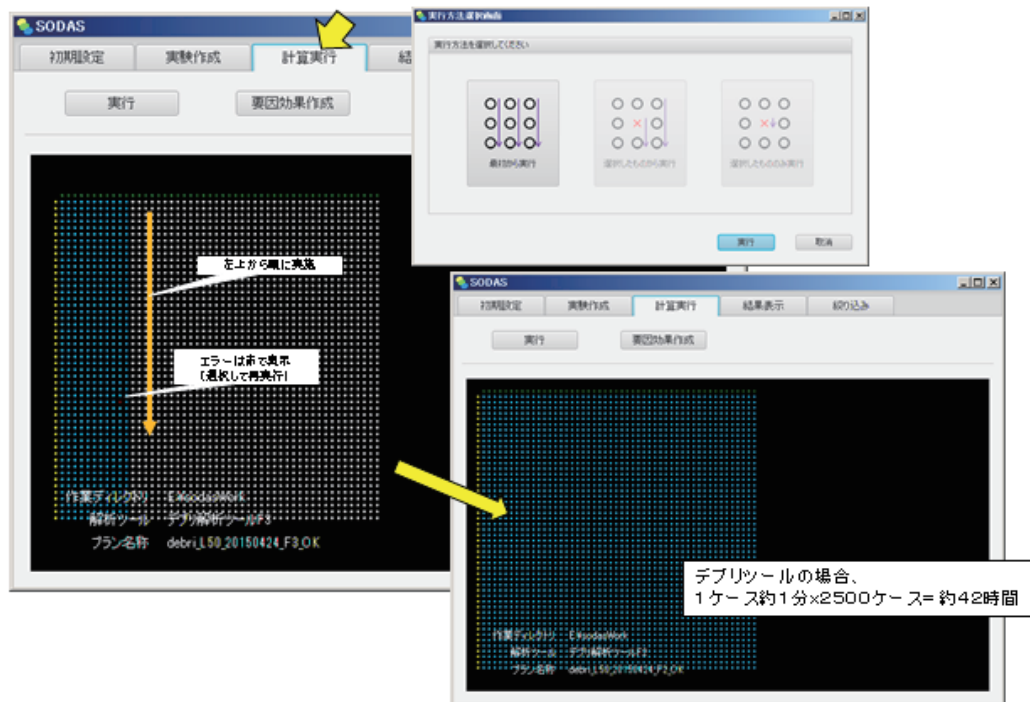


図 4 JIANT+SVA による「製品」と「運用」の総当たり計算

実験計画法は、表 1 で 3 因子(5 因子)とする場合、L36 (L50) 直交表を用いる。計算規模は L36 では解析回数 = $36 \times 36 = 1296$ 回、所要時間 = 1296 分(約 22 時間; 1 分/ケース)、L50 では、解析回数 = 2500 回、

所要時間 = 約 42 時間だった。2500 ケースを正常に終了したことをもって、パラメータの変化に対する機能のロバスト性を確保したことになる。

L50 直交表による 2500 回シミュレーションの際、

148 件エラーが生じ、表 1 の捕獲位置移動中の制御時間間隔 (catchCtrlInt) 30 秒の際に 76%, 移動速度 (catchVel) 0.8m/s の際に 53% ものエラーが集中した。エラー主要因は入力値の組合せだったが、ある値を選ぶと対象を通り過ぎ計算が収束しなくなるモデルとなっていて、ロバスト性に課題があり、バグ要因となると考えられた。

CPU はインテル (R) Xeon(R) E3-1281 v3 プロセッサ (クアッドコア 3.7GHz、8MB キャッシュ) である。

2500 ケースの解析とその結果の処理は、人力では簡単には整理し切れない量である。

5. JIANT+SVA 統合シミュレーション

シミュレーションの結果から、マクロな軌道評価を先に行うことが可能となる。

表 2 にシミュレーションのプロセスを示す。

表 2 シミュレーションのプロセス

- (1) 出力の評価項目を設定する。
 - ・軌道制御量 (ΔV)、運用時間 (T) を評価する。
- (2) フェーズを分割してシミュレーションする。
 - ・フェーズ 2 ホッピング だけ試す。
 - ・フェーズ 3 近傍 だけ試す。
- (3) 並行してフェーズを結合する。
 - ・フェーズ 2+3 (通し) を試す。
- (4) 解集合の範囲を確認する。
 - ・最初想定した解集合 A の外部も探索。
- (5) 軌道計画として解釈し直す。
 - ・品質工学結果から軌道計画に解釈し直す。

表 2 における評価の留意点は次の通りである。

- ・与えられた軌道計画がロバストかを評価する。
- ・任意初期値からの最適軌道の探索ではない。

表 2 に沿って、JIANT を用いて SVA 出力を後処理して図 5 ～ 図 7 をまとめた。各図は、縦軸：運用時間 T 、横軸： ΔV 量であり、1 ケースを 1 点とする解集合を示している。

図 5 は、フェーズ 2 のホッピングの解集合であり、離散的な集合が特徴である。運用時間は周年単位で変化するが、 ΔV 量はそう変わらない。

図 6 のフェーズ 3 の解集合も離散的である。

図 6 では ΔV 量は大きく変化するが、運用時間は変化せず、変化した項目は、図 5 ではその逆の傾向となっている。

以後、図 6 を解集合 A と呼ぶことにする。

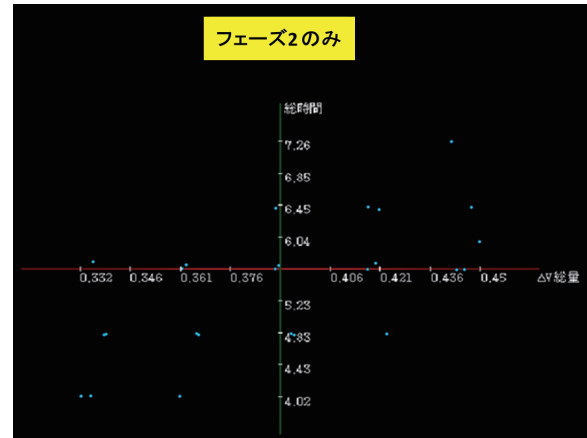


図 5 フェーズ 2 (ホッピング) の解集合
 $\Delta V = 0.33 \sim 0.45$ m/s (変化幅: 0.12 m/s)
 $T = 4.02 \sim 7.26$ h (変化幅: 3.24 h)

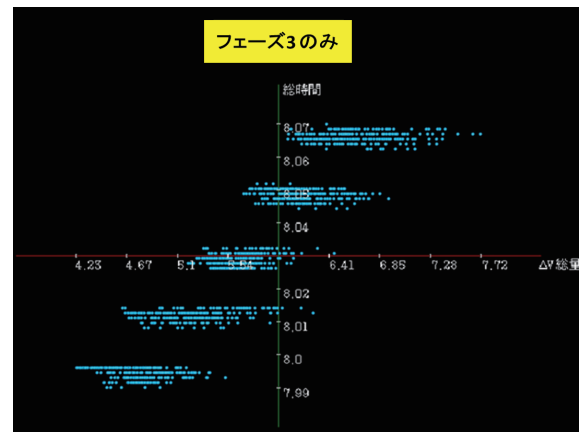


図 6 フェーズ 3 のみの解集合
 $\Delta V = 4.23 \sim 7.72$ m/s (変化幅: 3.49 m/s)
 $T = 7.99 \sim 8.07$ h (変化幅: 0.08 h)

図 7 はフェーズ 2+3 の解集合であり、図 6 の縦軸を圧縮したスケールで示している。

当初、目標を運用時間 < 16 h (2 チーム 1 シフトで運用)、 ΔV 量 < 5 m/s を基準としてシミュレーションして、運用時間に適合する解集合 $A \Delta V_{min} = 4.73$ m/s が適切(安全)と考えた。

しかし、範囲外の解集合も確認すると ΔV 量だけに注目すれば解集合 A の上に、最適な解集合 B があることが分かり、解集合 B では $\Delta V_{min} = 2.56$ m/s あれば捕獲できる可能性を示していた。

解集合 A と B は目的が異なる集合である。これらの目的は任意に選択してよいか、それともどちらかの目的を選択する適切な理由があるのだろうか。

運用を基準とするならば、解集合 A の方が余裕がある。搭載システムのリソースを基準とすれば ΔV 量を基準とする解集合 B を選択することが適切である。

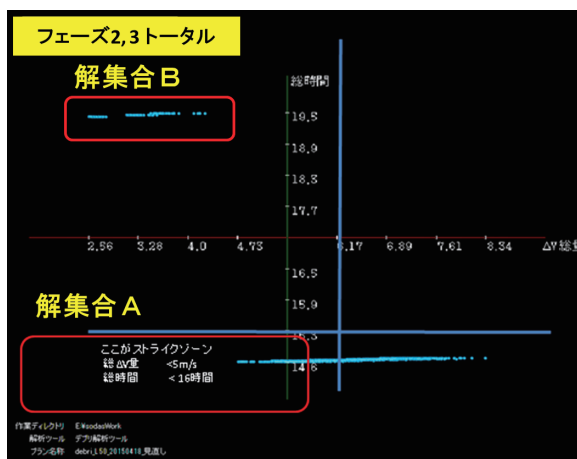


図 7 フェーズ 2～3 を通した解集合

6. 結果の評価

6.1 ΔV 量と運用時間 T の評価

表 3 に、図 5 ～ 図 7 の結果をまとめた。

軌道計画への品質工学ツールの応用可能性として得られた結果について評価する。

表 3 軌道シミュレーション結果の整理

評価指標	フェーズ 2 単独 (Hopping)	フェーズ 3 単独 (近傍)	フェーズ 2+3 通算
ΔV 量 (m/s)	0.33～0.45 [3.75]	4.23～7.72 [2.21]	B:2.56～4.50 A:4.73～8.34
時間 T (h)	4.02～7.26 [2.24]	7.99～8.07 [100.87]	A:14.5～ B:19.5

ΔV (フェーズ 2) + ΔV (フェーズ 3) = 4.56～8.17m/s.

T (フェーズ 2) + T (フェーズ 3) = 12.01～15.33 h.

[] 内は品質工学の S/N 比相当。小さいとばらつき大きいと安定している傾向を示す値である。

例 : [100.87] = 8.07/(8.07-7.99)

①分解したフェーズの重ね合せ評価の可能性

品質工学を応用するうえで、フェーズ 2 と 3 の結果が重ね合せ可能となるかは重要である。表 3 では、重ね合せはラフには成立しているが、厳密には成立していない。探索範囲を拡大するかフェーズの区切り方を見直せば可能になるかもしれない。

現時点では、フェーズの特長を確認するためにフェーズ 2 と 3 を個別に、全体の傾向はフェーズ 2+3 通算で評価することが適切であると考え。

②軌道計画の評価 (リソース)

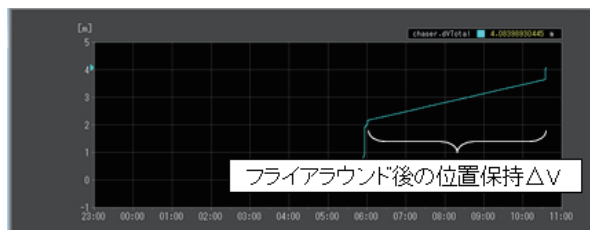
解集合 A と B の違いを軌道計画として解釈し直し、その違いを与える要因を評価した。要因は、待機

運用にあった。図 2 と図 3 の軌道形状だけでは待機運用は予測できないので、表 1 で出力として定義したパラメータから評価した。待機すれば ΔV 量も運用時間も増加するはずである。

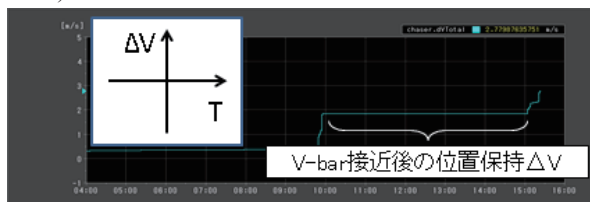
図 5 に少しその傾向はあるが、図 7 はそうならず ΔV 量だけが変化して水平方向に分布していた。

解集合 A, B で ΔV 量は変化するが運用時間はそう変わらない。解集合はフライア라운드開始前と捕獲前の待機時間の和が、ほぼ一定になる状況をとらえていた。軌道解析の見地からは当然ながら、除去衛星とデブリの高度差が原因となっていた。

図 8 a) に解集合 A の位置保持 ΔV を、b) に解集合 B の位置保持 ΔV を示す。フライア라운드直後高度差約 30m で捕獲前待機する場合は積算感度が大きい、V-bar 上で待機に要する位置保持 ΔV 量は高度差 = 0 のため積算感度は小さいので、解集合 A より解集合 B の方が安全で適切であるとした。



a) フライア라운드後に位置保持をした場合



b) V-bar 上で位置保持をした場合

図 8 位置保持の ΔV 量積算値

6.2 測位精度が軌道計画に与える影響

表 3 は、フェーズ 2 のホッピング軌道を多少変えても ΔV 量の感度は低く、運用時間は軌道周回単位で変化することを示した。ホッピング総距離を決めてこれを前提とすると、ホッピング回数(高さ)によらず総 ΔV 量が決まる。すると、その範囲は ΔV 量に依存せず対象に接近する軌道を選ぶこととなり、相対測位精度の良い軌道を設定できるのである。

この性質から、ホッピングは、対象の相対位置・姿勢(回転の状態)推定に向いており、対象の赤外光や可視光反射を識別できる時間帯と整合させてホッピング総距離を決めてよいことが分かる。

図 9 にこの関係を解析的に示す。フェーズ 2 以降

は、対象を測角航法から、物体として識別した航法に移行できるようにする領域である。ホッピングの初期に V-bar から高く跳ねて対象を大きな仰角(絶対値)で観測する軌道にすると測角測距精度が上がる。

この結果は次式により解析した。

$$\delta = dh / \tan \varphi - h d\varphi / \sin^2 \varphi$$

δ : 相対測距精度

dh : 相対高度推定精度, φ : 対象仰角

h : 相対高度, $d\varphi$: 測角精度

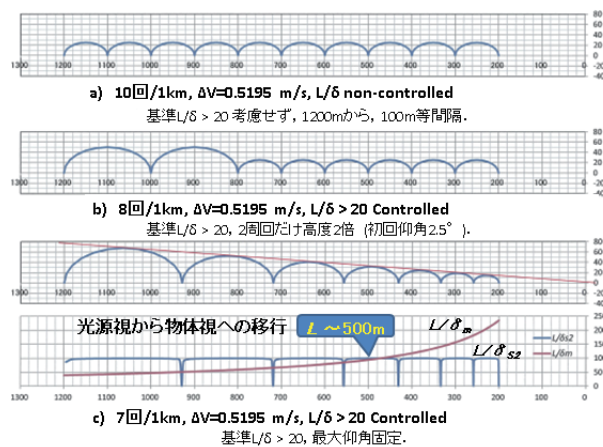


図 9 ホッピングの解析例 (デブリ除去衛星座標)

ホッピング高度が高ければ測角航法と物体視航法を併用した相対距離・姿勢推定が期待できるため、

より航法精度の高い安全な運用になる。

JIANT による感度解析の結果, V-bar 接近開始距離 200m, V-bar 接近終了距離 30m, ホッピング高変化率 0.85 で ΔV 総量を減らす効果があることが分かっている. V-bar 接近開始距離はパラメータの最大値 400m を越えるように考慮した。

3 節に示した通り, 測距精度指標として, 品質工学における S/N 比に相当する値として L/δ を定義した。

図 9 c) は, 画素数計測による物体視航法 L/δ_m が, 重心計測による測角航法 L/δ_{s2} を相対距離 500m 付近で上回るように軌道計画したものである。

V-bar 上は測角値 = 0 となり, 測角航法は使えなくなるので, 前もって対象を大きな画素数で識別して相対測位できる点として相対距離 500m を計画した。

これにより目標 L/δ は, 軌道計画に含めることが可能となる。つまり, 接近軌道の測距精度の概要を計画することが可能となる。

6.3 地上運用が軌道計画に与える影響

図 10 にフェーズ 3 について, 図 11 にフェーズ 2 と 3 の通算について, 地上軌跡のシミュレーション結果の例を示した。

図 11 から, フェーズ 2 と 3 は通算で確認する必要があることが分かる。表 4 に, 表 3 の ΔV 量と運用時間 T について軌道制御の内訳を示した。

重要イベントである捕獲開始は, 地上判断により運用したいとする要求が, 運用時間に影響することを SVA のシミュレーション結果が示していた。

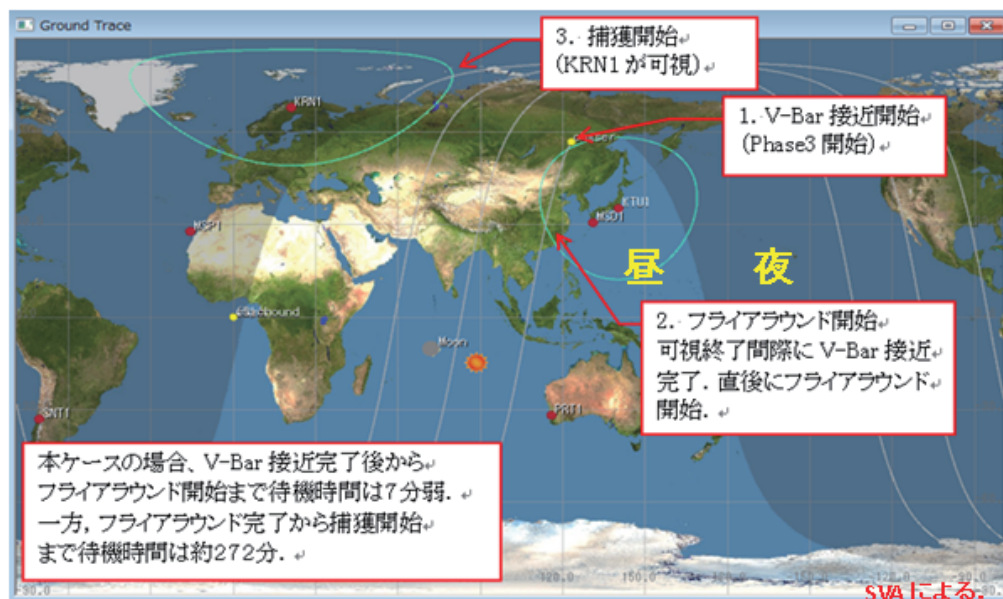


図 10 フェーズ 3 の軌道制御シーケンス

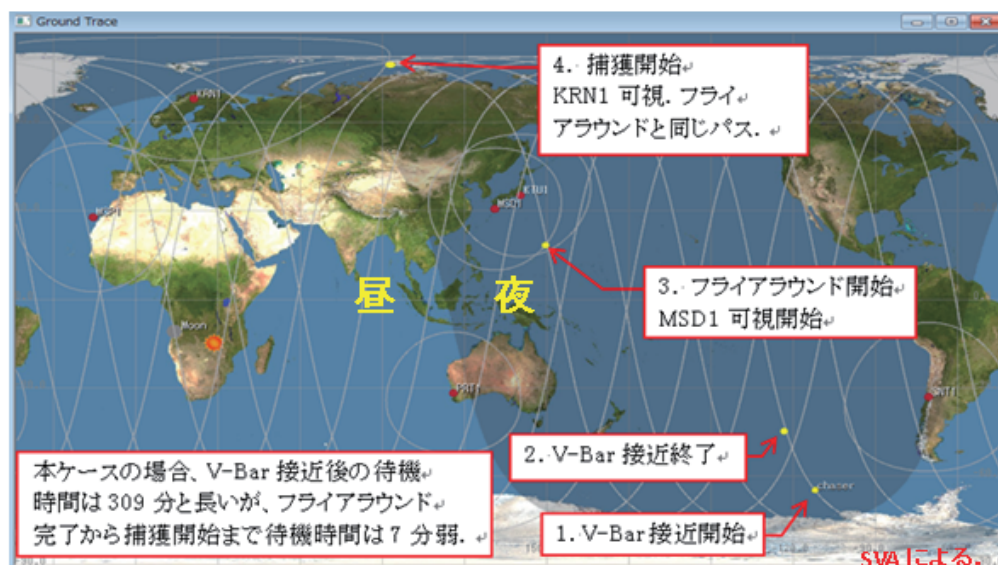


図 11 フェーズ 2+3 の軌道制御シーケンス

表 4 各軌道制御での ΔV 量と運用時間 T

軌道制御種別	ΔV [m/s]		運用時間 T [分]	
	フェーズ 2+3	フェーズ 3	フェーズ 2+3	フェーズ 3
ホッピング	0.376	0.000	829	N/A
V-bar 接近	1.037	1.094	5	7
FA 待機 (V-bar 上)	0.442	0.325	309	0
フライアラウンド (FA)	0.418	0.748	5	5
捕獲待機	0.069	1.478	12	273
捕獲	0.438	0.438	2	3
合計	2.780	4.084	1162	288

V-bar 上待機で時間調整し、捕獲待機を短くすると ΔV 量は減った (解集合 B)。
ホッピング開始距離と V-bar 接近距離は調整要。

6.4 軌道計画の評価 (フェーズの区切り)

図 2 に示したシミュレーション実施前後におけるフェーズ区切りの見方の変化について説明する。

フェーズを予め区切るという判断も、検証対象となるということである。

当初フェーズ 2 と 3 の区切りはホッピング終了とした。これはフライアラウンド開始としていたならば、結果をより早く理解できたかもしれない。

フェーズ 2 は V-bar 接近終了までに変更し、この目的を対象の相対位置・姿勢精度の高精度な推定として、フェーズ 3 の捕獲に必要な観測値を得る軌道計画をすることが適切と考える。

フェーズ 3 はフライアラウンド開始からとして、目的は最短時間または最小エネルギー問題と捉える。基本的には、運用者の判断によりフライアラウンド

を開始して、問題が無ければ捕獲まで一連の動作を中断せずに行うことが適切と考える。

捕獲開始時点のミッションの安全性を考慮すると、捕獲直前の判断余裕、または再チャレンジの余裕として、待機用 ΔV を設けると安全である。

つまり、計画は 4.73m/s として、実際は 2.56m/s で捕獲する。 ΔV の余裕は可視光カメラと照明を使い対象を捕獲する際の待機にも必要と考える。

7. JIANT+SVA 統合ツールの有効性と特長

本統合ツールにより、出力(ΔV 量, 時間)に対する設計パラメータの感度を確認できた。製品と運用のロバスト性を確保した軌道計画案は、ミッションのリスク低減を可能にすることができる。

- ①従来の軌道解析と異なる視点からのアプローチ
軌道計画に品質工学を応用する技法が得られた。
- ・解集合の分布を散布図や表で示すことで、各パラメータの特徴量を分析・把握できた。(6.1 節)
当初選択したパラメータは想定通り適切かを確認し易くなった。別途、5 因子の解析を実施した¹⁾。
パラメータ幅の想定が適切か検討できた。
 - ・各パラメータの感度を確認して、航法精度が良くなる軌道計画を選択することができた。(6.2 節)
 - ・最適解の候補を絞り込んだ後に、SVA により地上軌跡を確認して、重要イベントを運用する設計解の妥当性確認した。(6.3 節)
 - ・設計の妥当性確認として、フェーズ分けについての指針が得られた。(6.4 節)

②問題設定やモデル定義が適切か検証に繋がる情報 JIANT の利用を通じて確認された有効性としては以下がある。

- ・評価パラメータ(出力)の追加が容易なため、簡易で基本的なモデルから検討を始めて、徐々に詳細化させるアプローチが可能になった。
- ・パラメータ幅を広く取って、網羅的な検討を行うことで、「要求条件の見落とし」や「暗黙的な経験則による決めつけの有無」を確認する手掛りが得られた。
- ・解集合の分布と異なる目的の対応が可視化できるので、どの目的がどの様な理由で適切となるのかを比較して検討することができた。
- ・同時出力される他の結果と比較して、モデル定義の誤りやケース設定の不適切さを抽出できるのでバグ要因の絞込みや条件の見直しが容易である。
- ・解析結果に実験計画法を適用でき、またその様に結果を収納できる。結果をテキストデータ (CSV 形式) で出力できるので表計算ソフトなどによる再利用がし易く、表やグラフなどの作成が容易である。

8. まとめ

軌道計画への品質工学の応用可能性を探るために、非協力接近シミュレーションに品質工学手法を適用した。表 2 に示したシミュレーションのプロセスを以下の通り評価している。

- ・軌道制御量 (ΔV)、運用時間 (T) をリソースとして着目してマクロに評価した。
- ・軌道形状に陽に現れない待機運用の方針についてシミュレーションにより明確にした。そのうえで、リソースに対する感度が低くなる様に、高い測距精度を選択できる軌道(計画)を示した。

- ・フェーズを区切ることにより、ホッピングの様なパーツとなる区分軌道の特長を把握した。
- ・フェーズを結合することにより、リソースの選択について判断を見直した。
- ・予め設定したフェーズの区切り方について判断を見直した。

プロセスに一般性はあるか、軌道計画毎に異なるのかは興味深い課題である。

今後は規模を拡大して誘導制御を含めた品質工学を応用することが可能になるものと考えている。

また、他の軌道計画に应用することにより、共通課題を検証することは有効であると思われる。

その結果が、他分野の専門家の知見や経験と整合するか、使用方法も含めたツールの検証分野の課題となる。

参考文献

- 1) 角 有司, 池内 正之: 軌道計画立案における品質工学の応用可能性 –初期段階の仕様検討とロバスト設計解の取得–, 第 59 回宇宙科学技術連合講演会, 2015
- 2) 山元 透, 村上 尚美, 中島 悠, 山中 浩二: 軌道上デブリへの接近ストラテジ, 第 57 回宇宙科学技術連合講演会, 2013
- 3) 池内 正之, 角 有司: 軌道計画立案における品質工学の応用可能性 –非協力接近軌道計画における品質工学手法の導入と評価–, 第 59 回宇宙科学技術連合講演会, 2015
- 4) 井上 浩一, 平子 敬一, 河本 聡美, 大川 恭志, 鷲谷 正史, 壹岐 賢太郎, 導電性テザー実証実験計画, 第 57 回宇宙科学技術連合講演会, 2013
- 5) 加藤 貴昭, 田中 啓太, 池内 正之, 桑尾 文博, 山元 透, 中島 悠, 村上 尚美, 非協力ターゲットへの誘導制御シミュレーション, 第 58 回宇宙科学技術連合講演会, 2014
- 6) Space Visualization and Analysis tool, Spheresoft, <http://www.spheresoft.jp/index.html>, (Accessed 2015-7-30)

音声グラフ表示ソフトウェア **splot** を用いたスペクトル認識の可能性

～ 宇宙科学データ可聴化プロジェクト (4)～

宇野 伸一郎^{*1}, 外谷 渉^{*2}, 三浦 昭^{*3}

Possibility of spectrum recognition using a sound-based data-plotting program, **splot**.

～ Astronomical Data Sonification Project (4)～

Shin'ichiro UNO^{*1}, Sho SOTOYA^{*2}, Akira MIURA^{*3}

Abstract

Astronomical Data Sonification Project has been developed a sound-based data-plotting program, **splot**, to construct data analysis environment for visually impaired. Now we developed a sound-based user interface for spectral analysis tool named **splot-for-xspec**. Xspec is a widely used software for spectral analysis developed by NASA HEASARC. **Splot-for-xspec** is a tool to link **xspec** and **splot**. We also made a small survey in order to confirm the utility of the **splot-for-xspec** system, and we showed the possibility that one can recognize a spectrum without visual information. We are aiming to sonify the entire data analysis process.

概要

宇宙科学データ可聴化プロジェクトは、視覚障害者が X 線天文学を学び、実際のデータ解析をする環境を構築するため、音声グラフ表示ソフトウェア **splot** を開発してきた。今回我々は、スペクトルフィッティングを音声によって操作する環境 **splot-for-xspec** を構築した。**xspec** は、広く使われているデータ解析ソフトウェアで、NASA HEASARC が開発したものである。**splot-for-xspec** は **xspec** と **splot** の連携を行うツールとなる。さらに、我々は **splot** および **splot-for-xspec** によるグラフ認識の有効性を確認するため、小規模な調査を行った。その結果、音によるデータ比較・スペクトルフィットはひとつの手段となりうると考えられる結果が得られた。今後は音声化できる情報のみでデータ解析プロセス全体を行えるようにすることが課題である。

doi: 10.20637/JAXA-RR-16-007/0002

^{*} 平成 28 年 11 月 24 日受付 (Received November 24, 2016)

^{*1} 日本福祉大学健康科学部 (Nihon Fukushi University, Faculty of Health Science)

^{*2} 株式会社ラック (LAC Co., Ltd.)

^{*3} 宇宙航空研究開発機構宇宙科学研究所 (Japan Aerospace Exploration Agency, Institute of Space and Astronautical Science)

1 導入

視覚障害者の情報保障は、福祉工学の重要なテーマのひとつである。近年では STEM (Science Technology Engineering and Mathematics) 教育が重要視されており、障害者の教育においても、STEM 分野に力点が置かれてきている¹⁾。特に STEM 分野は、情報機器などを活用すれば障害者が独立して活躍できる領域が多く、そのための教育や環境整備が望まれ、様々な開発が行われている。

1.1 視覚障害者の図表認識

視覚に頼らないで図表を認識することは大きな困難が伴う。視覚障害者が図表を認識する方法としては、1. カプセルペーパーやドットプリンターなどを用いて紙上に立体パターンを作って手で触ることができるようにする方法や、2. 点字表示機能のあるピンディスプレイなどを用いて機械的に表現する方法などが用いられてきた。

しかし、紙にパターンを作成する方法は、作成に時間がかかったり、細かいものが表現しづらいといった問題がある。また、ピンディスプレイ等の機械は高価で、表示できる面積も狭い。さらにピンディスプレイは重いので可搬性に乏しく、専用のドライバーが必要となるため、特定のプラットフォーム上でしか稼働できないといった問題がある。

1.2 視覚障害者の情報利用環境

視覚障害者がパーソナルコンピュータを用いる場合、音声読み上げソフトウェアを用いている場合が多い。音声読み上げソフトウェアは主にテキストデータをコンピュータに音読させる機能をもっている。

また、音声読み上げソフトウェアは様々なアプリケーションも音読することができ、emacs などのエディタ、メール、ウェブ、ワープロ文書や、文字の埋め込まれた pdf、表計算文書などの読み上げを行う。同じように、データ解析を行ったり、コンピュータを管理する場合に必須となるコマンドターミナルも、音声読み上げソフトウェアによって読むことができる。

なお、コマンドターミナルなどでは大量のログが流れる場合があるが、これら全てを読み上げていると時間が浪費されるため、適宜必要なテキスト情報を視覚障害者自身が選択して読み上げさせている場合が多い。また、熟練した視覚障害者は非常に速い速度でこれらの音を聞き分けることができる例がある。

1.3 X 線天文学のデータ解析

X 線天文衛星のデータを解析する場合、衛星の取得した元データの整形、イメージや観測時間など様々な条件によるデータの取捨選択、検出器の応答関数のかけあわせ、理論モデルとの比較、など様々なプロセスを行う必要がある。

これらの解析を行うソフトウェアは、コマンドラインでのインターフェースと結果表示のグラフィカルなインターフェースを持っていることが多い。そのうち、スペクトル解析をする場合は、次に述べる qdp, xspec というソフトウェアを用いる。

1.4 qdp

QDP (The Quick and Dandy Plotter) は、アメリカ航空宇宙局 (the National Aeronautics and Space Administration: NASA) 高エネルギー天体物理学 科学アーカイブ研究センター (The High Energy Astrophysics Science Archive Research Center :HEASARC) ²⁾ が開発、配布するグラフ表示ソフトウェアである ³⁾。テキストファイルに書かれたデータを読み込み視覚的なグラフに表示する。一次元ヒストグラムだけでなく、散布図やコントアなども表示可能である。コマンドターミナルから、インタラクティブに操作することもできるし、一連のコマンドとデータを記したファイルを読み込んでグラフを作成することもできる。

1.5 xspec

xspec は、NASA HEASARC が開発・配布する、スペクトルフィッティングツールである ⁴⁾。X 線天文学におけるスペクトル解析ツールのデファクトスタンダードとなっている。2016 年 6 月現在、バージョン 12.9.0n がリリースされている。xspec は、複数のプラットフォームに移植されており、Linux, mac, windows 等で稼働させることができる。

一般的な使用方法をする場合、ユーザインターフェースは、主にふたつのウィンドウからなり、片方がコマンド入力等のためのテキストターミナル、もうひとつが、スペクトル等のグラフを画像表示するウィンドウである。操作は、主にテキストターミナルのコマンドラインからコマンドを入力することで行うが、一連の指示をファイルに保存しておいて、ファイルから指示を読み込むことも可能である。また、xspec はスペクトルフィットを行うためのツールであるが、画像表示部分は外部プログラムと連動する形でかかれており、デフォルトでは qdp が画像の整形などを行うようになっている。また、グラフデータファイルも、デフォルトでは qdp の形式で保存される。

1.6 宇宙科学データ可聴化プロジェクト

宇宙科学データ可聴化プロジェクトは、日本福祉大学 健康科学部 宇野研究室 と宇宙航空研究開発機構 宇宙科学研究所の 2 機関が共同で推進するプロジェクトで、宇宙科学データを視覚に頼らずに視覚に障害のある人達に伝えることを目指している。

本プロジェクトは、これまでに宇宙科学データを音声化する表現手法の検討を行い、X 線パルサーやバーストのデータの音声化 ⁵⁾ や、地磁気擾乱の程度を示す Kp 指数などの音声化を行った (<http://handy.n-fukushi.ac.jp/pub/uno/music/>, <http://www.isas.jaxa.jp/home/showcase/music/>)。また、プロジェクトは音声グラフ表示ソフト splot の試作を行った ^{6), 7)}。

1.7 音声グラフ表示ソフトウェア splot

splot は、宇宙科学データ可聴化プロジェクトが開発したグラフ音声化ソフトウェアである ⁶⁾。

splot は一次元ヒストグラムについて、x 軸を時間、y 軸を周波数 (音程) に対応させることで、グラフの情報を音として表現する機能を有している。通常のグラフプロットツールと同じように splot は指定された領域に含まれるデータを表現するが、ディスプレイ上に画像を表示する代わりに splot はデータを時間 (x) と周波数 (y) に置き換えた音を出力する。x 軸の表示範囲は出力音継続時間に相当し、デフォルトでは 5 秒に設定してある。この時間は可変となっており、大量のデータを聞く時は短く、細かいデータを聞き取る時は長く、と使い分けできるようになっている。データの y の値は周波数に相当し、デフォルトではピアノ中央

の3オクターブ分の音(最小値 $f_{min}=220$ Hz から最大値 $f_{max}=1760$ Hz)の間を使っている. f_{min} , f_{max} はユーザが設定可能で, 視聴覚の多重障害などの場合にも, 聞こえる周波数に音を絞れるように設計されている.

splot の入力 は X,Y 形式のテキストデータで, 複数カラムのデータも扱うことができる. splot の操作は コマンドラインからインタラクティブに行うこともできるし, 一連のコマンドとデータを記したファイルを読み込むこともできる. また, qdp とコマンドが一部互換性を保つように設計されている. このため, xspec など他のソフトウェアから生成されたファイルであっても, qdp 形式で書かれた一次元ヒストグラムであれば, splot で音声出力可能となっている(qdp はコントアなど一次元ヒストグラム以外のものも扱う).

1.8 視覚障害者の X 線天文学データ解析環境

視覚障害者が X 線天文学データを解析しようとした場合, データ選択, グラフ認識, 画像認識など, 様々な事を行わなければならない. データ解析の大きな部分を占めるスペクトル解析でも, xspec や qdp といったソフトウェアの出力するグラフを解釈することは必須となる. しかし, これらは音声化等の対応はされていない. splot は, テキストデータの音声化は可能であるが, xspec との連携はとれておらず, splot 単体ではスペクトル解析まで行うことはできない. 視覚障害者の X 線天文学データ解析環境を構築しようとする場合, スペクトル解析プロセスを音声化することが必須となる.

本研究では, X 線天文学データ解析ツール xspec と splot の橋渡しを担う, splot-for-xspec の試作を行い(本稿 第2章), さらに, これを用いて得られる音データの有効性の確認を行った(本稿 第3章). これらの結果を踏まえ, 本稿 第4章で, 視覚障害者の X 線天文衛星のデータ解析環境構築の可能性について論じる.

2 開発

2.1 splot-for-xspec

我々は, 視覚に依存しないスペクトルフィッティング環境を構築するため, 「xspec と splot を連動させるためのソフトウェア splot-for-xspec」を試作した.

splot-for-xspec は, perl スクリプトと xspec の設定ファイルからなり, 通常 xspec が画面表示するグラフを, splot により音声出力する. splot-for-xspec は, xspec の「ユーザー作成コマンド」のひとつとして機能する. このためユーザは, 音声出力の時だけ splot-for-xspec のコマンドを用い, それ以外の時は xspec 本体をそのまま使うことになる.

splot は, テキストデータを読み込み, x 軸を継続時間に, y 軸を音程にそれぞれ対応させて音を発生させる. データ点 (x_n, y_n) がある場合, x_n に対応する時刻 t_n は

$$t_n = \frac{(x_n - x_{min})}{(x_{max} - x_{min})} \times t_{max}$$

である. ここで, x_{min} , x_{max} はそれぞれ X 軸プロット範囲の最小値, 最大値, t_{max} は出力音継続時間である. また, 出力音の周波数 f_n は,

$$f_n = f_{min} \times 2^{\left(\frac{(y_n - y_{min})}{(y_{max} - y_{min})} \times \log_2 \frac{f_{max}}{f_{min}}\right)}$$

となる. ここで, y_{min} と y_{max} は y 軸の表示範囲, f_{min} と f_{max} は対応させる周波数である.

xspec を使う場合、コマンドを入力するコマンドターミナルと、結果を表示するグラフィカルなウィンドウの2つを使用する。視覚障害者が xspec を使用することを想定する場合、まず、コマンドターミナルについては、音声読み上げによる対応となる。コマンドターミナルの音声化自体は、複数の OS において複数のソフトウェアが用意されている。また、音声読み上げを別 PC (通常利用している Macintosh や Windows などのターミナルエミュレータなど) から稼働させ、リモートログインした unix 上で xspec / splot を動作させるということもありうる。このため、splot-for-xspec では、利用者は各自が使っているテキスト読み上げシステムをそのまま使う事を想定している。

xspec の画像表示ウィンドウは、コマンドターミナルから操作する。通常は、データの読み込みやエネルギーバンドの選定やフィッティングなど、一連の操作をした上で、画像を表示する手順となる。画像表示は、通常は「plot」というコマンドで行う。splot-for-xspec はこのコマンドを置き換えて音声表示を行う。ただ、xspec の視覚的なスペクトルフィットのグラフは、データ・モデル・残差が混在し、単純に音声化できる形にはなっていない。このため、これらを分解して音で認識できる形にするのが splot-for-xspec の役割のひとつである。

xspec のスペクトルフィットの表示は、主に上下2段にわかれた2つの画像からなり、上の段にデータとモデルが、下の段に残差が表示される。これらの組み合わせにより解析者は理論モデルとデータの特徴を素早く認識することができている。視覚障害者の場合、これら3つの情報を同時に得ることはできない。このため、データ、モデル、残差それぞれを別々に聞く必要がある。具体的には、splot-for-xspec は以下の各プロセスを半自動で行うように設計されている。

1. 表示されているスペクトルを qdp 形式のファイルとして保存する
2. qdp 形式のファイルから、「データ」「モデル」「残差」などのデータを抜き出して バックグラウンドで走る splot に渡す
3. splot で音声表示する。(場合によっては何度か繰り返す、部分拡大して表示するなど、複数の作業を行う)
4. splot を終了して xspec に戻る

用意したコマンドは以下の通りである。これらは、xspec の設定ファイルにより定義されている。なお、xspec では、複数の検出器、複数のモデル等々をひとつの画面に表示することに対応しているが、現時点では splot は対応していない。

- splot_model : モデルのみ発音
- splot_residual : 残差のみ発音
- splot_data : データのみ発音
- splot_all : データ・モデル・残差各々を順に発音

3 検証

音によるグラフ認識・フィッティングが現実的に可能であるか確認するために、我々は小規模な調査を行った。調査は4つの大問からなり、どの質問も、基準音となるグラフの音を聞いた後にもう一つの評価音を聞いて、その音が基準音とどう違うかを問うものとなっている。大問1,2では結果は5段階評価をするものと

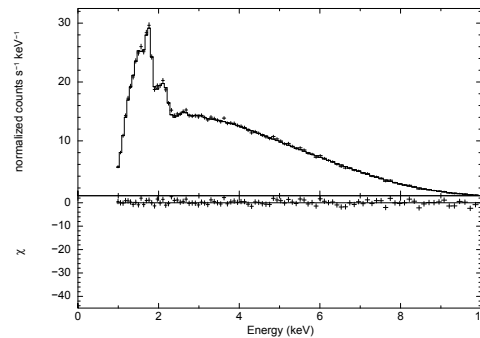


図 1: 基準とした べき 1.4 のスペクトルの例. 上段の実線がモデル関数. + 印がモデル関数から擬似的に生成されたデータ点. 下段はモデルとデータの差.

し, 大問 3,4 ではフリーハンドでグラフを描くものとした. 被験者はどのような音が表示されるのか知らされずに, 基準音と評価音を一度ずつ聞いて各問にこたえた. 全ての質問において音の継続時間は 5 秒とし, y 軸の値に相当する周波数は, 最小値, 最大値をそれぞれ $f_{\min} = 220\text{Hz}$ と $f_{\max} = 1760\text{Hz}$ とした.

音は主に X 線天文衛星すざくのスペクトルを模擬したものを xspec で作成し, splot-for-xspec を用いて音表示したものを用いた. また, この他に大問 3 では, $y = x^2$ などの単純な関数のものを splot で音表示した. 調査は 2016 年 6 月 8 日に日本福祉大学において晴眼の日本人学生を対象に行われ, 43 名から有効回答を得た (男 : 女 = 36 : 7, 平均年齢 19.6 ± 1.5 歳).

なお, 実際にどこまで精密に聞き取りができるかは, 個人差が大きく影響することが想定される. このため本調査では認識の限界を求めるのではなく, グラフの概形がある程度の範囲の人に認識できることを確認する事を目的とした.

3.1 調査結果 1: 差異の認識 1 (連続成分)

大問 1 では, グラフ中の x 軸の広い範囲での変動を聞き取れるか否かの確認をスペクトルの連続成分の変化を用いて行った. スペクトルは X 線天文衛星すざくのものを用い, モデルはべき関数に吸収のかかったものとした. スペクトルパラメータは, ノーマライゼーションと吸収水素柱密度は 1.0 に固定し, べきの値だけを変化させた. エネルギーバンドは 1.0 keV から 10.0 keV の範囲とした.

このうち, べき 1.4 のものは, 大問 1,2,4 で基準音に用いた. べき 1.4 の基準音のスペクトルを図 1 に示す. 上段がスペクトルのデータとモデル, 下段が残差を示している. スペクトルは X 線天文衛星すざくの応答関数が含まれる形で表示されるため, べき関数も単純な直線ではなく図 1 に示されるような形になっている. データ点はべき 1.4 のモデル (実線) から擬似的に生成したもので, 「+」記号で描かれている. スペクトルの基準音はモデルを, 残差の基準音は擬似生成したデータとモデルの差分を用いた.

通常 xspec は, データ・モデルは X 軸 Y 軸ともに対数軸で表示するが, 本調査では対数表示ではなく線形表示としたものを音声化した. これは, 調査の際, 特にグラフ記入時に対数表記での混乱を避けるための措置である. また, 後の図との比較のため残差のプロットの y 軸範囲は大きくとってある.

大問 1 は小問 a~d の 4 つの小問に分かれている. 小問 a 及び小問 c では, 基準音はべき 1.4 のスペクトルとし, 評価音は小問 a ではべき 1.3, 小問 c ではべき 1.0 のスペクトルとした. また, 小問 b 及び小問 d では, 小問 a 及び小問 c それぞれの残差のプロットの差異を問うた.

各問は、ひと組の音を聴き比べてその音に差があるか否かを5段階評価で答えてもらう形式とした。y軸の範囲は、スペクトルは0.68から32.5まで、残差は-45.0から3.0までで、このy軸の範囲は大問2,4も共通である。

小問a,bに用いたスペクトルの例を図2左に、小問c,dに用いたスペクトルの例を図2右に示す。

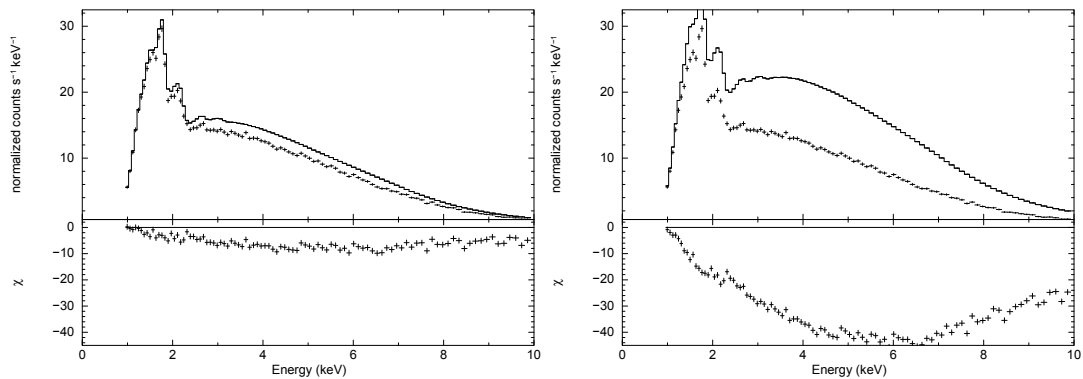


図2: べき1.3(左)と1.0(右)のスペクトルの例。上段の図の+印が基準音となるべき1.4のスペクトルデータ(図1と同じ)。実線がべき1.3と1.0それぞれのモデル関数。下段はデータとモデルの差。

3.1.1 結果

表1に大問1,2の集計結果を示す。a-d欄のa,bがべき1.4と1.3のスペクトルと残差を聴き比べた結果、c,dがべき1.4と1.0のスペクトルと残差を聴き比べた結果である。小問aで、べき1.4と1.3の違いを聴き比べた所、「同じ」もしくは「どちらかといえば同じ」との回答が21人(49%)、一方、「違う」もしくは「どちらかといえば違う」との回答が18人(42%)と、ほぼ拮抗していた。図2(左)程度のスペクトルの差異は、音表示では明確には聞き取り辛いことがわかる。

一方で、同じべき1.4と1.3のスペクトルも、小問bの残差を聞き比べた場合には、43人中42人(98%)が「違う」もしくは「どちらかといえば違う」と答えており、ほぼ全員が差異を認識できている。

べき1.4と1.0のスペクトルを聴き比べた所、こちらは35人(81%)が「違う」もしくは「どちらかといえば違う」と答え、残差を聞き比べれば43人全員が「違う」もしくは「どちらかといえば違う」と答えた。

3.2 調査結果2:差異の認識2(輝線)

大問2では、グラフ中の部分的な変動を聞き取れるか否かの確認をおこなった。質問形式は、大問1と共通とした。

大問2は小問e~hの4つの小問に分かれている。基準音は、大問1と共通の、べき関数に吸収のかかったモデルを用いた。べきは1.4のものをを用いた。評価音は基準音のスペクトルに、6.4keVにガウシアンを挿入した。ガウシアンは、 σ も0.1に固定し、ノーマリゼーションを0.005(小問e,f)と0.02(小問g,h)の2通りとした。小問e,fに用いたスペクトルの例を図3(左)に、小問g,hに用いたスペクトルの例を図3(右)に示す。

表 1: 基準音 (べき 1.4) ・ 評価音 (べき 1.3 および 1.0) の差異の評価
(人)

	大問 1				大問 2			
	a	b	c	d	e	f	g	h
1 違う	5	36	18	42	2	22	20	33
2 どちらかといえば違う	13	6	17	1	5	11	12	6
3 わからない	4	0	1	0	5	0	3	0
4 どちらかといえば同じ	14	1	4	0	15	5	5	3
5 同じ	7	0	3	0	16	5	3	1

a: べき 1.3 スペクトル, b: べき 1.3 残差, c: べき 1.0 スペクトル, d: べき 1.0 残差, e: ガウシアン (norm 0.005) スペクトル, f: ガウシアン (norm 0.005) 残差, g: ガウシアン (norm 0.02) スペクトル, h: ガウシアン (norm 0.02) 残差

3.2.1 結果

表 1 の e-h 欄のうち e,f がノーマリゼーション 0.005 のガウシアンを入れた場合のスペクトルと残差を聴き比べた結果, g,h がノーマリゼーション 0.02 のガウシアンを入れた場合ののスペクトルと残差を聴き比べた結果である。

ノーマリゼーション 0.005 のガウシアンが入ったスペクトルを聴き比べたところ 31 人 (72%) が「同じ」または「どちらかといえば同じ」と答えた。一方, 同じ 0.005 のガウシアン of スペクトルでも, 残差を聴き比べると 33 人 (77%) が「違う」または「どちらかといえば違う」と答えている。

また, 比較として, ノーマリゼーション 0.02 のガウシアンの聴き比べも行った。こちらはスペクトルを聞いて 32 人 (74%) が, 残差を聞いて 39 人 (91%) が「違う」または「どちらかといえば違う」と答えた。

3.3 調査結果 3: グラフの概形の認識

大問 3 は, 大問 1,2 と違い「聞いた音をフリーハンドでグラフに書く」ものとした。これにより差があるか否かだけでなく, どのくらいの差を感じているかということが, おおまかに評価される。

大問 3 は小問 i~k の 3 つの小問に分かれている。大問 3 では, 関数 $y = x$ を基準音として, 3 つの評価音 $y = \frac{1}{4}x$ (小問 i), $y = 2x$ (小問 j), $y = x^2$ (小問 k), の 3 つの関数をグラフ上に記入してもらった。各関数は $0 \leq x \leq 2$ の範囲の音を鳴らし, 記入欄のグラフ用紙は, $0 \leq x \leq 2, 0 \leq y \leq 4$, の範囲のものを用意した。ただし, 用紙にグリッドはあるが, 軸メモリには数値を記入していない。用紙には, 参照用に基準音の線が記入しており, 被験者はこれを見ながら評価音を書き込んだ。

3.3.1 結果

$y = \frac{1}{4}x$ を記入する小問 i では, 43 中 39 枚 (91%) の回答が $y = x$ よりも下に線を記入しており, そのうち直線は 36 枚 (84%) だった。ただし, 直線の傾きは相対的に大きく, 一番多かった回答は $y = \frac{1}{2}x$ の線の 22 枚 (51%) だった。また, 傾きが 1 を越えていた例は 4 例あり, 直線でなかった例は 3 例あった。

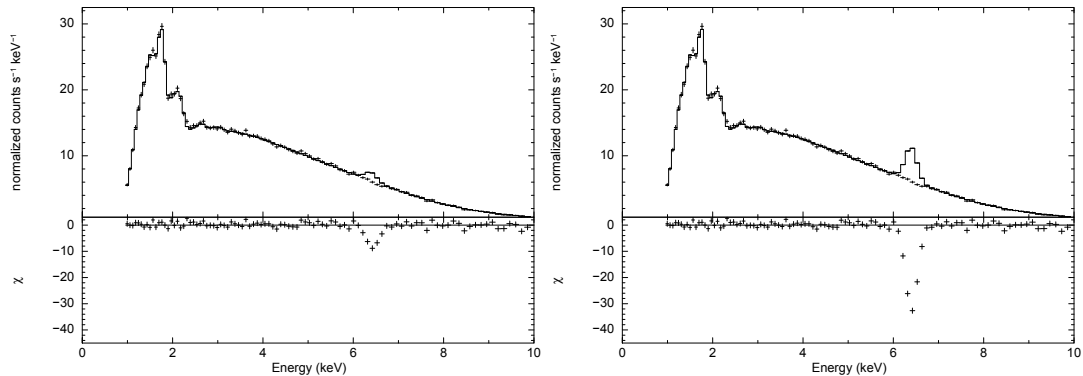


図 3: べき 1.4 のスペクトルにガウシアンを入れた例. ノーマリゼーション 0.005 (左) と 0.02 (右) 上段の図の + 印が基準音となるべき 1.4 のスペクトルデータ (図 1 と同じ). 実線がガウシアンをいれたモデル関数. 下段はデータとモデルの差.

$y = 2x$ を記入する小問 j では, 43 枚全ての回答が $y = x$ よりも上に線を記入しており, 直線は 42 枚だった. また, 直線の傾きも $y = 2x$ に近いものが多く, $y = 2x$ が 10 枚, $y = (7/4)x$ が 19 枚と, このふたつだけで 67% を占めた. 直線でなかった 1 例は下に凸の形をしていた.

$y = x^2$ を記入する小問 k では, 43 枚全ての回答で, グラフ右端で $y = x$ よりも上に線が到達していた. このうち, 直線を記入した例は 16 例 (37%) あり, 下に凸の曲線は 16 例 (37%), 下に凸の折れ線は 11 例 (26%) あった. 上に凸の回答はなかった.

グラフの終端 (右端) の y の値を見ると, $y > 2$: 4 例, $y = 2$: 12 例, $y = 7/4$: 9 例, $y = 3/2$: 12 例, $y = 5/4$: 5 例, $y = 1$: 1 例, で, 正解の $y = 2$ に到達している例が最大数ではあるものの, 回答には幅があるようすが見える.

全体として下に凸だということまでは聞き取ったが, 曲線であることは分かり辛かったことがわかる. また, y の正確な値は掴みづらく, 50% から 200% 程度の認識のずれがあることがわかる. 一方で, 回答の中には $y = x^2$ に近い例 (グラフ中央で $y = x$ と交差, グラフ用紙右上に到達, など) もあった. 図 4(左) に基準音 $y = x$ と評価音 $y = x^2$ を示す. 比較用のため軸メモリをつけていないが, x 軸の範囲は 0 から 2, y 軸の範囲は 0 から 4 である. また, 図 4(右) に $y = x^2$ (小問 k) に対する解答例を示す.

3.4 調査結果 4: スペクトルの概形の認識

大問 4 では, 複雑なスペクトルが聞き取れるかの確認を行った. 基準音は, `splot-for-xspec` の出力を使い, モデルはべき 1.4 のべき関数に吸収のかかったものとした. べき関数のべきは 1.4 とし, ノーマリゼーション, 吸収は両方とも 1.0 とした. これらのパラメタは大問 1, 2 と同じである.

評価音は, 基準音のスペクトルに輝線構造を想定したガウシアンをいれたものとした. ガウシアンのパラメタは, 中心エネルギー 6.4 keV, $\sigma = 1.0$, ノーマリゼーション 0.05 とした. これに加え, 吸収の水素柱密度を 1.0 から 2.0 へと変化させた. これにより, 1 keV 近辺の y 軸の値と, 高エネルギー側 (6.4 keV) にある輝線の y 軸の値が同程度となった. 図 5(左) に, 基準音, 評価音を示す.

スペクトルは x 軸 $1 \leq x \leq 10$, y 軸 $0.68 \leq y \leq 32.5$ の範囲のデータで, これは, 大問 1, 2 と共通としてい

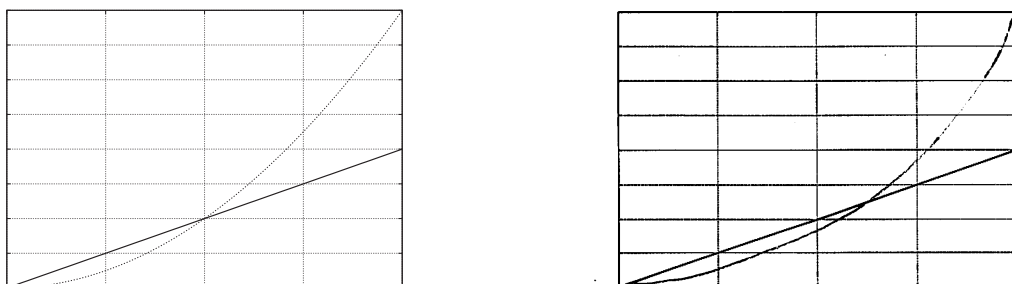


図 4: 左: 大問 3-小問 k の基準音 (実線 $y = x$) と評価音 (点線 $y = x^2$) 右: 大問 3-小問 k ($y = x^2$) を手書きで書き込んだ回答例. $y = x$ の直線は最初から解答用紙に印刷されている.

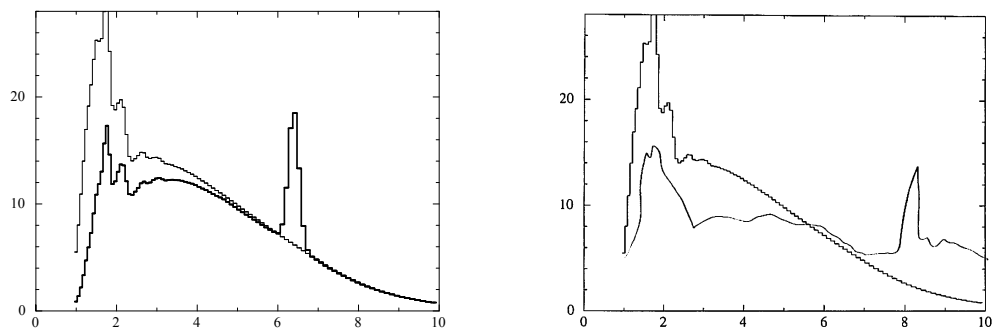


図 5: 左: 大問 4 の基準音 (実線 べき 1.4, 吸収 1.0) と評価音 (太線 べき 1.4, 吸収 2.0, ガウシアン追加) 右: 回答例. べき 1.4 の基準音の線は最初から解答用紙に印刷されている.

る. 用紙には, 参照用に基準音の線が記入してあり, 被験者はこれを見ながら評価音を書き込んだ.

3.4.1 結果

大問 4 の問題のスペクトルは, 一次関数等に比べ複雑な形をしているため, 回答には大きなばらつきがでた. これらのうちいくつかの回答では, グラフの概形をよく掴んでいるものがあつた. 回答を特徴別にまとめると, 表 2 のようになった. ひとつの回答が複数の特徴を有している場合もあり, これらは別々にカウントしている. また, 図 5(右) に回答の一例を示す.

表 2: 複雑なグラフの聞き取り結果

	人 (%)
2山構造 (1.8keV 6.4keV 近辺) をしている	34 (79 %)
1.8keV のピークは基準音の約半分の高さである	20 (47 %)
1.8ekV と 6.4 keV のピークの高さはほぼ同じである.	12 (28 %)
4kev 以下で基準音より評価音の方が低くなっている	18 (42 %)
ガウシアン of ピークは 6.4 ± 0.5 の範囲にある	16 (37 %)
グラフ上でガウシアンの中は 1.0 keV くらいに書かれている	23 (53 %)
ガウシアン of ピークの高さは連続成分の 2-3 倍程度である	12 (28 %)
7.0 keV 以上では基準音と評価音の線はほぼ一致している	11 (26 %)

4 考察

4.1 スペクトルの差分の認識

データプロットの比較 (大問 1) で べき 1.4 と 1.0 では差があると、ほぼ全員が回答していることから、今回サンプルにしたスペクトル程度で べき 1.4 と 1.0 の違いであれば、特に訓練がなくても聞き取りが可能ということがわかった。一方で、べき 1.4 と 1.3 で「違う」と答えたのは半数以下であり、べきの差が 0.1 程度だと聞き取りが難しい可能性がある。ただし、残差の音を聞いた場合は、97% の被験者が「差がある」と答えている。

同じように、大問 2 の小問 e, f で問うたような小さな構造 (ガウシアン, norm=0.005) も、データだけを聞いて比較した場合は差を認識できたのは 16% に過ぎなかったが、残差を聞いた場合 77% が「違う」もしくは「どちらかといえば違う」と答えている。

視覚によってグラフを見る場合と同様のことはあるが、残差を聴き比べることにより、より正確なグラフの差分を認識することができるし、それは音声プロットの場合も同様であると言える。残差をプロットするのは xspec の標準機能ではあるが、splot でこれを実現したことで、視覚障害者も細かい図の聞き分けができる可能性がある。

大問 4 にあるように、複雑なグラフでも一部の人にとってはよく特徴をつかめている場合もある。個人差や慣れの問題もあるため一概には言えないが、音によってスペクトルの概形を理解できる可能性があることは、今回の調査によって確認できたと思われる。本調査は、音によるデータ比較・スペクトルフィットはひとつの有効な手段となりうることを示したと言える。

4.2 判別しづらいグラフ

調査の中で手書きでグラフを答えた問題では、設問の関数とは違う回答が多かった間が複数あった。ひとつは、大問 3 の小問 k で、 $y = x^2$ の音を問うたもので、誤回答には、 $y = 2x$ などの一直線のものや、途中で折れ曲がった 2 本の直線などが多くあった。

もうひとつは 大問 4 で、x 軸のどのあたりにガウシアンのピークがあるか、4 keV 以下の連続成分は、基準音と評価音のどちらが上か、などに多くのばらつきがあった。

解決方法の一例としては、splot の「グリッド」機能を用いることが考えられる。splot の「グリッド」は、音でグラフを表現する際、一定時間ごとにシグナルをいれることにより X 軸の位置を認識しやすくする機

能である。また、晴眼の解析者がデータを何度もプロットしなおしてその特徴を見極めるように、splot でも (データ範囲などの) 条件を変えて複数回プロットすることで、図形の特徴をつかめるようになることを期待する。

4.3 STEM と X 線天文衛星のデータ解析

宇宙科学データ可聴化プロジェクトは、将来は科学衛星のデータ解析を、視覚障害者自身で行うことができる環境を構築することを目指している。これは STEM 分野で視覚障害者が活躍できる環境を作ろうとする近年の潮流にも合致するものである。

本研究では、splot-for-xspec によって視覚に頼らずにスペクトルフィットを行うことができる可能性を示した。しかし、スペクトルフィットはデータ解析プロセスの一部に過ぎない。X 線衛星のデータを解析するには、データの取得や、観測時間・衛星の姿勢・イメージ領域など様々な条件での取捨選択、応答関数の準備、その他多くの作業を行わなければならない。これらフィッティング以外のデータ解析プロセス全体のバリアフリー化が今後の課題である。

5 結論

宇宙科学データ可聴化プロジェクトでは、音声によるグラフ表示ソフトウェア splot を X 線スペクトルフィッティングツール xspec から動作させる splot-for-xspec を試作した。また、splot-for-xspec を用いてスペクトルの聞き分けができるかを調べる小規模な調査を実施した。ここから、音によるデータ比較・スペクトルフィットはひとつの有効な手段となりうることを示した。認識精度の向上や、フィッティング以外のデータ解析プロセス全体のバリアフリー化などが今後の課題である。

謝辞

本研究の調査・分析において、日本福祉大学の吉原智恵子准教授に有益な助言をいただいた。ここに謝意を表する。

参考文献

- [1] Hwang, Jiwon and Taylor, Jonte C. (2016) "Stemming on STEM: A STEM Education Framework for Students with Disabilities," Journal of Science Education for Students with Disabilities: **Vol. 19: Iss. 1**, Article 4., pp.39-49 Available at: <http://scholarworks.rit.edu/jsesd/vol19/iss1/4>
- [2] "HEASARC website" <https://heasarc.gsfc.nasa.gov/>
- [3] "The QDP/PLT User's Guide" <https://heasarc.gsfc.nasa.gov/docs/software/ftools/others/qdp/qdp.html>
- [4] Arnaud, K. A., : "XSPEC: The First Ten Years", *Astronomical Data Analysis Software and Systems V eds. Jacoby G. and Barnes J*, **Vol. 101**, pp.17 (1996)

- [5] 宇野 伸一郎, 亀山 哲也, 堀畑 昌希, 浅野 仙久, 海老沢 研, 田村 隆幸, 笠羽 康正, 篠原 育, 宮下 幸長, 三浦 昭, 松崎 恵一, 村上 弘志, 古澤 (秋元) 文江: “宇宙科学データ可聴化プロジェクト～プロジェクト立ち上げと初期データ公開～”. 日本福祉大学 健康科学論集, 第 10 巻, pp.1-9 (2007)
- [6] 宇野 伸一郎, 外谷 渉, 三浦 昭, 海老沢 研, : “宇宙科学データ可聴化プロジェクト (2) ～音声グラフ表示ソフトウェアの試作～”. 日本福祉大学 健康科学論集, 第 14 巻, pp.1-9 (2011)
- [7] 宇野 伸一郎, 外谷 渉, 三浦 昭, 海老沢 研, : “宇宙科学データ可聴化プロジェクトの現状”. 宇宙科学情報解析論文誌, 第 1 巻, pp.7-11 (2012)

Web 型データ解析ツール ERGWAT の開発と今後の発展計画

梅村 宜生^{*1}, 瀬川 朋紀^{*1}, 宮下 幸長^{*1}, 桂華 邦裕^{*1}, 三好 由純^{*1}, 堀 智昭^{*1},
小路 真史^{*1}, 田中 良昌^{*2}, 関 華奈子^{*3}, 篠原 育^{*4}

Development and Future Plans of the Web Based Data Analysis Tool ERGWAT

Norio UMEMURA^{*1}, Tomonori SEGAWA^{*1}, Yukinaga MIYASHITA^{*1}, Kunihiro KEIKA^{*1},
Yoshizumi MIYOSHI^{*1}, Tomoaki HORI^{*1}, Masafumi SHOJI^{*1}, Yoshimasa TANAKA^{*2},
Kanao SEKI^{*3} and Iku SHINOHARA^{*4}

Abstract: This paper reports the web-based interactive data analysis system ERG Web Analysis Tool (ERGWAT) developed by the ERG Science Center. The ERGWAT provides an environment for visualization of various kinds of geospace data including Akebono satellite and ground-based observations as well as simple time-series analysis. Interactive Data Language (IDL) / Space Physics Environmental Data Analysis Software (SPEDAS) is used as an engine of ERGWAT to load the data via internet and to make plots for data. All commands for ERGWAT are realized as tool bars, pull-down menus on the web browser, so that users do not have to be familiar with IDL/SPEDAS commands. The ERGWAT provides not only plots on the web browser but also postscript files that can be used for papers and presentations. It is expected that ERGWAT is a useful resource for the capacity building for solar-terrestrial physics community. The possible link between ERGWAT and Inter-university Upper atmosphere Global Observation NETwork (IUGONET) metadata database is a future subject, which provides both the plot results and the information of data itself.

Keywords: ERG, ERG Science Center, ERGWAT, Space Physics, Data Analysis Tool, Web Platform

概要

本論文では、宇宙航空研究開発機構宇宙科学研究所と名古屋大学宇宙地球環境研究所統合データサイエンスセンターによって運営されている宇宙科学連携拠点 ERG サイエンスセンターが開発・公開しているデータ解析ツール ERGWAT (ERG Web Analysis Tool) のシステム概要、提供内容、今後の発展計画について述べる。ERGWAT は、ウェブブラウザをフロントエンドとして用い、Interactive Data Language (IDL) ルーチン群である Space Physics Environment Data Analysis Software (SPEDAS) がバックエンドで動くことで、ウェブブラウザ上の操作のみで各種データの可視化や時系列解析等を行える機能を有しており、IDL ユーザでなくても簡単に各種データの可視化や解析を行うことが可能であ

doi: 10.20637/JAXA-RR-16-007/0003

* 平成 28 年 11 月 24 日受付 (Received November 24, 2016)

^{*1} 名古屋大学宇宙地球環境研究所 (Institute for Space-Earth Environmental Research, Nagoya University)

^{*2} 国立極地研究所 (National Institute of Polar Research)

^{*3} 東京大学大学院理学系研究科 (School of Science, Tokyo University)

^{*4} 宇宙航空研究開発機構宇宙科学研究所 (Institute of Space and Astronautical Science, Japan Aerospace Exploration Agency)

る。本稿執筆時点(2016年9月), ERGWAT では, ACE, GOES, THEMIS, Van Allen Probes などの科学衛星のデータに加え, ERG サイエンスセンターが独自に開発した IDL ルーチンにより, あけぼの衛星データ, フラックスゲート磁力計, 誘導磁力計, SuperDARN レーダー, EISCAT レーダーなどの地上観測機器のデータのプロットや解析も可能である。ERGWAT は, 画面に表示されたボタンやプルダウンメニューを操作するのみで, 単純な時系列プロットの作成のみならず, パワースペクトル変換やウェーブレット変換などの解析を行うことも可能である。また, ERGWAT は, プロット画像を Postscript 形式で出力する機能, ERGWAT で実行した一連のコマンドを IDL の実行プログラム形式のファイルとして出力する機能も持つなど, 操作が非常に容易でありながら, 実用的な解析も可能としている。ERGWAT は, このような利便性の高さから, 太陽地球系科学におけるキャパシティビルディングへの貢献も期待されている。今後は, IUGONET のメタデータ・データベースと連携してプロット画像とデータの詳細情報を一元的に提供することを目指すとともに, 他プロジェクトと協調しながらシステムの利便性の更なる向上を図る計画である。

1 はじめに

太陽地球系物理学分野における観測データは, 科学衛星, 地上観測機器などにより取得されており, 各コミュニティが用いる解析ソフトウェアのベースも MATLAB, IDL など複数に渡っている。太陽地球系科学においては, フレアや CME など太陽で発生した物理現象が, 惑星間空間, 磁気圏, 電離圏, 大気圏, 地表圏にどのような形でエネルギーを伝え, どのように相互作用してその場の環境が変化するかなど, 太陽地球系をシームレスなシステムとして総合的に捉える必要がある反面, 多様な観測データを解析するためには, それぞれの観測データに合わせて異なる言語やソフトウェアのベースを使う必要があるなど, 必ずしも総合的に解析するための環境が実現されているとはいえなかった。研究者が複数のソフトウェアを使いこなすことは容易ではなく, 統合解析を行う際のデータハンドリングの敷居の高さが問題となっていた。

現在, 日本では, ジオスペース探査(ERG)プロジェクトが進行している¹⁾。ジオスペース探査(ERG)プロジェクトは, 宇宙航空研究開発機構宇宙科学研究所(JAXA/ISAS)によって打ち上げられる ERG 衛星, 国内の各種大学・機関によって展開・運用されている連携地上観測, 統合解析・モデリングの 3 つの研究チームによって推進されている。また, プロジェクトの多様なデータを一元的にアーカイブ・公開するとともに, 統合解析環境を提供するため, JAXA/ISAS と名古屋大学宇宙地球環境研究所によって宇宙科学連携拠点 ERG サイエンスセンターが設置, 運営されている。ERG サイエンスセンターでは, 観測データは原則として CDF 形式でアーカイブ²⁾, 公開するとともに, 衛星軌道と地上観測点の位置関係を簡単に表示できる Conjunction Event Finder (CEF) の開発・運用も行っている³⁾。また, 統合解析環境として, IDL 言語で記述されている Space Physics Environment Data Analysis Software (SPEDAS)⁴⁾と呼ばれる解析ツールを使用している。SPEDAS は, 米国 NASA の人工衛星と地上観測プロジェクトである THEMIS プロジェクトのデータ解析のために開発された THEMIS Data Analysis Software (TDAS) を前身としており, UCLA, UCB を中心に, 様々な国の研究機関や大学がそれぞれのプロジェクトに対応する解析のためのプログラムである IDL ルーチンをプラグインとして提供している。ERG サイエンスセンターでは, これまで IUGONET (Inter-university Upper atmosphere Global Observation

NETwork)⁵⁾⁶⁾ と協力して、ERG 連携地上観測データを解析するための IDL ルーチンを開発し SPEDAS への実装を行ってきた⁷⁾。このように SPEDAS に多くのプロジェクトが参画していることで、科学衛星、地上観測による多種・多拠点のデータの可視化と解析が実現されており、SPEDAS は、その内容と規模を拡大している。

SPEDAS は、有償である IDL ライセンスを利用できる環境下であれば IDL ルーチンを自由に開発してコンパイル・実行することができるコマンド操作方式(CUI)と、IDL ライセンスを必要とせずともコンパイル済の実行ファイルをインストールするのみで予めパッケージ化された IDL ルーチン群に対する観測機器のデータを IDL バーチャルマシン上で解析することのできる画面操作方式(GUI)の 2 つのインターフェースを持つ。特に後者は、画面を使った直感的な操作が可能であること、IDL ライセンスを必要としないことから、日常的に IDL を使っているユーザのみならず、普段 IDL を用いていないユーザであっても容易に使用することが可能である。実際に、インド、インドネシア、ペルー、アフリカ諸国等の開発途上国にて GUI 方式を用いた解析講習会が開催されるなど国際的にも広く普及しつつあり⁸⁾、太陽地球系科学のキャパシティビルディングにも貢献している。

ERG サイエンスセンターでは、この SPEDAS が扱う観測データの多様性に着目し、SPEDAS の操作性を更に向上させて解析の利便性を高めることを目的とした新しいツールの開発を進めている。このツールは ERGWAT (ERG Web Analysis Tool) と呼ばれるもので、フロントエンドにウェブブラウザを使用し、ウェブブラウザに入力された情報をもとにバックエンドで処理エンジンとして動作する CUI 方式の SPEDAS を操作することで、各種データの可視化・解析を行うことができるものである。ERGWAT を使うことで、IDL を使用したことがないユーザでも直観的な操作が可能であり、IDL と SPEDAS をインストールする必要もなく、ウェブブラウザを搭載したパソコンやモバイル端末を操作するのみで、データを”Quick”に可視化・解析することが可能である。ERGWAT は 2009 年に開発が開始され、順次機能改良を加え、現在はバージョン 4 に至っている。本論文では、この ERGWAT のシステムの詳細と、実際の科学データ解析例について紹介し、最後に今後の発展計画について述べる。

2 ERGWAT の構成

2.1 ERGWAT の画面構成

図 1 に、ERGWAT を使い、九州大学国際宇宙天気科学・教育センターが運用、名古屋大学宇宙地球環境研究所がデータを公開している 210 度地磁気観測データの 1 分値データ⁹⁾をプロットした例を示す。

ERGWAT の画面は、プロットを作成するための日時、観測機器、パラメータを設定するフォームと、プロット画像を表示する領域で構成される。基本的な操作手順は、(1)プロットしたい時間範囲を設定する、(2)プロットしたい観測機器とその観測所、観測モード等のパラメータを選択する、(3)Load ボタンを押す、(4)描きたい物理量を選んで Plot ボタンを押す、のように単純なステップで構成される。コマンド操作型の解析ツールのように専門用語を入力する必要はなく、画面に示されるパラメータに従い操作すれば希望するプロットを描くことができる。なお、この ERGWAT は、同時利用数を制限するためにユーザ認証制を導入しており、ID とパスワードでログインした後に図 1 の画面へと移行する。

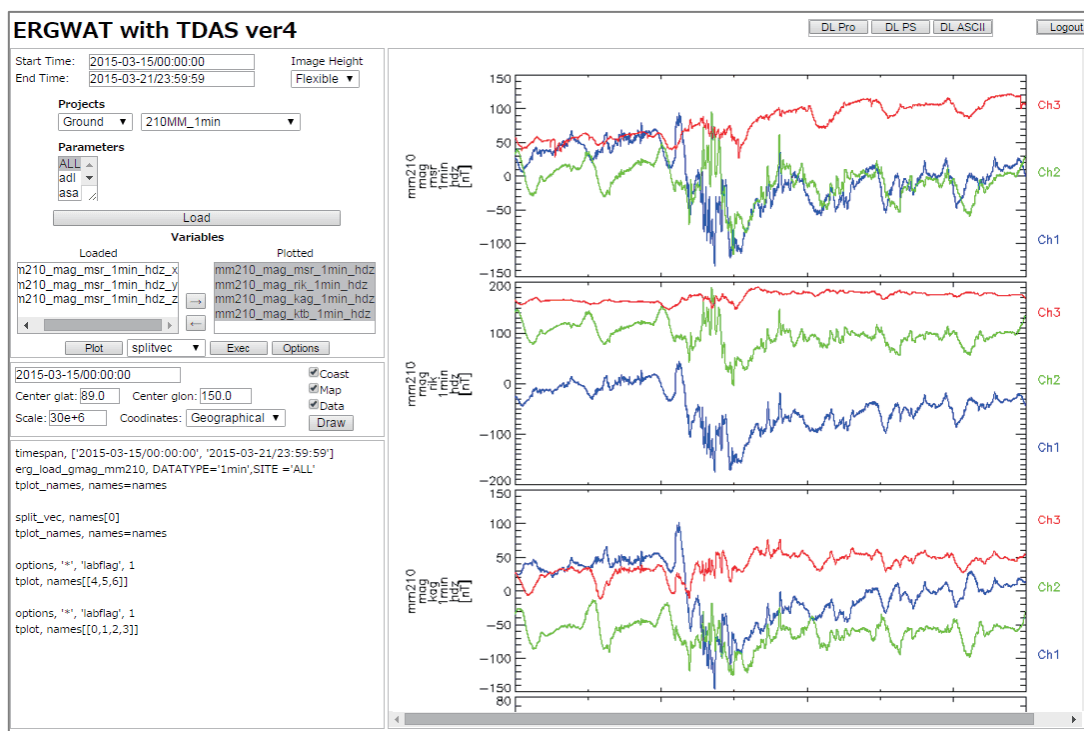


図 1 ERGWAT 操作画面。画面の左側に、時間範囲、観測機器、観測所、パラメータ等を選択するためのフォームと実行ボタン、右側にプロット画像を表示する領域を持つ。

2.2 ERGWAT のシステム構成

図 2 に ERGWAT のシステム構成図を示す。ERGWAT は Linux で動作し、画面を構成してウェブブラウザと IDL ルーチン/IDL を中継するクライアントサイド・サーバーサイドスクリプト(Perl, JavaScript Ajax), 実行時に必要な IDL ルーチン名, 実行に必要なパラメータ等を格納する RDBMS (MySQL), IDL ルーチンのファイルを格納するための環境変数 IDL_PATH を通したディレクトリ, IDL プロセスを監視して実行結果をユーザに返却するためのデーモンである TDASD, 実際の実行プログラムである IDL ルーチン, IDL ルーチンの実行環境である IDL で構成される。

ERGWAT では、バックグラウンドで IDL ルーチンを実行するための情報は RDBMS に格納されており、RDBMS は、図 3 に示すように、観測機器グループ名称、観測機器名称、対応する IDL ルーチンの名称を格納するための 3 つのプロジェクト系テーブル(eg_proj1~3)と、IDL ルーチン実行時に付与するオプション(データ種、観測所情報、観測モード等の引数)の名称とその値を格納するためのプロジェクト系テーブル(eg_param1~4)で構成されている。例えば図 1 の最上段の 210 度地磁気観測データ北海道・母子里観測所のデータの場合、eg_proj1 テーブルに観測機器グループ名称を示す文字列 Ground, eg_proj2 テーブルに観測機器名称を示す文字列 MM210_1min, eg_proj3 テーブルに IDL ルーチンの名称 erg_load_gmag_mm210 を登録する。また、同データの IDL ルーチンを実行するためのコマンドは SPEDAS の CUI と同等の erg_load_gmag_mm210, DATATYPE='1min', SITE='msr'であり、この場合、eg_param1 テーブルに DATATYPE='1min', SITE と msr を登録する(ここで、データ種を示す DATATYPE='1min'は固定パラメータであるために、観測所を示す SITE パラメータ名と連結して、1 つのカラムに登録している。また、SITE パラメータが示す観測所の種類に応じてレコードを作成する)。

他の観測機器で必要があれば `eg_param4` までのテーブルを使う．次に，実際の IDL ルーチンのファイル `erg_load_gmag_mm210.pro` を，`IDL_PATH` を通したディレクトリに配置する．IDL ルーチンは，**ERGWAT** 用に変更する必要はなく，**SPEDAS** の CUI で動作するものでよい．上記の設定を行うことにより，**ERGWAT** 画面において，観測機器グループ，観測機器，実行時のオプション等，プロット画像を作成するためのパラメータの絞り込みが可能となる．

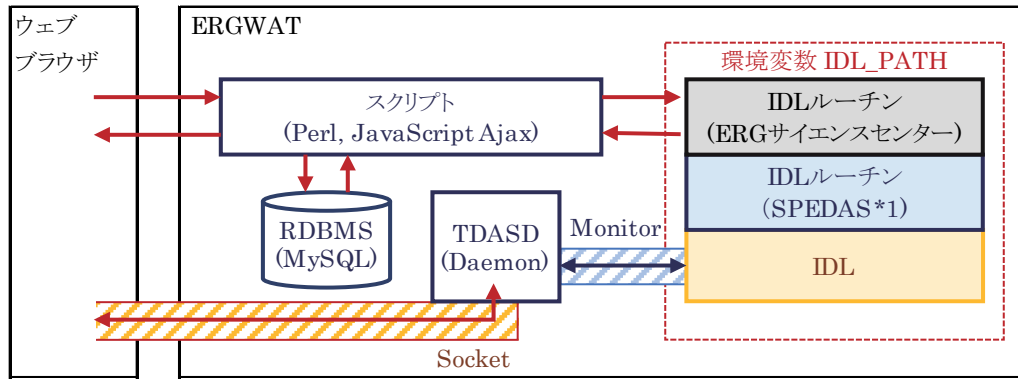


図2 ERGWAT のシステム構成．ユーザのウェブブラウザとバックエンドのIDLの処理は，サーバに設置されたスクリプトにより中継される．また，IDL ルーチンの実行に必要な情報は RDBMS に格納されている．なお，搭載している SPEDAS のバージョンは 1.00 である(*1)．

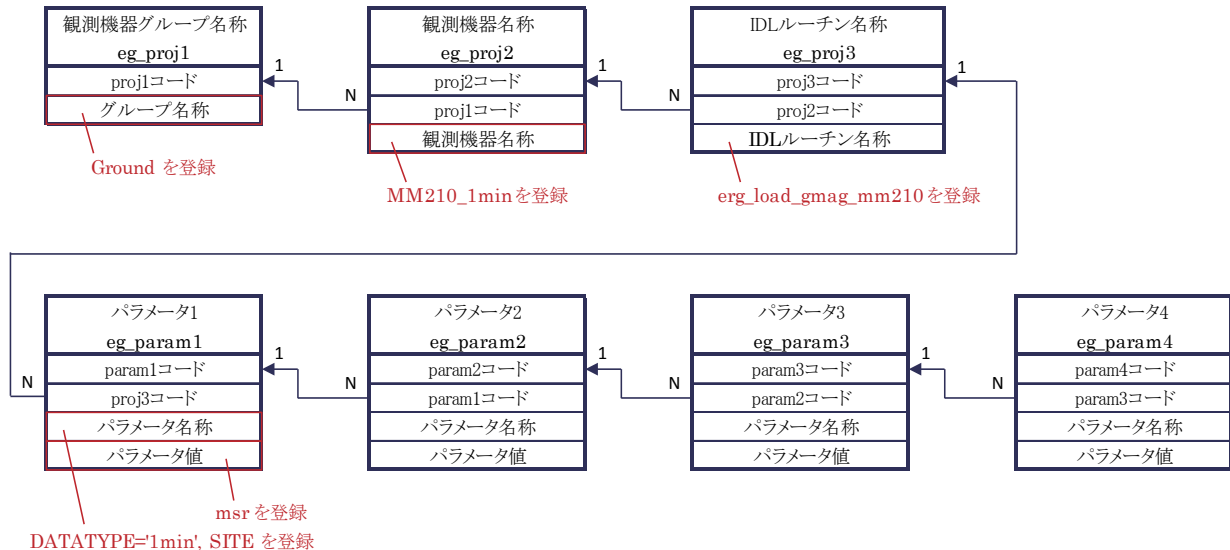


図3 ERGWAT におけるテーブルレイアウト．観測機器グループ名称のテーブルを起点に，観測機器名称，IDL ルーチン名称，パラメータ 1～4 のテーブルに向けて，1:N ($N>0$) で正規化されている．

図4に，ユーザによるパラメータの選択からプロットが作成されるまでのシーケンスを示す．ユーザにより観測機器グループが選択された場合，クライアントサイドスクリプトにて操作内容を取得し，サーバーサイドスクリプト，RDBMS と通信して，指定された観測機器グループに含まれる観測機器名を取得して画面に表示する．例えば，図1の Projects の左側のプルダウンメニューにて地上観測を示す

Ground を選択した場合、右側のプルダウンメニューには、`eg_proj2` テーブルに格納されている値のうち、`210MM_1min`⁹⁾、`SuperDARN`¹⁰⁾、`STEL_INDUCTION`(誘導磁力計)¹¹⁾、`MAGDAS`¹²⁾といった地上観測機器名を一覧表示する。同様に、右側のプルダウンメニューにて `MM210_1min` を選択した場合、`eg_param1` テーブルに格納されている `MM210_1min` の実行時のオプションである `DATATYPE='1min'`、`SITE=`に対する値である `ALL`、`adl`、`asa`、`msr`、`zgn`、`zyk` 等の観測所の名称を取得し、`Parameters` フィールドに一覧表示する。このようにユーザの操作に合わせて動的に観測機器名、観測所名等を表示し、ユーザにより選択され `Load` ボタンが押されると、再びサーバサイトスクリプトに入力値が渡り、それに基づく `IDL` ルーチン名、実行オプションを `RDBMS` の `eg_proj3`、`eg_paramN`(`N` は 1..4) テーブルから取得して実行に必要なコマンドを生成し、コマンドを `IDL` ルーチンに引き渡す。`IDL` は、`SPEDAS` を `CUI` で実行する場合と同様に、初期化コマンドの `thm_init`、時刻設定コマンドの `timespan`、`IDL` ルーチンコマンド(図 1 の例では `erg_load_gmag_mm210`、`DATATYPE='1min'`、`SITE='ALL'`)を実行し、観測データをダウンロードした後、処理結果である物理量を `tplot` 変数と呼ばれる `IDL` の構造体変数に格納する。`tplot` 変数のデータがセットされると、`tplot` 変数名は、`TDASD` プロセスからソケットを経由してウェブブラウザに戻されて `Variables` 欄に表示され、ユーザによる選択が可能となる。この後、ユーザにより可視化したい `tplot` 変数名が指定され `Plot` ボタンが押されると、再びサーバサイドスクリプト、`IDL` に命令が渡り、選択した `tplot` 変数に対するプロット画像を作成してユーザに提供する。このとき、可視化に必要な観測データのファイルは `ERGWAT` が動作しているサーバにダウンロードされる。従って、ユーザは、プロット画像のみを可視化・解析結果として得る場合は、手元に観測データをダウンロードする必要はない。第 3.2 章に後述するが、実際の観測データの中身を見たい場合にのみ、観測データをダウンロードすればよい。

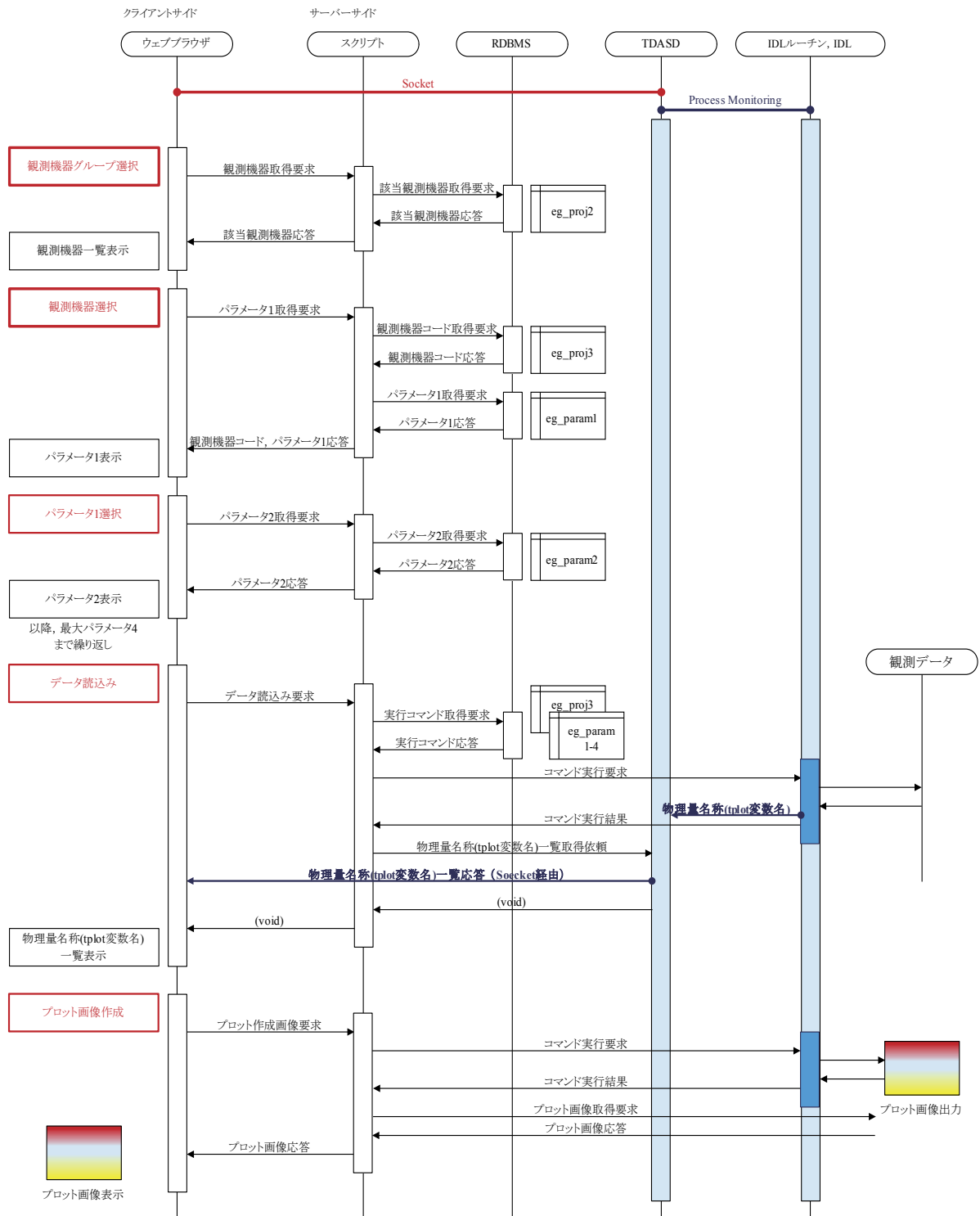


図4 ERGWATのシーケンス. 矢印は通信の流れを示す. ユーザの操作によりサーバーサイドのスクリプト・RDBMSと通信を行い, 観測グループから, 観測機器名, 実行パラメータの絞り込みを行う. データ読み込み時もサーバーサイドスクリプトを中継してバックエンドのIDLに要求が渡り, 結果はIDLを監視しているTDASDからソケットを経由してユーザに戻される(図中青線). また, プロット画像作成要求時もサーバーサイドスクリプトからIDLに要求が渡り, IDLにより作成されたプロット画像をサーバーサイドスクリプトが取得して, ユーザに戻す.

3 ERGWAT が果たす ERG プロジェクトの科学成果への貢献と期待

3.1 ERGWAT で解析することができる観測データ

本稿執筆時点(2016 年 9 月時点)では、表 5 に示すように、SPEDAS, ERG サイエンスセンター, IUGONET が提供している IDL ルーチンを ERGWAT に登録している。第 1 章で述べた通り, ERG サイエンスセンターと IUGONET が開発している IDL ルーチンは SPEDAS に標準パッケージとして組み込まれているために, ERG プロジェクトに関わるデータも, 実行に関する情報を RDBMS に登録するのみで, ERGWAT 上で可視化・解析することが可能である。また, 米国 Van Allen Probes の一部観測装置, あけぼの衛星など, ERG サイエンスセンターが独自または先行開発している IDL ルーチンも随時追加しており, 解析対象となるプロジェクトやデータは増加している。今後, ERG 衛星が打ち上がり, データの取得が始まった際には, ERGWAT 上でも迅速に可視化・解析できるよう準備を進めていく予定である。

表 5 ERGWAT にて解析することができる観測データ(2016 年 9 月現在)。SPEDAS に含まれる ACE, GOES, THEMIS 等の科学衛星データに加え, SPEDAS には未収録であるが, ERG サイエンスセンターによる IDL ルーチンの独自開発により, Van Allen Probes の一部観測装置(*1), 先行開発しているあけぼの衛星データについても解析が可能である(*2)。また, IUGONET にて開発が進められている MU レーダー, オーロラ全天カメラほか全 39 種類の地上観測機器についても, RDBMS へのレコード登録を行えば解析できる状態にある(*3)。

IDLルーチン提供元	観測機器区分	観測機器名称 (プロジェクト名称)
SPEDAS	科学衛星	ACE
		GOES
		THEMIS
		Van Allen Probes
	その他	OMNIデータ
ERG Science Center	科学衛星	Van Allen Probes, ECT *1
		あけぼの *2
	地上観測機器	210度地磁気
		フラックスゲート磁力計
		誘導磁力計
		SuperDARN
		MAGDAS
		EISCAT
		OMTI全天カメラ
IUGONET *3	地上観測機器	MUレーダー
		オーロラ全天カメラ
		ほか全39種類

3.2 ERGWAT を用いた高度な解析

図 1 に示した 210 度地磁気観測データの 1 分値データのうち、北海道・母子里観測所の観測データに対し、時系列解析の一例として、各成分に分解して x 成分値のみを取り出し、フーリエスペクトル変換、ウェーブレット変換を施したものを、図 6 に示す。ERGWAT 画面には成分分解、フーリエスペクトル変換、ウェーブレット変換を行うためのボタンを実装しており、各ボタンを押して表示される子画面(図 6 左)にて変換したい物理量名称(`tplot` 変数名)を選択し[OK]ボタンを押すのみで、バックグラウンドの IDL ルーチンに実行命令が出され、変換結果の新しい `tplot` 変数を受け取ることができる。変換後の `tplot` 変数を選択し、Plot ボタンを押すことによって、図 1 では視覚的に捉えることが困難であった地磁気変動の周波数特性を詳細に読み取ることが可能となる。

筆者らの知る限り、太陽地球系科学分野で利用されている類似のデータ解析ツールとしては、米国 NASA が開発・公開している CDAWeb (Coordinated Data Analysis [Workshop] Web)¹³⁾がある。CDAWeb は、プロット画像を作成する前に簡易なノイズ除去フィルターを適用できる程度であるが、ERGWAT は、プロット画像を作成した後であっても、解析対象を絞り込み、詳細に解析したい部分を取り出す、取り出した部分に対して解析を加えるなど、対話方式での操作が可能である。また、図 5 左の Advanced setting 欄に、ERGWAT に登録されていない追加の実行オプションを入力することで、フーリエスペクトル変換における窓関数の指定、時刻範囲の再設定、別変数への結果の格納といった詳細な解析や計算を施すことも可能である。さらに、ERGWAT では、画面上部の DL PS ボタンの押下によりプロット画像を PostScript 形式としてダウンロードすること、DL ASCII ボタンの押下によりプロットに用いた観測データを ASCII 形式に変換してダウンロードすることに加え、画面左下部にログインからその時点までに実行した一連のコマンドを表示するとともに DL Pro ボタンの押下によりその一連のコマンドを IDL の実行プログラム形式のファイルとしてダウンロードすることも可能である。実際の SPEDAS のコマンドを確認しながら解析を進められるとともに、ERGWAT で実行した内容を保存して手元の IDL 環境で再現させたり、その内容を編集して手元で解析を継続したりといったことも可能である。また、IDL ライセンスを所有していない場合や IDL ユーザでない場合、特に他分野の研究者であっても、実際の解析の手法を知ることができ、前述したデータハンドリングの敷居の問題を低減できるものとして有意である。このように、ERGWAT は簡易な操作性を有しつつ、研究者にとり実用的な機能を有している。

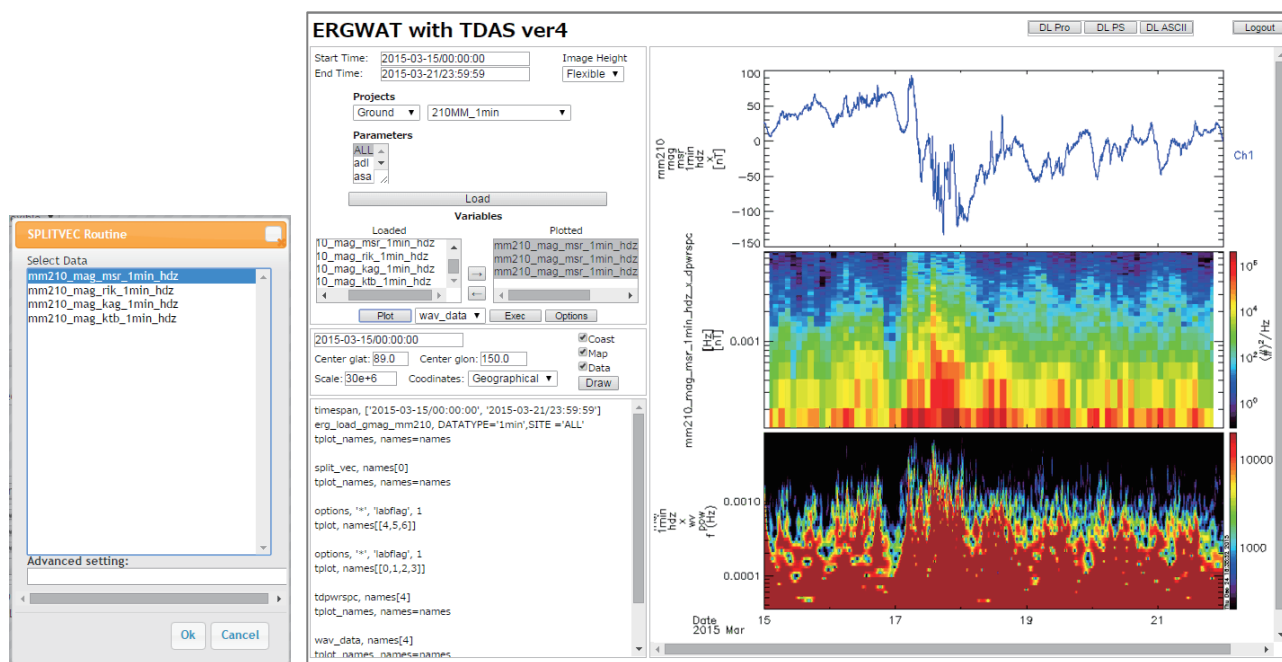


図6 210度地磁気観測データのうち北海道・母子里観測所の観測データに対して、 x 成分のみ取り出し(上段)、フーリエスペクトル変換(中段)、ウェーブレット変換(下段)を適用した例。ERGWAT画面に表示されている各変換用のボタンを押すと子画面(左図)が表示され、変換対象としたい物理量(tplot 変数名)を選択するのみで各変換が実行される。

3.3 ERGWAT の今後の開発計画

今後、ERG 衛星の打ち上げも控え、ERG サイエンスセンターでは ERGWAT の機能を一層拡大していくことを計画している。具体的には、登録する観測データの拡大に加え、以下に挙げる機能の拡張を予定している。

3.3.1 ERGWAT 画面に表示する値への別名の付与

現在の ERGWAT の画面に表示される IDL ルーチン名などは、そのデータを専門とする研究者以外にとり必ずしも分かり易いとはいえない。今後、画面に表示される観測機器グループ名、観測機器名、IDL ルーチン名、パラメータ名、tplot 変数名などに別名を付与することで、説明語句や文章も参照しながら、プロット画像の作成と解析をより容易に行うことができることを目指している。

例えば図1の場合、Project 欄の「MM210_1min」を「MM210_1min (Geomagnetic Observation using the Magnetometer Network, 210 Magnetic Data)」, Parameters 欄の「msr」を「msr (Moshiri, Hokkaido, Japan)」, Variables 欄の「mm210_mag_1min_hdz」を「mm210_mag_1min_hdz (Magnetometer 1min Resolution Data)」のように併記することを考えている。第2.2章で述べた通り、画面に表示する値はデータベースで管理していることから、データベースのテーブルに別名用のカラムを加え、値を追加し、スクリプト(Perl, JavaScript Ajax)にその変数に対する表示処理を付け足すのみで対応が可能である。また、別名は、運用時であってもそのカラムを書き換えることで随時変更することができる。この機能は2016年度より順次取り入れる予定である。

3.3.2 IUGONET メタデータ・データベースとの連携

IUGONET メタデータ・データベース¹⁴⁾と ERGWAT を連携させる予定である。具体的には、IUGONET のメタデータの画面に ERGWAT へアンカーを設けてプロット画面へ遷移する、ERGWAT で扱うデータセットに IUGONET メタデータ・データベースの URL をセットして観測データに関する詳細な情報のページに遷移することを目指している。このように、ERGWAT の特徴であるウェブブラウザでの可視化・解析の機能と、IUGONET が開発してきたデータ説明の機能を接続することで、プロット画像と観測データの情報を一元的に提供することが可能となる。今後、ERG サイエンスセンターが独自に扱うデータセットも IUGONET に登録するとともに、表 5 に挙げた IUGONET の IDL ルーチンも ERGWAT で解析できるようにして利便性を向上させることで、多くのユーザによる ERGWAT の利用が期待される。

4 まとめ

宇宙科学連携拠点 ERG サイエンスセンターでは、ウェブ方式で動作する解析ソフトウェア ERGWAT を 2009 年から開発・公開している。ERGWAT は、フロントエンドにウェブブラウザ、バックエンドに IDL ルーチン (SPEDAS, ERG サイエンスセンター独自開発分) を用いており、ユーザの画面操作により IDL ルーチンをサーバ側で実行し、観測データの可視化結果をウェブブラウザ上に表示する。現時点では SPEDAS に含まれる ACE, GOES, THEMIS 等の科学衛星のデータに加え、ERG サイエンスセンターが独自に開発した IDL ルーチンにより、フラックスゲート磁力計、誘導磁力計、SuperDARN レーダー、EISCAT レーダー等の地上観測データ、あけぼの衛星のデータのほか、ERG 衛星に先行して打ち上げられた米国の放射線帯探査衛星 Van Allen Probes による一部のデータについても ERGWAT 上での可視化が実現されている。

今後は、ERGWAT 画面に表示する名称に分かりやすい別名を付与してより使いやすいものとするとともに、IUGONET のメタデータ・データベースと連携してプロット画像と観測データについての情報を一元的に提供することを計画している。このように、2016 年度に計画されている ERG 衛星の打ち上げと ERG 衛星による観測データの取得の開始に向けて、統合解析のための研究環境を整備していくとともに、衛星観測、地上観測、シミュレーションモデリング研究からなる ERG プロジェクトの成果の拡大、さらに太陽地球系科学コミュニティにおけるキャパシティビルディングに貢献するものとして計画を進めている。

謝辞

ERG サイエンスセンターは、ジオスペース探査(ERG)プロジェクトのサイエンスセンターとして活動しており、宇宙航空開発研究機構・宇宙科学研究所と名古屋大学宇宙地球環境研究所による宇宙科学連携拠点として設置・運営されています。ERGWAT の開発については、SPEDAS 開発チームならびに IUGONET 開発チームから多くの助言と支援を受けています。

参考文献

- 1) Miyoshi, Y., T. Ono, T. Takashima, K. Asamura, M. Hirahara, Y. Kasaba, A. Matsuoka, H. Kojima, K. Shiokawa, K. Seki, M. Fujimoto, T. Nagatsuma, C.Z. Cheng, Y. Kazama, S. Kasahara, T. Mitani, H. Matsumoto, N. Higashio, A. Kumamoto, S. Yagitani, Y. Kasahara, K. Ishisaka, L. Blomberg, A. Fujimoto, Y. Katoh, Y. Ebihara, Y. Omura, M. Nose, T. Hori, Y. Miyashita, Y. Tanaka, T. Segawa, and ERG working group, The Energization and Radiation in Geospace (ERG) Project, in Dynamics of the Earth's Radiation Belts and Inner Magnetosphere, Geophys. Monogr. Ser., vol 199, edited by D.Summers, I.R. Mann,D.N.Baker,and M.Schulz, pp.103-116, AGU, Washington, D.C. doi:10.1029/2012BK001304, 2012.
- 2) Hori T, N. Nishitani, Y. Miyoshi, Y. Miyashita, K. Seki, T. Segawa, K. Hosokawa, A. S. Yukimatu, Y. Tanaka, N. Sato, M. Kunitake, and T. Nagatsuma, An integrated analysis platform merging SuperDARN data within the THEMIS tool developed by ERG-Science Center (ERG-SC), Advances in Polar Science, 24, 1, 69-77, doi:10.3724/SP.J.1085.2013.00069, 2013.
- 3) Miyashita, Y., I. Shinohara, M. Fujimoto, H. Hasegawa, K. Hosokawa, T. Takada, and T. Hori, A powerful tool for browsing quick-look data in solar-terrestrial physics: "Conjunction Event Finder", Earth Planets Space, Vol. 63 (No. 1), pp. e1-e4, doi:10.5047/eps.2011.01.003, 2011.
- 4) Space Physics Environment Data Analysis Software,
<http://themis.ssl.berkeley.edu/software.shtml> .
- 5) Hayashi, H., Y. Koyama, T. Hori, Y. Tanaka, S. Abe, A. Shinbori, M. Kagitani, T. Kouno, D. Yoshida, S. UeNo, N. Kaneda, M. Yoneda, N. Umemura, H. Tadokoro, T. Motoba, and IUGONET project team, Inter-university Upper Atmosphere Global Observation Network (IUGONET), Data Science Journal, 12, WDS179-WDS184, doi:10.2481/dsj.WDS-030, 2013.
- 6) Tanaka, Y.-M., A. Shinbori, T. Hori, Y. Koyama, S. Abe, N. Umemura, Y. Sato, M. Yagi, S. UeNo, A. Yatagai, Y. Ogawa, and Y. Miyoshi, Analysis software for upper atmospheric data developed by the IUGONET project and its application to polar science, Advances in Polar Science., 24, 231-240, doi: 10.3724/SP.J.1085.2013.00231, 2013.
- 7) Hori, T., Y. Miyashita, Y. Miyoshi, K. Seki, T. Segawa, Y.-M. Tanaka, K. Keika, M. Shoji, I. Shinohara, K. Shiokawa, Y. Otsuka, S. Abe, A. Yoshikawa, K. Yumoto, Y. Obana, N. Nishitani, A. S. Yukimatu, T. Nagatsuma, M. Kunitake, K. Hosokawa, Y. Ogawa, K. T. Murata, M. Nose, H. Kawano, and T. Sakanoi, CDF data archive and integrated data analysis platform for ERG-related ground data developed by ERG Science Center (ERG-SC), Journal of Space Science Informatics Japan, JAXA Research and Development Report, 4 (JAXA-RR-14-009), 75-89, ISSN 1349-1113, 2015.
- 8) Yatagai, A., Y. Sato, A. Shinbori, S. Abe, S. UeNo and IUGONET Team, "The capacity-building and science-enabling activities of the IUGONET to the solar-terrestrial research community", Earth, Planets and Space, 67, doi:10.1186/s40623-014-0170-2, link, 2015.
- 9) Yumoto, K., Y. Tanaka, T. Oguti, K. Shiokawa, Y. Yoshimura, A. Isono, B. J. Fraser, F. W. Menk, J.

- W. Lynn, M. Seto, and 210 (deg) MM magnetic meridian observation group, Globally coordinated magnetic observations along 210 (deg) magnetic meridian during STEP period: 1. Preliminary results of low-latitude Pc 3's, *J. Geomag. Geoelectr.*, 44, 261-276, 1992.
- 10) R. A. Greenwald, K. B. Baker, J. R. Dudeney, M. Pinnock, T. B. Jones, E. C. Thomas, J.-P. Villain, J.-C. Cerisier, C. Senior, C. Hanuise, R. D. Hunsucker, G. Sofko, J. Koehler, E. Nielsen, R. Pellinen, A. D. M. Walker, N. Sato and H. Yamagishi, DARN/SUPERDARN, *Space Science Reviews*, Volume 71, Issue 1-4, pp. 761-796, doi: 10.1007/BF00751350, 1995.
- 11) Shiokawa, K., R. Nomura, K. Sakaguchi, Y. Otsuka, Y. Hamaguchi, M. Satoh, Y. Katoh, Y. Yamamoto, B. M. Shevtsov, S. Smirnov, I. Poddelsky, and M. Connors, The STEL induction magnetometer network for observation of high-frequency geomagnetic pulsations, *Earth Planets Space*, 62(6), 517-524, doi:10.5047/eps.2010.05.003, 2010.
- 12) Yumoto K. and the MAGDAS Group, MAGDAS project and its application for space weather, *Solar Influence on the Heliosphere and Earth's Environment: Recent Progress and Prospects*, Edited by N. Gopalswamy and A. Bhattacharyya, ISBN-81-87099-40-2, pp. 309-405, 2006.
- 13) Coordinated Data Analysis Web, <http://cdaweb.gsfc.nasa.gov/>.
- 14) Abe, S., N. Umemura, Y. Koyama, Y. Tanaka, M. Yagi, A. Yatagai, A. Shinbori, S. UeNo, Y. Sato, N. Kaneda, Progress of the IUGONET system - metadata database for upper atmosphere ground-based observation data, *Earth, Planets and Space*, 66, doi:10.1186/1880-5981-66-133, 2014.

SVM による深発月震分類のための有効な特徴量の分析

加藤 広大^{*1} 菊池 栞^{*1} 山田 竜平^{*2} 山本 幸生^{*3} 廣田 雅春^{*4}
横山 昌平^{*5} 石川 博^{*1}

Analysis of the effective features to classify the deep moonquake sources using the SVM

Kodai Kato^{*1}, Shiori Kikuchi^{*1}, Ryuhei Yamada^{*2}, Yukio Yamamoto^{*3}, Masaharu Hirota^{*4},
Shohei Yokoyama^{*5}, Hiroshi Ishikawa^{*1}

Abstract

NASA had obtained the moonquake data for about 7 years. The data is available to study the lunar internal structure and the focal mechanisms of moonquakes. Classification of sources of the deep moonquakes is one of important issues. The conventional method to classify deep moonquake sources is mutual comparison among waveforms. Recent machine learning approach enables us to improve the detection of moonquake, and classification of the sources. In this paper, we investigate the effective features to classify the moonquake sources. As a result, we showed that power spectral density of moonquake, and distance between the moon and the earth are effective features to classify the deep moonquake sources using the machine learning approach.

Key Words: Moonquake, Machine Learning, SVM

概 要

NASA が行ったアポロミッションによって、約 7 年半分に及ぶ月地震データが収集された。これらのデータは、月の内部構造の推定や、月震の発生メカニズムの解明において非常に重要であり、現在も解析が行なわれている。月震は、その発生要因から、深発月震、浅発月震、熱月震、隕石衝突などに分類されており、検出された月震の約半数は深発月震である。月震解析のひとつに、深発月震の震源分類がある。深発月震の震源分類は従来、波形の相互比較から分類が行われてきた。近年の研究によって、これまでの月震波形の検出手法や、震源分類の分類基準を、機械学習を用いて改良可能である事が示唆されている。震源の分類基準を改良するためには、それぞれの震源を分類するのに有効な特徴量を発見することが必要であるが、これらは従来、専門家の知見に基づくものであった。そこで、本研究では、機械学習を用いて、半自動的に震源を分類するのに有効な特徴量を発見する。結果として、月震のパワースペクトルや、月-地球間の距離が震源に適した特徴量であることを示し、深発月震の分類基準を再考するための特徴量発見に機械学習を用いることの有用性を示した。

doi: 10.20637/JAXA-RR-16-007/0004

^{*} 平成 28 年 11 月 24 日受付 (Received November 24, 2016)

^{*1} 首都大学東京大学院システムデザイン研究科 (Graduate School of System Design, Tokyo Metropolitan University)

^{*2} 国立天文台 RISE 月惑星探査検討室 (Research of Interior Structure and Evolution of solar system bodies, National Astronomical Observatory of Japan)

^{*3} 宇宙航空研究開発機構宇宙科学研究所 (Institute of Space and Astronautical Science, Japan Aerospace Exploration Agency)

^{*4} 大分工業高等専門学校情報工学科 (National Institute of Technology, Oita College)

^{*5} 静岡大学情報学部 (Faculty of Informatics, Shizuoka University)

1 はじめに

NASA が行った月への有人宇宙飛行計画であるアポロミッションによって、月の観測装置である、Apollo Lunar Surface Experiments Package (ALSEP) が月面に設置された。その ALSEP に含まれる地震計を用いて、Passive Seismic Experiment (PSE) と呼ばれる、月の地震（以後、本論文では月震と表記する）の観測が行われた。観測されたデータは、長周期地震計の 3 軸成分と短周期地震計の上下動成分で構成されている。また、データは全て Web 上で公開され、データの閲覧、取得が可能となっている [11]*1。

PSE によって取得されたデータは、月の研究において貴重なデータである。アポロミッションから約 40 年が経過した現在でも解析が行なわれており、月震の発生メカニズムの解明や、月の内部構造の推定に利用されている [5, 3]。

これまでの解析から、月震は、地球の地震と大きく異なる性質を持つ事がわかってきている。まず、地球での地震の発生要因となるプレート運動は月に存在しない。そして、月震は、発生要因から、深発月震、浅発月震、隕石衝突、熱月震などに分類されている。特に、発生頻度の高い深発月震は、同一の震源から繰り返し発生しており、Weber らは [9]、月-地球間に働く潮汐力と関連した発生原因の解明の研究も行っている。また、同一の震源から発生する月震波形が類似していることが知られている。一般に月震イベントの震源は、月震波が震源から観測点まで到達する時間である走時データから、その位置が決定され、月面に設置された複数の地震計の走時データの時間差を利用することで、深発月震の震源位置が推定された。しかし、振幅の小ささやノイズの影響で、走時データを利用した震源推定ができない月震は、月震波形の類似性から震源位置の違いが判別されている。ALSEP によってデータが取得された 1970 年代当初は、同一の震源から発生する月震波形が類似することに着目し、特に目視で分類を行っていた [6]。

その後、計算機の発達に伴い、Nakamura ら [4] による、月震波形の相互相関係数を用いた最短距離法に基づく階層的クラスタリングを用いた月震分類が行われた。これが、現在の震源分類基準となっている。そして、深発月震の震源には、震源ラベルが付与されている (A1, A6 のように Axx とラベリングされる)。しかし、月震データには、非常に多くのノイズが含まれており、ラベルを付与できないものが存在する。また、誤ったラベルが付与される可能性もある。最近の研究では、月震波形の前処理を改良することで、Bulow ら [1] は、新しい A1 震源の月震を多数発見している。また、Endrun ら [2] は、隠れマルコフモデルを用いた手法で、大規模な月震データベースから、これまでに発見されていなかった新たな月震を検出することに成功している。これらの研究により、これまで発見されていなかった月震の検出、未分類であった月震の分類がなされている。また、後藤ら [10] による月震の可視化システムでは、従来の、月震波形の相互相関係数ではなく、周波数成分を特徴量として、教師なし学習のひとつである SOM による深発月震の分類を可視化した。SOM の結果では、複数の震源が混合したクラスタリング結果が得られている。また、ノイズによって波形の性質が埋もれている波形が集まるセルも見られた。これらの結果から、機械学習により、これまでに目視で発見、分類されていた従来の月震波形の検出手法や、震源の分類基準を改良することが可能であると考えられる。

震源の分類基準に着目すると、現在の分類基準は、波形の相互相関のみを考慮した基準であり、これまでに発見されていない震源に起因した要素が存在する可能性がある。つまり、震源の分類基準を再考するために、震源の分類に寄与する新たな特徴量を発見する必要がある。しかし、従来、これらの特徴量を発見することは、専門家の知見に基づく必要があった。本研究では、機械学習を用いることで、人手に頼らず、震源の分類に寄与する新たな特徴量の発見を行う。新たな特徴を発見することは、専門家の月震解析の補助、震源の分類基準の再考、未知の震源の発見につながる。

本論文は、以下の構成に従う。2 章では、深発月震の震源分類に寄与する特徴量の評価手法について述べる。3 章では、実験結果について述べる。4 章では、考察について述べる。5 章では、本論文のまとめを述べる。

*1 DARTS at ISAS/JAXA <http://darts.jaxa.jp>

2 提案手法

本研究では、機械学習における代表的な教師あり機械学習手法のひとつである Support Vector Machine (SVM)[8] を適用し、特徴量の検証を行う。SVM は、正例、負例の 2 つのクラスのデータを分類する手法であり、与えられたデータを 2 クラスに分離する超平面を生成することで、未知データがどちらのクラスに属しているかを判定する。SVM は 2 クラスの分類しかできないため、SVM で多クラス分類を行う方法として、one-versus-one 法と one-versus-the-rest 法がある。 n クラス分類において、one-versus-one は、ある特定のクラスに入るか、別の特定のクラスに入るかを判定する分類器を $n(n-1)/2$ 個利用するものであり、one-versus-the-rest は、ある特定のクラスに入るか、他の $n-1$ 個のクラスのいずれかに入るかを判定する分類器を n 個利用するものである。

検証方法として、従来の震源ラベルを正解ラベルとし、月震から生成した特徴量に基づいて、従来の震源ラベルが SVM によって再現可能かを確認する。このとき、分類性能の検証には k -分割交差検定を用いる。 k -分割交差検定とは、分類モデルの妥当性を検証する手法である。 k -分割交差検定の手法を以下に示す。

1. データセットを、 k 個に分割する
2. 分割されたデータセットの 1 つめテストデータとして、残りの $k-1$ 個を訓練データとする。訓練データを用いて学習した分類モデルで、テストデータに対する分類性能を評価する。
3. 次に、分割されたデータセットの 2 つめをテストデータとして、残りの $k-1$ 個を訓練データとする。2. と同様に分類性能を評価する
4. この作業を、 k 回繰り返し k 個のテストデータに対する分類性能の平均を最終的な分類性能とする

交差検定を用いる事で、分類モデルの汎用性と妥当性を評価できる。

本論文では、5 交差検定の結果を機械学習の結果とする。その結果として、SVM による分類結果と、従来の震源ラベルの多くが一致していれば、その特徴量による機械学習を適用することで、従来の分類結果をおおよそ再現が可能である。つまり、その特徴量は、震源を分類するのに適切である。

2.1 特徴量

本研究では、機械学習に用いる特徴量として、PSD, J2000 座標系における月-地球間の距離、月の位置 (x , y , z)、月の速度 (v_x , v_y , v_z) を用いる。

PSD は、周波数毎の振幅強度を計算したものである^{*2}。PSD 計算の際に用いたセグメント長は 2048 点とし、窓関数に hanning 窓を用いた。約 15 分間の月震波形データを、セグメント長の半分ずつシフトして計算した PSD の平均値を導出する。導出された PSD をベクトル長が 1 になるよう正規化したものを最終的な特徴量とする。また、月震は、地震と違い、1 時間以上揺れが続くものも存在する。従って、PSD の計算において、震源ごとの特徴の差異が現れやすい時刻を、切り出し位置を推移することで検証する。

J2000 座標系は、地球を中心とし、J2000.0 における平均赤道と平均春分点を基準とした座標系である。ここで、J2000 は 2000 年 1 月 1 日正午であり、その時刻における平均春分点の方向を x 軸とする^{*4}。

J2000 座標系における月-地球間の距離、位置、速度は、月震発生時刻に基づいて、SPICE^{*3}を用いて計算を行う。本論文では、これまでの研究で用いられてきた波形から得られる特徴量を用いずに、震源分類が再現可能であるかを検討する。各特徴ごと、平均値が 0、標準偏差が 1 となるよう正規化し、距離、位置、速度それぞれの組み合わせと分類性能を比較する。

SVM の実装には、scikit-learn[7] の SVC を用いた。多クラス分類法には、one-versus-the-rest を用いた。

表 1 データセットの内訳

震源	A1	A6	A7	A8	A9	A10	A18	A21	A33	合計
データ数	138	24	19	24	51	33	36	26	28	379

また、本論文で扱うデータセットは、震源ごとにデータ数が異なる。SVM の学習の際に、データ数のばらつきは、過学習の発生につながる。これらの不均衡データに対応するため、データ数に応じて自動で重みをつけるよう class-weight パラメータを “balanced” に設定した。また、パラメータについては、各条件ごとに rbf と linear カーネルを用いてグリッドサーチを行い、それぞれ一番分類性能のいいパラメータを選択した。

3 実験結果

3.1 データセット

表 1 に、実験に用いたデータセットに含まれる震源の種類と深発月震のイベント数を示す。本データセットに用いる月震イベントは、アポロ 12 号のミッションで取得されたものである。本データセットは、震源全体の中で比較的月震イベントが多く観測され、解析が進んでいる 9 震源から、人手でノイズの少ない月震を選定している。前処理として、平均引き、トレンド引き、0.3-1.5Hz のバンドパスフィルタ処理、スパイク除去処理を行った長周期地震計の Z 軸成分 (LPZ) のデータを実験に用いた。

また、ALSEP において、月震データの観測は Peaked モードと Flat モードが存在する。これらの観測モードによって、周波数応答が異なる。そのため、本論文では、Peaked モードで観測された波形のみを用いる。

3.2 評価指標

本論文では、震源の分類の性能を評価するための評価指標として、適合率、再現率、F 値を用いる。例として、A1 震源における適合率、再現率、F 値は以下の式で表される。

$$\text{適合率} = \frac{\text{A1 と予測されて正解だった数}}{\text{A1 と予測された数}}$$

$$\text{再現率} = \frac{\text{A1 と予測されて正解だった数}}{\text{全体の A1 の数}}$$

$$F \text{ 値} = \frac{2 * \text{適合率} * \text{再現率}}{\text{適合率} + \text{再現率}}$$

適合率は、分類の正確性を測る指標であり、再現率は、分類の網羅性を測る指標である。適合率と再現率はトレードオフであるため、適合率と再現率の調和平均である F 値は、適合率と再現率のバランスを考慮した評価指標である。

3.3 PSD を用いた場合の結果

本節では、PSD が、震源を分類するのに有効な特徴量であるかを評価するための実験結果について述べる。PSD は約 15 分間の月震波形から計算を行う。月震波形の切り出し開始点は、各月震の P 波到達時刻を基準

*2 <http://www.cygres.com/OcnPageE/Glosry/SpecE.html>

*3 <http://naif.jpl.nasa.gov/naif>

*4 <http://www2.nc-toyama.ac.jp/mkawai/lecture/radionav/tfcoordinate/coordinate.html>

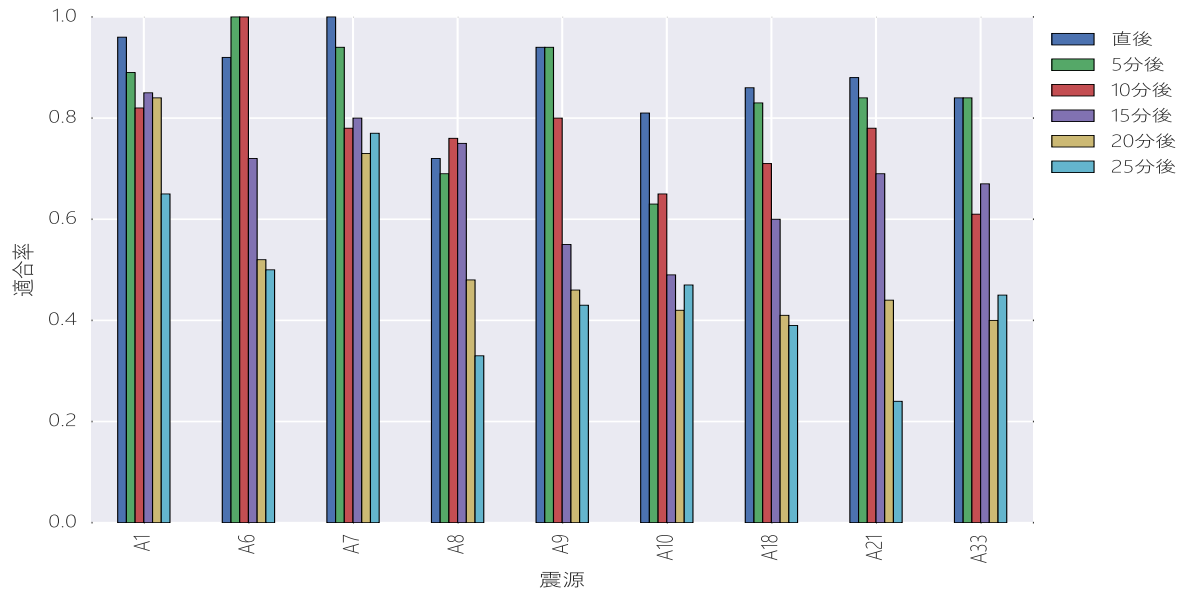


図1 P波到達時刻からの切り出し時刻と適合率

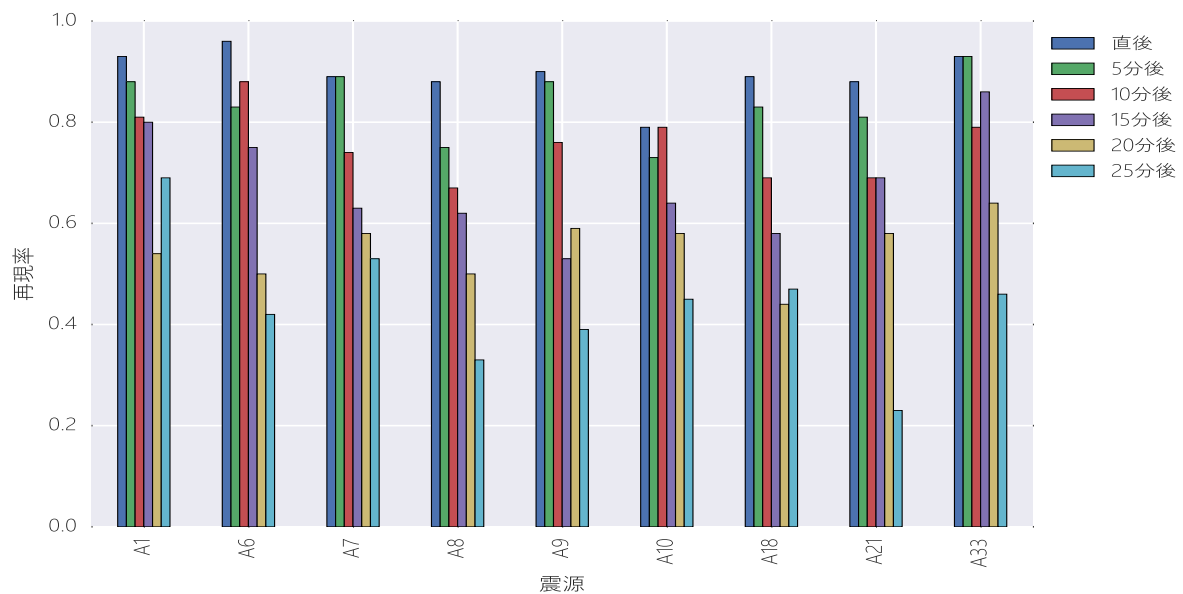


図2 P波到達時刻からの切り出し時刻と再現率

とし、5分毎に切り出し開始点をずらした、切り出し毎のSVMの分類性能を計算し、それぞれの適合率、再現率、F値で評価する。

図1, 2, 3は、PSDを特徴量としたときに、P波到達時刻からの切り出し開始時刻の推移に伴う、適合率、再現率、F値の推移を示したものである、横軸は震源名、縦軸は適合率、再現率、F値の値である。

図1, 2, 3より、適合率、再現率、F値の全ての評価指標において、P波到達直後から15分間の波形のPSDを特徴量とした場合の分類性能が高く、P波到達時刻からの差が大きくなるにつれて、SVMの分類性能が低下していることが示された。

また、P波到達直後から15分間の波形を使用した場合の適合率、再現率、F値の全体平均はそれぞれ0.91, 0.90, 0.90である。

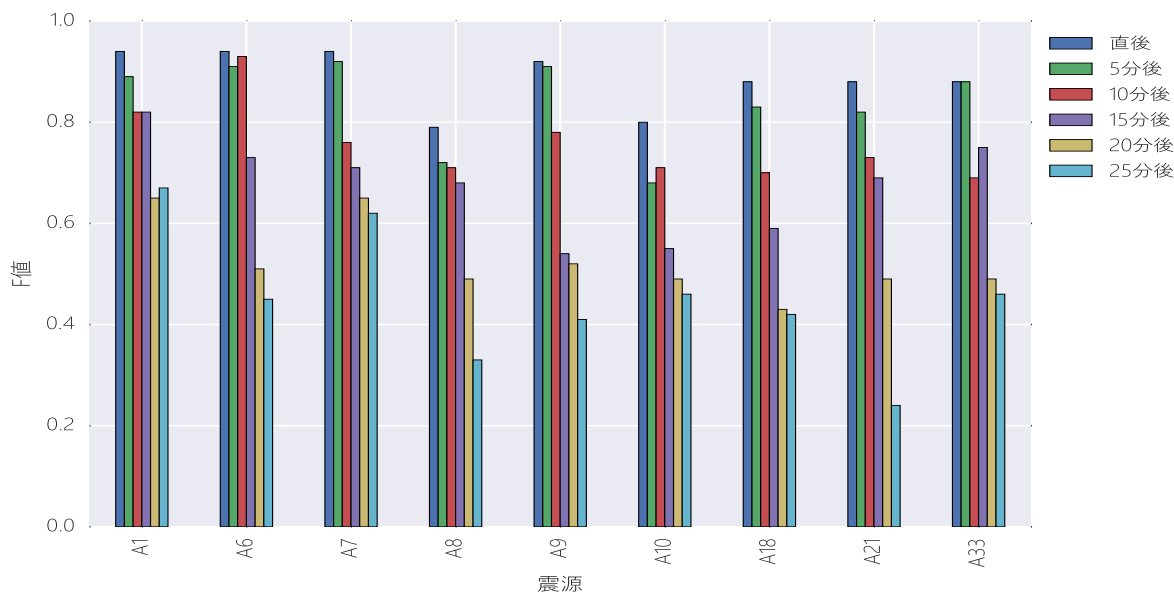


図3 P波到達時刻からの切り出し時刻とF値

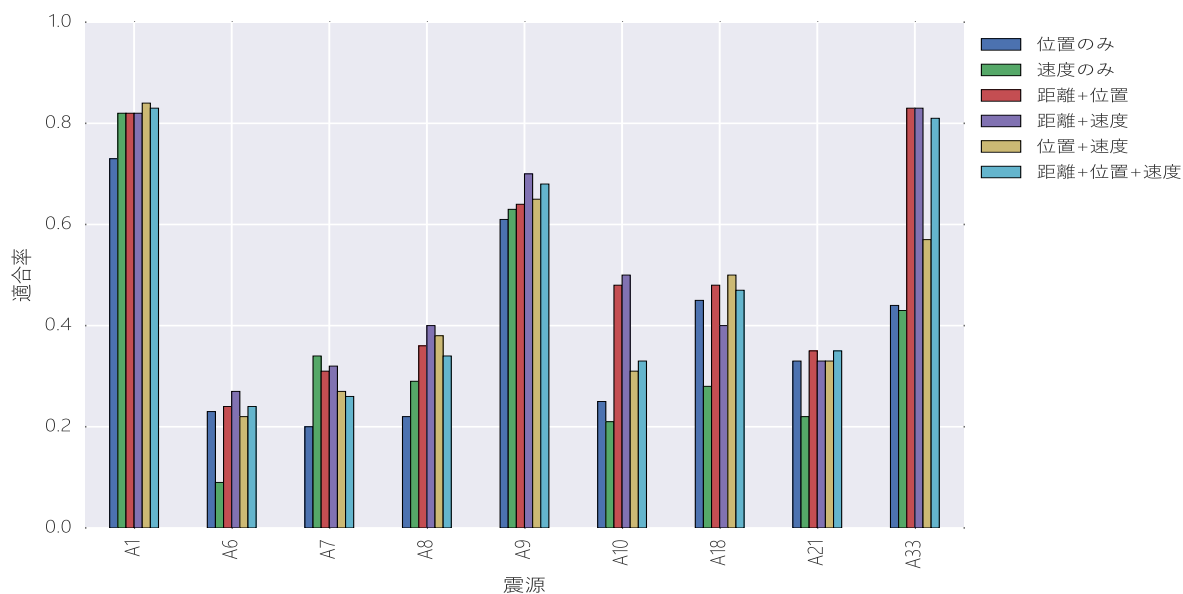


図4 距離、位置、速度それぞれの組み合わせと適合率

3.4 月-地球間の距離、位置、速度を用いた結果

図4, 5, 6は、月-地球間の距離、位置、速度を特徴量としたときの、震源別のSVMの適合率、再現率、F値を示したものである。横軸は震源であり、縦軸はそれぞれの分類性能である。はじめに、特徴量として距離、位置 (x, y, z 全てを特徴量とした3次元ベクトル), 速度 (vx, vy, vz 全てを特徴量とした3次元ベクトル) の3要素を特徴量とし、位置、速度、距離+位置、距離+速度、位置+速度、距離+位置+速度の

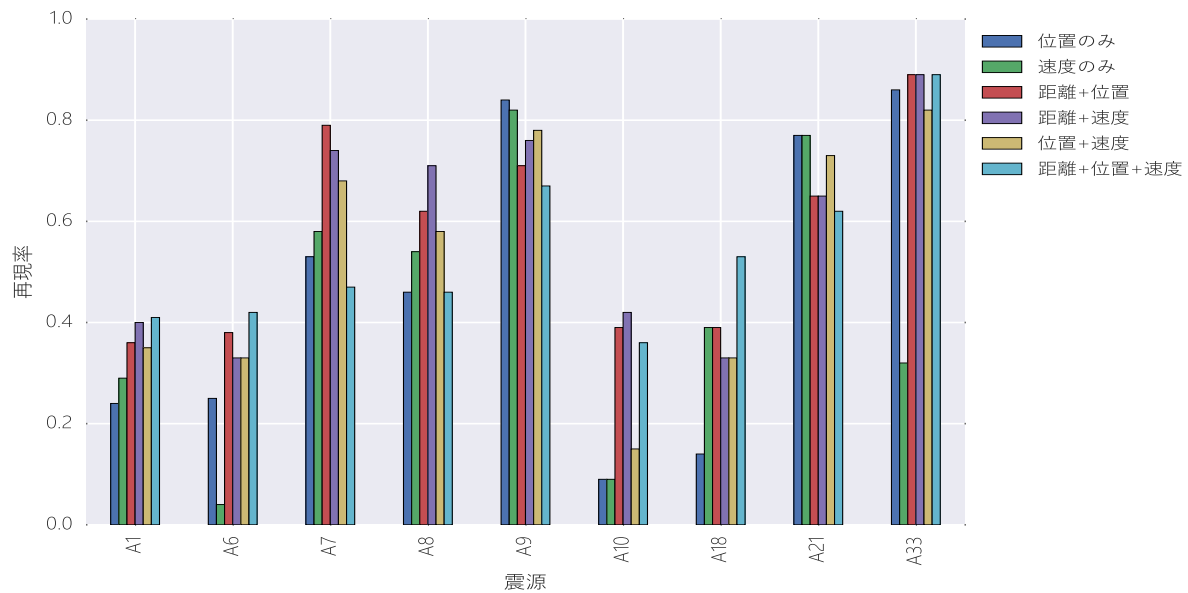


図5 距離, 位置, 速度それぞれの組み合わせと再現率

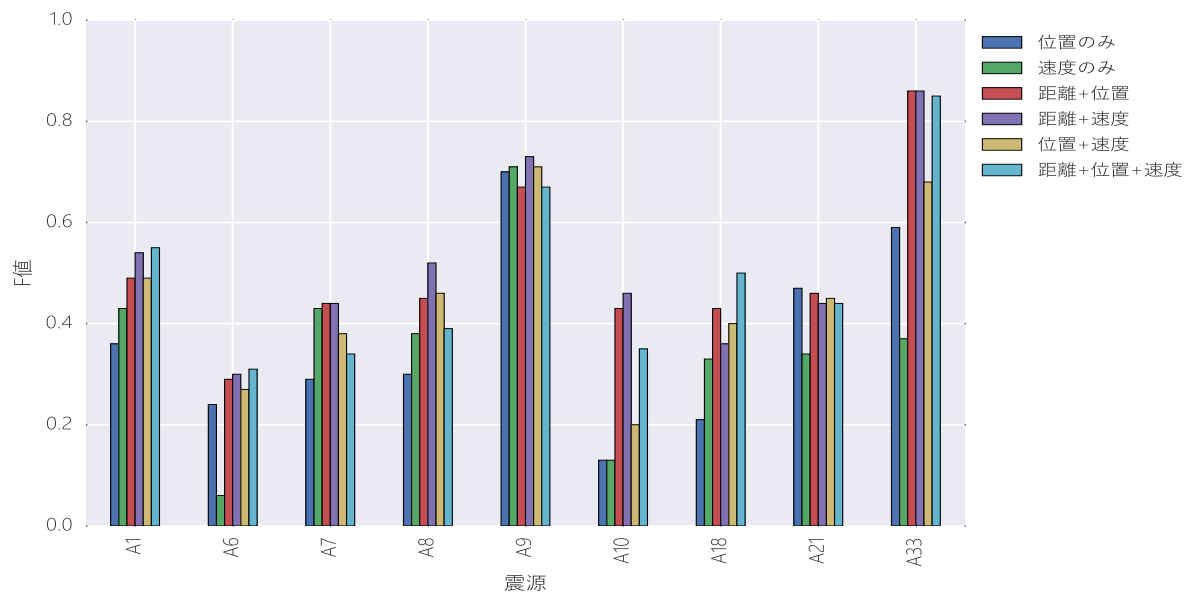


図6 距離, 位置, 速度それぞれの組み合わせと F 値

6 種類の組み合わせにおいて, それぞれ分類性能を求めた.

図4より, A1, A9, A33 は, 距離を特徴量のひとつとして用いた場合に, 他の震源に比べて適合率が高いことが示された. また, A1, A9 は, 位置や速度のみを用いた特徴量において, 他の震源に比べて適合率が高い. 結果として, これらの特徴量を用いる事で, 一部の震源について分類に有効な特徴量である事が示された.

図7は, 月-地球間の距離, x座標の位置, y座標の位置, z座標の位置, x座標の速度, y座標の速度, z座標の速度の7つの特徴量から, 2つを特徴量として組み合わせた場合の, SVMの震源全体における適合率, 再現率, F値を示したものである. 横軸は特徴量の組み合わせであり, 縦軸はそれぞれの分類性能である. 図4, 5, 6の結果で, 距離を特徴量とした場合に分類性能が高いことが示されたが, 図7で, 距離+x軸の速度を特

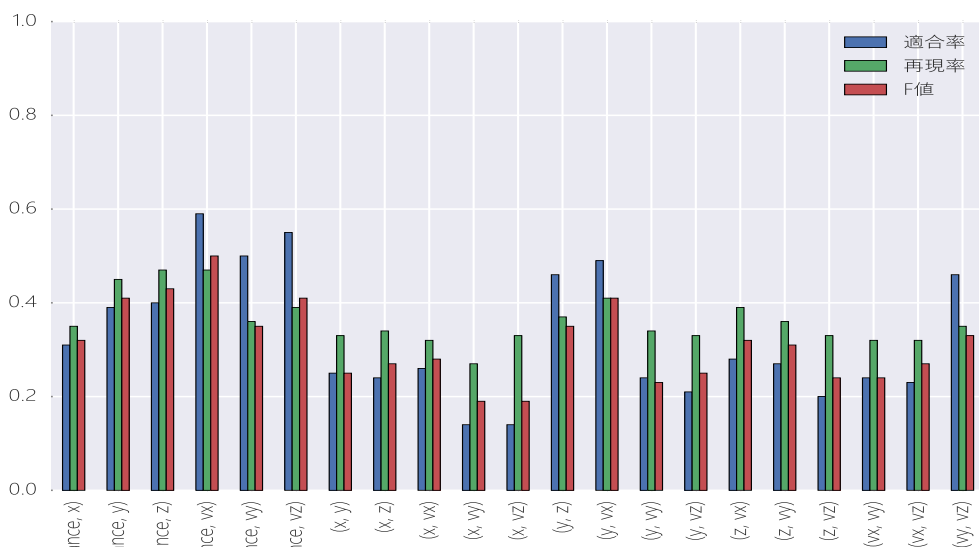


図7 距離、位置 (x, y, z)、速度 (vx, vy, vz) から2つを特徴量としたときの分類性能

表2 距離 +x 軸の速度を特徴量としたときの SVM の分類性能

震源	適合率	再現率	F 値
A1	0.81	0.38	0.51
A6	0.17	0.29	0.22
A7	0.08	0.21	0.12
A8	0.29	0.38	0.33
A9	0.69	0.73	0.70
A10	0.45	0.39	0.42
A18	0.43	0.64	0.52
A21	0.38	0.38	0.38
A33	0.81	0.89	0.85

表3 距離 +z 軸の速度を特徴量としたときの SVM の分類性能

震源	適合率	再現率	F 値
A1	0.86	0.31	0.46
A6	0.23	0.50	0.32
A7	0.13	0.42	0.20
A8	0.34	0.71	0.46
A9	0.38	0.10	0.16
A10	0.09	0.03	0.05
A18	0.34	0.39	0.36
A21	0.25	0.62	0.36
A33	0.53	0.68	0.59

微量としたときに最も分類性能が高いことが示された。また、y 軸の位置 +z 軸の位置、y 軸の位置 +x 軸の速度、y 軸の速度 +z 軸の速度を特徴量としたとき、他の位置、速度の組み合わせと比較して適合率が高い。

図7において、分類性能が高い組み合わせである距離 +x 軸の速度、距離 +z 軸の速度、y 軸の位置 +z 軸の位置、y 軸の位置 +x 軸の速度における震源別の分類性能を表2, 3, 4, 5に示す。

4 考察

4.1 PSD について

本節では、3.3 節の結果について述べる

PSD を特徴量とした機械学習を用いて、震源ラベルを十分に再現できていることから、PSD は震源分類に

表 4 y 軸の位置 +z 軸の位置を特徴量としたときの SVM の分類性能

震源	適合率	再現率	F 値
A1	0.75	0.28	0.41
A6	0.11	0.08	0.09
A7	0.19	0.74	0.30
A8	0.17	0.38	0.23
A9	0.62	0.80	0.70
A10	0.23	0.09	0.13
A18	0.30	0.28	0.29
A21	0.33	0.77	0.46
A33	0.12	0.04	0.06

表 5 y 軸の位置 +x 軸の速度を特徴量としたときの SVM の分類性能

震源	適合率	再現率	F 値
A1	0.69	0.31	0.43
A6	0.00	0.00	0.00
A7	0.21	0.79	0.33
A8	0.21	0.38	0.27
A9	0.60	0.71	0.65
A10	0.38	0.30	0.34
A18	0.50	0.39	0.44
A21	0.25	0.35	0.29
A33	0.41	0.64	0.50

有効な特徴量であることがわかる。また、分類性能を低下させる要因として、

- 特徴量と手法の組み合わせが適切でない
- 従来の分類基準と SVM による分類基準が一部異なる
- PSD で捉えることができない特徴が月震に含まれている

と考えられる。また、PSD において、深発月震の P 波到達直後が、最も震源ごとに特徴の差が現れていると推測される。これは、時間の経過とともに、月表面で揺れが散乱するため、震源ごとの特徴が埋もれてしまうためと考えられる。

4.2 距離、位置、震源について

本節では、3.4 節の結果について述べる。

図 4, 5, 6 より、震源ごとに分類性能にばらつきがあり、A1, A9, A33 は他の震源に比べて、分類性能が高い。距離 + 位置、距離 + 速度の比較を行う。A1, A8, A9, A10 は、距離 + 速度を特徴量とした時の分類性能が、距離 + 位置の分類性能と比べて高い。しかし、A18, A21 は、距離 + 位置の分類性能が距離 + 速度の分類性能に比べて高い。これらの結果から、位置が大きく影響する震源と、速度が大きく影響する震源が異なり、A1, A8, A9, A10 は速度に特徴が現れる震源であり、A18, A21 は、位置に特徴が現れる震源であることが推測される。また、すべての震源は、距離を特徴量のひとつとした時に分類性能が高い。

- 月-地球間の距離そのものが月震発生に寄与している
- 月-地球間の距離と、位置もしくは速度との関係に、月震発生に寄与しているものが存在する

といった推測が可能である。この事は、先行研究 [9] で述べられているよう、深発月震が潮汐力の寄与で発生している事に関連していると考えられる。特に、A1, A9, A33 は、月-地球間の距離や、位置、速度が大きく関わっていることが推測される。しかしながら、4.1 節の分類性能と比較すると、距離や速度、位置のみを用いた SVM の分類性能は十分であるとは言えない。

ここで、距離を特徴量とした場合に分類性能の高さに着目し、図 8 に、震源別の距離の分布を示す。横軸は月-地球間の距離であり、縦軸は、各震源ごと、36 万キロメートルから 41 万キロメートルまでを 5000 メートルずつ分割した際のイベント数を示している。分類性能の高かった A33 の月震は、40 万キロメートル付近の限られた距離で発生している。一方、その他の震源の距離の分布において、震源ごとの距離の分布に多少の差

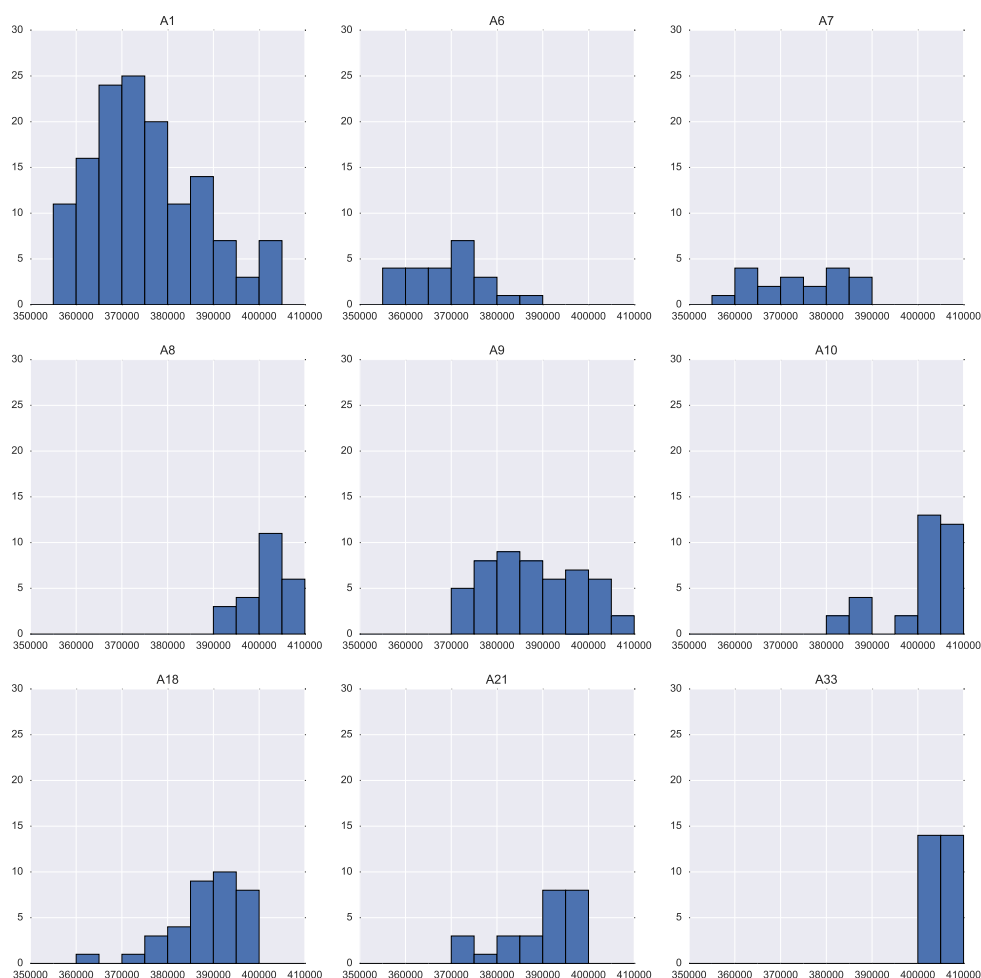


図8 震源別の距離の分布

があることは推察できるが、目視によって震源固有の特徴を発見することは難しい。しかし、機械学習の結果では、距離を特徴量として用いたときに分類性能が高くなっていることから、分類性能に影響する要因として、距離、位置の x , y , z 成分、速度の x , y , z 成分の組み合わせによる震源との相関が存在する可能性があり、震源ごとの相関が現れる要因の分析を行う必要がある。

そこで、震源ごとの要因の分析を行う前段階として、図7の分類性能を比較すると、特徴量の組み合わせによって、分類性能が変化していることがわかる。全体の分類性能が最も高い組み合わせである表2の結果では、A9のF値が0.70、A33のF値が0.85である。また、A1の適合率は0.81である。しかし、表3では、A1の適合率は0.86であるが、A9のF値は0.16であり、A33のF値は0.59である。これは、距離+ z 軸の速度を特徴量としたときに、A9やA33固有の特徴を表現できないと考えられる。また、表4では、A9の分類性能が最も高く、F値が0.70であるのに対し、A33のF値は0.06である。従って、A9は表4の特徴量の組み合わせでも震源の特徴を表現できているのに対し、A33は表4の特徴量の組み合わせでは震源の特徴を表

現できないことが示されている。表 5 では、A33 の分類性能が、表 2 の分類性能と比べて向上していることから、A33 は、x 軸の速度に震源の特徴を表現出来る要因が含まれていることが推測される。これらの結果から、震源によって最適な特徴量が異なると考えられ、今後、震源ごとの有用な特徴量の分析を行う必要がある。

5 まとめ

本研究では、機械学習を用いて、震源を分類できる深発月震の特徴量の半自動的な発見を行った。手法として、SVM を用いた教師あり学習とその精度を用いて、震源分類結果の再現が可能かを考察した。その結果、月震は開始直後に震源ごとの特徴が現れる傾向が見られること、月震の PSD、月-地球間の距離は震源分類に有効な特徴量であり、月-地球間の距離は特定の震源の推定に利用が可能であることを示した。これらの特徴量を考慮した月震分類基準を作成することで、未だに分類がなされていない深発月震の分類や、未知の震源の発見に貢献できると考えられる。本論文では、先行研究で示されていたスペクトルや、潮汐力と関係がある月-地球間の距離を用いたが、今後は、特徴量間の相関の発見や、これまで言及されていない太陽や他の惑星等の様々な要因を考慮した分析を行う予定である。

謝辞

本稿の執筆にあたって多数の有益な助言を頂いた、首都大学東京特任助教の莊司 慶行氏に感謝いたします。本研究 (の一部) は傾斜的研究 (全学分) 学長裁量枠戦略的研究プロジェクト戦略的研究支援枠「ソーシャルビッグデータの分析・応用のための学術基盤の研究」による

参考文献

- [1] R. C. Bulow, C. L. Johnson, and PM Shearer. New events discovered in the apollo lunar seismic data. *Journal of Geophysical Research: Planets (1991–2012)*, Vol. 110, No. E10, 2005.
- [2] B. Knapmeyer-Endrun and C. Hammer. Identification of new events in apollo 16 lunar seismic data by hidden markov model-based event detection and classification. *Journal of Geophysical Research: Planets*, Vol. 120, No. 10, pp. 1620–1645, 2015.
- [3] P. Lognonné, J. Gagnepain-Beyneix, and H. Chenet. A new seismic model of the moon: implications for structure, thermal evolution and formation of the moon. *Earth and Planetary Science Letters*, Vol. 211, No. 1, pp. 27–44, 2003.
- [4] Y. Nakamura. New identification of deep moonquakes in the apollo lunar seismic data. *Physics of the Earth and Planetary Interiors*, Vol. 139, No. 3, pp. 197–205, 2003.
- [5] Y. Nakamura, G. V. Latham, and H. J. Dorman. Apollo lunar seismic experiment—final summary. *Journal of Geophysical Research: Solid Earth (1978–2012)*, Vol. 87, No. S01, pp. A117–A123, 1982.
- [6] Y. Nakamura, G. V. Latham, H. J. Dorman, and J. Harris. Passive seismic experiment long-period event catalog. *Galveston Geophysics Laboratory Contribution*, Vol. 491, , 1981.
- [7] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, Vol. 12, pp. 2825–2830, 2011.
- [8] Vladimir Naumovich Vapnik and Vlamimir Vapnik. *Statistical learning theory*, Vol. 1. Wiley New York, 1998.
- [9] RC Weber, BG Bills, and CL Johnson. Constraints on deep moonquake focal mechanisms through analyses of tidal stress. *Journal of Geophysical Research: Planets*, Vol. 114, No. E5, 2009.
- [10] 後藤康路, 山田竜平, 山本幸生, 横山昌平, 石川博. 階層型 SOM に基づいた大規模月地震波形の可視化システム. 第6回データ工学と情報マネジメントに関するフォーラム, E1-4, 2014.
- [11] 山田竜平, 山本幸生, 桑村潤, 中村吉雄. アポロ月地震データ公開システムの開発. 宇宙科学情報解析論文誌宇宙航空研究開発機構研究開発報告, No. 1, pp. 121–131, 2012.

深発月震源分類に適した機械学習法の研究

菊池 栞^{*1} 加藤 広大^{*1} 山田 竜平^{*2} 山本 幸生^{*3} 廣田 雅春^{*4}
横山 昌平^{*5} 石川 博^{*1}

Study of the suitable machine learning techniques for classification of deep moonquake sources

Shiori Kikuchi^{*1}, Kodai Kato^{*1}, Ryuhei Yamada^{*2}, Yukio Yamamoto^{*3}, Masaharu Hirota^{*4},
Yokoyama Syohei^{*5}, Hiroshi Ishikawa^{*1}

Abstract

Deep moonquake occurs at depth of about 1000km, and it is most frequent lunar seismic event. Due to considerable noises and low amplitudes in deep moonquake waveforms, we have some difficulties to locate the sources when we use conventional method. In this paper, we have investigated suitable machine learning methods to classify the deep moonquake sources based on similarity among the waveforms. The machine learning-based method is more useful to reduce the computational time to classify the sources compared with the conventional method, and it has also advantage to apply the some types of effective features such as power spectral density for the classification. We compared performances of some machine learning methods in order to find the suitable method. Then, the deep events misclassified from some machine learning methods has been also investigated, and we have analyzed causes of misclassification (e.g., mislabel, outlier) for improvement of the classification. It can be expected that the suitable machine learning method will enable us to create a new moonquake event catalog and identify the unclassified events.

Key Words: Moonquakes, Machine Learning, Neural Network

概 要

深発月震とは、深さ約 1,000km で発生し、観測された数が最も多い月震である。これまで深発月震の震源の推定が行われてきたが、深発月震の波形にノイズが多く含まれることと、振幅が小さいために、これまでの手法では、震源の推定が困難な月震も多く存在している。そこで、我々は、波形の類似性に基づいて、月震の震源推定を行うための、最適な機械学習法について研究を行なった。震源の分類に機械学習を用いることで、これまでの推定とは異なる新たな特徴を用いた震源の推定や、推定を行う際の 解析処理時間の削減が期待できる。本研究では、深発月震の震源を高精度に推定する手法を発見するため、複数の機械学習の推定の性能を比較し、検討する。その後、分類性能の向上のために、誤って推定されたイベントに着目し、複数の手法や特徴量による震源の推定結果の比較を行なった。その結果から、月震ラベルの違いや、震源における外れ値などを視野に入れながら、誤って推定される原因の分析を行った。本研究で選択した最適な機械学習法による分類を行うことで、新しい月震イベントカタログの作成や、これまで未分類であったイベントの同定が可能となることが期待される。

doi: 10.20637/JAXA-RR-16-007/0005

^{*} 平成 28 年 11 月 24 日受付 (Received November 24, 2016)

^{*1} 首都大学東京大学院システムデザイン研究科 (Graduate School of System Design, Tokyo Metropolitan University)

^{*2} 国立天文台 RISE 月惑星探査検討室 (Research of Interior Structure and Evolution of solar system bodies, National Astronomical Observatory of Japan)

^{*3} 宇宙航空研究開発機構宇宙科学研究所 (Institute of Space and Astronautical Science, Japan Aerospace Exploration Agency)

^{*4} 大分工業高等専門学校情報工学科 (National Institute of Technology, Oita College)

^{*5} 静岡大学情報学部 (Faculty of Informatics, Shizuoka University)

1 はじめに

NASA のアポロミッションによって、Apollo Lunar Surface Experiments Package (ALSEP) と呼ばれる観測装置群が設置された。ALSEP を用いた実験のひとつである Passive Seismic Experiment (PSE) は、ALSEP に含まれる地震計を用いて月面上で月震の連続観測を行う実験である。この実験では、5つの地震計が月面に設置されており、その内、1-4台の地震計を使用して約7年半に渡る月震の記録が取得された。これらのデータは、長周期地震計の3軸成分と短周期地震計の上下動成分で構成されている。

PSE によって得られた月震の観測データは、現在も解析が行われており、月震の発生原因、活動度、および月の内部構造の予測について多くの知見が得られている [10],[8]。月震は、発生深度や、発生要因をもとに、熱月震、人工衝突の信号、隕石衝突、浅発月震、および深発月震の5種類に分けられている。

深発月震は、深さ約 1,000km で発生する月震であり、これまでに観測された数が最も多い月震である。また、深発月震は、同一震源から周期的に発生し、同じ震源で発生した月震の波形^{*6} は類似していることがわかっている [7],[2]。観測された深発月震の一部には震源を表すラベルが付与されており、月震イベントカタログ^{*7} として公開されている [17] ^{*8}。ラベルは、A1, A2, ... と名付けられており、109 のグループに分類されている。特に Apollo では、複数の点で観測されたイベントが少なく、波形の振幅も小さくノイズも多いため、カタログでは、波形の類似性を用いた震源推定とラベルの付与も行われている [9]。一方、これまでの推定手法では、そのノイズの多さなどから、既存の震源に分類することが困難な月震も多数存在する。この問題に対して、震源が未知の深発月震の震源推定には、従来の方法で見えなかった、震源ごとに共通する新たな特徴の発見や、新たな分類手法が必要となる。そこで、本研究では、深発月震源の新たな分類手法として、機械学習に着目する。特に、ラベルが付与されていない大量のデータに人手でラベルを付与することは、解析処理時間が膨大にかかってしまうが、機械学習を用いることで、それが自動化できると考えられる。そこで、我々は、月震の震源推定を多クラス分類問題と捉え、従来の分類結果を学習データとして、機械学習の1つである教師あり学習を行うことで、自動的に深発月震の震源を分類し、分類結果をラベルとして深発月震に付与することを目指す。クラス分類とは、幾つかの震源を選択し、その震源で発生したイベントを学習器に学習させ、震源が未知であるイベントを、選択された震源のうちどの震源に属するのが適切かを予測することである。

教師あり学習を深発月震に適用した研究に、後藤ら [16] の研究が挙げられる。後藤ら [16] の研究では、Support Vector Machine (SVM)[13] を深発月震分類に用いており、その有効性が示されている。しかし、SVM 以外での分類が行われていないため、SVM 以外に深発月震分類に適した手法が存在する可能性がある。そこで本研究では、代表的な教師ありの学習の手法を用いて深発月震の分類を行い、分類性能を比較する。比較する手法は、Neural Network[5], SVM, ロジスティック回帰 [3], アンサンブル学習から Random Forest[1] とブースティングの5つを用いた。SVM では Linear カーネルと、RBF カーネルを、ブースティングでは AdaBoost[4] を用いた。そして、手法ごとに分類性能を比較することで、深発月震に適した機械学習手法を検討する。また、分類性能の向上を目指すために、本研究では誤分類されたイベントに着目した。誤分類される原因には、手法や特徴量が、震源の特徴をうまく捉えることができていないことや、ラベルの違いなどが考えられる。このラベル間違いが疑われるイベントをデータセットから除くことで、分類性能が向上すると考えられる。そこで、複数の手法や特徴量で、誤分類されるイベントを比較し、誤分類される原因について考察を行なった。手法は分類性能の高い手法を用いて、特徴量は分類性能の比較で用いた特徴量と、時系列の月震波形を用いた。

本論文の構成は次の通りである。2章では、機械学習手法の分類性能を比較する手法とその結果を述べる。3章では、誤分類となる確率の高いイベントの決定を行う手法と、結果を述べる。4章では、2章と3章の結

表 1 実験に用いた震源ごとの月震イベントの数

震源名	A1	A6	A7	A8	A9	A10	A18	A20	A21	A33
イベント数	173	32	29	39	62	35	36	48	26	28

果から得られた考察を述べる。5 章では、本論文のまとめと今後の課題について述べる。

2 機械学習手法の分類性能の比較

本章では、月震源分類における機械学習手法の分類性能の比較を行う。

2.1 実験手法

本章では、実験に用いるデータセット、特徴量と、比較する代表的な教師あり学習を行う、機械学習の手法と、その評価方法について述べる。

2.1.1 データセット

本論文ではイベントの数が特に多い深発月震のみを扱う^{*9}。月震データは、X 軸、Y 軸、Z 軸の 3 方向の成分で記録されている。本研究では、長周期地震計のデータを用いて、X 軸、Y 軸、Z 軸の月震の 3 成分をそれぞれ LPX, LPY, LPZ として表記する。また、月震の波形は、多くのノイズを含むため、本研究では、前処理を適用した波形を用いて震源の分類を行う。前処理として、平均引き、トレンド引きを行った後、0.3-1.5Hz のバンドパスフィルタを適用し、スパイク除去処理を行った。1 つの波形に対して、3 成分ごとに前処理を行った後の、月震波形の例を図 1 に示す。図 1 において、横軸は時間を示し、縦軸は振幅を示す。図 1 より、1 つのイベントにおいて、成分の違いによって波形が異なることがわかる。本研究では、筆者らが行った先行研究 [15] より、LPZ を用いた場合に各分類手法の分類性能が高かったため、本研究でも LPZ のデータを使用する。

本研究では、最も観測期間が長く、イベント数が多い Apollo12 号地点で観測された月震データのみを用いる。分類の対象とする震源は、従来手法でラベルが付与されているイベント数が 20 以上ある 10 の震源を用いた。各震源のイベント数を表 1 に示す。また、震源によってイベント数が異なるが、筆者らの先行研究 [15] より、イベント数にばらつきがある場合でも分類性能が高かったため、イベント数を均衡にするなどの前処理は行わない。

本研究では、分類に使用するイベントの、継続する長さを、筆者らの先行研究 [15] において分類性能が高く、かつデータ量が少なくなる 15 分とした。

2.1.2 比較する手法

イベントの特徴を学習するために用いる、代表的な教師あり学習を行う、機械学習手法について述べる。

Neural Network

Neural Network は、様々な研究者達によって構築された機械学習のアルゴリズムである [5]。2012 年

^{*6} 以後、特に明示のない場合は、波形は月震の波形を表す。

^{*7} 以下、カタログとする。また、このカタログに合わせて、発生した、深発月震が強く疑われる個々の現象を、イベントと呼称する。

^{*8} DARTS as ISAS/JAXA <http://www.darts.isas.jaxa.jp>

^{*9} 以後、特に明示のない場合は、深発月震を月震と呼称する。

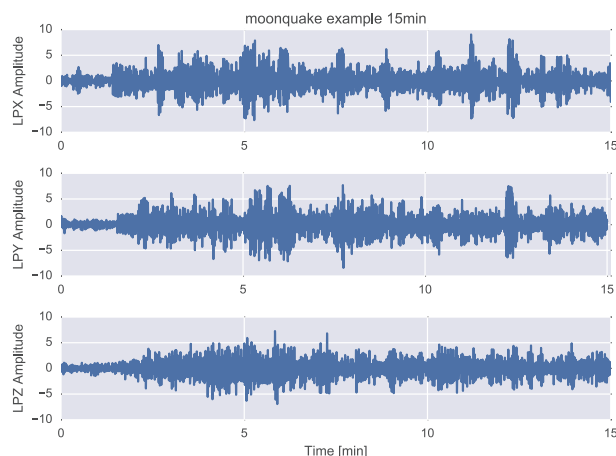


図1 月震波の例

の画像コンテスト^{*10}において、Hinton らがこの手法を用いて、その有用性に注目が集まった。本研究では、Neural Network の一種である、多層パーセプトロンを用いる。入力データを各ニューロンで、重みや活性化関数を用いながら値の変更を行っていく。そして、出力層での出力と、正解データとを比較し、その誤差を逆伝播させることで重みを更新し、学習を行う。多層パーセプトロンは、入力層と出力層の2層のみでは線形分類と等価であるが、中間層を追加すると、非線形分類が可能になる。本研究は、多クラス分類を行うため、出力層の誤差関数にはクロスエントロピーを用い、活性化関数には、ソフトマックス関数を用いた。ソフトマックス関数とは、出力の値の総和を1とすることで、分類されるクラスを確率分布にするための関数である。出力層のニューロン数は、本研究で行う分類のクラス数である10とした。

Random Forest

アンサンブル学習の一種である Random Forest は、2001 年に Leo Breiman が提案した機械学習のアルゴリズムである [1]。Random Forest は、ブートストラップサンプルから複数の決定木で学習する。そして、それぞれの決定木による分類結果を統合し、最終的な分類結果とする。この手法の特徴として、学習が高速であることや、データの欠損や偏りがある場合でも性能の低下が少ないことが挙げられる。

SVM

SVM は 1995 年頃に Vladimir N. Vapnik が提案したパターン識別用の機械学習のアルゴリズムである [13]。与えられた学習データから識別境界面を設定し、識別境界面に最も近いデータ点の識別境界面からの距離を最大化し、未知データがどちらに分類されるかを判定する。この手法の特徴として、データの次元数が大きい場合でも分類の性能が高いことや、設定するパラメータが比較的少ないことが挙げられる。本研究では、線形分類を行うために用いられる、線形カーネルと、非線形分類を行うために用いられる、RBF カーネルをそれぞれ用いた。本論文では、それぞれを SVM-Linear と SVM-RBF と表記する。

ロジスティック回帰

ロジスティック回帰は David Cox が提案したアルゴリズムである [3]。ロジスティック回帰は、ロジスティック関数を用いて、与えられた説明変数を基に、ある要素がどのカテゴリに属するかを条件付き確

^{*10} ImageNet Large Scale Visual Recognition Challenge
<http://image-net.org/challenges/LSVRC/2012/index>

率を予測するアルゴリズムで、分類モデルとして利用することができる。出力が 0 から 1 の間であること以外は線形回帰と同じ振る舞いをするが、ロジスティック回帰では閾値を設けてクラスの予想を行っている。

AdaBoost

ブースティングの一種である AdaBoost は、Yoav Freund と Robert E. Schapire が提案した機械学習のアルゴリズムである [4]。ブースティングとは、学習データを用いて学習を行い、その学習結果を踏まえて逐次に重みの調整を繰り返すことで複数の学習結果を求め、その結果を組み合わせることで性能を向上させる手法である。そのブースティングの代表的な手法である AdaBoost は、分類精度が 50 % を少し超える程度の、弱い分類器を大量に用意し、その 1 つ 1 つに重みを付けることで強い分類器を作成し、分類を行う。本研究では弱い分類器として決定木を用いた。また、本研究ではマルチクラス分類を行うため、AdaBoost のアルゴリズムとして、SAMME.R[14] を用いた。この手法の特徴として、ノイズが多いデータや異常値に影響を受けてしまうものの、他のブースティング手法より比較的過学習が起きにくいことが挙げられる。

2.1.3 評価指標

手法の分類性能の評価を行うために、交差検定を行なう。まず使用するデータを学習データと評価データに分割する。学習データを用いて学習器のモデルを作成し、そのモデルを用いて評価データの分類を行う。その分類結果を用いて、本研究では、評価の指標として、適合率と再現率の調和平均である F 値を用いる。適合率とは、分類の結果の中に、どれだけ正解が含まれているかの割合を指し、再現率とは、分類の結果として出てくるべきもののうち、実際正しく分類されたものの割合を指す。この適合率と再現率はトレードオフの関係のため、両方の値をバランスよく評価するために、調和平均を計算した F 値を求め、それを用いて分類性能の評価を行う。例として、イベントを分類した際の、A1 の適合率、再現率、F 値は以下の式で定義される。

$$\begin{aligned}\text{適合率} &= \frac{A1 \text{ と予測して正解した数}}{A1 \text{ と予測した数}} \\ \text{再現率} &= \frac{A1 \text{ と予測して正解した数}}{\text{全体の } A1 \text{ の数}} \\ F \text{ 値} &= \frac{2 * \text{再現率} * \text{適合率}}{\text{再現率} + \text{適合率}}\end{aligned}$$

他の震源においても以上の式を用いて、F 値を計算する。

2.1.4 本研究で用いる特徴量

本研究では、月震の波形を用いて Power Spectral Density (PSD) を計算したものを特徴量として用いる。PSD は、周波数成分毎に振幅の強度を計算したものであり、時系列データの時間相関分析などに用いられる。本研究では、Welch 法を用いて、PSD の推定を行なっている。その際に、FFT を用いてスペクトルの計算を行なっているため、2 の累乗となるように、サンプリング数を決定する。本実験では、筆者らの先行研究 [15] により、分類性能が最も高かった 2048 をサンプリング数とした。

2.2 実験結果

本章では、2.1 節で述べたそれぞれの手法を用いて、月震の分類を行い、その分類性能を評価する。

表 2 Neural Network のハイパーパラメータ

ニューロン数 (一層目)	ニューロン数 (二層目)	活性化関数	最適化関数	Dropout
60	40	tanh	Adam	あり

表 3 各手法における適合率, 再現率, F 値

手法名	適合率	再現率	F 値
Neural Network	0.97	0.97	0.96
Random Forest	0.79	0.69	0.66
SVM-Linear	0.92	0.92	0.92
SVM-RBF	0.85	0.84	0.84
ロジスティック回帰	0.87	0.87	0.87
AdaBoost	0.13	0.34	0.19

2.2.1 ハイパーパラメータの決定

本研究で月震の分類に用いる各手法には、それぞれハイパーパラメータという任意に決定できる値が存在する。分類性能を上げるためにはこれらのハイパーパラメータを調節し、データセットに最適な分類器を構築する必要がある。

Neural Network については予備実験を行い、ハイパーパラメータである中間層のニューロン数、層の数、活性化関数、最適化関数、Dropout の有無を決定する。本研究では中間層のニューロン数、層の数、活性化関数、最適化関数、Dropout の有無を変更し分類を行い、それぞれの分類結果について、F 値を求めることで適切なハイパーパラメータを決定する。はじめに、1 つ目の中間層のニューロン数を決定する。ニューロン数を増加させながら分類を行い、ニューロン数の増加によって、分類結果の F 値の変動がほとんど起こらなくなる時点でのニューロン数を本研究で用いるパラメータとする。

中間層の数と、追加される中間層のニューロン数を決定する。中間層の層の数を増加させ、分類結果の F 値の変化を確認し、F 値が上昇した場合、層の数を増加させる。このとき、新たに追加した層について、ニューロン数を 1 つ目の層の場合と同様に決定する。これを、分類結果の F 値が上昇しなくなるまで繰り返す。

各層に適用する各種関数を決定する。活性化関数 sigmoid, tanh, ReLU のうち、分類を行なった際に最も F 値が高くなる関数を各層に適用する。

Neural Network においては、学習中にランダムに選んだ一部のニューロンの働きを停止させることによって、頑健な学習を実現する、Dropout と呼ばれる手法の有効性が示唆されている。Dropout を無効にした状態で分類を行なった結果、過学習を起こしていることが観測されたため、Neural Network の中間層には Dropout を適用した。

最後に、最適化関数を決定する。最適化関数として Adam[6], AdaGrad, AdaDelta, SGD のそれぞれを用いて分類を行い、分類結果の F 値が最も高くなる最適化関数を本研究で用いるパラメータとする。チューニングを行なった結果の、Neural Network に適用するパラメータを表 2 に示す。

Neural Network については、Python のモジュールである Chainer[12] を用いた。その他の手法については scikit-learn[11] を用いて、グリッドサーチによるハイパーパラメータの決定を行なった。

2.2.2 各手法の分類結果

2.2.1 項のチューニング後、月震のデータセットに対して、6 つの手法を用いて分類を行い、それぞれの分類性能の比較を行う。その分類の結果得られた適合率、再現率、F 値を表 3 に示す。これらの値は、震源ごとに

計算したものを平均した値である。表 3 より、Neural Network を用いて分類を行なった場合に、適合率、再現率、F 値共に最も高くなった。また、F 値は 0.96 となり、月震を高精度に分類することができた。その次に SVM-Linear、ロジスティック回帰となり、線形分類を行う手法の分類性能が高い結果となった。逆にアンサンブル学習である Random Forest と AdaBoost は低い分類性能となった。

3 誤分類となる確率の高いイベントの決定

分類性能の向上を目指すため、本章では誤分類されたイベントに着目する。誤分類にはいくつかの原因が考えられるが、ラベルの間違いなどが疑われるイベントをデータセットから除くことで、分類性能の向上が見込まれる。本章では、誤分類となる確率の高いイベントを発見し、その誤分類の原因についての分析を述べる。

3.1 実験手法

分類手法や、特徴量などを変化させた複数の分類法で誤分類されたイベントを比較し、誤分類される確率の高いイベントの発見を行う。機械学習の分類性能を算出する際、全体のデータを学習データと評価データに分割するが、全ての分割パターンで分類を行うことは膨大な計算コストがかかるため、今回は学習データと評価データの分割をランダムに変更したもので、複数回分類を繰り返し、誤分類されたイベントをカウントした。本研究では、計算時間が長時間にならない 1500 回の分類を行なった。そして、各イベントの誤分類された回数を、そのイベントが評価データとして用いられた回数で割ることで、そのイベントがどの程度の確率で誤分類されるかを計算する。これらの実験に加え、誤分類されたイベントが、どういった原因で誤分類されるかを確かめるための実験も行なった。これまでの具体的な流れを以下に示す。

1. 複数の手法を用いた分類を複数回行う。その中で、イベントごとに誤分類された回数をカウントし、確率を計算する。
2. 複数の特徴量を用いた分類を複数回行う。その中で、イベントごとに誤分類された回数をカウントし、確率を計算する。
3. 各震源におけるイベントの外れ値を検出し、誤分類される確率の高いイベントが、その外れ値でないかを確認する。
4. 1, 2 より、手法と特徴量ごとに誤分類される確率の高いイベントをランキング形式にし、ランキング同士で相関係数を計算する。

1 は、手法ごとに誤分類されたイベントを比較することで、手法の違いに関わらず誤分類されるイベントを発見する。2.1.2 項の中で、F 値が 0.80 を超えた手法を用いて分類を行い、誤分類イベントをカウントする。

2 では、特徴量ごとに誤分類されたイベントを比較することで、特徴量の違いに関わらず誤分類されるイベントを発見する。特徴量には、PSD の他に、月震波形のデータを用いた。月震波形は、PSD と同様に LPZ を用いて、イベントの継続する長さを 15 分とした。

3 は、誤分類の要因が外れ値でないかを確認する。外れ値の確認は、マハラノビス距離を用いて、外れ値の判定を行う。マハラノビス距離とは、統計学で用いられる距離を表す手法の 1 つで、多次元のデータが相関を持つ場合に使用される。平均が $\mu = \{\mu_1, \mu_2, \mu_3, \dots, \mu_p\}^T$ 、共分散行列が Σ であるような多変数ベクトル $x = \{x_1, x_2, x_3, \dots, x_p\}^T$ で表される行列に対するマハラノビス距離は、下記の式のように定義される。

$$D_M(x) = \sqrt{(x - \mu)^T \Sigma^{-1} (x - \mu)}$$

具体的には、 μ には誤分類イベントの属する震源のイベントの平均を入力し、 x には誤分類イベントを入力する。このマハラノビス距離を用いて、誤分類イベントの距離を各イベントのものと比較し、その誤分類イベン

表 4 各手法と各特微量における誤分類確率の平均（上位 5 件）

(NN : Neural Network, SVM-L : SVM-Linear, SVM-R : SVM-RBF, LR : ロジスティック回帰)

イベント	PSD				月震波形				平均
	NN	SVM-L	SVM-R	LR	NN	SVM-L	SVM-R	LR	
A18-29	1.000	1.000	1.000	1.000	0.997	1.000	1.000	1.000	1.000
A21-12	0.454	0.814	1.000	0.978	1.000	1.000	1.000	1.000	0.906
A20-33	0.285	0.994	0.977	1.000	0.843	1.000	1.000	0.895	0.874
A10-33	0.986	0.994	1.000	1.000	0.837	0.820	0.550	0.693	0.860
A21-4	0.230	0.732	0.891	0.841	0.995	1.000	1.000	1.000	0.836

表 5 各手法と各特微量における分類同士の相関係数 (NN : Neural Network, SVM-L : SVM-Linear, SVM-R : SVM-RBF, LR : ロジスティック回帰)

特微量	手法	PSD				月震波形			
		NN	SVM-L	SVM-R	LR	NN	SVM-L	SVM-R	LR
PSD	NN		0.700	0.620	-0.662	-0.281	0.008	0.062	-0.240
	SVM-L			0.722	-0.860	-0.257	-0.015	0.174	-0.214
	SVM-RBF				-0.791	-0.079	0.153	0.397	-0.077
	LR					-0.217	0.083	0.162	-0.205
月震波形	NN						0.724	0.127	-0.779
	SVM-L							-0.309	-0.451
	SVM-RBF								-0.144
	LR								

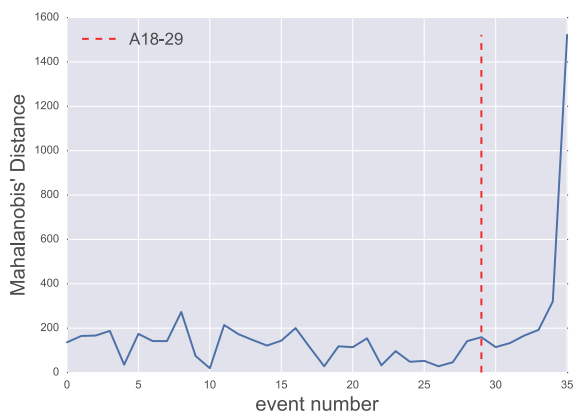


図 2 A18 における各イベントのマハラノビス距離

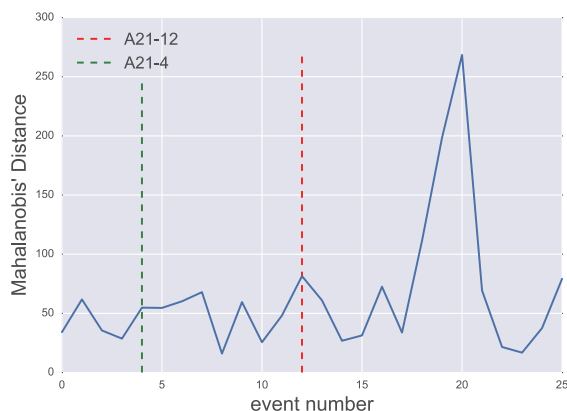


図 3 A21 における各イベントのマハラノビス距離

トが、属する震源の外れ値でないかを判定する。

4 は、1 と 2 による、特微量と手法ごとの誤分類確率による順位を用いて、各分類結果間の相関係数を計算する。この相関係数によって、こういった要因で誤分類されるイベントが異なるのかを考察する。相関係数には、ピアソン相関を用いた。

3.2 実験結果

月震分類における分類性能の高かった、Neural Network, SVM-Linear, SVM-RBF, ロジスティック回帰を用いて、誤分類イベントの決定を行う。特微量には、PSD と、PSD を計算していない時系列の月震波形を用いて分類を行なった。月震波形を用いた分類では、Neural Network, SVM-Linear, SVM-RBF, ロジス

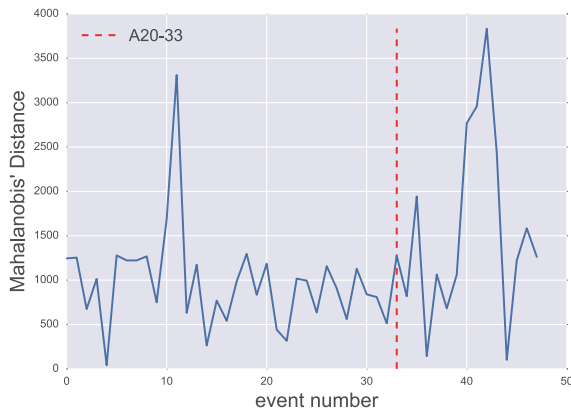


図4 A20における各イベントのマハラノビス距離

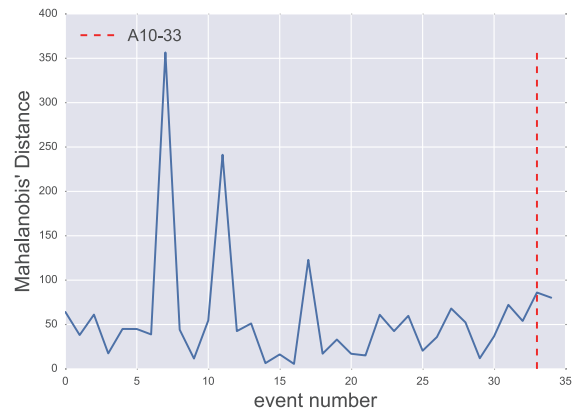


図5 A10における各イベントのマハラノビス距離

ティック回帰のどの手法においても F 値が 0.80 以上となり, PSD を用いた分類よりは低いものの, 比較的高い F 値となった. これらの各手法と各特徴量を用いて, 学習データと評価データの分割をランダムに行いながら 1500 回の分類を行い, 各イベントの誤分類された確率の平均を求めた. その確率の高いイベントの上位 5 件の各特徴量と各手法による誤分類確率と, その平均を表 4 に示す. この順位の 6 位以下は, 誤分類確率の平均は 0.8 以下となった. 表 4 のイベントの番号は, 震源の名前と, その震源ごとに観測された日時の早い方から番号を設定している. また, 表 4 は, 小数点第 3 位までを表示しているが, 表中の誤分類確率が 1.000 のものは, 1500 回の試行で一度も正しく分類されなかったことを意味している.

表 4 のイベントがその震源の外れ値でないかを確認するために, 各震源における各イベントのマハラノビス距離を計算した. 各震源における各イベントのマハラノビス距離を, 図 2, 図 3, 図 4, 図 5 に示し, 表 4 のイベントをそれぞれの図中の縦線で表す.

PSD を用いて分類した誤分類イベントと, 月震波形を用いて分類した誤分類イベントの誤分類確率による順位を用いて, 各手法と各特徴量の相関を求めた結果を表 5 に表す. 表 5 より, PSD を用いた分類では, 相関係数が比較的高い結果となった. また, 特徴量を変えた場合では, 相関係数が低くなる傾向が見られた.

4 考察

本章では, 2 章と 3 章の結果について考察について述べる.

4.1 機械学習手法の分類性能の比較の考察

表 3 より, Neural Network を用いて分類を行なった場合に, 適合率, 再現率, F 値共に最も高くなった. 本研究で用いた Neural Network は中間層が存在し, 非線形の分類を行うことのできるモデルで, その中間層において, 適切に特徴量が抽出できたため, イベント数の少ない震源を含んだデータセットでも分類性能が高くなったと考えられる.

表 3 より, Neural Network の次に分類結果の F 値が高くなったのは, SVM-Linear を用いて分類を行なった場合であり, 次に F 値が高くなったのは, ロジスティック回帰であった. SVM-Linear は, SVM で線形分類を行う場合である. 同様に, ロジスティック回帰も線形分類を行う手法である. これらの手法は, 非線形の分類を行う SVM-RBF よりも高い F 値を観測した. 一般的に, 次元数の高いデータにおける高次元空間では, データの分離に複雑な識別平面を要さない事が多いため, 線形分類と非線形の分類性能に, あまり違いが

出ないことが知られている。しかし、線形分類を行う手法は、非線形分類を行う手法に比べて、境界面を決定する際のハイパーパラメータが少ないため、今回のような学習データが少ない場合に、分類性能が高くなったと考えられる。これらのことから、月震データは線形識別器でも高精度で分類可能であると考えられる。

表3より、Random Forest と AdaBoost は、他の4つの手法に比べ分類結果のF値の値が低くなった。決定木をベースとした AdaBoost は決定木自体がデータ集合の細部に対して非常に敏感であり、過学習を起こしやすいことが知られている。また、決定木を組み合わせた Random Forest は、学習データが少ない場合に過学習を起こしやすいことが知られている。これらの手法については、過学習が発生したため、分類性能が低くなったと考えられる。

これらより、Neural Network が月震の分類において、分類性能が高い手法であることがわかった。

4.2 誤分類となる確率の高いイベントについての考察

表4より、誤分類される確率の高いイベントを示した。誤分類確率の平均が最も高い A18-29 では、表4より、月震波形を用いた Neural Network を除いた全ての分類方法で、正しく分類されることは一度もなかった。また、月震波形を用いた Neural Network でも高い確率で誤分類していることから、A18-29 は平均して 0.9996 という高い確率で誤分類された。また、図2より、他のイベントのマハラノビス距離と値が近いいため、外れ値が原因で誤分類が起こっている可能性は低いと考えられる。よって、A18-29 は手法、特徴量に関わらず、誤分類される確率が高いイベントであることが示唆された。このイベントが誤分類される理由としては、震源の主な特徴と類似しない波形であったため、または震源のラベルが間違っていることなどが挙げられる。

表4の A18-29 以外のイベントについても考察する。図3、図4、図5より、A18-29 と同様に、その震源における他のイベントのマハラノビス距離と値が近いため、これらのイベントの誤分類は外れ値が原因で起こっている可能性は低いと考えられる。A21-12 は誤分類確率の平均が高かったものの、PSD を用いた Neural Network では、誤分類される確率が低かった。同様に、A20-33 も特徴量に関わらず、誤分類される確率が高いが、PSD を用いた Neural Network による誤分類確率は低く、この分類法では誤分類されることが少なかった。よって A21-12 と A20-33 のイベントは、PSD を用いた Neural Network によって正しく分類できる確率が高い。そのため、これらのイベントの誤分類確率が高い原因は、手法によるものだと考えられる。A10-33 は、PSD を用いた分類では誤分類確率が1に近かったが、月震波形を用いた分類では誤分類確率は0.5から0.8程度となった。A21-4 は、月震波形を用いた分類では誤分類確率が1に近かったが、PSD を用いた分類では誤分類確率が0.8程度となり、PSD を用いた Neural Network では、0.230 となった。よって、A10-33 と A21-4 は、全体的に誤分類確率が高いものの、特徴量によって分類に成功する場合もある傾向となった。そのため、これらのイベントの誤分類確率が高い原因は、特徴量によるものだと考えられる。

表5では、PSD を用いた分類のそれぞれで相関が高い傾向となった。同様に、月震波形を用いた分類でも、PSD よりも少ないものの、同じ傾向が見られた。これにより、月震分類は、特徴が原因で誤分類が起こる可能性があることが示唆された。PSD を用いた場合と、月震波形を用いた場合の、手法ごとの相関を比較した場合では、あまり相関が見られなかった。これらのことより、月震分類において、学習のされ方と、どのイベントが誤分類されるかは、手法よりも特徴量に大きく依存すると考えられる。

5 まとめ

本論文では月震分類に適した分類手法を検討するため、機械学習の複数手法の分類性能の比較を行なった。その結果、Neural Network による分類の分類性能が、本研究で用いた手法の中では最も高くなった。この手法を用いる事で、ラベルが付与されていないイベントに高精度でラベルを付与できることが期待される。併せて、Neural Network は分類のクラスが確率で出力されるため、確率付きの月震カタログを作成も可能である。

また、高頻度に誤分類されるイベントに着目し、その理由の究明を行い、震源の特徴に類似しないまたは、震源ラベルが間違っているイベントを示唆した。そして、分類性能は特徴量の種類に大きく起因していることも示唆された。今後の課題として、分類に適した特徴量の発見が挙げられる。これによって、さらに高精度な分類が可能になると考えられる。また、ノイズを多く含んだイベントを含めての実験も挙げられる。本実験はノイズの少ないイベントのみで実験を行なったため、ノイズを多く含んだイベントも含めることで、震源が未知のイベントの推定も可能になると考えられる。

謝辞

本稿の執筆にあたって多数の有益な助言を頂いた、首都大学東京特任助教の莊司 慶行氏に感謝いたします。本研究（の一部）は傾斜的研究（全学分）学長裁量枠戦略的研究プロジェクト戦略的研究支援枠「ソーシャルビッグデータの分析・応用のための学術基盤の研究」による。

参考文献

- [1] BREIMAN, L. Random forests. *Machine learning* 45, 1 (2001), 5–32.
- [2] BULOW, R. C., JOHNSON, C. L., BILLS, B. G., AND SHEARER, P. M. Temporal and spatial properties of some deep moonquake clusters. *Journal of Geophysical Research: Planets (1991–2012)* 112, E9 (2007).
- [3] COX, D. R. The regression analysis of binary sequences. *Journal of the Royal Statistical Society. Series B (Methodological)* (1958), 215–242.
- [4] FREUND, Y., AND SCHAPIRE, R. E. A decision-theoretic generalization of on-line learning and an application to boosting. In *Computational learning theory* (1995), Springer, pp. 23–37.
- [5] HAYKIN, S., AND NETWORK, N. A comprehensive foundation. *Neural Networks* 2, 2004 (2004).
- [6] KINGMA, D., AND BA, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).
- [7] LAMMLEIN, D. R. Lunar seismicity and tectonics. *Physics of the Earth and Planetary Interiors* 14, 3 (1977), 224–273.
- [8] LOGNONNÉ, P., GAGNEPAIN-BEYNEIX, J., AND CHENET, H. A new seismic model of the moon: implications for structure, thermal evolution and formation of the moon. *Earth and Planetary Science Letters* 211, 1 (2003), 27–44.
- [9] NAKAMURA, Y., LATHAM, G., DORMAN, J., AND HARRIS, J. Passive seismic experiment long-period event catalog. *Galveston Geophysics Laboratory Contribution* 491 (1981), 314.
- [10] NAKAMURA, Y., LATHAM, G. V., AND DORMAN, H. J. Apollo lunar seismic experiment—final summary. *Journal of Geophysical Research: Solid Earth (1978–2012)* 87, S01 (1982), A117–A123.
- [11] PEDREGOSA, F., VAROQUAUX, G., GRAMFORT, A., MICHEL, V., THIRION, B., GRISEL, O., BLONDEL, M., PRETTENHOFER, P., WEISS, R., DUBOURG, V., VANDERPLAS, J., PASSOS, A., COURNAPEAU, D., BRUCHER, M., PERROT, M., AND DUCHESNAY, E. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.
- [12] TOKUI, S., OONO, K., HIDO, S., AND CLAYTON, J. Chainer: a next-generation open source framework for deep learning. In *LearningSys Workshop on Machine Learning Systems at Neural Information Processing Systems (NIPS)* (2015).

- [13] VAPNIK, V. N., AND VAPNIK, V. *Statistical learning theory*, vol. 1. Wiley New York, 1998.
- [14] ZHU, J., ZOU, H., ROSSET, S., AND HASTIE, T. Multi-class adaboost. *Statistics and its Interface* 2, 3 (2009), 349–360.
- [15] 菊池栞, 山田竜平, 山本幸生, 横山昌平, AND 石川博. 月震分類に適した機械学習手法の検討. **第8回データ工学と情報マネジメントに関するフォーラム**, E4-1 (2016).
- [16] 後藤康路, 山田竜平, 山本幸生, 横山昌平, AND 石川博. 波形の類似性を考慮した大規模月地震データの som による可視化システム. **第6回データ工学と情報マネジメントに関するフォーラム**, E1-4 (2014).
- [17] 山田竜平, 山本幸生, 桑村潤, AND 中村吉雄. アポロ月地震データ公開システムの開発. **宇宙科学情報解析論文誌 宇宙航空研究開発機構研究開発報告**, 1 (2012), 121–131.

れいめい衛星搭載磁力計 GAS のデータベース作成・整備：

沿磁力線電流の推定のために

益岡 葵^{*1}, 高田 拓^{*1}, 平原聖文^{*2}, 浅村和史^{*3}

Database production and management of magnetometer GAS onboard Reimei

Satellite: Estimation of field-aligned current

Aoi MASUOKA^{*1}, Taku TAKADA^{*1}, Masafumi HIRAHARA^{*2}, Kazushi ASAMURA^{*3}

Abstract

Reimei satellite contributes to the developments of auroral sciences, by the simultaneous observations of auroral emissions and auroral particles. Within the limits of a small spacecraft mission, the main mission instruments are only the Multi-spectral Auroral imaging Camera (MAC), the Electron Energy Spectrum Analyzer/Ion Energy Spectrum Analyzer (ESA/ISA), and the plasma CurRent Monitor (CRM). The data analysis of the Reimei satellite is primarily conducted by using the two mission instruments such as MAC and ESA/ISA. The Geomagnetic Aspect Sensor (GAS) is onboard the spacecraft for attitude control and it can make output of high-time resolution magnetic field data. Because of lack of calibration it has not been possible to use GAS data for scientific discussion. The noise level of the data is quite high, compared to the variation of magnetic fields in space, because the magnetometer is not in an optimal position onboard the spacecraft. We calibrated the GAS data using coordinate transformation, averaging, and offset subtraction. As the magnetic field variation of calibrated GAS data corresponds to the characteristics of auroral particles well, it is considered to be affected by field-aligned currents (FACs). We can now estimate the FAC from the magnetic field variations from GAS data. Based on the database of calibrated GAS data, Reimei data analysis is expected to make progress, where three instruments can clarify aurora behaviors and the auroral particles, monitoring FAC effects. Finally, we note that calibrated GAS data can detect even the effects of a weak FAC in spite of the large noise level of GAS raw data. This suggests that magnetometers onboard other similar microsatellites may also be used for FAC studies.

Keywords: magnetometer, field-aligned current, Reimei Satellite

概要

れいめい衛星は、オーロラと降り込み粒子の同時観測を行うことで、オーロラ科学の発展に寄与してきた。超小型衛星計画であるため、理学ミッション機器は多波長オーロラカメラ、オーロラ粒子観測器、ラングミュアプローブの3つのみであり、主にオーロラカメラと粒子観測器の2つのミッション機器による高性能で高時間分解能のデータ解析が中心であった。一方、姿勢制御用の磁力計 GAS (Geomagnetic Aspect Sensor) を用いることで、高時間分解能の磁場データを扱うことができる。ただし、磁力計データの補正等が行われておらず、そのままでは科学の議論ができる状態ではなかった。また、磁力計は衛星本体からの電磁波を受けやすい場所に設置されており、宇宙空間の磁場変動に比べて大きなノイズ成分を含んだデータとなる。座標変換、平均化、オフセット調整などのデータ補正を行うことで、磁力計データの磁場変動が、オーロラ粒子データと非常によく対応し、沿磁力線電流による影響を顕著に表していることが確かめられた。これにより、磁場変動から沿磁力線電流を推定することができる。現在、GAS の補正データのデータベースの作成・整備を行っている。GAS のデータベースを利用することで、オーロラとプラズマ粒子に加えて、沿磁力線電流の影響も同時に確認することができ、れいめい衛星データ解析が大きく前進することが期待される。また、ノイズ成分の大きな磁力計であっても、沿磁力線電流の影響を精度よく測ることができており、超小型衛星等への磁力計搭載について、重要な意味を持っている。

doi: 10.20637/JAXA-RR-16-007/0006

^{*} 平成 28 年 11 月 24 日受付 (Received November 24, 2016)

^{*1} 高知工業高等専門学校 (National Institute of Technology, Kochi College)

^{*2} 名古屋大学宇宙地球環境研究所 (ISEE, Nagoya University)

^{*3} 宇宙航空研究開発機構 宇宙科学研究所 太陽系科学研究系 (Department of Solar System Sciences, ISAS, JAXA)

1. はじめに

れいめい衛星計画は、オーロラ粒子観測器 ESA/ISA (Electron Energy Spectrum Analyzer/Ion Energy Spectrum Analyzer) によって、高時間分解能のオーロラ粒子を観測しつつ、多波長オーロラカメラ MAC (Multi-spectral Auroral imaging Camera) により、降り込み粒子が光らせるオーロラを撮像するミッションである[1, 2, 3]。超小型衛星であるため、高精度な磁場や電場の測定はミッションに組み込まれていないが、オーロラ粒子とオーロラの関係に関する重要な成果を挙げてきた[4, 5]。さらに、沿磁力線電流を推定できる磁場変動のデータがあれば、大きな系の中での微細なオーロラの振る舞いについて、踏み込んだ解析が可能である。オーロラとプラズマ粒子の同時観測にあたって、れいめい衛星の姿勢を特定するため磁力計 GAS が搭載されており、衛星位置での地球固有磁場の方向をもとに、姿勢制御を行っている。GAS は、200 Hz サンプリングで高時間分解能による磁場計測を行っているが、衛星本体からの誘導磁場の影響を受けやすい位置に搭載されており、宇宙空間磁場に比べて大きなノイズ成分が含まれている。また、GAS データに関しては、衛星座標系での取得値が公開されているのみで、容易に科学的な議論を行うことができる状況にない。

オーロラ帯と地球磁気圏の赤道面は、沿磁力線電流によって大きな電流回路を構成しており、その電流系を理解することは、オーロラ現象のメカニズムの全容解明につながると考えられる。れいめい衛星高度に沿磁力線電流が流れていれば、衛星上で比較的大きな誘導磁場が測定される。この沿磁力線電流は、プラズマ粒子の流れによって生成されており、衛星で測定されたプラズマ粒子データとも対応する。一般に、磁力計で測定される磁場変動から沿磁力線電流の大きさが推定できる。れいめい衛星においては、プラズマ粒子データにより磁場変動が実際に沿磁力線電流によるものかを確認することが可能である。この利点を活かして、姿勢制御用の磁力計による沿磁力線電流の推定の可否を調べるができる。

本研究では、れいめい衛星に搭載された GAS データを科学データとして扱いやすい形式に変換し、GAS データベースの作成・整備を行った。GAS データの補正処理を行い、プラズマ粒子データと比較することにより、沿磁力線電流が適切に推定されていることを確認している。

2. れいめい衛星と磁力計 GAS

れいめい衛星は、工学的な新規技術の実証と理学目的であるオーロラの観測を行うために開発された超小型科学衛星である。表 1 に、れいめい衛星の機体データの概要を示す。2005 年に打ち上げられ、高度約 650 km の極軌道を周回し、2012 年度まで運用された。多くのオーロラと降り込み粒子の高時間分解能の同時観測データを蓄積しており、それらのデータの価値は高い。主要搭載機器として、姿勢制御のためのいくつかの機器が搭載されている。GAS は、磁力線方向を認識して、姿勢制御するための機器として搭載された。理学ミッション機器としては、MAC[2]、ESA/ISA[3]、ラングミュアプローブ (CRM) が搭載されている。これまでのオーロラ科学に関する研究では、主に MAC と ESA/ISA から得られたデータ解析が中心だった。

表 1: れいめい衛星の機体データ

名称	れいめい (INDEX)	
打上	日時	2005 年 8 月 24 日 6 時 10 分 (日本標準時)
	場所	カザフスタン共和国
構造	質量	約 72 kg
	形状	約 62×62×72 cm
軌道	高度	近地点 610 km 遠地点 654 km
	傾斜角	97.8 度
	種類	略円軌道
	周期	97 分
主要搭載機器	(1) スタートラッカ (STT) (2) スピン/ノンspin型太陽センサ (SSAS/NSAS) (3) 磁力計 (GAS) (4) 三軸の光ファイバジャイロ (FOG) (5) リアクションホイール (RW) (6) 磁気トルカ (MTQ) (7) 多波長オーロラカメラ (MAC) (8) オーロラ粒子観測器 (ESA/ISA) (9) ラングミュアプローブ (CRM)	



図1：れいめい衛星搭載磁力計（GAS）

GAS は、3 軸のフラックスゲート磁力計である。図 1 に、れいめい衛星に搭載された GAS の写真を示す。GAS は 200 Hz の時間分解能で測定を行っているが、数万 nT の地球固有磁場を測ることが目的であるため、目標とする測定精度が高くなく、衛星本体からの誘導磁場による影響を受けやすい位置に搭載されている。GAS は、大きなノイズ成分を含んだデータを出力しているが、科学ミッションのために搭載された測定器ではないため、補正等を行われてこなかった。GAS の測定値には、地球固有磁場、沿磁力線電流による誘導磁場、衛星関連の誘導磁場等が含まれており、一般的には、宇宙空間の磁場変動を観測するのに適した観測器ではない。実際に GAS によって測定されたデータの一例を図 2 に示す。

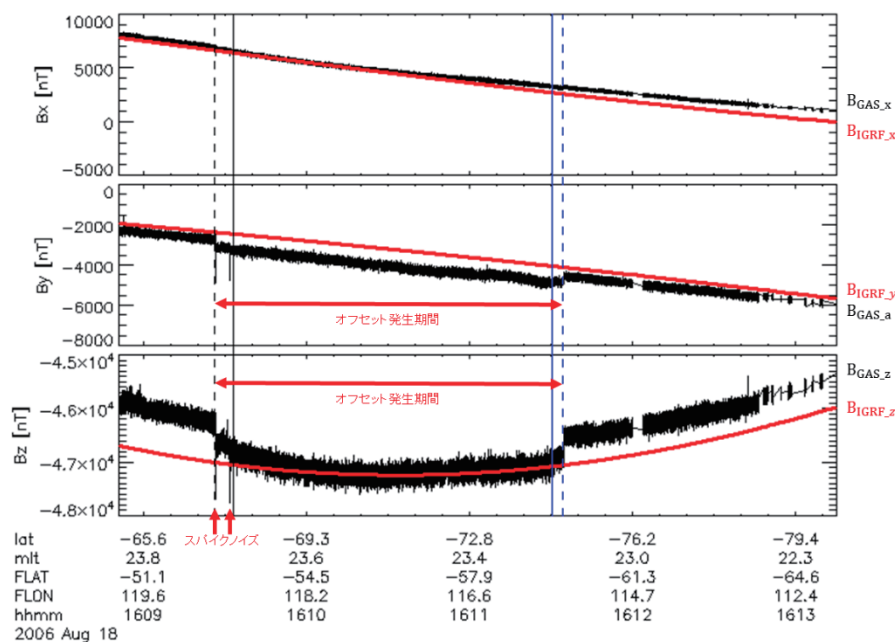
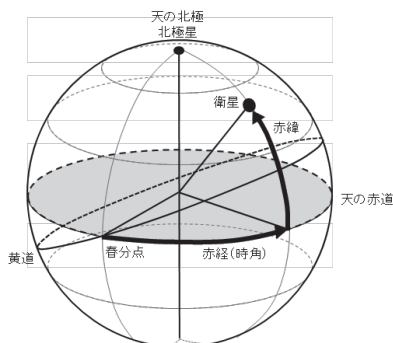


図2：GAS によって測定された磁場データ：衛星座標系での磁場の3成分（黒線）と、IGRF モデルによる地球の固有磁場（赤線）。縦線は、MAC のコマンド実行時刻（順に、起動、測定開始、測定終了、停止）。

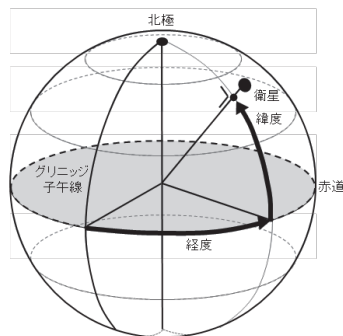
図 2 では、黒線で示された磁場の 3 成分が、赤線で示された IGRF (International Geomagnetic Reference Field) 10 モデルによる地球固有磁場の値と同程度の値を示している。磁場データの黒線は太い帯状に見えるが、これは短時間での磁場変動が大きく、大きなノイズ成分を含んでいるためである。また、MAC の起動時と測定開始時には、縦の矢印で示すようなスパイク状のノイズが見られ、MAC の測定中は、横向き矢印で示す区間に階段状の磁場変動（オフセット）が見られる。MAC の起動中（縦の点線の間の区間）に誘導磁場の影響を受けていることが分かる。MAC 起動から測定開始の間（16:09:26-33）には、スパイク状のノイズに加えて、オフセットノイズが時間的に変動している。MAC 測定終了から MAC 停止の間（16:11:30-16:11:34）では、スパイク状のノイズはないが、オフセットノイズの変動は見られる。オフセットノイズの時間変動を補正するのは難しいので、今回は、それらの区間の補正は行っていない。沿磁力線電流に伴う誘導磁場成分を評価するためには、磁場変動のノイズ等をできるだけ軽減し、オフセットノイズを調整する必要がある。

3. 磁力計データの補正

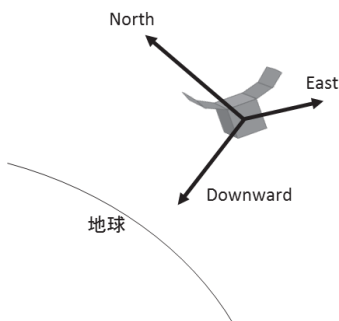
a. 赤道座標系 (J2000, TOD)



b. 測地系



c. NEZ座標系 (North/East/Downward)



d. 衛星座標系

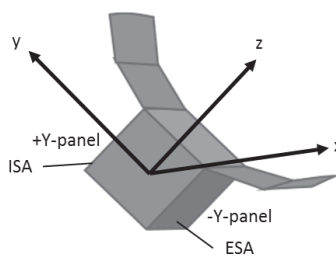


図3：データ処理で用いた座標系

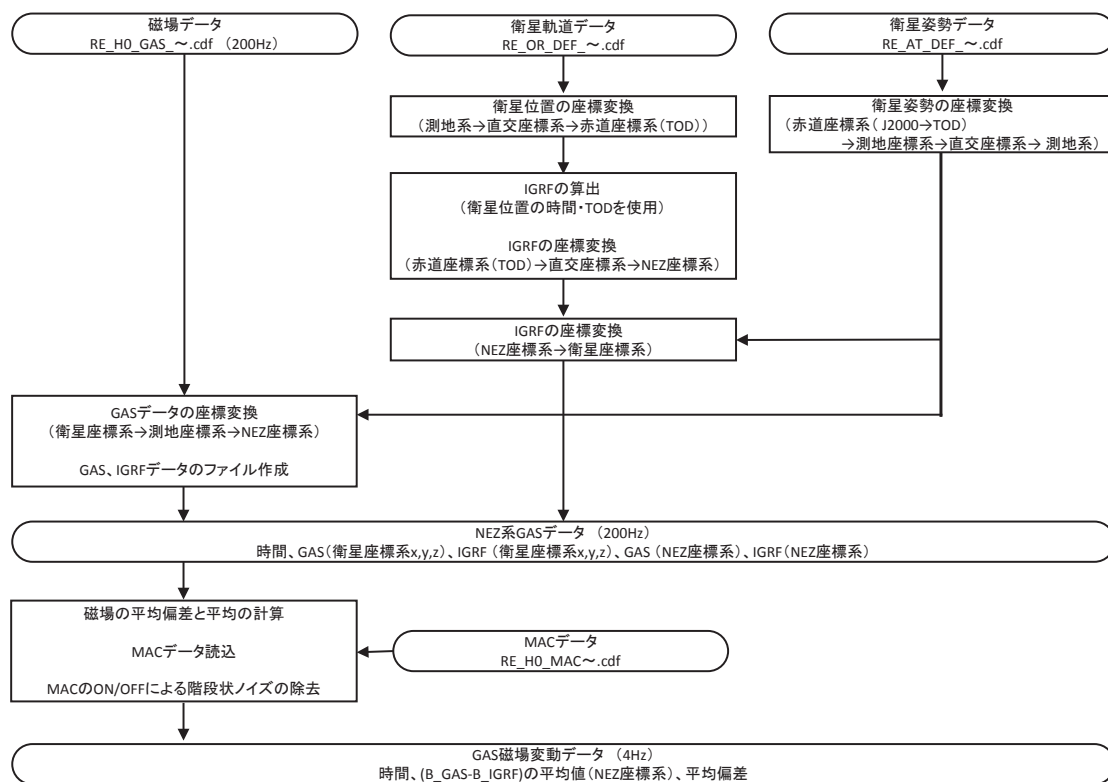


図4：GASデータの補正処理の流れ

GAS データを補正するためには、測定された GAS データの座標変換および、地球固有磁場との差分、ノイズ成分の除去が必要となる。観測値は、衛星座標系で与えられており、衛星の軌道データおよび姿勢データを用いて、座標変換される。使用する 4 つの座標系を図 3 に示す。図 3a の赤道座標系で Mean of J2000.0 赤道座標系 (J2000) と True of Date 赤道座標系 (TOD) を用いている。J2000 では地球時 2000 年 1 月 1.5 日 (J2000) における平均赤道面、平均春分点を基準としており、TOD では瞬時の真の赤道面、真の春分点を基準とする。図 3b の測地系は、経緯度および標高によって位置を示している。図 3c の NEZ 座標系は北、東、地球への向きを各々正とした座標系であり、オーロラ帯の現象を扱う際に、視覚的に理解しやすく、よく使われる座標系である。図 3d の衛星座標系は、衛星の構造に準拠した座標系であり、GAS が測定を行うときに使用している。

図 4 に、GAS データの補正処理の概要を示す。200 Hz サンプリングの GAS データを、最終的に 4 Hz サンプリングの補正データとし、データベース化する GAS データに加えて、衛星軌道データと衛星姿勢データの情報をを用いて、各種の座標変換を行う。また、IGRF10 モデルに衛星の軌道データおよび時刻を入力することで、当該時刻および当該衛星位置における地球固有磁場を導出する。座標変換に関しては、最終的に NEZ 座標系に変換し、東向き成分、北向き成分、地球向き成分の磁場とする。座標変換のみを行ったデータに関しても、IGRF 磁場と合わせて、ASCII データとして保存している (NEZ 座標系 GAS データ : 200 Hz)。GAS データと IGRF の差分を求めることで、宇宙空間の磁場を導出でき、強い沿磁力線電流が近傍に流れている領域では、この磁場変動が沿磁力線電流の誘導磁場に相当すると考えられる。ノイズレベルを小さくするため、200 Hz データに対して、1 s 間の平均値と平均偏差を計算して記録してある。沿磁力線電流の構造は数 10 km 以上と大きく、細かな磁場変動は、機器的なノイズ成分と考えられる。また、MAC の起動に伴って、一定値のオフセットノイズが含まれていることに関しては、次のような処理を行っている。MAC データより、MAC の測定開始および測定終了時刻を読み込み、運用手順から、MAC 測定開始の 6.8 s 前に起動、測定終了後 4 s 後に、MAC 停止となることから、MAC の起動に伴って影響を受ける時間帯を特定する。MAC の測定開始前と測定終了後は磁場が不安定になるため、補正処理を行っていない。このことにより、起動時と測定開始の間のスパイクノイズも無視していることになる。MAC 測定中のオフセットノイズは、測定区間の端の値の平均を用いて、オフセット値を算出し、測定区間の磁場変動から、オフセット値を引いている。これらの補正処理を行った磁場データを、サンプリングレートを 4 Hz に落として、データ保存している (GAS 磁場変動データ)。サンプリングレートを落としている理由は、処理の過程で 1 s 間の平均値を取っていることもあり、物理的に意味のある高周波の変動を再現できるとは考えられないためである。

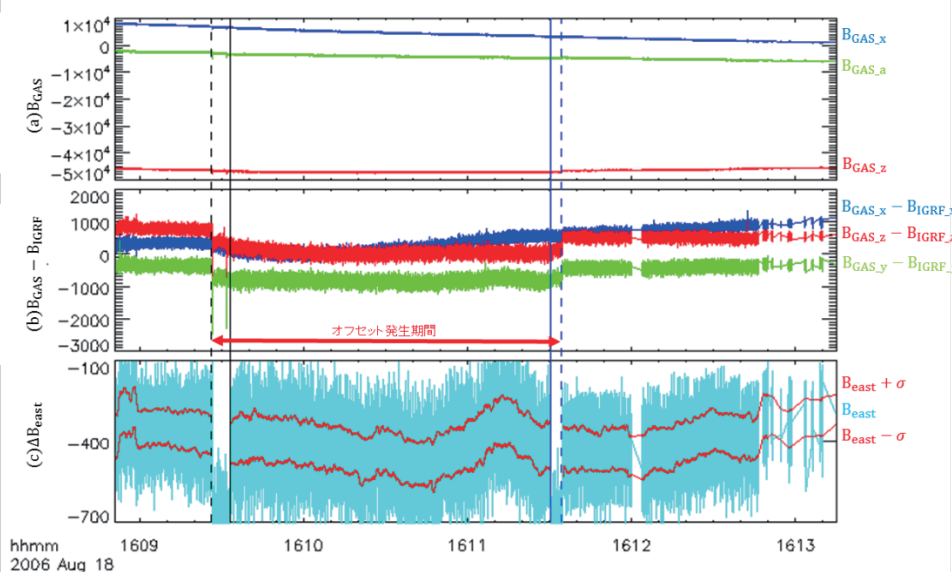


図 5 : (a) GAS の元データ, (b) IGRF との差分データ, (c) 東向き成分の磁場変動 : 補正 4 Hz データ (赤線) とオフセットのみ補正 200 Hz データ (水色)

図 5 に、補正処理を行う際の各データを示す。図 5(a) では、衛星座標系の GAS データが示されており、地球固有磁場程度の大きな値を示している。図 5(b) では、GAS と IGRF との差分データを示しており、宇宙空間の磁場および衛星機器関連の誘導磁場を合わせた磁場の値となっている。MAC の動作の影響によって図 2 で説明したオフセットノイズが含まれていることが顕著に見られる。本イベントでのオフセット値は、 $(B_x, B_y, B_z) = (-146, -386, -504)$ であった。図 5(c) では、補正データの東向き成分の磁場変動データを示している。水色が 200 Hz のデータで、赤線は、平均値±平均偏差の値を 4 Hz サンプリングで示している。MAC 測定中 (16:09:33-16:11:30) の磁場変動の値 (縦の実線の間の区間) が、MAC 起動前後の値とほぼ同じレベルに補正されていることが分かる。

4. データ処理の妥当性の検討

本節では、補正データの磁場変動が沿磁力線電流による影響を反映しているかを確認する。図6に、GASの補正データによる東向き磁場変動成分、ピッチ角ごとの電子・イオンのエネルギー—時間スペクトラムを示す。補正前のGASデータ（水色：オフセットのみ考慮）は、ノイズ成分が非常に大きい。補正データである赤線は、沿磁力線電流による緩やかな変動を明確に示している。れいめい衛星が南半球にいて、上向きの沿磁力線電流を北から南へと横切るとき、沿磁力線電流の北側では西向きの誘導磁場を感じ、南側では東向きの誘導磁場を感じる。東向きを正とすると、磁場変動の東向き成分が増加傾向にあれば、沿磁力線電流は上向き、減少傾向にあれば下向きとなると考えられる。また、衛星速度はほぼ等しいので、増減区間が短ければ、沿磁力線電流の空間スケールも小さく、増減区間が長ければ、空間スケールが大きいことを示している。図6の上部には、磁場変動の増減と時間長さを踏まえて、沿磁力線電流の向きと空間スケールを模した矢印を描いている。一方、観測範囲のプラズマ粒子が主な電流の担い手である場合、電子の降り込み量が多ければ、電流は上向きに、イオンの降り込み量が多ければ、電流は下向きに流れているはずである。例えば、16:11:00前後のD区間では、電子の降り込み（ピッチ角120°-180°）が強く、上向きの電流があることが分かるが、それに対応して磁場の東向き成分は増加しており、上向き電流であることが示唆している。16:10:22~16:10:30のB区間では、0.8 keV程度とエネルギーが低く、フラックス量の小さな電子の降り込みがあり、弱い沿磁力線電流であることが推測されるが、小さいながらも明確な磁場変動の増加が見られた。一方、16:10:43~16:11:13のE区間では、強いイオンの降り込み（ピッチ角120°-180°）があり、電子の降り込みは強くないため、下向き電流であることが分かる。最後に、ESA/ISAの観測終了後およそ15秒後の16:12:48~16:12:52のI区間に、15秒程度の微小な磁場変動の増加・減少が見られている。これは、ESA/ISAの停止後に常に見られる機器由来の磁場変動である。図6では、沿磁力線電流に対応する電子とイオンが確認できる。

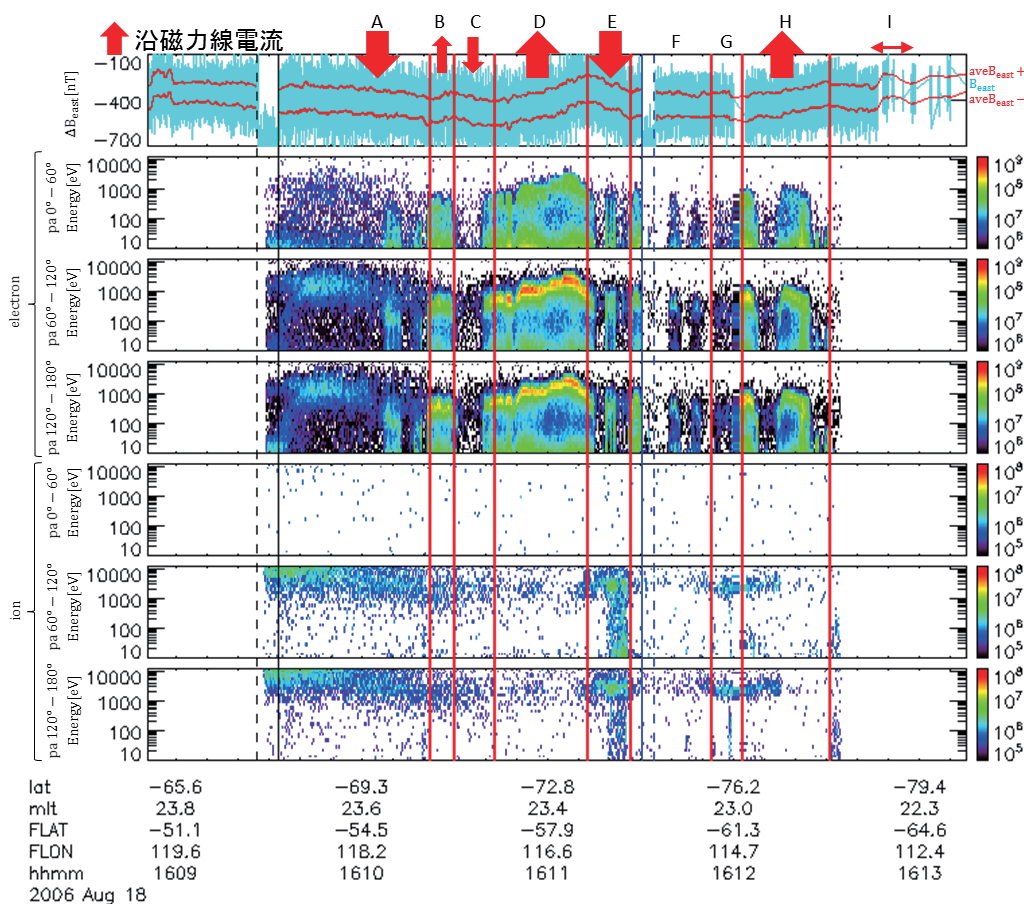


図6：補正後のGASデータ（東向き成分）とピッチ角ごとの電子・イオンのエネルギー—時間スペクトラム

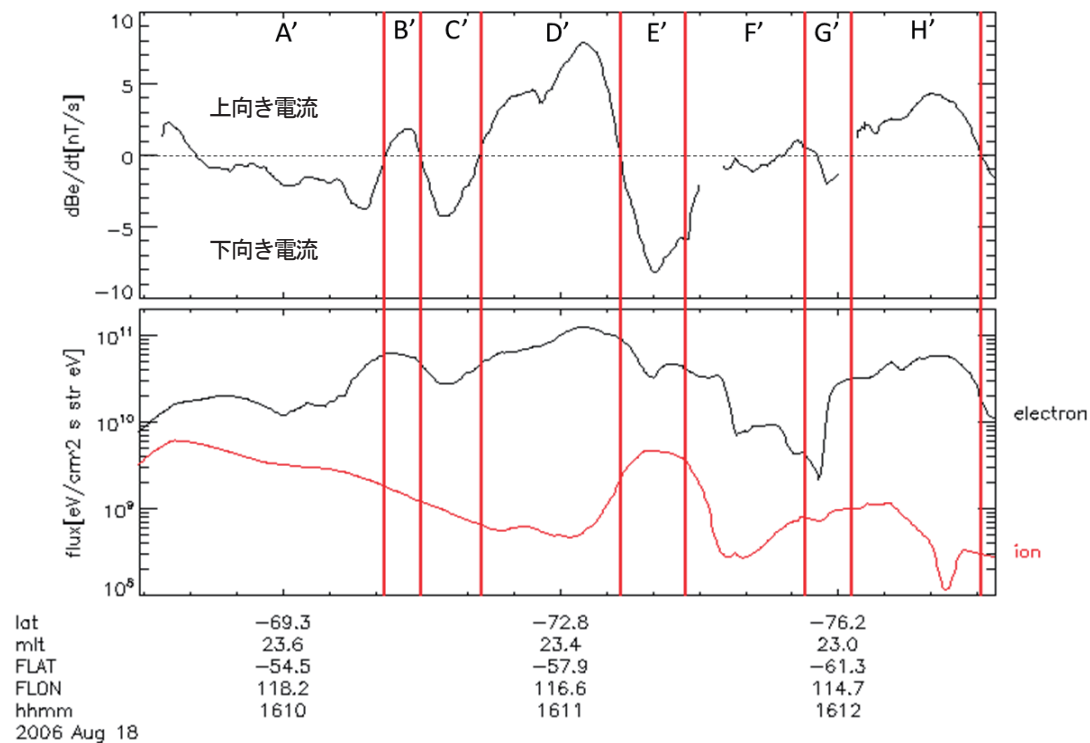


図 7: GAS データ (東向き成分) の傾きと電子・イオンのフラックスの対応

図 7 では, GAS, ESA/ISA から得られた沿磁力線電流に対応する物理量を比較する. GAS データは, 磁場の勾配に相当する dB_{east}/dt を時間幅 15 秒での最小二乗法により求めた. この時間幅を使用することで, 約 100 m 以下の微細な電流構造は考慮していないことになる. dB_{east}/dt が正のとき上向き電流, 負のとき下向き電流となる. ESA/ISA データでは, エネルギー方向と時間幅 15 秒間のフラックスを足し合わせた. また, ESA では, 正味の下向きフラックスを求めるため, 下向きフラックス (ピッチ角 $120^\circ - 180^\circ$) から上向きフラックス (ピッチ角 $0^\circ - 60^\circ$) を差し引いている. ISA では, 上向きフラックスがほぼ 0 のため, 下向きフラックスの値のみを示している. 電子とイオンフラックスの比較に関しては, 観測器のエネルギー範囲の制限や光電子の影響などのため, 定量的な比較はできない. 電子フラックスがピークをもつ区間 B', D', H' では, dB_{east}/dt が正で上向き電流となっている. 一方, イオンフラックスが多い区間 A', E', G' では, 電子フラックスも多くなっているため, dB_{east}/dt は負で, 下向き電流となっている. 区間 C' では, イオンフラックスはそれほど多くないが, 電子フラックスが減っており, 下向き電流となっている. 区間 F' は, 電子・イオンフラックスとも少なくなっており, dB_{east}/dt は微弱な変動を示している. これらのことより, 磁場変動の傾きは電子・イオンのフラックスから予想される沿磁力線電流とよく対応する. このように, GAS の補正データの磁場変動は, 粒子観測機で確認できる沿磁力線電流の向きやスケールの大きさと, 非常によく一致を示している.

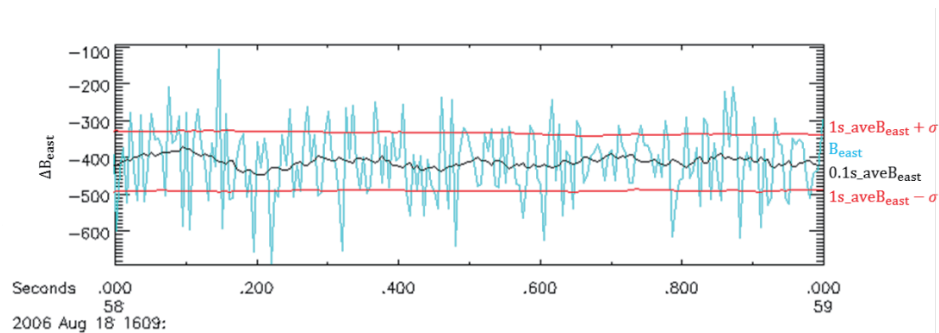


図 8 : 短時間 (2006. 08. 18 16:09:58~16:09:59) の GAS 磁場変動 ΔB の東向き成分 :
200 Hz (水色線), 0.1 秒平均値 (黒線), 1 秒平均値 \pm 平均偏差 (赤線)

表2. : 補正データのノイズレベルのまとめ (2006.08.18 16:09:55~16:10:05)

	時間分解能	平均偏差 [nT]	対応する電流密度 [$\mu\text{A}/\text{m}^2$]
元データ	200 Hz	75.5 ~ 92.7 (1 秒間)	7.94 ~ 9.72
補正データ	4 Hz	6.94 (10 秒間)	0.72

次に、GAS の補正データの磁場変動がほとんどない区間を選び、磁場のノイズ成分の影響について考察する。図8は、1秒間の磁場変動であり、赤線は4 Hz サンプリングの補正データの磁場変動を示しているが、ほとんど変化がないことが分かる。このとき、200 Hz の磁場データ（水色）は、平均偏差を超える磁場振動を周期的にしている。また、黒線は、200 Hz データを0.1秒平均したものであるが、0.2秒程度の緩やかな振動成分がある。ほぼ一定の衛星速度を考慮すると、平均時間である1秒以上の磁場変動は数 km 以上の空間スケールを持つ沿磁力線電流に対応するため、それらの沿磁力線電流の推定には、補正データが対応可能であると考えられる。

次に、磁場変動の大きさに対する評価を行う。表2に補正前後のノイズレベルをまとめた。200 Hz の元データでは、1秒間の平均偏差は75-93 nT であり、宇宙空間の磁場変動と比べて非常に大きい。変位電流を無視したアンペール・マクスウェルの式

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{j}$$

を用いて、磁場変動を電流密度に換算すると、約8-10 $\mu\text{A}/\text{m}^2$ となり、実際の沿磁力線電流（約1-2 $\mu\text{A}/\text{m}^2$ ）と比べても、非常に大きいことが分かる。一方、4 Hz の補正データでは、10秒間の平均偏差は約7 nT であり、電流密度に換算すると1 $\mu\text{A}/\text{m}^2$ 以下であり、かなり弱い沿磁力線電流も検知可能なデータとなっている。電流密度の計算に関しては、れいめい衛星の速度が7.585 km/s であると仮定して単位時間当たりの変位を計算し、平均偏差分の磁場変動に対応する電流密度を導出した。補正データは、結果的に、沿磁力線電流の見積もりに十分な精度の磁場変動データとなっており、その磁場変動は、プラズマ粒子データの傾向とよく合う。今後、非常に微弱な沿磁力線電流の例などに対して、プラズマ粒子データとも合わせて、電流密度の大きさの見積もりとその妥当性を検討したい。

5. まとめ

れいめい衛星には、MAC, ESA/ISA と CRM の3つの理学ミッション機器が搭載されており、降り込み粒子データとオーロラ撮像画像による微細なオーロラ解析が行われてきた。本研究では、姿勢制御用の磁力計 GAS のデータを処理し、沿磁力線電流の大きさや向きについて議論できるかどうかの評価を行った。座標変換、機器依存の高周波のノイズ成分の平均化、および、MAC の起動に伴う磁場変動の除去等の補正を行い、データベースとして整備した。また、ESA/ISA データと比較することで、GAS の補正データによる磁場変動が、沿磁力線電流による変動とよく合うことを確認した。これらの補正データは、科学論文でも利用され、沿磁力線電流の根拠として使われている[5]。一方で、補正データの磁場変動は、比較的規模の小さな沿磁力線電流に対しても、非常によく対応することが分かった。れいめい衛星の ESA 観測器という高時間分解能で高精度な電子データとの比較により、本来であれば GAS データのノイズレベルより小さな磁場変動であっても、優位な変動が見られている。今後の課題として、微弱な沿磁力線電流に対してのさらなる定量的な比較が必要であるが、現時点では、多くの沿磁力線電流を適切に検知できる磁場変動のデータベースとなっている。データベースは公開予定である。

謝辞

本研究に際して、国立極地研究所研究員の福田陽子氏、名古屋大学宇宙地球環境研究所研究員の元場哲郎氏には、GAS の補正データの使用等に関して、ご協力を頂いたので、ここに感謝いたします。宇宙航空研究開発機構 宇宙科学研究所の齋藤宏文教授、坂井真一郎准教授をはじめとする、れいめい衛星チームの皆さまには、衛星機器や運用等に関して、お世話になりましたので、感謝いたします。本研究は、平成28年度名古屋大学宇宙地球環境研究所「データベース作成共同研究」の助成を受けたものです。また、1名の査読者の方からは有意義なご助言をいただきましたので、感謝いたします。

参考文献

- [1] Saito, H., et al., “INDEX: Piggy-back satellite for aurora observation and technology demonstration”, Acta Astronautica, 48, 5-12, 723-735, (2001).
- [2] Sakanoi, T., et al., “Development of the multi-spectral auroral camera onboard the index satellite”, Adv. Space Res., 32, 3, 379-384, (2003).
- [3] Asamura, K., et al., “Auroral particle instrument onboard the index satellite”, Adv. Space Res., 32, 3, 379-384, (2003).

- [4] Asamura, K., et al., “Sheared flows and small-scale Alfvén wave generation in the auroral acceleration region”, *Geophys. Res. Lett.*, 36, 5, L05105, (2009).
- [5] Fukuda, Y., et al., “Electron properties in inverted-V structures and their vicinities based on Reimei observations”, *J. Geophys. Res.*, 119, 5, 3650-3663, (2014).

Application of data science techniques to disentangle X-ray spectral variation of super-massive black holes¹

S. Pike^{*1}, K. Ebisawa^{*1}, S. Ikeda^{*2}, M. Morii^{*2}, M. Mizumoto^{*1}, and E. Kusunoki^{*1}

^{*1}*Institute of Space and Astronautical Science, Japan Aerospace Exploration Agency, Japan*

^{*2}*Research Center for Statistical Machine Learning, Institute of Statistical Mathematics, Japan*

Abstract

We apply three data science techniques, Nonnegative Matrix Factorization (NMF), Principal Component Analysis (PCA) and Independent Component Analysis (ICA), to simulated X-ray energy spectra of a particular class of super-massive black holes. Two competing physical models, one whose variable components are additive and the other whose variable components are multiplicative, are known to successfully describe X-ray spectral variation of these super-massive black holes, within accuracy of the contemporary observation. We hope to utilize these techniques to compare the viability of the models by probing the mathematical structure of the observed spectra, while comparing advantages and disadvantages of each technique. We find that PCA is best to determine the dimensionality of a dataset, while NMF is better suited for interpreting spectral components and comparing them in terms of the physical models in question. ICA is able to reconstruct the parameters responsible for spectral variation. In addition, we find that the results of these techniques are sufficiently different that applying them to observed data may be a useful test in comparing the accuracy of the two spectral models.

1 Introduction

Narrow-line Seyfert 1 galaxies (NLS1), a particular class of super-massive black holes, are known to exhibit high X-ray luminosity (Boller et al., 1995) as well as a broad iron fluorescence line and edge feature around 6.4 keV (Fabian et al., 2000). Their X-ray spectral variation has been primarily explained by two different physical models.

The first is the relativistic “disk-line model” (e.g. Tanaka et al., 1995). According to this model, X-ray spectral variation is the result of changes in the geometry of the X-ray emitting region in the very vicinity of the central black hole. In particular, this model claims that changes in the height of the very compact X-ray source (a.k.a. “lump-post”) account for spectral variability, and the broad iron features result from gravitational redshift and Doppler shift of emission lines originating via fluorescence in the innermost part of the accretion disk (Fabian et al., 1995). Mathematically, the observed X-ray flux predicted by this model may be written as the sum of the observed X-ray flux originating from the compact X-ray source, represented as a power law, the flux originating from the accretion disk, represented by a multicolor blackbody distribution, and the flux originating from reflection off of the disk:

$$F(E, t) = A_I(E)(N_P(t)P(E) + N_B B(E) + N_R(t)R(E)) \quad (1)$$

Where A_I is the effect of interstellar absorption, $N_P(t)$, N_B , and $N_R(t)$ are normalization factors, $P(E)$ is the power-law component, $B(E)$ is the blackbody component, and $R(E)$ is the disk-reflection component. Observed spectral variation is mostly explained by changes in the normalization of the power-law normalization, $N_P(t)$, and that of the reflection component, $N_R(t)$.

The second model, known as the “variable double partial covering (VDPC) model”, instead posits that the characteristic spectral shape of NLS1 results from partial absorption by warm intervening absorbers, presumably composed of two layers of different ionization levels: an optically thicker low-ionized inner layer and an optically thinner high-ionized envelope (Mizumoto et al., 2014). According to the VDPC model, observed spectral variation

doi: 10.20637/JAXA-RR-16-007/0007

^{*} 平成 28 年 11 月 24 日受付 (Received November 24, 2016)

^{*1} Based on observations obtained with XMM-Newton, an ESA science mission with instruments and contributions directly funded by ESA Member States and NASA.

is the result of changes in the partial covering fraction, α , which quantifies the extent to which the X-ray emitting region is partially occulted by the intervening clouds. Because the model predicts that X-rays originating from the emission region are affected by variable absorbers, this model, unlike the disk-line model, is multiplicative. It may be written as:

$$F(E, t) = A_I(E)(1 - \alpha(t) + \alpha(t)W_n(E))(1 - \alpha(t) + \alpha(t)W_k(E))(N_P(t)P(E) + N_B B(E)) \quad (2)$$

Where $W_n(E)$ and $W_k(E)$ are the effects of the optically-thinner high-ionized absorbers and the optically-thicker low-ionized absorbers, respectively. Distributed, one can see that the model can be written as a linear combination of non-independent spectral components:

$$F(E, t) = A_I(E)((1 - \alpha(t))^2 + \alpha(t)(1 - \alpha(t))(W_n(E) + W_k(E)) + \alpha^2(t)W_k(E)W_n(E))(N_P(t)P(E) + N_B B(E)) \quad (3)$$

In this formulation, there is an uncovered component with coefficient $(1 - \alpha(t))^2$, a partially covered component with coefficient $\alpha(t)(1 - \alpha(t))$, and a fully covered component with coefficient $\alpha^2(t)$. In this model, most spectral variation is explained by the variable partial covering fraction, $\alpha(t)$, as well as the normalization of the power-law component, $N_P(t)$.

Observed static X-ray spectra of NLS1 fit both models well. Therefore, the models must be compared via methods other than simple spectral model fitting. One such method is to probe the mathematical structure of observed spectral variation. Certain features, such as the number of spectral components necessary to reproduce the observed variation or the shapes of such components may differ depending on the X-ray production mechanisms involved.

The spectral data may be represented by 2D matrices with row indices corresponding to time and column indices corresponding to energy. Thus, there exist basis vectors which may be linearly combined in order to reproduce the original data matrices. In the case of AGN spectra, these basis vectors are spectral components whose combinations approximate the original spectra. In other words, the problem of determining spectral components may be approached as a two-dimensional matrix factorization problem.

Among many matrix factorization methods, Nonnegative Matrix Factorization (NMF), Principal Component Analysis (PCA), and Independent Component Analysis (ICA) are most widely used in astronomical problems, and each is known to have advantages and disadvantages for different problems (Ivezic et al., 2014). For this reason, we try NMF, PCA and ICA to solve our X-ray astronomy problem.

In order to better understand these advantages and disadvantages and to determine how these techniques may perform when applied to the spectra produced via the two physically different models, we produced simulated spectra and applied the above data science techniques. Below we describe the methods of simulation, provide an overview of each of the techniques, and discuss the results of each of the data science techniques.

2 Data Simulation

In order to understand the behavior of the data science techniques when applied to astronomical data, we simulated data according to the disk-line and VDPC models. First, the model parameters were chosen to fit the average spectra of the NLS1 MCG-6-30-15 obtained during simultaneous observation by NuSTAR (Harrison et al., 2010) (ID: 60001047002, 60001047003, 60001047005) and XMM-Newton (ID: 0693781201, 0693781301, 0693781401) between January 29 and February 2, 2013. In addition to the models represented by (1) and (2), a static full-covering ionized absorber, which produces many absorption lines, was also required to fit the spectra. During simulation, all but two parameters were fixed to be the average of their respective fitted values. In the case of the disk-line model,

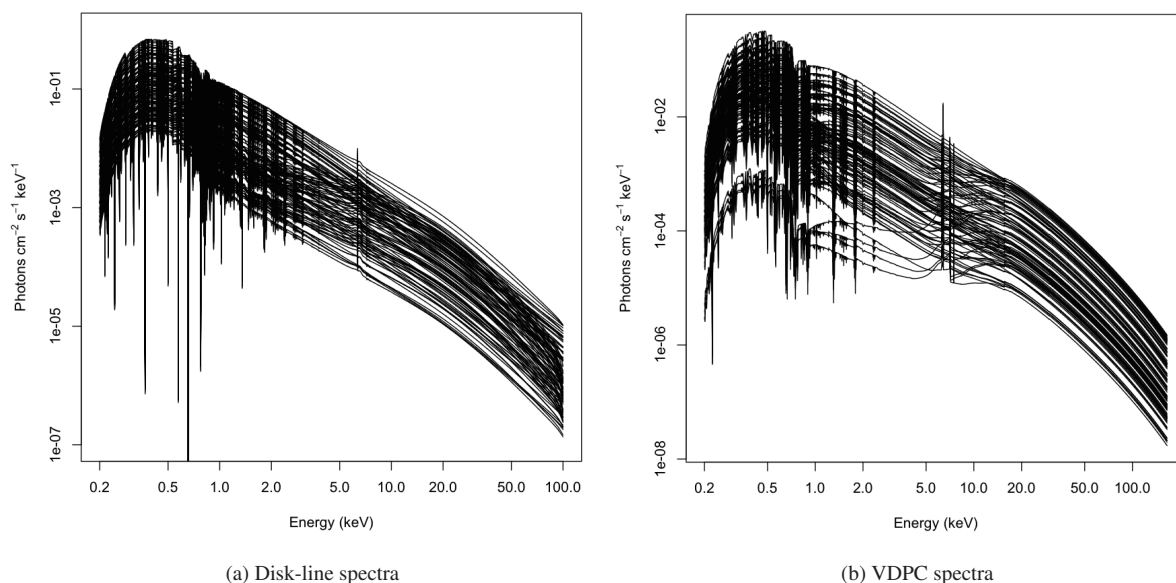


Figure 1: X-ray spectra simulated according to the disk-line and VDPC models.

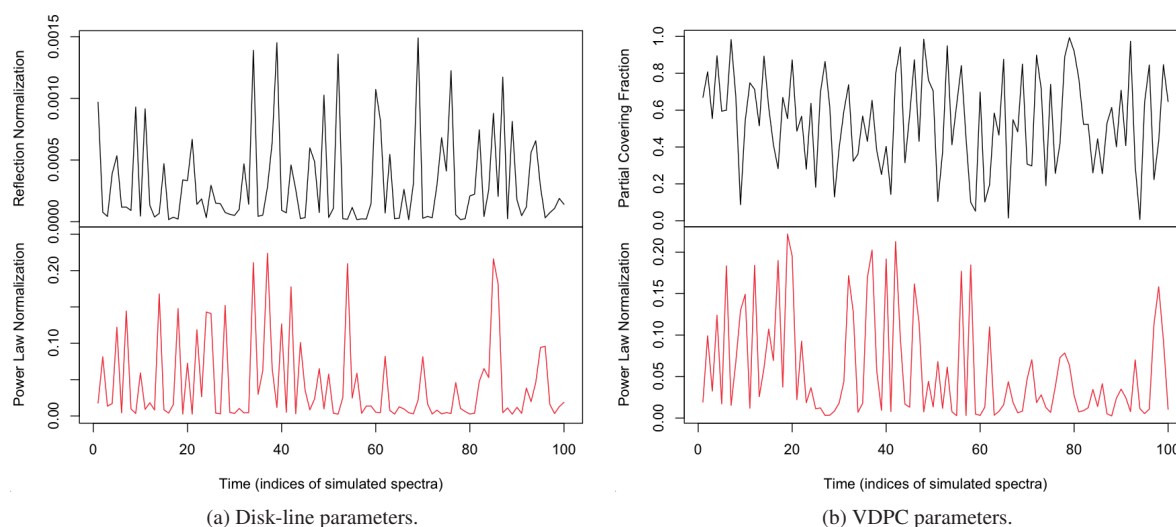


Figure 2: Parameters varied during simulation of X-ray spectra according to the disk-line and VDPC models.

the two parameters which were varied were the normalizations of the power-law and reflection components. All fitting and modeling were done using `xspec` (v.12 Arnaud, 1996).

In the case of the VDPC model, the normalization of the power-law component as well as the partial covering fraction were varied. Normalization factors were varied by choosing random values from logarithmically uniform distributions between 10 and 0.1 times their average values determined during fitting. The partial covering fraction was varied by choosing random values from a uniform distribution between 0 and 1.0, exclusive. One hundred spectra were simulated for each model (Figure 1). Note that spectra simulated according to the VDPC model exhibits more variability in the soft X-ray regime and the iron edge and line features around 6.4 keV are more pronounced, while the absorption lines in the soft X-ray regime are far more pronounced in the spectra simulated according to the disk-line model. In Figure 2 we plot the parameters varied during simulation.

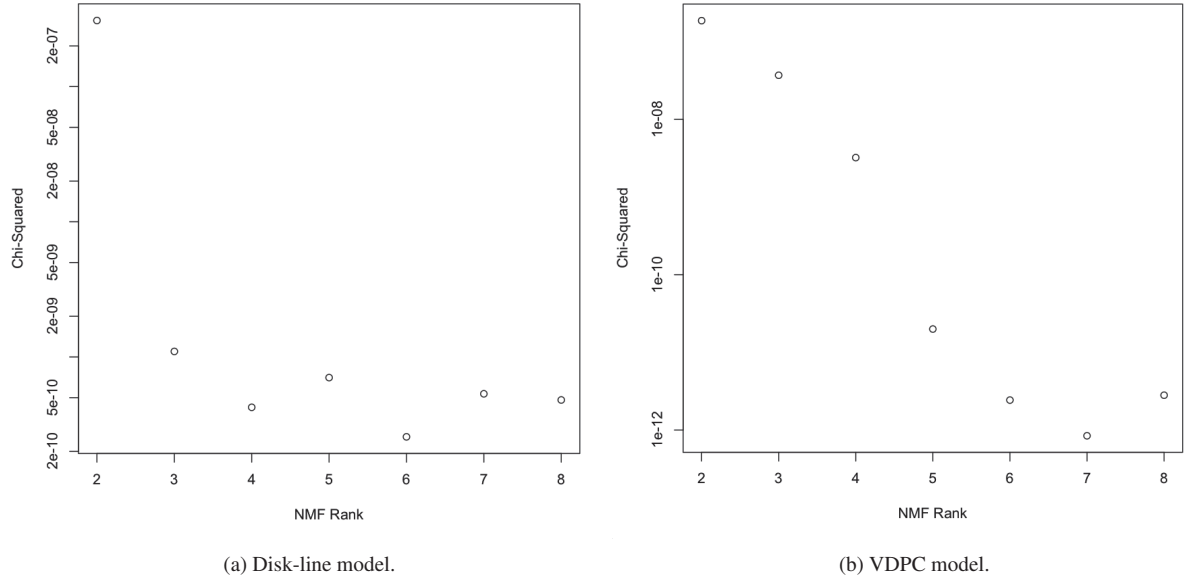


Figure 3: Quality of factorization of simulated spectra by NMF.

3 Nonnegative Matrix Factorization

3.1 Overview of the method

The goal of Nonnegative Matrix Factorization is to factor a matrix into two matrices with nonnegative entries (Paatero & Tapper, 1994):

$$X = WY \quad (4)$$

Where X is the original data with n rows corresponding to variables and m columns corresponding to samples or vectors. All the matrix elements of X have non-negative values, as they correspond to the original photon counts (e.g., not taking logarithm) per spectral bin. W has r columns which may be thought of as basis vectors, while Y consists of coefficients which specify how these basis vectors should be combined in order to reproduce the original matrix. In general, the goal of NMF is to factor X such that $r < m$, thereby reducing the dimensions of the data. NMF is particularly well-suited to the analysis of spectra and other physical quantities due to its lack of negative values. This constraint produces results whose physical meaning can be interpreted in the case of observations such as photons counts which are inherently nonnegative, and the technique has been used in studies of subjects ranging from audio decomposition (Brown, 2003) to X-ray spectral decomposition of neutron stars (Degenaar et al., 2016). Ideally, a spectral decomposition method would break observed spectra into independent components which could be compared to spectral models, but the basis components extracted by NMF are not guaranteed to be independent. This means that the components may be linear combinations of independent sources, such as blackbody radiation from the accretion disk and power-law radiation from the compact source.

3.2 Application to the simulated datasets

We applied Lee & Seung's (Lee & Seung, 2001) NMF algorithm to the simulated disk-line and VDPC data using the R package, "NMF" (Gaujoux & Seoighe, 2010). The algorithm was applied for ranks two through eight, with the number of runs set to 30 for each rank. The quality of each factorization is measured by a chi-squared metric

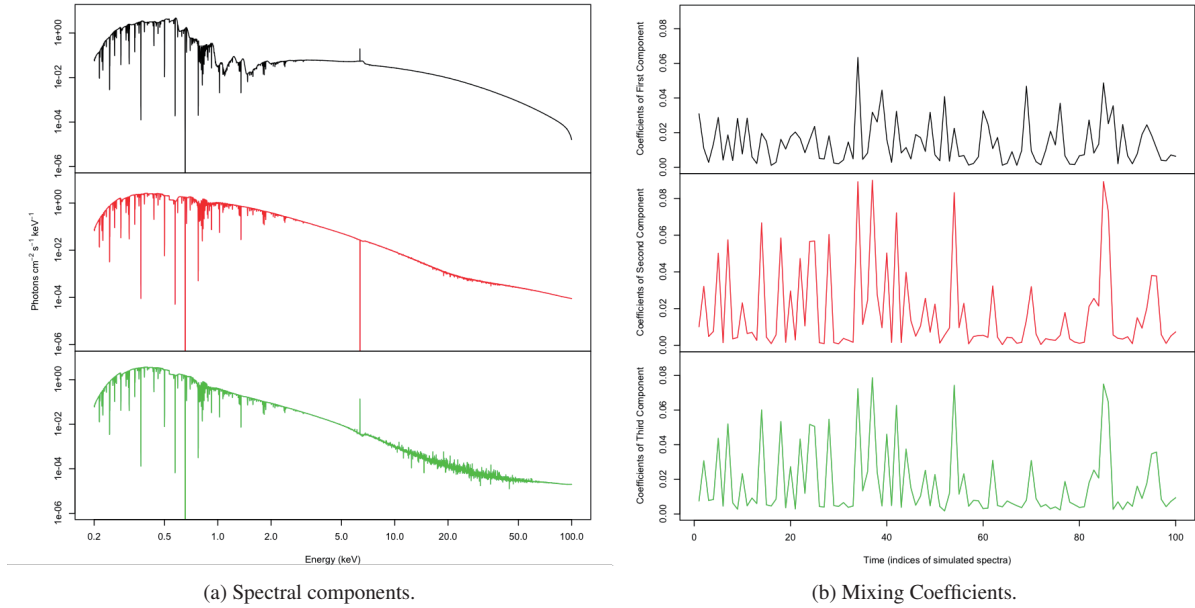


Figure 4: Spectral components of disk-line spectra and corresponding mixing coefficients as extracted by rank 3 NMF.

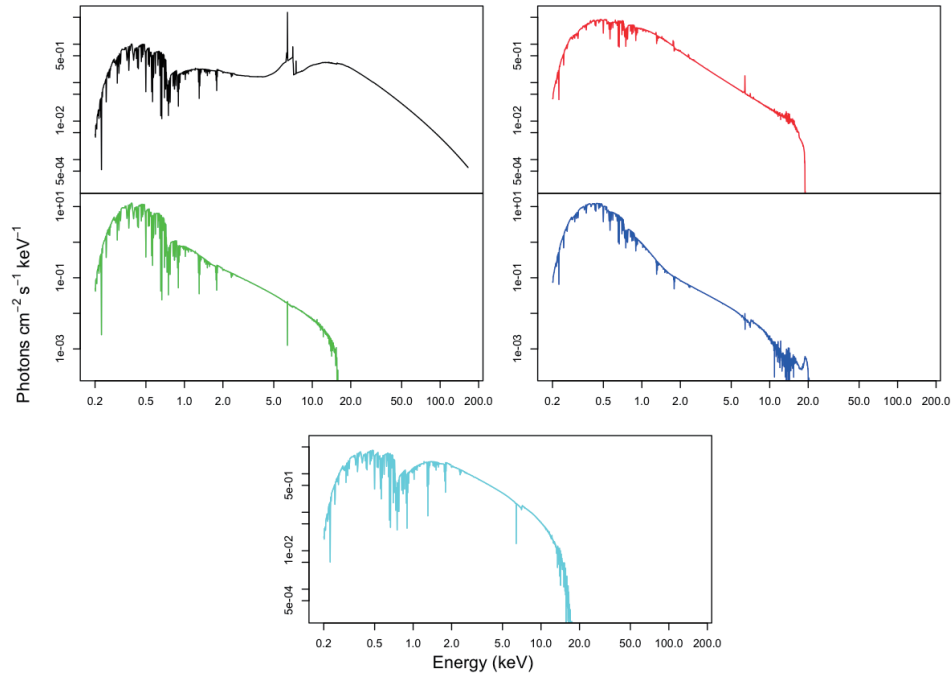
given by

$$\chi^2 = \frac{\sum_i^n (\sum_j^m (X - WY)_{ij}^2)}{nm} \quad (5)$$

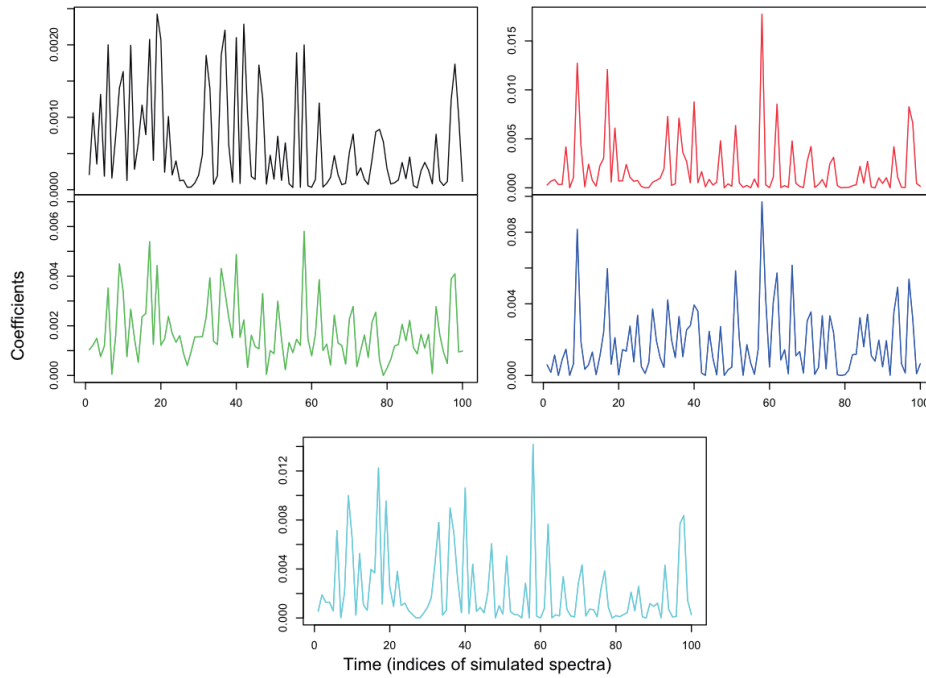
The dimensionality of the input data can be estimated using NMF by observing the evolution of this metric with increasing rank. If a “knee” is observed, where increasing rank results in only small decreases in the metric, then the data can be said to be described well by, at minimum, the number of components which corresponds to the location of the knee (Koljonen, 2015). In addition, the spectral components determined by NMF may be useful in analyzing the mechanisms of X-ray production.

3.2.1 Application to the disk-line model

When applied to data simulated according to the relativistic disk-line model, the results of NMF show a clear increase in the quality of reproduction when the rank of the resulting matrices is increased from two to three (Figure 3a). The chi-squared metric used to measure the quality of factorization decreases by more than two orders of magnitude between ranks 2 and 3, but when the rank of the resulting matrices is increased from 3 to 4, the increase in factorization quality is much smaller. Past rank 4 NMF, changes in factorization quality remain relatively small. Therefore, NMF indicates that data simulated according to the disk-line model may be accurately reproduced by at least three spectral components. Figure 4 shows the spectral components and mixing coefficients resulting from rank 3 NMF. Importantly, NMF does not specify the order of its resulting components, so the order in which we list the components has no significance. The first component, shown in black, appears to be dominated by the blackbody component, which makes up the hump at energies below 1 keV, and the disk reflection component, which makes up the wider hump at energies above 1 keV. Furthermore, an iron edge and line, features explained by reflection off the accretion disk according to the disk-line model, are relatively prominent around 6.4 keV in the first component. The second and third component, in contrast, appear to be dominated by the power-law component with slightly differing contributions by the blackbody component in the soft X-ray regime. These interpretations are further solidified by the resemblance between the mixing coefficients and the simulation



(a) Spectral components.



(b) Mixing Coefficients.

Figure 5: Spectral components of VDPC spectra and corresponding mixing coefficients as extracted by rank 5 NMF.

parameters. It is important to note that although we varied only two components with time, the results of NMF suggest that three spectral components are necessary to fully reproduce the input spectra. This may be due to the third, blackbody component of the model. Since it does not vary in time, it is independent from the other two terms, meaning that it cannot simply be absorbed into the other two components. Thus NMF may require a

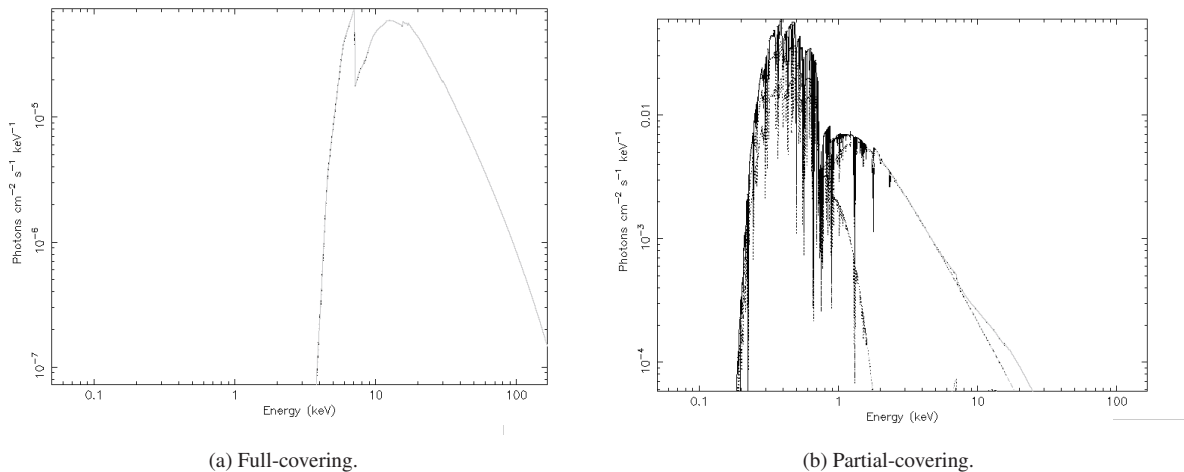


Figure 6: The full-covering and partial-covering terms of the VDPC model.

third component so that, when combined, the total blackbody contribution remains constant. Because NMF has no information regarding the components which produced the input spectra, it is unlikely to separate out a constant blackbody component among all possible factorizations.

3.2.2 Application to the VDPC model

The results of the application of NMF to spectra produced according to the VDPC model are more difficult to interpret. First, in plotting the chi-squared metric of the NMF results (Figure 3b), it is difficult to determine the location of a knee. There is a relatively large increase in the quality of the factorization when the rank is increased from 4 to 5, but the metric continues to decrease until the rank reaches 7, the location of the lowest value of the chi-squared metric. However, the decrease in the metric between ranks 6 and 7 is small compared to the changes between lower ranks. Therefore, the minimum number of essential components appears to be five or six. In Figure 5 we show the results of rank 5 NMF. The terms of the model do not appear to be cleanly split among the components. The first component, shown in black, contains features which resemble both the “full-covering” term with coefficient α^2 as well as the “partial-covering” term with coefficient $\alpha(1 - \alpha)$, shown in Figure 6. Specifically, the strong iron edge feature as well as the hump visible at energies just above the iron edge are features which belong to the full-covering term. The hump occurring at energies above 1 keV on the other hand is a feature visible in the partial-covering term. This feature is also prominent in the fifth component, shown in cyan, and visible but less prominent in the third component, shown in green. All of the components feature the blackbody distribution, dominant at energies below 1 keV, and the second, third, and fourth components appear to be dominated by the power-law component at higher energies. This mixing of terms is also reflected in the mixing coefficients. All of the light curves have similar features, such as a dip around $t = 27$ or the large spike near $t = 58$. While the light curve of the first component does resemble the power-law normalization, the other curves do not appear to directly correspond to either of the parameters. This behavior can be attributed to the fact that the VDPC model is multiplicative while NMF decomposes the input spectra into additive components. Thus one would expect the VDPC spectra to be represented as a linear combination of associated spectral components, as in (3).

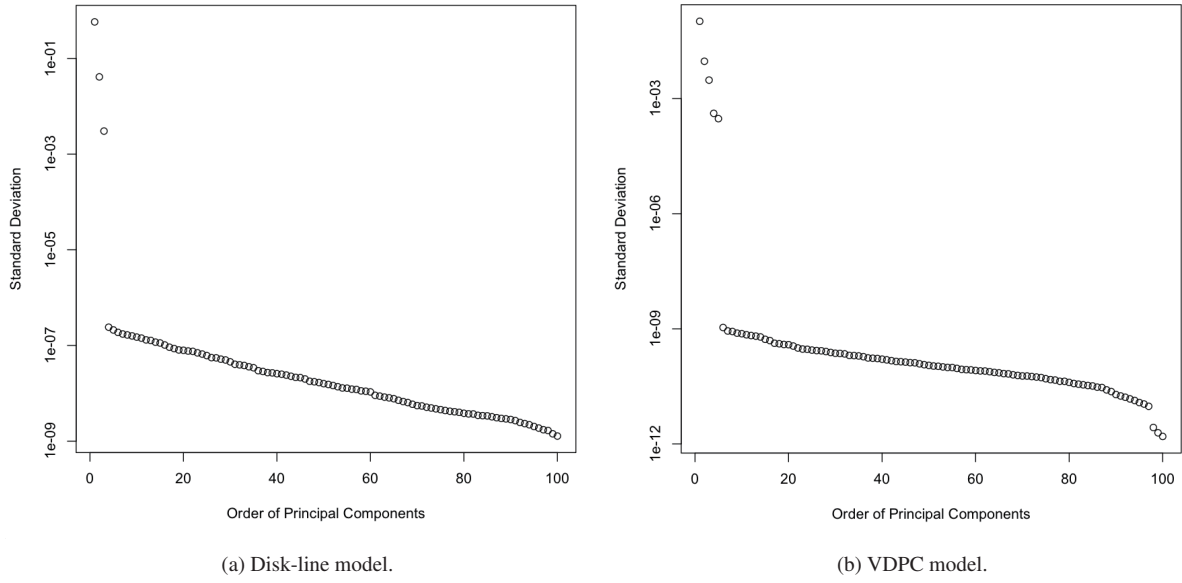


Figure 7: Standard deviation of simulated spectra with respect to principal components.

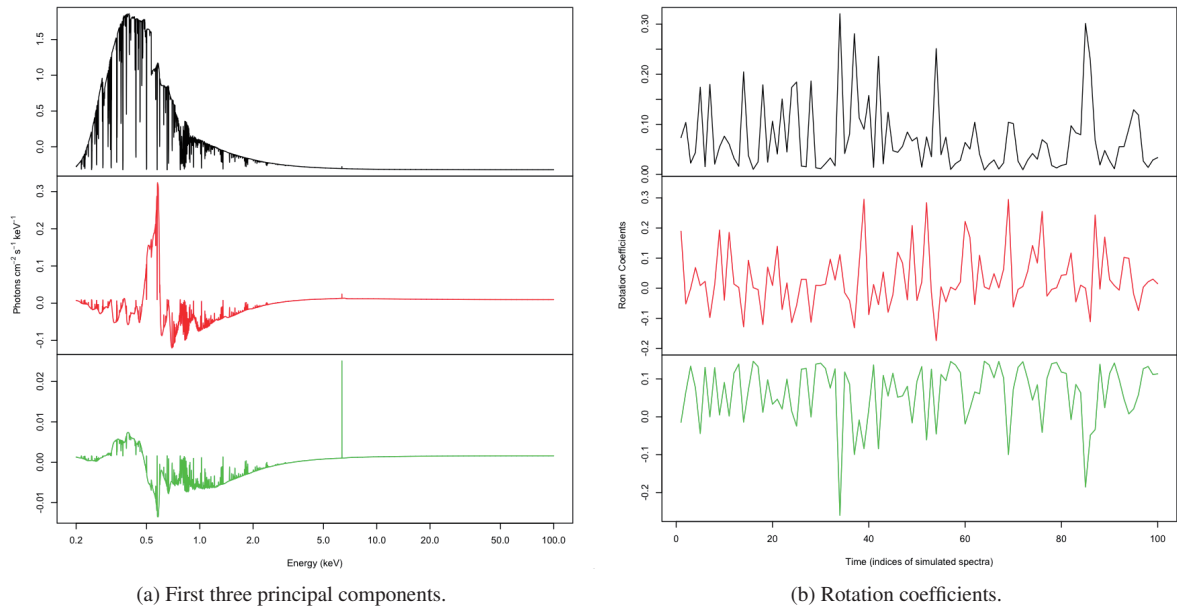
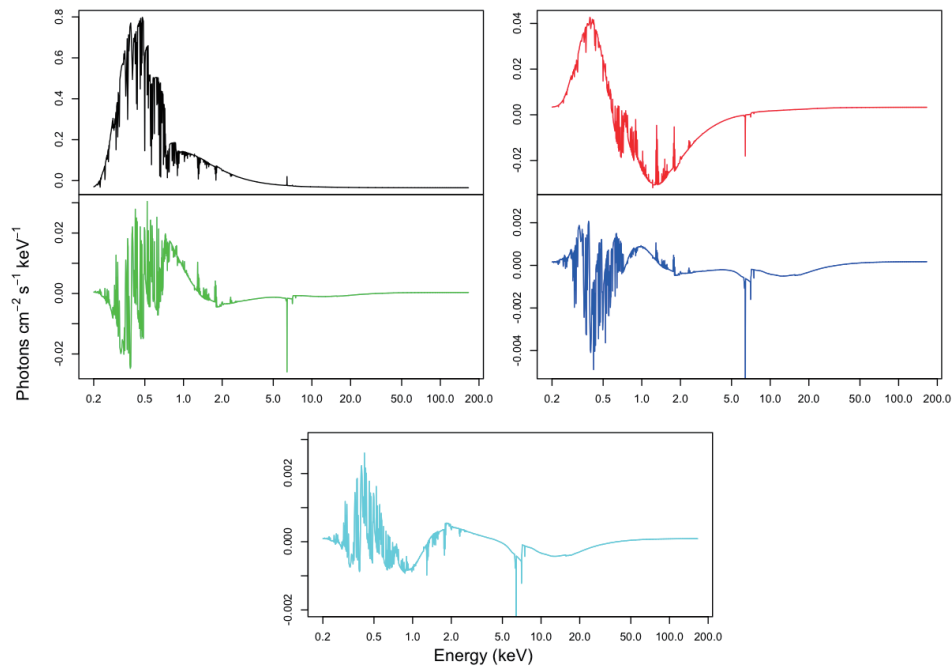


Figure 8: First three principal components of the disk-line spectra and corresponding rotation coefficients.

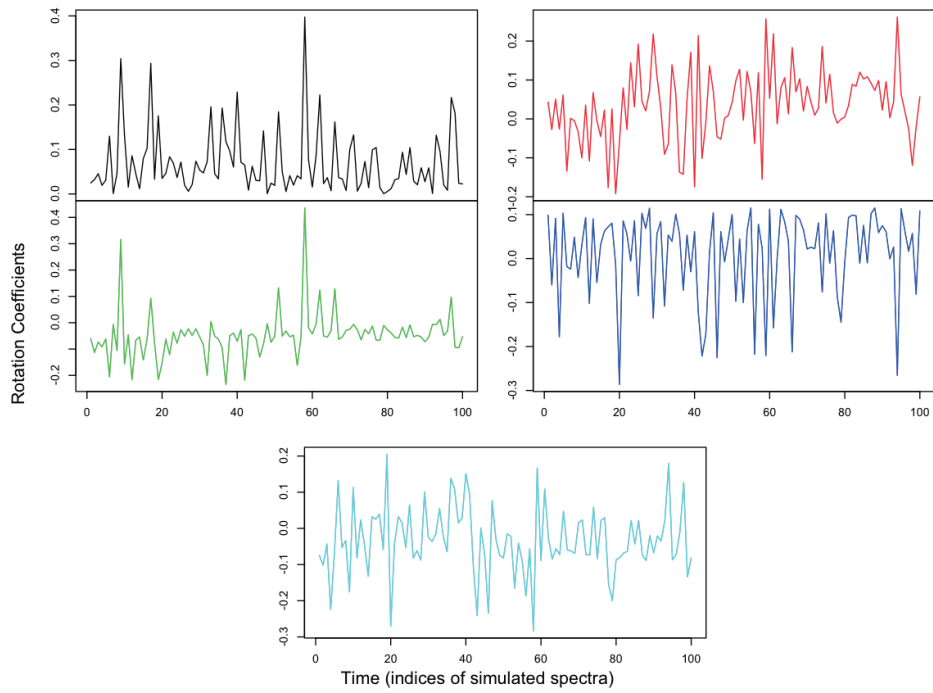
4 Principal Component Analysis

4.1 Overview of the method

Principal Component Analysis is one of the oldest and most widely used techniques of dimension reduction (Jolliffe, 2002). Whereas NMF aims to simply factor a matrix, the goal of PCA is to rotate the coordinate axes of a given data set such that the variance along the resulting axes is maximized (Ivezic et al., 2014, pg. 292). In other words, PCA attempts to rotate the data in such a way as to compress information into as few orthogonal



(a) First five principal components.



(b) Rotation coefficients.

Figure 9: First five principal components of the VDPC spectra and corresponding rotation coefficients.

components as possible. The technique may be written as

$$U = XV \quad (6)$$

Where X is the input data, the columns of which are individual vectors, V is the rotation matrix, and U is the rotated data, the columns of which are the new coordinate axes, or principal components. Unlike the matrices

produced by NMF, U has the same dimensions as X . Instead of determining arbitrary components, PCA ranks its resulting orthogonal axes according to the standard deviation of the data vectors with respect to them. In terms of spectral data, if the input spectra exhibit a high variance with respect to a specified axis or spectral component, then that component may be said to explain much of the variation in the original data. Thus, the dimensionality of the data can be estimated in addition to the relative contributions of each component. Because PCA does not constrain the rotated axes to nonnegative values, it is able to estimate the number of dimensions more accurately than NMF, but the resulting axes may not resemble real spectra. Therefore the spectral components produced by PCA may be more difficult to interpret than those of NMF.

4.2 Application to the simulated datasets

Because of the drawback explained above, PCA's most important function in comparing the disk-line and VDPC models is in determining their dimensionality. If PCA rotates observed spectra into the same number of components as it does data simulated by one of the models, that may be an indication that the model correctly describes the mechanisms of X-ray production in NLS1.

We applied PCA to the simulated data sets using the R function, “prcomp” (R Core Team, 2015). It is important to note that the input spectra are centered before the data is rotated. In addition, the first component returned by PCA is the mean of the centered spectra. As such, higher order components may be considered measurements of deviation from the mean.

4.2.1 Application to the disk-line model

Similarly to NMF, PCA reflects the fact that the disk-line model consists of three components. Again, the presence of the constant blackbody term in the model constrains the results so that two components are not sufficient. As shown in Figure 7a, the standard deviations corresponding to the lowest order three principal components lie between unity and 0.001, while the standard deviations of all higher order principal components lie below 10^{-6} . The fact that the number of the principle components is “three” corresponds to that there are three independent additive terms in equation (1).

While the spectral components are difficult to interpret, other than the prominent emission line near 6.4 keV, due to the fact that large portions of the values are negative, we can look to the rotation coefficients to determine whether certain components correspond to specific terms of the model (Figure 8a). The coefficients corresponding to the first order principal component resembles the power-law normalization, while the coefficients corresponding to the third order principal component resemble the reflection normalization, inverted (Figure 2a). The second order coefficients, however, are not immediately recognizable.

4.2.2 Application to the VDPC model

In contrast to the results of NMF, when applied to the VDPC spectra, PCA clearly indicates that the model has five principal components. The five lowest order components have corresponding standard deviations between 10^{-4} and around 0.1, while all the higher order components have standard deviations around 10^{-9} and below, with a difference of five orders of magnitude between the fifth and sixth order standard deviations (Figure 7b). The fact that the number of the principle components is “five” corresponds to that there are five additive terms in equation (3).

As with the disk-line results, the spectral components, shown along with their corresponding rotation coefficients in Figure 9 do not resemble real spectra because they largely consist of negative values, but certain features are clearly visible. For example, the iron line and edge features are prominent in the fourth and fifth components,

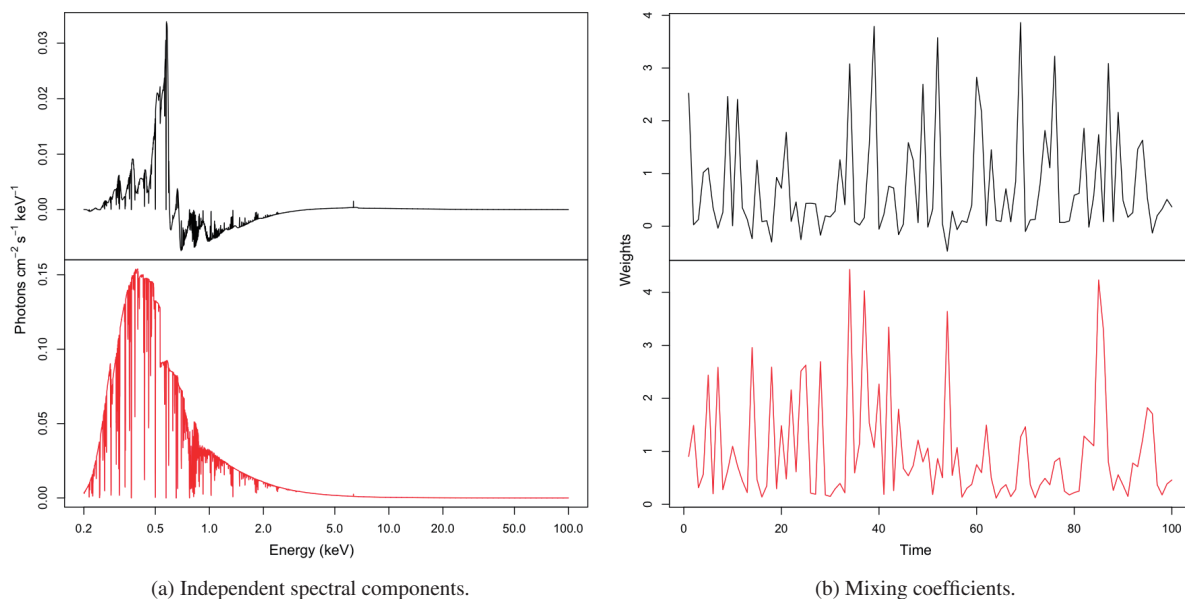


Figure 10: Independent components of the disk-line model determined by ICA.

shown in blue and cyan, respectively. The rotation coefficients do not immediately associate any of the components with specific parameters, providing further indication that the terms of the model are not cleanly separated by PCA.

5 Independent Component Analysis

5.1 Overview of the method

Independent Component Analysis is similar to PCA in that it too rotates a matrix into a new set of coordinates. Unlike PCA however, the goal of ICA is to rotate a set of data samples such that they are statistically independent from one another. Given the matrix, X , the columns of which are individual data samples, ICA attempts to determine matrices A and Y such that the columns of Y are statistically independent and

$$X = AY \quad (7)$$

In addition, ICA computes the inverse of A , call it W , such that

$$Y = WX \quad (8)$$

W and A are referred to as the unmixing matrix and the mixing matrix, respectively, because the observations, X , are considered to be mixtures of statistically independent sources, Y . Statistical independence here is defined by probability densities of the samples. Consider two random variables, y_1 and y_2 , with probability densities $p_1(y_1)$ and $p_2(y_2)$. If the joint probability density, $p(y_1, y_2)$ can be written as the product of the individual probability density functions, then the variables are said to be independent (Hyvärinen & Oja, 2000).

In the case of spectral analysis, ICA cannot be applied directly. The goal is not to determine spectral components whose shapes are independent from one another, but instead to determine components whose light curves, or relative contributions, are independent from one another. In other words, changes in the contribution of one

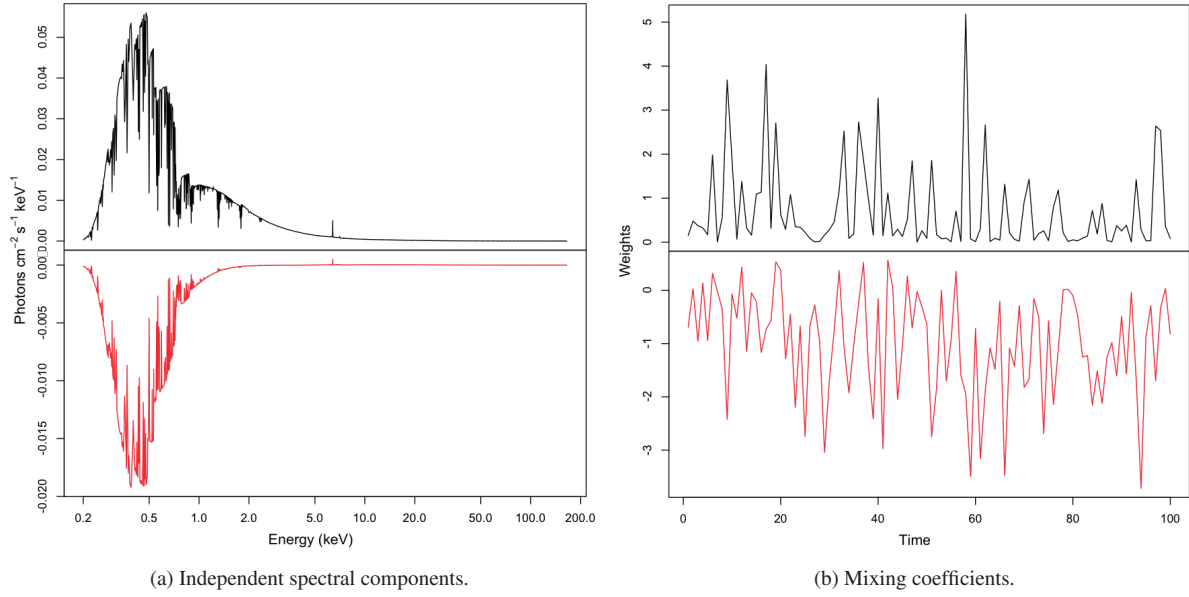


Figure 11: Independent components of the VDPC model determined by ICA.

component should provide as little information as possible about changes in the contributions of the other components. Thus, light curves must be determined by a different method, then these time series may be rotated into independent components. A good first step then is to find the principal components of the spectra via PCA. The rotation matrix returned by PCA may be interpreted as a set of light curves which may be rotated by ICA. In terms of the notation used here, the columns of the rotation matrix V (alternatively, the rows of the transpose of V) are time series which specify how the principal components of X are to be mixed in order to reproduce the observed data. The transpose of V may be rotated into independent sources such that the results of PCA may be rewritten as

$$X = UAY \quad (9)$$

Now, the columns of the product UA contain the independent spectral components which, when mixed by Y , produce the observed spectra contained in the columns of X .

5.2 Application to the simulated datasets

By this method, the results of PCA may be improved upon in order to probe the independent mechanisms of X-ray production in NLS1. ICA may be particularly useful in determining whether the disk-line or VDPC model better explain observed spectra because the VDPC model predicts that the spectral components are not independent. The uncovered component, for example, is by definition directly anti-correlated with the fully covered component. In contrast, the disk-line model is a simple sum of independent components. Therefore, we expect ICA to be able to identify unambiguous independent components in observed data if the disk-line model is correct. ICA should, however, be less successful at determining independent components of the VDPC model.

We first applied PCA to each of the sets of simulated spectra. In order to compare the results of ICA to the simulation parameters, the first two principal components were centered and rotated using the “icafast” function provided in the R package “ica” (Helwig, 2015). After rotation, the results were uncentered. This method results in two independent components whose light curves may be compared to the independent parameters which were used in simulation. As Figures 10 and 11 illustrate, the spectral components determined by ICA are difficult to compare

to the input spectra. Several of these components cross the x-axis, a feature that is unphysical. In addition, the first component of the disk-line results has a shape, like that of the second principal component, which does not resemble the simulated spectra. However, we find that the coefficients corresponding to each of the independent components are in very good qualitative agreement with the simulation parameters. This can be seen quite clearly in the case of the disk-line model, where the first independent component appears to correspond to the normalization of the reflection component, while the second component corresponds to the power-law normalization (Figures 2a and 10b). In the case of VDPC, the coefficients resulting from ICA also resemble the original parameters. Although not as striking as the results in the disk-line case, it is clear that variation in the first independent component corresponds to changes in the power-law normalization, while variation in the second independent component corresponds to changes in the partial covering fraction (Figures 2b and 11b). In other words, with some limitations, ICA is able to reconstruct the parameters which are responsible for spectral variation.

6 Discussion

Our analysis makes clear some of the advantages and disadvantages of PCA, NMF, and ICA with regards to spectral analysis. While the spectral components resulting from NMF are more easily interpreted and spectral features are more visible than in those resulting from PCA, the latter is better for estimating the dimensionality of the data being analyzed. ICA, on the other hand, produces spectral components which are difficult to interpret, but whose mixing coefficients provide a good estimate of the variable parameters in each spectral model.

Determining the number of essential components via NMF involves choosing an appropriate metric to measure the quality of factorization as well as careful consideration of the evolution of this metric with increasing rank². Both of these steps can introduce bias, whereas PCA ranks the principal components by standard deviation and is able to more accurately factor the data due to the allowance of negative values in its results. ICA inherits the latter of these qualities, but is unable to rank components because each independent component is a mixture of principal components.

In addition to comparing the results of NMF analyses of different ranks, the chi-squared metric may also be used to compare the quality of PCA's results and those of NMF. We find that the disk-line spectra reproduced using only the first three principal components have a chi-squared value of 4.2×10^{-15} while the spectra reproduced using the results of rank 3 NMF have a chi-squared value of 1.1×10^{-9} . In the case of the spectra simulated according to the VDPC model, the spectra reproduced with the first five principal components had a chi-squared value of 8.3×10^{-20} and the spectra reproduced with the results of rank 5 NMF had a chi-squared value of 2.0×10^{-11} . In other words, PCA consistently factors the data more accurately by several orders of magnitude.

In addition, the spectral components produced by NMF sometimes exhibit nonphysical behavior in areas where the rank of the input spectra decreases. For example, the shape of the VDPC model at energies above 20 keV may be described entirely by the first component. Accordingly, the other components show zero flux above 20 keV. In the case of the disk-line model, although the first component shows no unexpected behavior, the second and third components returned by NMF exhibit a relatively high level of variability between around 5 and 50 keV. In fact, in some runs of NMF we have observed results in which two of the spectral components show no flux in this region only to return to a level of flux comparable to that of the first component above 50 keV. It may be possible to remedy this behavior by utilizing an NMF algorithm that searches for smooth components (e.g. Yokota et al., 2015), but this may decrease the accuracy of factorization, especially in the soft X-ray regime where multiple absorption and emission lines are observed.

²We have noticed that there are attempts to choose the most desirable NMF rank based on objective statistical criteria, such as Akaike's Information Criterion (AIC) (e.g., Watanabe et al., 2012).

7 Summary and Conclusions

In order to gain a better understanding of data science techniques in the context of X-ray astronomical spectral analysis, and to determine their usefulness in comparing spectral models, we applied Nonnegative Matrix Factorization (NMF), Principal Component Analysis (PCA), and Independent Component Analysis (ICA) to time-series of simulated X-ray energy spectra. Simulation is made according to the competing “disk-line model” and the “variable double partial covering (VDPC) model”, both of which can explain observations of Narrow-line Seyfert 1 galaxies (NLS1) equally well (Figure 1). Only two spectral parameters are made independently variable, such that in the former the variable spectral components are additive (1), while in the latter they are multiplicative (2).

When we applied NMF and PCA to the simulated data we found that PCA is able to clearly determine the dimensionality of the models (Figure 7), while determining the minimum number of spectral components using NMF is more difficult and requires more careful consideration of the results (Figure 3). NMF, however, is able to produce spectral components which are more easily interpreted due to its nonnegativity constraint (Figures 4a and 5a). Our results also indicate that ICA may be used to probe evolution of the independent parameters which drive spectral variability (compare Figures 2a and 10b, and Figures 2b and 11b).

In the case of the disk-line model, PCA gives number of the principle components correctly as three (two variable components and one constant component), and NMF, assuming the rank three, reproduce the individual spectral components. On the other hand, in the case of the VDPC model, PCA requires as much as five principle components while only two parameters are made variable in the simulation, and the results of NMF are more difficult to match to components of the model due to its multiplicative structure. This difference may make NMF and PCA particularly well suited to testing the comparative viability of the additive (disk-line) and multiplicative (VDPC) models. If observed data require less than five principal components by PCA, and the results of NMF resemble the reflection and power-law components, then the additive disk-line model should be favored. If observed data instead requires at least five principal components and the components produced by NMF resemble the partial covering or full covering terms, then the multiplicative VDPC model should be favored.

Before application to observed data, however, the data science techniques should be applied to simulated data which more closely resembles observed data. In order to understand how the techniques respond to noisy data, error should be included in the simulated spectra. As such, our next goal is to produce simulated spectra which reflect the predictions of each model as accurately as possible, in order to gain a thorough understanding of how the above techniques should be expected to perform when applied to real data. Given the promising results we have obtained thus far, our eventual goal is to apply NMF, PCA, and ICA to observed data in order to determine whether the disk-line model or the VDPC model better describes the mechanisms of X-ray production in NLS1.

8 Acknowledgements

The authors would like to acknowledge the Japanese Aerospace Exploration Agency for allowing Sean Pike to undertake this project as a technical trainee under the supervision of Ken Ebisawa. We thank the anonymous referee for several useful comments to improve the paper.

References

- Arnaud, K. A. 1996, *Astron. Data Anal. Softw. Syst.* V, 101
- Boller, T., Brandt, W. N., & Fink, H. 1995, *Astron. Astrophys.*, 305, 53
- Brown, J. C. 2003, in *IEEE Work. Appl. Signal Process. to Audio Acoust.*, 177–180

- Degenaar, N., Koljonen, K. I. I., Chakrabarty, D., et al. 2016, *Mon. Not. R. Astron. Soc.*, 456, 4256
- Fabian, A. C., Iwasawa, K., Reynolds, C. S., & Young, A. J. 2000, *Publ. Astron. Soc. Pacific*, 112, 26
- Fabian, A. C., Nandra, K., Reynolds, C. S., et al. 1995, *Mon. Not. R. Astron. Soc.*, 277, L11
- Gaujoux, R., & Seoighe, C. 2010, *BMC Bioinformatics*, 11, 367
- Harrison, F. F. A., Boggs, S., Christensen, F., et al. 2010, *Sp. Telesc. Instrum. 2010 Ultrav. to Gamma Ray*. Ed. by Arnaud, 7732, 77320S
- Helwig, N. E. 2015, *ica: Independent Component Analysis*, r package version 1.0-1
- Hyvärinen, A., & Oja, E. 2000, *Neural networks*, 13, 411
- Ivezic, Z., Connolly, A. J., VanderPlas, J. T., & Gray, A. 2014, *Statistics, Data Mining, and Machine Learning in Astronomy: A Practical Python Guide for the Analysis of Survey Data* (Princeton Series in Modern Observational Astronomy) (Princeton University Press)
- Jolliffe, I. 2002, *Principal Component Analysis* (Springer Series in Statistics) (Springer)
- Koljonen, K. I. I. 2015, *Mon. Not. R. Astron. Soc.*, 447, 2985
- Lee, D., & Seung, H. 2001, *Adv. Neural Inf. Process. Syst.*, 556
- Mizumoto, M., Ebisawa, K., & Sameshima, H. 2014, *Publ. Astron. Soc. Japan*, 66, 112
- Paatero, P., & Tapper, U. 1994, *Environmetrics*, 5, 111
- R Core Team. 2015, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria
- Tanaka, Y., Nandra, K., Fabian, A. C., et al. 1995, *Nature*, 375, 659
- Watanabe, K., Hidaka, A., Otsu, N., & Kurita, T. 2012, *PLOS ONE*, 7, 1
- Yokota, T., Zdunek, R., Cichocki, A., & Yamashita, Y. 2015, *Signal Processing*, 113, 234

宇宙科学資料データベースの開発

川上 修司^{*1*2}, 本田 秀之^{*1*2}, 小野 縁^{*1*2}, 長木 明成^{*1}, 斎藤 宏^{*1*2}, 平井 寿美子^{*1}, 周東 三和子^{*1}, 竹島 敏明^{*1}

Developments of the space science technical documents, photos, and films database

Shuji Kawakami^{*1*2}, Hideyuki Honda^{*1*2}, Yukari Ono^{*1*2}, Akinari Choki^{*1}, Hiroshi Saito^{*1*2}, Sumiko Hirai^{*1}, Miwako Shuto^{*1}, Toshiaki Takeshima^{*1}

Abstract

We have continued to develop the space science technical documents, photos, and films database since 2007, for the purpose of preventing the dissipation of valuable technical materials that has been accumulated in Institute of Space and Astronautical Science (ISAS) and promoting the share and the effective use of the materials. We report developments in the database and the materials that have been registered in the database.

Keywords: データベース, 宇宙科学, DSpace.

概要

宇宙科学研究所 (ISAS) に蓄積されている歴史的に価値のある貴重な資料を含む技術資料の散逸を防止し、情報の共有化と有効活用を図ることを目的として、2007 年度より宇宙科学資料データベースのシステム構築及びデータ登録作業を続けてきている。宇宙科学資料データベースの構築作業及び登録されている資料について報告する。

1 はじめに

宇宙航空研究開発機構 (JAXA) 宇宙科学研究所 (ISAS) には、JAXA 宇宙科学研究本部、文部省 宇宙科学研究所、東京大学宇宙航空研究所、そして東京大学生産技術研究所時代に遡る約 60 年間以上に及ぶ資料が保管されている。その中には、糸川博士をはじめとして歴代の教官が残した資料など、歴史的な意味をもつ貴重な資料がある。これらの資料のほとんどが、書類は紙ベースで、画像はプリントされアルバムに綴じ、さらにネガフィルムの形態で保存されている。映像の多くは、16mm フィルムとして、フィルム缶に個別に収納されて保存されている。

ISAS の草創期のものから近年のものまで幅広く、膨大な資料を整理するきっかけになったのが、次期固体ロケットの開発研究に向けて、M-V ロケットの関連資料を整理し如何に有効活用を進めるかという課題であった。同時に、ISAS に残された資料を如何に後世に継承するかという点と、所蔵している資料の内容を理解できる人が年々少なくなって来ている点が問題であった。さらに、団塊の世代の大量退職はその問題を更に深刻なものにした。このため、資料の破損・劣化に伴い記録されている情報が失われていくのを防止し、古い資料の整理・保存のため、様々な資料をデジタル方式で記録することし、メタデータ (付加情報, キーワードなど) を付与してデータベース技術を用いて蓄積し、ネットワークを通

doi: 10.20637/JAXA-RR-16-007/0008

^{*} 平成 28 年 11 月 24 日受付 (Received November 24, 2016)

^{*1} 宇宙航空研究開発機構 宇宙科学研究所 科学衛星運用・データ利用ユニット

(Science Satellite Operation and Data Archive Unit, Institute of Space and Astronautical Science, Japan Aerospace Exploration Agency)

^{*2} 技術情報システムチーム (Technical information systems team)

して必要とする資料を閲覧・研究することができるようデジタルアーカイブ[1]を開発することにした。

以下に保有資料概観を示す。これらは、科学衛星運用・データ利用ユニット(C-SODA)で把握している代表的なもので、主に宇宙工学側の資料である。

○ データセンタ保管資料

実験計画書・報告書 A4 ファイル 約 500 冊, A4, B5 冊子 約 360 冊

退官教授資料 ダンボール箱 約 200 箱 (80cm 幅棚 12 段分)

設計会議資料, 承認図等多数

○ 映像記録班保管資料

ネガフィルム 35mm 版 約 39 万コマ

ネガフィルム ブローニー版 約 26 万コマ

検索用写真ベタファイル 約 500 冊

デジタル写真 約 20 万枚

○ M計画室保管資料

M-V 関連資料

M-3 S II 関連資料

観測ロケット関連

その他

○ 対外協力室保管資料

16mm 映画 540 本

16mm 記録フィルム原版多数

○ 高速度カメラ(光学観測等)記録資料

16mm 計測カメラデータ他 約 1,200 缶(フィルム缶入り)

その他 35mm ロケット追跡記録等

これらの資料は、文書・画像・映像の3つの種別に分けられる。文書としては、実験計画書・報告書をはじめとして、会議資料・議事録や承認図、それに退官教授による寄贈資料などがある。画像は、ロケット・衛星・大気球実験各々の組立て・試験・打上げを撮影したものをはじめ、総計 84 万枚程度(スチール約 64 万枚, デジタル約 20 万枚)に上る。映像も、記録映画作品や貸出し用ビデオのほか、未使用の撮影素材も多数保有している。

この他にプロジェクト保管資料や個々の研究室にて保管されている資料も沢山有ると考えられる。いずれこれらの資料もなんらかの手段で管理すべき資料であるので、今後、所として調査を実施し、目録を作成した後、保有量に応じた保管の手段を検討し、対応をする必要がある。

2 宇宙科学資料データベース

これらの資料を限られた期間で可能な限り活用しやすい形に整理・保存することが必要である。資料を整理し、デジタル化し、データベース化して有効活用することが当面の課題であった。更に将来に渡って利用可能な資料管理システムの構築を行うことを考慮した。基本的な方針としては、各種資料のデジタル化を進め、そのデジタル・データを DSpace[2] というソフトウェアを組込んだサーバーに格納して、検索できるシステムを採用した。DSpace とは、内外の大学・研究機関等において幅広く用いられているリポジトリ構築のための標準的ソフトウェアの一つである。日本語化の問題や保守管理の継続性も考慮して、専用のコンピュータに DSpace をインストールして、宇宙研向けに設定を施した状態で導入し、ソフトウェアの年間保守を付けて維持することにした。映像関係資料を先行してデジタル化を行った。映像記録班管理のフィルム(スチール写真)は5年計画でデジタル化を実施した。対外協力

室管理の記録映画関連は3年計画でデジタルビデオ化を行い順次関係者へ公開した。難題は、デジタル化した映像資料、つまり1枚1枚の画像をチェックして検索キーワードを付ける作業であった。この膨大な作業は内容に精通した当時の関係者を招聘して行った。文書データの検索キーワード付けに関して同様の作業を行った。構築しているデータベースのトップページ及びコンテンツを図1及び図2に示す。

保有する文書・画像・映像全ての種類のコンテンツをDSpaceに格納するにあたっては、使い勝手を確保するための工夫が必要となった。紙ベース文書に関しては、OCR機能のついたスキャナで読み取ったPDFファイルを格納し、DSpaceの機能を利用してインデックスを作成して全文検索ができるようにした。画像については、42枚を上限としたブロックに分けたサムネイルで確認し、目的とする画像を概ね特定してから、ダウンロードして解凍するような仕組みにした。映像もサムネイルを設定して表示し、的確な内容説明（キーワード）を付すことによって、スムーズにデータを入手できるようにした。

実験計画書・報告書やM-V関連の会議議事録等についてはDSpaceにほぼ格納を完了している。しかし、各種図面やM-3SII関連資料については、デジタル化が適切か書誌情報のみでよいのか、費用対効果や紙ベース保管でのスペース等を考慮しなければならず、検討を要するため未登録である。画像のデジタル化は完了し、Mロケットと人工衛星関連（嚙合せ・総合試験～フライトオペまで）や大気球・観測ロケットの記録画像等の記録を継続している（図2）。映像データについては、作品化されているものは全てDSpaceに格納済みである。なお、一般向けISASホームページのTV@ISAS[3]やJAXAデジタルアーカイブ[4]でも視聴可能である。

このように資料のデジタル化や検索システムの構築を進める一方、内外の研究機関等におけるアーカイブ・リポジトリ構築をめぐる動向を調査してきた。近年では、海外はもとより国内でも文書の保存管理を推進する動きが幅広くみられる。全国の大学・研究機関の多くは、図書館の業務の一環としてリポジトリ構築に取り組むだけでなく、専任の教職員を置いてアーカイブの整備を進めている。また、すでに関連学会として日本アーカイブズ学会や記録管理学会が活動しており、アーカイブの運営を司るアーキビスト養成のための課程を設置している大学もある。このような動きの背景には、国全体で説明責任に関する意識が高まってきたことが挙げられる。2009年6月には、「公文書等の管理に関する法律」も定められた。同法は、各省庁における「公文書」の管理方法を定めたものであり、詳細なガイドラインが制定されている。ISASにおいても、このような動向をフォローしつつ、このデータベースを運用していかなければならないと考えている。



図1 宇宙科学資料データベーストップページ

宇宙科学資料データベースコンテンツ

- **SES刊行物**
 - [1. 衛星・探査機／打上げロケット](#) [294]
 - [2. 観測ロケット](#) [324]
 - [3. 地上燃焼試験](#) [196]
 - [4. 再使用ロケット実験機](#) [16]
 - [5. 大気球](#) [200]
 - [6. 衛星・ロケット等実験概要](#) [92]
 - [7. 装置・施設関連](#) [6]
 - [8. その他](#) [40]
- **記録映像**
 - [作品\(一般公開可\)](#) [277]
 - [素材](#) [583]
 - [光学観測](#) [51]
 - [その他](#) [0]
- **記録写真**
 - [衛星・探査機／衛星打上げロケット](#) [8887]
 - [観測ロケット](#) [9234]
 - [地上燃焼試験](#) [3882]
 - [再使用ロケット実験機](#) [485]
 - [大気球](#) [9651]
 - [SFU](#) [77]
 - [SEPAC](#) [195]
 - [ペネトレータ書入実験](#) [752]
 - [有翼飛翔体](#) [516]
 - [ロクーン](#) [129]
 - [道川ロケット実験](#) [975]
 - [施設](#) [885]
 - [チャンパー製作過程](#) [30]
 - [ペンシルロケット50周年\(2005年\)](#) [141]
 - [その他](#) [14]
- **技術文書**
 - [1. M-V](#) [85]
 - [2. M-3SII](#) [43]
 - [3. その他:所内チーフ会議\(1985/11~2000/11\)](#) [79]

登録済み資料件数

SES刊行物(1168冊)
 記録映像(910件) うち作品(277件)
 記録写真(約65.5万枚)
 技術文書(207件)

10

図2 宇宙科学資料データベースコンテンツ

3 宇宙科学資料データベースシステムの構成

本システム立上げ当初(2007年)は独自サーバーに DSpace とツール群を一つのシステム(以下 DSpace システムと呼ぶ)として導入した。その後、2013 年の科学衛星データ処理システム(DPSS)換装に併せて、ハードウェアと OS、及びデータバックアップ用ディスク装置を DPSS の一部として組み込み、DSpace のバージョンアップ(V1.4 から V1.7 に)と登録データの移行を実施した。それらの作業は、将来的に保守が継続的に実施できるように業者に依頼して行った。DSpace サーバーは公開系と試験系の2台構成とした。さらに、資料・画像データの前処理用 PC と利用者が個別に DSpace 上でそれらを検索可能とするような専用クライアント PC を用意した(図3)。

持ち込まれた資料は、一旦前処理用 PC のディスクにコピーされ、ファイル名の調整やメタデータ(登録するデータの関連情報:ここでは、撮影日、撮影場所、映像に写っている物・人物の情報等)の作成、登録データの作成と試験系サーバーへのアップロードを行う。ここでメタデータが所定の場所に表示できているか、間違いがないか等の確認を行い、問題なければ公開系のサーバーにアップロードする。整理を終えた資料はオリジナルとして保存され、また公開系と試験系サーバのコンテンツや作業用 PC のデータは、定期的にバックアップが取られるようになっている。

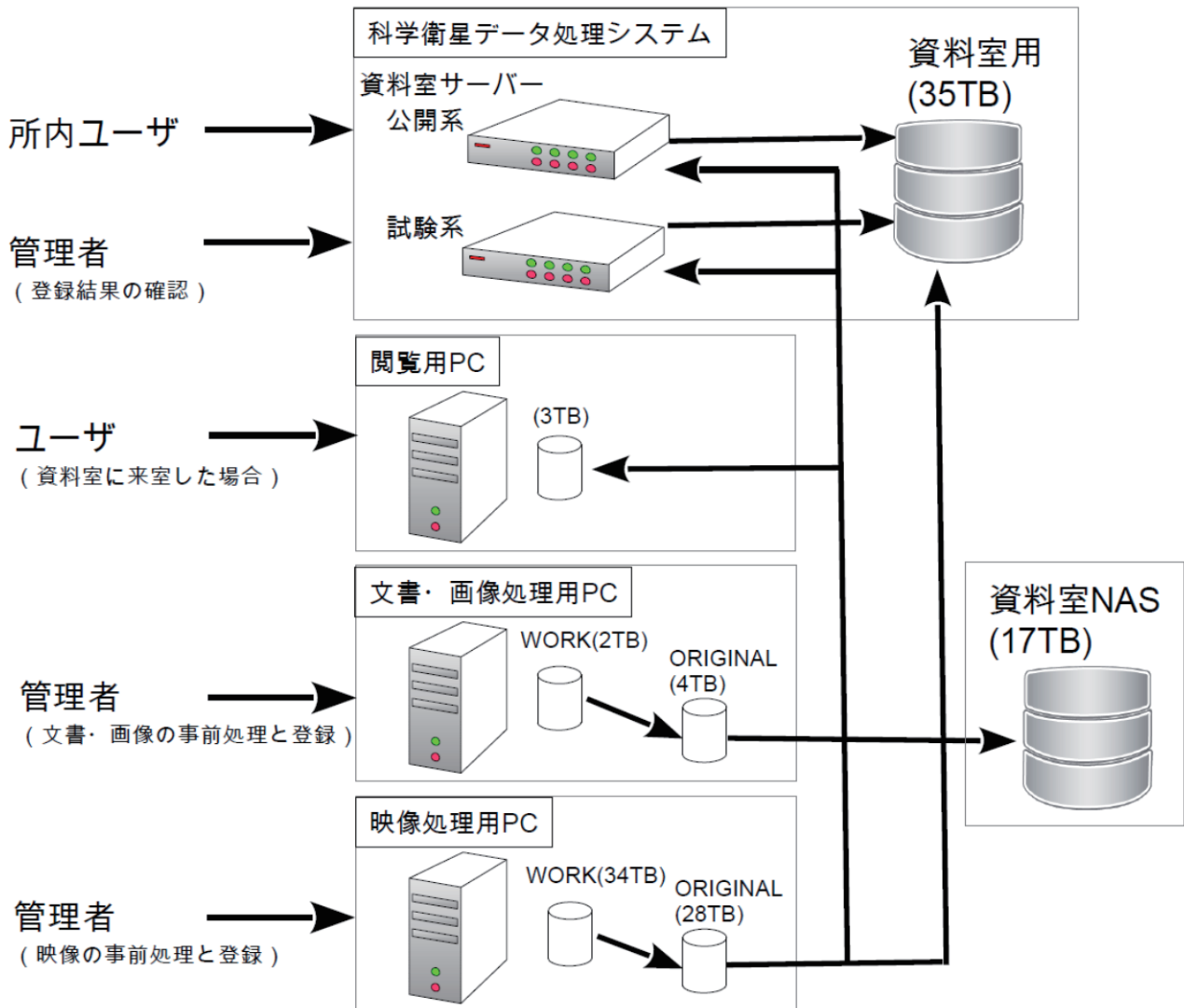


図3 システム構成とデータの流れ

3.1 情報を整理格納するにあたっての事前検討

3.1.1 登録すべき情報

DSpace のコンテンツは、投稿・レビュー・アクセスといったアイテム（登録の最小単位）を管理する単位である“コミュニティ”毎に登録される。各コミュニティには、サブコミュニティ、コレクションを作成でき、アイテムを登録できる。Dspace 本来の使い方としては、コミュニティは、研究室、研究センターのようなコンテンツを管理する単位を念頭に置いてあるが、ここでは、コミュニティ＝大分類、サブコミュニティ＝中分類、コレクション＝小分類という使い方をしている。DSpace にアイテムを登録する場合には、定められた書式でメタデータを整理しファイル化しておく必要がある。この中には必須の項目とシステムが自動的に付加する項目、その他を合わせて計 87 項目ある。購入した Dspace システムにはウェブ経由で登録作業を容易に行えるようにしたアイテム一括登録プログラムが提供されており、先に述べた項目から選択された一般的に必要な項目（計 35 項目）だけが登録できるようになっていた。この中には利用者が DSpace で情報を閲覧する場合に直接確認できる項目とそうでないものがあり、また検索対象に含めることができるものとそうでないもの等があった。そこで、各コミュニティ（登録の大分類）間で登録に使用する項目を一致させ、またそこに書くべき内容の調整を行い、全体でバランスのとれた表示内容や検索結果が得られるように考えた。また全てのアイテムは画像・文書・映像にかかわらず、各々の特徴を表す JPEG 画像（映像であればタイトル画面、文書であればその表紙）

も併せて登録し、DSpace の自動サムネイル作成機能を利用してアイテム一覧表示画面に表示されるようにした。

3.1.2 トップ画面の表示内容

トップ画面は5つのコミュニティ（大分類：当サイトの使用法について、SES (science and engineering of space) 刊行物、記録映像、記録画像、技術資料）とその配下の幾つかのサブコミュニティ（中分類）のみの2レベルまでしか表示されないようにプログラムを一部変更し、利用者がサイト全体の構成を把握しやすいようにした。なお、「当サイトの使用法について」は、アクセス制限はなく誰でも見ることができ、ユーザ登録の方法や効率の良い検索方法の説明、オリジナル資料へのアクセス方法などを提供している。

3.1.3 アクセス権

本来 DSpace は情報を広く公開するためのアプリケーションであり、登録時の既定の権限には Anonymous READ（誰もが参照可）が付与される。また、アイテム単位で細かく参照可否を設定するやり方は運用や保守の負担が大きく、現実的ではない。そのため、管理者権限を持つ利用者を除いて、サーバー上の各アイテムに参照権限のみを持つ Guest グループを作成し、登録利用者全てをそれに属させることにした。従って、新規利用者には個別アカウントを作成すると同時に Guest グループに属させることにより、容易に各アイテムへの参照権限を付与できる。現在登録されているコンテンツは、既に一般公開されている記録映像の一部が Anonymous READ になっているものを除いて、他は全て Guest READ に設定して登録利用者のみが参照できるようになっている。

3.1.4 記録画像の登録事前作業

記録画像は、SES 刊行物・記録映像・技術資料とは異なり、1画像1アイテムとしてメタデータ付きで登録することは非効率であり、また特に古い画像であれば、1枚1枚の画像には検索で一度に多数の画像が抽出されるようなメタデータしか付加できないことが予想された。そこでメタデータを作る過程で工夫し、必須と決めた項目が同一であればそれらをひとくくりにして登録し、検索時の利用者の利便性に配慮すると同時に登録点数の削減による作業性の向上に寄与するような方法を考えた。具体的には、撮影日付（刊行日付）と撮影内容の計5項目（号機/衛星名、実験名/施設名、撮影項目大分類、撮影項目中分類、撮影項目小分類）が同一の場合には最大42枚を一つにまとめる方法で、これらはいずれも利用者画面に表示される内容である。

この作業を開始する前に、画像のファイル形式は JPEG で統一した。また全ての画像にユニークな名前をつけることにした。これは既存のベタ焼きアルバムやオリジナル画質の画像との対応をつけ易くするため、「プロジェクト名_アルバム番号_画像番号.jpg」の3つの要素でファイル名を表すことにした。プロジェクト名は MV5 や S310 などとし、作業や画像利用者にとって分かりやすい分類名としたが、メタデータには正式名称の M-V-5 や S-310 とした。アルバム番号は既存のベタ焼きアルバムの番号に一致させ、そのない古い画像やデジタルカメラで撮影した画像は、新規にアルバム番号を付与して台帳管理することにした。画像番号は原則既存のファイル名をそのまま採用したが、特殊記号などが含まれている場合は処理プログラムの動作に支障のない文字に変更した。またアルバム番号や画像番号は前に'0'を付加するなどして桁数を統一し、画像の番号順と配列時の名前の並び順が同一になるようにして作業性に配慮した。

DSpace 上で配布する画像は、通常必要とされる解像度を考慮し、一律 600×400 ピクセルにリサイズした（なお資料室においては、閲覧用 PC でオリジナル画質の画像の閲覧や取得が可能である）。また Exif 情報は撮影日時 (DateTimeOriginal) 以外の内容を削除し、著作権表示として Comment に "C-SODA/ISAS/JAXA" を追加した。これらの処理の後、メタデータの内容によりひとくくりまとめて zip 圧縮する処理を行い、各画像のサムネイルをまとめた JPEG ファイル及びアイテム一括登録プログラム用データファイルも同時に作成した（先に述べたように、この JPEG ファイルは DSpace に登録されると DSpace の機能によって再度サムネイル化され、アイテム表示画面にサムネイルとして表示される）。このように作業者がメタデータ作りに専念できるよう、また単純ミスなどが発生しないように、機械的にできるデータの書式チェックやマッピング、変換等を行う一連のプログラムを作成した。

3.1.5 アイテムの一括登録

購入した Dspace システムにはアイテム一括登録ツールが付属していたが、これは一つのコレクションに複数のアイテムをまとめて登録することができるだけである。実際の作業では、必ずしも一つのコレクション単位で事前準備や登録作業をしているわけではなく、関連する複数のコレクションの写真を同時並行して処理している場合が多いので、それらの処理結果を一括で登録できると作業効率が上がることが期待できる。そこで、一括登録プログラムにかぶせて、コレクションをまたがってアイテムを登録可能にするプログラムを作成した。

3.2 コミュニティやコレクションの構成

最上位のコミュニティ配下のサブコミュニティ構成は、5つのコミュニティ（当サイトの利用について・SES 刊行物・記録映像・記録画像・技術資料）で粒度を同じ程度にし、また特に SES 刊行物と記録画像は似たような分類内容になるので、題目や並び順が同じになるようにして、利用者の利便性を高めた。

3.3 セキュリティー上の課題

3.3.1 公開範囲

当初宇宙科学資料サーバーはネットワーク構成上公開系領域に設置されていたが、2010 年 7 月の所内の情報システム運営委員会で ISAS 内限定公開にすべきとの議論の結果を受けて、機構内向けのネットワークに移動した。またその後の ISAS 内の議論を受けて、登録すべきアイテムは、機構内限定の情報として扱い、公開範囲は ISAS 内とすることになった。ただし、30 年以上経過した資料は、一般情報として扱えることにした。ネットワーク上の制限も ISAS の IP アドレス範囲内のみとしている。また管理者として必要な ssh でサーバーにアクセスできる IP も個別に制限している。このような事情から、ハーベスティング（外部公開のための書誌情報の収集）機能には対応しないこととした。

3.3.2 アイテム個別の対応

コレクション単位あるいはアイテム単位のアクセス制限は、運用管理が煩雑になるため行わない方針である。しかしアイテムの登録はそれぞれのコミュニティ担当者によって手動で行われるため、アイテム・コレクション・コミュニティの作成時やアイテムの登録時に設定ミスが発生する可能性があるため、これを防止するためにコレクションやコミュニティ作成時に権限も自動的に付与するプログラムを作成した。念のため、最終的にサーバー上の全アイテム・コレクション・コミュニティをスキャンして、不適切な権限設定が見つければ修正している。さらにコンテンツがアカウント保持者の手を離れ外部に流出するようなことがあった場合の追跡を可能にするために、電子透かしや権利表示等を導入することも検討されたことがあるが、現時点では実現に至っていない。

3.3.3 ユーザ管理

試験的に個別アカウントを付与した際には、アカウント保持者限定の公開である旨を周知して利用をおこなっていた。所内公開になってからも、利用したい方のアカウント作成要望を担当職員が受け付けて手動で作成し、所内限定公開である旨を周知して使用させている。これらのアカウントは Guest グループに所属し、全てのコンテンツへの参照権限のみを持たせている。

3.4 DSpace の使用感

DSpace はサーバーの換装に対しては当初からよく考えられており、比較的簡単に新バージョンに移行ができる。因みに試験的に V1.7 から V5.5 にデータの移行を行ってみたことがあるが、インターネット上の情報を利用することで、問題なく実施できた。また、当初から多言語に対応できていたり、画面表示内容はかなりの部分がカスタマイズ可能で、しかも設定ファイルの書き換えのみで良いように作られているため、換装や移行も容易に行なうことができる。但し設定はかなり複雑で、適切でない場合は表

示が崩壊することもある。

DSpace は登録されたアイテムを効率的に公開するシステムであり、ここで示すような登録ユーザのみに限定した使い方は少数派ではないかと考えられる。このため注意してアイテムを登録しなければ、既定の設定 (Anonymous READ) が適用されて誰にでも見えてしまうことにもなるので、先に述べたように定期的に権限の設定状況の確認が必要と考えられる。

コミュニティとコレクションの表示順は、DSpace の既定のままであるとアルファベット順になり、何らかの設定に依って任意の順には変更できない。項目の表示順はサイトの事情に依って任意に並べ替えたいので、先頭に付加する英数字に依って制御するように工夫しているサイトも有るが、それも画面上に表示されるため煩わしい。そこで、ここでは参考文献[5]に解説してある方法を取り入れ、先頭に数字を付加して表示順を制御するが、画面上には表示させないようなプログラムの変更を行っている。

画像のように多数のアイテムを効率よく登録する手段、その中から目的の画像を素早く見つける方法、登録語句の検索機能等、登録の手間と利用者の便を図ることの両立が必要となる。DSpace そのものの検索機能はバージョンが上がるに連れ改善されてきており、何らかの形で情報が登録されていれば、それを効率よく検索することが可能となることが期待できる。但し検索のためのインデックスを作成する方法には日本語ならではの問題があり、形態素解析プログラムを使用するか N-gram を使うか、或いは両者の利点を利用する方法をとるかによって検索性能に違いが出てくる場合がある。形態素解析プログラムを使う方法は、我々の場合のように比較的限定された特殊な用語を使用する場合に、辞書を維持管理することによって日本語として意味のある最小単位にまで正確に分解可能となることが期待できるが、新しい用語が出てくるたびに辞書を改定しインデックスを再作成しようとする、その手間は小さくない。一方 N-gram 方式では、文字列を N 文字ずつに機械的に分割してインデックスを作成するため辞書を必要としない利点があるが、インデックスが大きくなることや意味的には間違った検索結果が得られる場合が出てくることが欠点である。なお、形態素解析プログラムは DSpace V1.4 に導入されており、N-gram は N=2 として DSpace V1.7 で使用されている。V1.4 の時にはしばしば検索にかからない用語が出るがあったため、既登録用語を抽出して辞書登録を試みたことがあった。V1.7 では比較的適切な検索結果が出てくるように感じられてはいるが、利用者が増加した場合に同様に感じられるかは定かではない。

コンテンツを整備して行く段階で、ISAS 内の有識者 (コンテンツ提供者を含む) にアカウントを付与したり、また頻繁に専用の検索端末で画像等を借り出す人にもアカウントを付与し、試験的に使ってもらい使用感を聞くことも行ってきた。試用を依頼した ISAS 内の有識者からいくつかのコメントがあった。その中で対応したのが次の2点である。

- (1) トップページに「当サイトの利用について」というコミュニティを作り、その中に利用者ガイドを登録しているが、DSpace の既定の動作では各コミュニティやコレクションのトップページに移動しても、その状態ではアイテムの一覧は表示されないようになっている。表示させるためには中央部のブラウザメニューから表示内容 (キーワード・タイトル・著者・発行日、撮影日のいずれか) を指定する必要がある。初めて利用する場合にはそれが分からずとまどう人が多いため、各トップページの上部に文字色を変えて簡単な操作ガイドを表示するようにした。
- (2) DSpace 上に表示される画像関係のアイテムとオリジナルアルバムの画像やネガとの対応をつけやすくするため、画面上に表示されるタイトル部分に撮影場所とアルバム番号も表示させるようにした。

4 まとめと今後の課題

DSpace を用いて宇宙科学資料データベースのシステム構築に取り組み、資料の格納作業と平行して、一部試験運用を開始した後、ISAS 内の利用者へ公開できるようになった。同時に、ISAS が収集・保存・公開すべき資料を選別するガイドラインの検討が行われ、一部は所内に周知される状況になりつつある。DSpace の運用開始に伴い、資料の継続な受け入れ手順や、公開方法、公開範囲等については、ガイドラインを作成しつつ運用を実施する予定である。

このように、ISAS 内の資料の扱いについては、徐々に解決されつつあるが、課題はまだ残されている。例えば、教職員の退職時に資料が提供された場合などに、どのように受け入れていくか等、現時点では決まった方針がない。ISAS に蓄積される資料は ISAS の資産であり、その資産を有効に利活用していくため、決して散逸してしまうことのないよう、システムに格納すると同時にカタログを作成して

ISAS 内部でアクセスできるようにするプロセスを確立する必要があると考える。今後、ISAS 内で幅広く議論しつつ、宇宙科学資料の保存及び広い活用を促す活動を展開する必要があると考えている。

謝辞

資料の取得・提供・保存していただいた皆様，特に，データセンター，映像記録係，M計画室，宇宙科学企画情報解析センター，宇宙科学資料室（宇宙科学本部システム開発部）の歴代の方々に感謝いたします。また，現在の宇宙科学資料データベース構築にかかわった新倉克比古氏，吉山京子氏，佐藤靖氏，平田玲子氏，中西 功氏に感謝いたします。

参考文献

- [1] デジタルアーカイブの構築と技法 谷口知司 編著 晃洋書房
- [2] <http://www.dspace.org/>
- [3] http://www.isas.jaxa.jp/tv_isas/index.html
- [4] <http://jda.jaxa.jp/>
- [5] http://d.hatena.ne.jp/zuki_ebetsu/20120429/1335669965

地球観測衛星データ処理におけるJAXAスパコン活用の効果検証

齋藤 紀男^{*1}, 上田 陽子^{*1}, 中西 功^{*1}, 仁尾 友美^{*1}, 小西 利幸^{*1}, 斎藤 進^{*1},
田中 誠^{*2}, 井口 茂^{*2}, 井上 淳一^{*3}, 鳥居 雅也^{*3}

Verification of Extreme Time Reduction for Earth Observation Satellite Data Re-processing with JAXA Supercomputer

Norio Saito ^{*1}, Yoko Ueda ^{*1}, Isao Nakanishi ^{*1}, Tomomi Nio ^{*1}, Toshiyuki Konishi ^{*1}, Susumu Saito ^{*1},
Makoto Tanaka ^{*2}, Shigeru Iguchi ^{*2}, Junichi Inoue ^{*3}, Masaya Torii ^{*3}

Key Words : Earth Observation, Data Processing, Supercomputer

Abstract

Observation by satellites orbiting around the Earth is an important tool for scientific study of global changes. Observation data are reprocessed almost every year according to progress in computation model to obtain physical information from observation. The amount of observation data grows day by day. This yields longer and longer time to reprocess the data. However, data latency is an important subject because quick access to earth observation data considerably encourages research into global environment changes and their effects to our society. On this point, JAXA Supercomputer System Generation 2 (JSS2) helps us to solve the problem since it can enhance throughput of data processing by using a number of computing nodes. First, we have utilized JSS2 for reprocessing of IBUKI (GOSAT: Greenhouse gases Observing Satellite) observation data. IBUKI was launched on Jan. 23rd, 2009, and the data during 6.5 year period has been reprocessed in Nov. 2015. We obtained incredibly high throughput; 30 nodes (360 cores) of Pre-Post system of JSS2 were assigned and resulted in 33 times faster latency. Next, we have used JSS2 for GPM (Global Precipitation Measurement) Core Observatory that was launched on Feb. 28th, 2014. The data during 2.2 year period has been processed and 4 times faster latency was achieved in June 2016 by using 3 nodes at a maximum. As a conclusion, it is shown that JSS2 is a powerful tool to improve data latency of Earth observation data.

概要

人工衛星による地球観測は、地球環境の変化を科学的に研究するための重要なツールである。観測データは、1年に1回程度、物理量を算出するための計算モデルが改訂されることに伴って再処理される。観測データの量は日に日に増加するため、再処理に要する時間は増加し続ける。したがって、観測データをユーザへ提供するまでのレイテンシが重要な課題となっている。なぜならば、地球観測データをできるだけ早く入手できることは、地球環境変動の研究の進展に大きく寄与できるからである。この点で、JAXAが有するスーパーコンピュータシステム（JSS2: JAXA Supercomputer System Generation 2）を活用することが有効である。なぜなら、JSS2の計算ノードを大量に使用することで、データ処理のスループットを大きく改善できるからである。初めに、温室効果ガスの全球観測を行っている衛星「いぶき」（GOSAT: Greenhouse gases Observing Satellite）におけるJSS2利用の成果を報告する。2009年1月23日に打ち上げられた「いぶき」の6年半に渡る観測データを2015年11月に再処理した。JSS2プレポストシステムの30ノード（360コア）を使用し、従来の手法で再処理した場合に比して33倍のスループットが達成できた。次に、GPM主衛星（GPM: Global Precipitation Measurement, 2014年2月28日打上げ）でのJSS2利用について述べる。2016年6月に2年3ヶ月の観測データのレベル1処理を最大3ノードを使用して実施し、従来の4倍のスループットを達成した。JSS2は地球観測データの提供時間を大幅に短縮するために極めて効果的なツールとなることが示された。

doi: 10.20637/JAXA-RR-16-007/0009

^{*} 平成 28 年 11 月 24 日受付 (Received November 24, 2016)

^{*1} 宇宙航空研究開発機構 第一宇宙技術部門 衛星利用運用センター

(Satellite Applications and Operations Center, Space Technology Directorate I, Japan Aerospace Exploration Agency)

^{*2} 日本電気株式会社 (NEC Corporation)

^{*3} 富士通株式会社 (FUJITSU Limited)

1 はじめに

人工衛星による地球観測では、日常の定常的な観測を遂行する他に、取得データの解析アルゴリズムを改良し、精度・確度の高い情報をユーザーへ提供する取り組みを日々行っている。アルゴリズム改良に伴い、観測データを解析し直す作業（再処理という）が行われるが、観測データは日々蓄積され増加し続け、その結果、再処理に要する時間は観測期間に比例して増大する。

温室効果ガス観測技術衛星「いぶき」

（GOSAT: Greenhouse gases Observing Satellite）を例にとり、人工衛星から地球を観測する様子を図1に示す。衛星の進行方向を指してアロングトラック方向という。

「いぶき」に搭載された2種類のセンサーは、それぞれ異なる動作をしている。TANSO-CAIという可視画像センサーは進行方向に垂直な方向（クロストラック方向）に帯状に連続観測を行う。一方、フーリエ分光器であるTANSO-FTSは、図1に楕円印で示すように、飛び飛びに複数点を観測する。CAIによる一日分の観測データの例を図2に示す。

次に、いくつかの地球観測データの例を示す。図3-aは「いぶき」が観測した大気中の二酸化炭素濃度の月平均を示し、年々、その濃度が上昇していることを示す。次に図3-bは、GPM主衛星に搭載された二周波降水レーダ（DPR; Dual frequency Precipitation Radar）の観測例である。最後に、図3-cに全球降水観測計画（GPM: Global Precipitation Measurement）から提供して

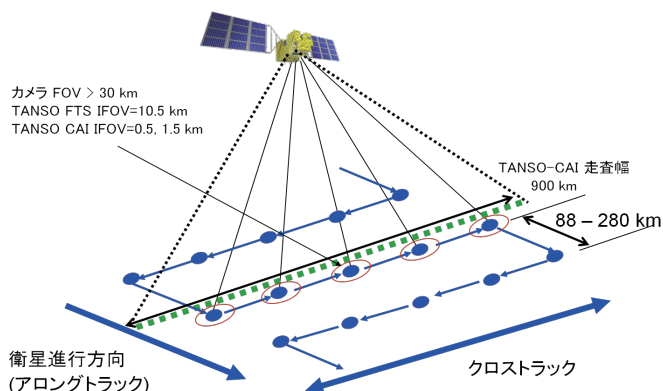


図1 いぶき(GOSAT)による地球観測の様子

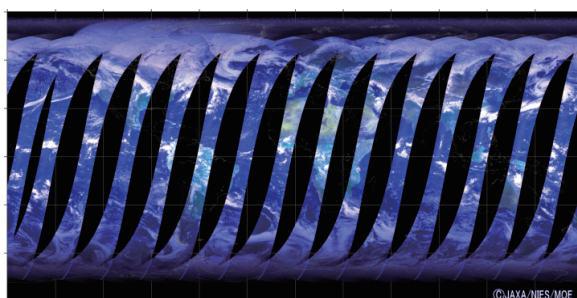


図2 TANSO-CAIの観測例（日照域）
（JAXA/EORCのホームページより）

いるGSMaP（世界の雨分布を示すデータ）の例を示す。地球全体の降水量のデータを、ほぼリアルタイムに提供している。

これらの例のように、地球全体の観測を継続的に実施することにより地域ごとに時間経過を的確に把握できるようになり、また、局所的な短時間の気象変化といった、地上からの観測では得られない知見を得ることができる。これらのデータは地球の気候変動等の理解のために用いられている。

衛星観測データは、各々のプロジェクトを推進している機関のホームページから無償で公開されている。「いぶき」については国立環境研究所から、GPM

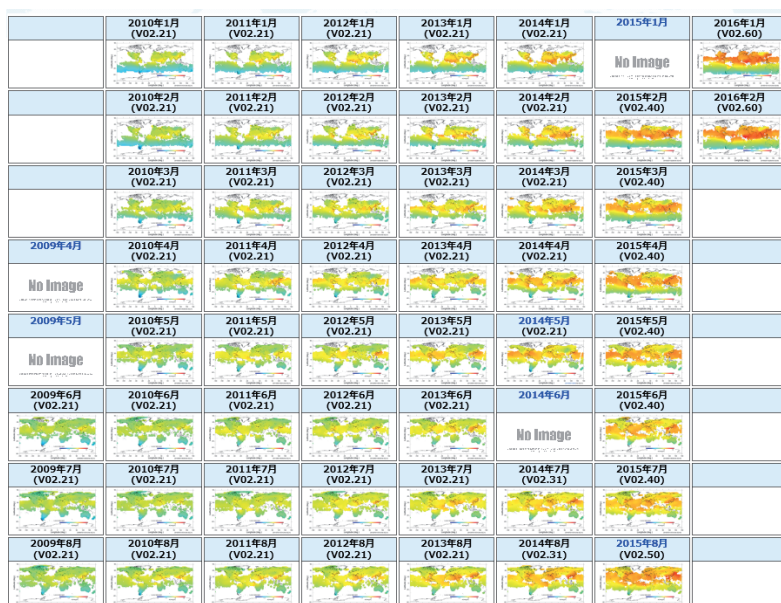


図3-a 「いぶき」観測による二酸化炭素分布の変化（国立環境研究所のホームページより）

関連についてはJAXA第一宇宙技術部門の地球観測研究センター (EORC) から容易に入手できる。また、地球観測データ提供システム (G-Portalという) では、観測時刻や場所をキーにプロダクトを検索し、一般ユーザへ提供するサービスを行っている。

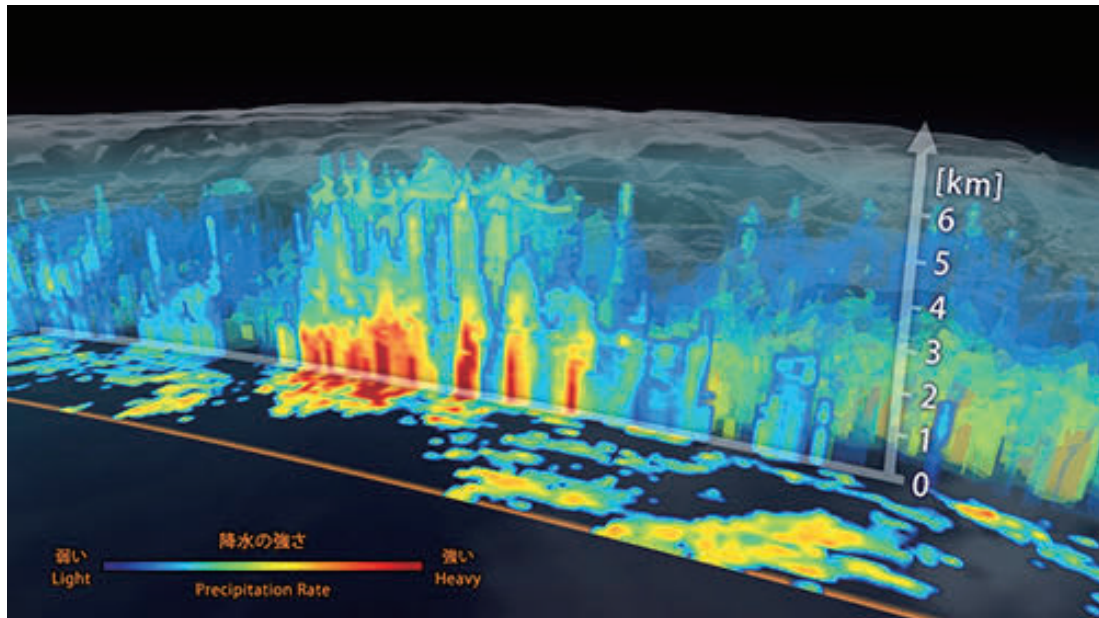


図3- b 二周波降水レーダ (DPR) の観測例 (JAXA/EORCのホームページより)

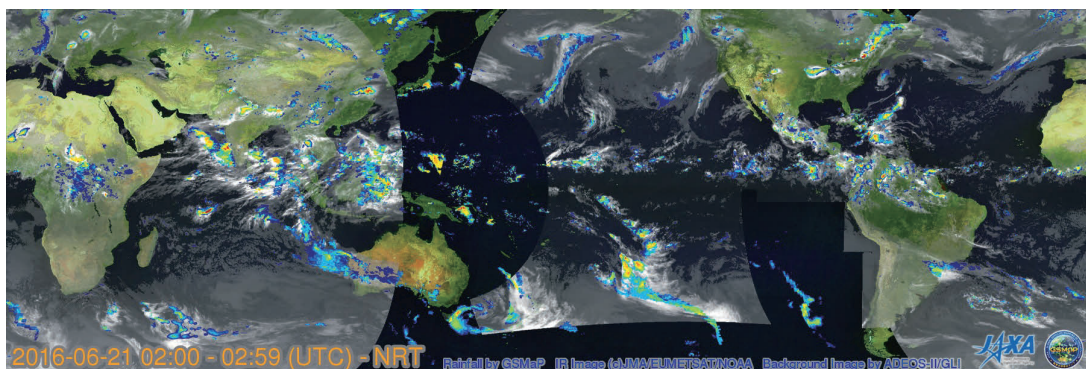


図3- c GSMaPの例 (JAXA/EORCのホームページより)

2 ゴールと背景

地球観測データは、衛星の軌道周回の単位にデータファイルを作り、各種のプロダクトとして提供される。これは定常運用という。図4に、その概念図を示す。

定常運用においては、衛星の各周回で得られたデータを、それぞれの周回単位にいかに早くユーザへ届けられるか、ということが重要である。

観測データの蓄積に従い、データの解析処理アルゴリズム (物理的なモデルに基づく) を改善したり、また、処理パラメータの改訂 (センサの校正係数の改訂など) が行われる。センサーの経年変化に対応するためのアルゴリズム、パラメータの変更も行われる。その際には、蓄積された観測データの全てを、新しいアルゴリズム、パラメータで再度の処理を行う。これを再処理運用と言っている。

定常処理や再処理に先立って、処理アルゴリズムが正しいかどうか、センサの校正係数が妥当かなど、処理の結果に関する妥当性を検証する必要があり、そのために行うデータ処理を試験処理ということがある。特に長期間に渡って蓄積されたデータの全数を処理して検証する場合には長期試験処理という。その結果、データが妥当であると判断されて初めて再処理の実施とユーザへのプロダクト提供が行われる。表1に処理の種類と概要の例を示す。ただし、正確な定義は衛星プロジェクトによって異なることが有ることに注意する必要がある。

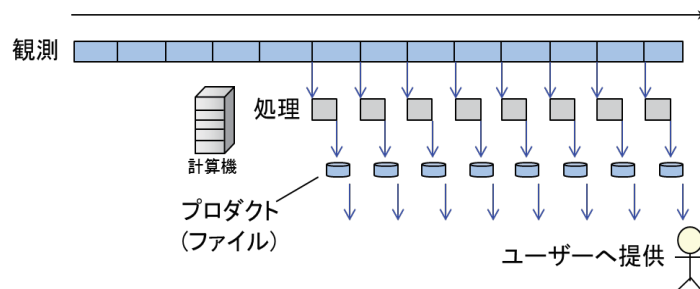


図4 定常運用とプロダクト提供の概念図

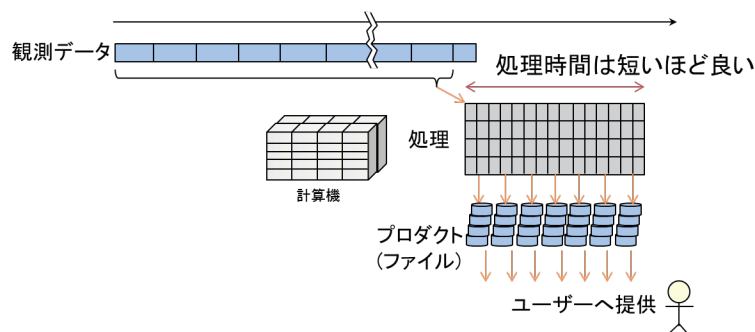


図5 再処理の概念図。処理期間は短い方が良い。

表1 処理の名称と概要

名称	概要	頻度
準リアルタイム処理 (NRT)	観測データを周回単位に地上へダウンリンクし、即時処理してユーザへ提供する。	1回／軌道周回
標準処理 (STD)	軌道データ、姿勢データ等の確定後に再度、データ処理し、ユーザへ提供する。このプロダクトを標準プロダクトとし、保存・管理する。	軌道データ等の受領後
再処理	アルゴリズムの改訂、パラメータ(校正係数等)の改訂時に累積した観測データを全数、処理し、ユーザへ提供する。	1回／年
長期試験処理	再処理に先立ち、アルゴリズム評価のために実施する処理。再処理と同等の長期間の処理を実施する。	1回／年

再処理は年に1回あるいはそれ以上、行われる。観測が長期化するとデータ量も膨大になり、全データの再処理を完了するまでに長い期間を要するようになる。例えば、降水量の観測を担ってきた衛星TRMMは2015年4月に観測を終了し、その観測期間は17年間以上に及ぶ。他にも衛星Aquaに搭載されたAMSREは10年間のデータを蓄積した。

定常運用は筑波宇宙センター（以下、『TKSC』という）に整備されたデータ処理設備を用いて実施されるが、その計算資源には限りがある。その理由はコスト、及び、設置場所の制約である。その制限により、再処理のために大きな設備を抱えるわけにはいかず、必要最小限の設備を整備してきた。そのため、従来は再処理の開始からプロダクトの提供までのレイテンシが1年以上に達したことがあった。その結果、データを必要とする研究者等のユーザには長期間の待ちを余儀なくさせる状況になっていた。

我々の目的は、JSS2を活用することにより、再処理のスループットを大幅に向上させることにある。目標として、従来は年単位の時間を要した再処理を数日から十数日へ短縮させる。これにより、

データ提供を抜本的に早め、防災・減災などへの社会貢献や、地球環境変動の研究利用に対するさらなる貢献を果たすことがゴールである。

現在、再処理にJSS2を活用することを計画している衛星、センサーについて表2に示す。もちろん、今後打上げを計画する衛星についてもJSS2活用が計画される。

表2 再処理でのJSS2利用対象

衛星／センサー	観測期間（2016年6月現在）
TRMM/PR	17年
Aqua/AMSR-E	10年
GCOM-W（しずく）/AMSR2	4年以上（運用中）
GOSAT（いぶき）/FTS	7年以上（運用中）
GPM/DPR	2年以上（運用中）
GSMaP （TRMM, GPM等の複数マイクロ波放射計データの複合プロダクト）	18年以上

なお、以下に記述する「いぶき」に関する成果については、JAXAが主体となって、JAXAのセキュリティ・情報化推進部スーパーコンピュータ活用課の協力を得て、日本電気株式会社に作業を委託¹⁾して得られたものである。GPM/GSMaPに関しては、同様に富士通株式会社に作業を委託²⁾したものである。

3 JSS2利用の方法

TKSCでの運用設備とJSS2では計算機の構成が異なる。したがって、処理エラーが起きないことの確認と計算結果であるデータの妥当性を検証する必要が有った。また、再処理の効率を最大にできるように、必要な課題を洗い出し、必要な対策を施す準備を進めた。以下、次の項目に沿って今回のJSS2利用の成果を示す。

- 処理プログラムの移植とプロダクト検証
- 運用方式の設計と環境整備

3. 1 処理プログラムの移植とプロダクト検証

(1) 「いぶき」処理プログラムの場合

TKSCで利用しているシステムとJSS2の主な相違点を表3に示す。

表3 プログラミング・実行環境の比較（「いぶき」のケース）

項目	JSS2 (SORA-PP)	TKSC運用環境
CPU アーキテクチャ	Intel Xeon E5-2643 v2	Intel Xeon E5640
OS	RHEL 6.4	RHEL 5.5
コンパイラ	gcc4.4.7	gcc4.1.2

この他、ディレクトリ構成の差異などにより各種PATHの設定やMakefileの修正が必要であった。また、「いぶき」の運用開始時期（2009年）以降に言語規格に発生した変更に対応するためプログラム修正があった。ただし、アルゴリズムに影響するようなプログラム改変は必要なかった。

プロダクトの妥当性については、gcc(GNUコンパイラ群)のバージョンの違いによると考えられる計算結果の微小な差異が出たものの、数値計算の誤差の程度と判断され、プロダクト品質への影響は無いと判断した。GOSAT(いぶき)/FTSのポインティング機構が2014年末から2015年初めに掛けてA系からB系に切り替えられ、レベル1処理の校正係数が変更されたので、精度検証に当たっては、A系の運用期間のうち1カ月分とB系の1カ月分を対象とした。この検証は、JAXA/EORCの協力を得て実施した。

(2) GPM/GSMaP処理プログラムの場合

「いぶき」の場合と同様に、TKSCで利用しているシステムとJSS2の主な相違点を表4に示す。

表4 プログラミング・実行環境の比較 (GPM/DPR, GSMaPのケース)

項目	JSS2 (SORA-PP)	JSS2 (SORA-MA)	TKSC運用環境
CPU アーキテクチャ	Intel Xeon E5-2643 v2	SPARC64XI fx	Intel Xeon E5640
OS	RHEL 6.4	Linux	RHEL 6.4
コンパイラ	Intel Cluster Studio XE 2013	Technical Computing Suit (富士通製)	Intel Cluster Studio XE 2013

JSS2のプレポストシステムという位置づけのSORA-PPを利用する場合は、OS、コンパイラ共に製品とバージョンがTKSCの運用環境と同一である。ディレクトリ構成の差異などにより各種PATHの設定やMakefileの修正が必要となったのは「いぶき」の場合と同様であるが、コンパイラが同一だったので、作成されるプロダクトは完全に一致した。このため、検証にあたってJAXA/EORCの協力を得るには及ばなかった。なお、降水に関する物理量は季節による変動が大きいことから、2014年の4月、7月、10月、12月の連続する各3日間のデータを対象とした。比較ツールとしては、地球観測で用いられることの多いHDFファイルの比較には“hdfcmp”を、それ以外のファイルの比較には“diff”を用いた。

一方、メインの計算システムであるSORA-MAは京コンピュータの後継とされるSPARCプロセッサを用いたシステムであり、主たる言語環境は富士通製である。そのためプログラム移植が必要となり、一部のアルゴリズムについてSORA-MAへの移植を試みた。このケースでは言語環境の変更への対応とともに、バイナリデータのエンディアンの違いに伴うプログラム修正が必要となり、異環境への移植に当たっての教訓、知見を得ることができた。今回、移植とデータ検証の作業規模や期間を考慮して、再処理にはSORA-PPを使用することを選択した。

以上により、「いぶき」、GPMともにJSS2上で再処理するためのプログラム移植・動作確認とプロダクト検証に成功したので、次に、大量のデータの再処理とユーザへのプロダクト提供をスムーズに行うための環境整備に取り組んだ。

3. 2 運用方式の設計と環境整備

(1) 「いぶき」の場合

レベル1処理においては、異なる衛星周回のデータは互いに独立に処理できる。したがって、まず処理アルゴリズムの実行時間をプロダクト一つ毎に単体で測定しておけば、目標とする処理完了までの時間に応じて必要な計算リソースを見積もることができる。プロダクト一つの平均的な処理時間は数分であったため、今回、10日間前後の処理期間を目標としてSORA-PPの30ノード(360コア)を占有利用することとした。

次に課題になるのは、処理の対象となるプロダクトの数が衛星の周回数に比例して増大し、6年半の間の蓄積数は数百万のオーダーとなり、ジョブ投入などの操作をマニュアルで行うのが実質的に不可能なことである。その操作を自動化するとともに、処理プログラムの実行状況に対応してジョブ投入を制御し、スループットの最大化を図れるような手段が必要である。その対策として、ジョ

ブの実行状況の把握、ジョブキューの蓄積状況の把握を行いながら、処理対象のプロダクトを選択しジョブ実行スクリプトの生成と投入を実現できるシェルスクリプトを開発した

これらの準備作業は2015年10月までに完了し、また、TKSCから調布航空宇宙センター（以下、『CAC』という）のJSS2へデータ移行等の作業を行って、2015年11月に6年半分の観測データに対する再処理を実施した。

その結果、TKSCのシステムを利用する従来通りの再処理運用では約1年間を要すると見積もられた再処理を11日間、約9.5万コア時間（コア数×H）で完了した。図6にこの成果を示す。

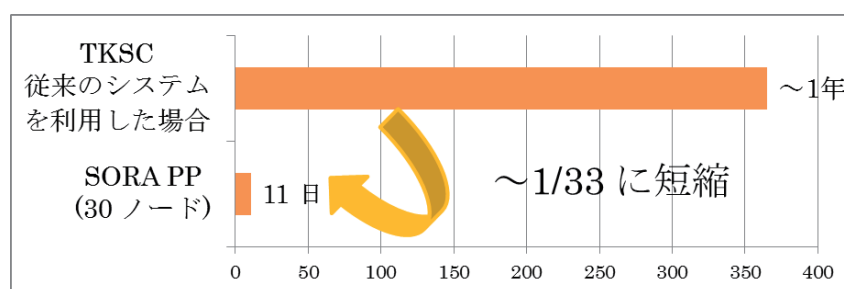


図6 「いぶき」データの再処理期間の改善

計算ノードを占有したことによりジョブの実行待ちもなく、大幅なレイテンシの改善が達成できた。2016年1月には国立環境研究所等の機関ユーザに対してJSS2上から直接プロダクトを提供している。

その後、センサーの運用状況に応じてパラメータの見直しが行われ、約7年分の再処理が2回実施された。今回は、再処理の実行スクリプトに次のような改善を施している。

- 処理の進捗状況が容易に確認できること。
- 処理エラーが生じたケースについて、その内容を容易に確認・識別できること。
- エラーが生じた処理を容易に識別し、再投入できること。
- 処理ログ等の中間ファイルの自動削除。

2回の再処理の合計で約35.7万コア時間を要する再処理を、シェルスクリプトの改善効果もあって最大の効率で実施できた。すなわち、占有した計算ノードの利用効率はほぼ100%に達した。今後の再処理運用においても、JSS2の利用によって従来の運用を凌駕する効率的な再処理が実施できることが期待される。

(2) GPM/GSMaP処理の場合

「いぶき」のケースと同様に、処理プログラムの単体性能測定を行い、SORA-PPのノード数30程度を計算リソースの必要量の目安とした。

一部の処理アルゴリズムには並行実行ができないという制約が有る。まず、GPM主衛星に搭載されたセンサーDPRは処理レベル1からレベル3へ向かい順に処理する必要がある。ただし、同一レベルのデータ同士は処理の単位毎に独立している。すなわち、レベル1と2のデータは衛星周回毎に独立であり、レベル3データは日、または、月毎に独立である。そのため、同一レベルの処理においては並行実行が可能であり、したがって、割り当てるノード数に反比例して処理期間を短縮することが可能である。

さらに、JSS2には、MPIライブラリを活用して複数の周回の観測データを同時に処理起動することにより、全体のスループットの向上を図る機能が整備されている。図7はこの機能によるスループット向上の効果を示す。左右どちらのグラフも、横軸は時刻、縦軸は実行中のジョブ数を示す。

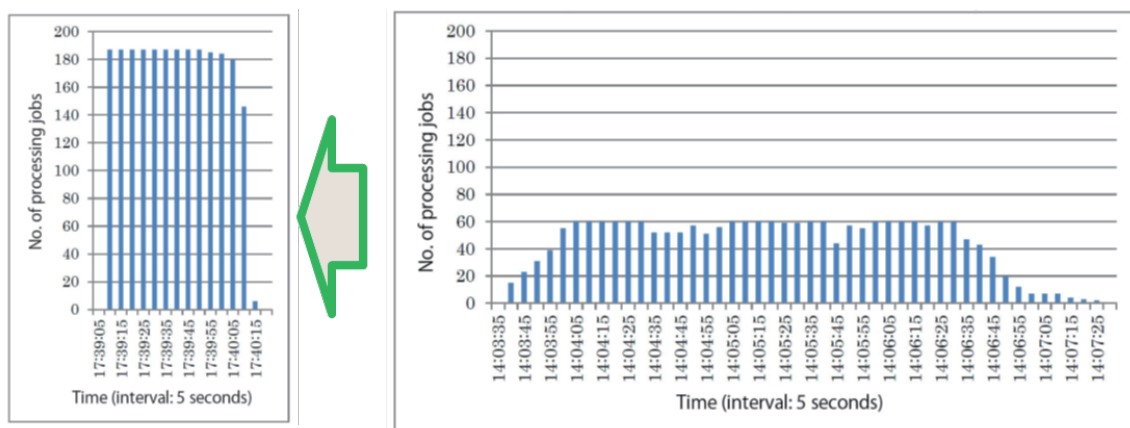


図7 MPI同時起動機能によるスループット向上

通常通りジョブ起動した場合、処理の時間が分程度と短いために、多数のジョブが立ち上がる前に処理完了してしまう。そのため、図7の右図に示したように、同時に稼働するコア数が伸びないという現象が生じた。このままでは多数のノードを同時利用できないというJSS2の効果が発揮できないが、同時起動機能を活用することにより、図7の左図に示すように同時に多数のコアを活用した処理が可能となる。これにより、計算リソースの利用効率が大幅に向上する。

ここでも同様に膨大な量のプロダクトの処理を自動化する手段が必要である。処理レベル間のデータの依存性を考慮しながら実行ジョブを自動生成し、実行状態の管理や、エラーが生じた場合の再実行をサポートする機能として、JSS2に備わる衛星データ処理ワークフロー制御機能を使用した。この機能を使用するためには、予め処理の依存性を定義する情報をデータベースに登録するとともに、TKSCでの定常処理の記録（ワークオーダと呼ばれる情報とその関連情報）をJSS2に複写する必要がある。

2016年6月、これらの環境を活用し、GPM主衛星打上げ（2014年2月末）後の2年3ヶ月間の観測データの長期試験処理を実施し、再処理用アルゴリズムを評価するためのデータセットを作成した。今回はレベル1のみの試験処理であることから計算ノードの占有は行わず、2～3ノードを使用し、処理期間は二日間かかった。

今後、11月には全ての処理レベル（レベル1、2、及び3）の長期試験処理が計画されている。使用する計算リソースとしては、今回の処理性能から30ノード程度を利用する計画である。これにより、従来よりデータレイテンシを1/20程度に短縮できる見通しである。

4. 考察

再処理運用において同等のスループットを達成するためにTKSCで単独に計算機調達を行うとすると、そのコストは観測期間に比例して大きくなる。詳細は省くが、JSS2への移植作業やデータ移行のコストなどを考慮しても、計算機コストは従来の10倍以上に達すると考えられ、JSS2を活用する効果は大きい。

今後数年間の再処理予定によれば、TKSCとCAC間で輸送を要するデータ量は表5に示す通りである。なお、入力データは多少時間がかかっても事前にTKSCからCACへ輸送ないし伝送しておくことが可能なので、ここではユーザへ早く提供すべき再処理結果である出力データの量に注目している。

昨年度までは、TKSCとJSS2間のデータの経路として利用可能なものはJAXAの社内ネットワークであるJAXAnetのみであり、低速（帯域幅100 Mbps）で送受信しなければならなかった。この場合、処理時間は短縮できてもデータ転送に時間を要し、ユーザへの提供を早くするという目的は果たせない。表5のデータ量では少なくとも約2 Gbpsの帯域幅が必要である。これに対する対

表5 今後数年の再処理予定と出力データ量

衛星, センサ	FY27	FY28	FY29	FY30
GOSAT	55TB	>60TB	未定	未定
TRMM/PR	—	>40TB	>150TB	未定
GPM/DPR	—	>50TB	未定	>90TB
AMSR-E, AMSR2	12TB	15TB	9TB	未定

策として本年5月初めに、TKSCの地球観測システムとJSS2とをSINET5を介して高速に接続する環境を整備した。現在、TKSCとJSS2は10 Gbpsの帯域を持つネットワークで接続されている。これにより、GPM/DPRの試験処理では、TKSCからユーザへのデータ提供に要するオーバーヘッドを従来の約1/10に短縮できた。このデータ伝送路の高速化により、JSS2からTKSCへの伝送時間を含めても、JSS2で処理した方が、TKSCの既存設備で処理するより、大幅な時間短縮が可能となった。

さらに、一般ユーザは対象ではないが、「いぶき」の機関ユーザ（国立環境研究所など）に対して、JSS2からプロダクトを直接提供する方式を採用した。従来は必ずTKSCからユーザへ送付しているため、新たなデータの経路を設けたことになる。外為法に基づく手続きを経る必要があったが、今後、定常的なデータ提供手段として検討する価値は非常に大きい。

また、昨年度より、衛星データ処理はJSS2の重点利用課題となった。その効果は大きく、以下の通り地球観測業務にとって大いに助けとなっている。

- 優先度が最高レベルに設定されているため、ジョブ投入後の待ちが少ない。
- 申請によりノードを占有することが可能である。この場合、完全に滞りなくジョブ実行できる。

こうした環境が無ければ、ジョブ実行の待ちにより、スループットの大きな低下が避けられない。

TKSCにおけるJSS2利用は、AMSR-E、並びに、GCOM-W衛星に搭載のAMSR2へも拡大される。6月にはAMSR-Eの10年分のレベル1処理を完了した。さらに、2016年中にAMSR-E及びAMSR2の全処理レベルの再処理を計画している。

他の機関の事例も参考となる。「いぶき」と同様に温室効果ガスを観測する衛星にOCO-2 (Orbiting Carbon Observatory-2, 2014年7月2日打上げ) がある。これはNASAが打ち上げ、運用を行っている。昨年、NASAのAmes Research CenterにあるスーパーコンピュータPleiadesを利用して10か月分の観測データの再処理を行った。観測データの60%の再処理にHaswell Node (2088ノード, 4176CPU)のうち500ノードを利用し、140万CPU時間を費やしたとのことである。NASAのように潤沢な予算を誇る機関であっても、スーパーコンピュータを利用できることの利点は大きく、この規模の計算を短期間に実施できたのはその効果であると、NASA担当者も認めている。

5. 結論

JSS2利用の効果は大きく、GOSAT/TANSO-FTSの再処理を従来の33倍のスループットで実現できた。

GPM/DPRの試験処理では、データ提供を従来の約4倍のスループットを達成できた。11月の長期試験処理や、2017年度に計画されている再処理では、計算ノードを増やすことにより処理期間の大幅な短縮が達成できる見通しである。

JSS2の利用効果は、AMSR-E、及び、AMSR2においても期待される。

2018年に打上げ予定のGOSAT-2は、JSS2の利用を前提として運用することを考えており、そのためのシステム設計、構築を進行中である。

以上の通り、TKSCでの衛星データ処理の業務では、JSS2は業務効率化とユーザの利便性を両立させて実現するための必須のツールであると考えられる。

参考文献

- 1) 日本電気株式会社, 「温室効果ガス観測センサ(FTS)処理固有部 ソフトウェアの移植」 (JX-PSPC-421815) 成果報告書 (GOSAT-H27移植-提005), 2015, 日本電気株式会社
- 2) 富士通株式会社, 「全球降水観測計画 (GPM) プロダクト再処理時のJAXA Supercomputer System 2 (JSS2) 利用検討 (その2)」 (JX-PSPC-410301) 成果報告書 (FJ-GPM-JSS2利用検討-14-020), 2015, 富士通株式会社
- 3) Tsengdar Lee, Climate Data Downscaling Projects at NASA, Presented at SC15 Conference, Nov 16 – 20, 2015

謝辞

調布のスーパーコンピュータJSS2の遠隔利用において、移植時のプログラム相談やジョブ実行の効率化対策などにおいてJSS2運用窓口の協力を得た。紙面を借り謝辞を申し上げる。

火星大気データを地球大気解析ツールで解析できるようにするためのデータ整備 —MRO/MCS データの netCDF 化と GrADS による可視化—

野口克行^{*1}, 林寛生^{*2}

Conversion of the MRO/MCS data into netCDF format and gridding of them for analysis
and visualization by the use of GrADS

by

Katsuyuki Noguchi^{*1} and Hiroo Hayashi^{*2}

Abstract

This paper presents a method to process data obtained by Mars Climate Sounder (MCS) onboard Mars Reconnaissance Orbiter (MRO). The method includes two steps; the first step is the format conversion of original text data into netCDF (Network Common Data Form), and the second step is gridding of the data to be available for use by Grid Analysis and Display System (GrADS). The method proposed here will contribute to multidisciplinary use and/or educational use of such data for the Martian atmospheric science.

Keywords: netCDF, GrADS, Mars, atmospheric science, visualization

概要

本論文では、Mars Reconnaissance Orbiter (MRO) 搭載の Mars Climate Sounder (MCS) による観測データの利用を広めるために行った環境整備について紹介する。元データのデータ量が膨大であり、データ構造もそのままでは解析等において取り扱いにくいいため、まずデータフォーマットを Network Common Data Form (netCDF) に変換することで、データの可搬性と可読性を高めた。次に、地球の大気・海洋科学分野で広く使われている Grid Analysis and Display System (GrADS) で扱えるようにデータの格子化を行い、解析と可視化が容易にできるようにした。このようにデータの取り扱いにおける敷居の高さを下げることによって、他分野の研究者の参入や学生教育におけるデータ利用への貢献が期待できる。

キーワード: netCDF, GrADS, 火星, 大気科学, 可視化

doi: 10.20637/JAXA-RR-16-007/0010

^{*} 平成 28 年 11 月 24 日受付 (Received November 24, 2016)

^{*1} 奈良女子大学研究院自然科学系 (Faculty of Science, Nara Women's University)

^{*2} 富士通エフ・アイ・ピー株式会社 (FUJITSU FIP CORPORATION)

1. はじめに

近年、日本において火星探査の機運が高まっている。日本が 1998 年に打ち上げた火星探査機「のぞみ」¹⁾は残念ながら火星周回軌道への投入には成功しなかったものの、その後に複数の火星探査計画の検討が精力的に行われている^{2) 3) 4)}。火星探査機の本体や搭載機器の開発を進めることはもちろん重要だが、すでに諸外国によって複数の探査機が火星に送り込まれている以上、それらの探査機によって得られている既存データの解析を進めることで次期探査の目標とするサイエンスを深化させていくことも重要である。特に、米国によって取得された観測データは積極的に公開されており、誰でも容易に入手してデータ解析研究を進めることが可能である。

本論文で着目するデータは、2006 年以来観測を続けている米国の火星探査機 Mars Reconnaissance Orbiter (MRO) に搭載された赤外放射計 Mars Climate Sounder (MCS)⁵⁾の観測データである。MCS で取得された火星大気 of 赤外スペクトルから、気温や水蒸気量、氷量などが導出されている⁶⁾。MCS データに含まれる火星大気の物理量は、地球の大気科学における手法を応用して解析ができるものも多い。そのため、データ解析環境さえ整えば、地球の大気科学研究者による利用の機会を増やすことも可能だと思われるが、現在までに蓄積された MCS データの観測高度プロファイル（観測物理量の高度分布）の数は数百万本と膨大な上、公開されているデータのディレクトリ及びファイル内の構造も扱いやすいとは言えない。データはいわゆるテキスト形式³⁾であるため、データの中身をテキストエディタ等ですぐに見ることのできる利点はあるものの、可搬性やデータ解析・可視化における取り扱い易さについては改善の余地があると思われる。筆者の研究グループは当初、このデータをテキスト形式のまま個々の高度プロファイルに分離することでデータ解析に使用していたが、ファイル数が膨大なためにファイルの検索効率やコンピュータ間の可搬性が悪く、また可視化等の処理を行う際にもデータの読み書きに時間が掛かり、取り扱いに困難を感じていた。火星を含む惑星大気観測では、通常、観測装置が受信した信号データを物理量に変換したデータは観測ミッションより提供されるが、幅広いサイエンス研究や教育に役立つような加工データ（例えば、全球格子化データ）は個々の研究者が作成しているのが現状である。また、可視化や解析に関わるツールは必ずしも観測ミッション側から提供されるとは限らず、MCS データに関しても観測ミッション側からそのようなツールは提供されていない。

本論文では、この MCS データの取り扱いを少しでも容易にし、広く研究や教育で利用できるようにするため、二つの手法を提案する。まず、地球の大気・海洋科学分野で広く使われている Network Common Data Form (netCDF)⁷⁾というバイナリ形式に変換する。テキスト形式と比べて、バイナリ形式にすると読み書きに必要な計算機環境の整備（ライ

¹⁾ NASA/PDS 及び ESA（欧州宇宙機関）の惑星科学データ部門（PSA）における大気科学者は、大気データの高度プロファイルを主にテキスト形式で扱っていたため、現行データにもテキスト形式が採用された。

MROM_2xxxディレクトリ(yyy=001~110)

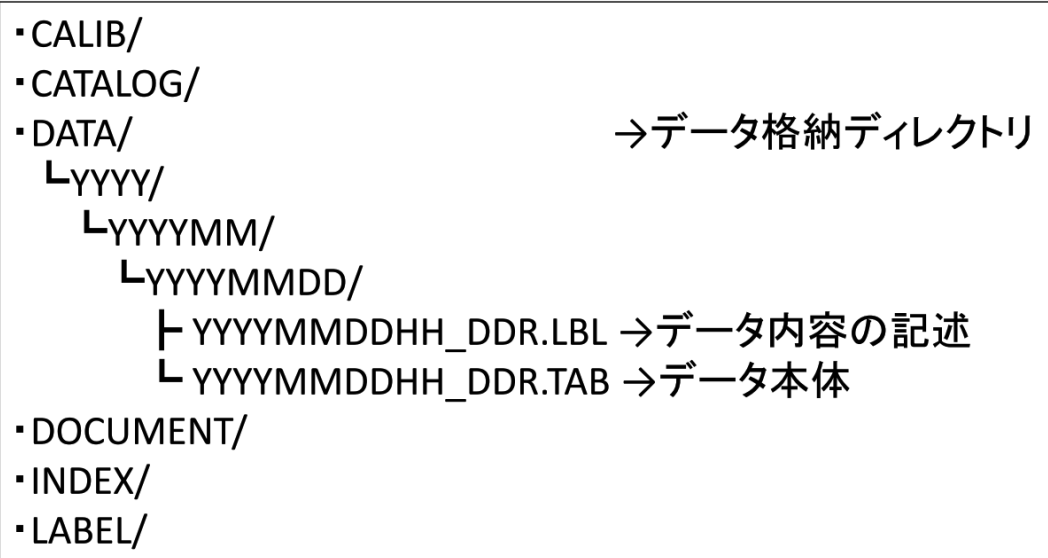


図 1 公開されている MCS データのディレクトリ構造. 拡張子 LBL ファイル, TAB ファイルともにテキスト形式である.

ブラリ等のインストール) が必要になるものの, そのデメリット以上にデータの可搬性や解析における取り扱いやすさというメリットが得られる. netCDF の使用は無償であることもメリットとして挙げられる. 次に, MCS データを格子化して, 解析・可視化ツールの一つである Grid Analysis and Display System (GrADS)⁸⁾ による処理を可能にする. GrADS もやはり地球の大気・海洋科学分野においてよく用いられている無償のソフトウェアである. 他分野も含めてユーザ数が多く, インストールが簡単でインタラクティブに扱えるので初心者でも使いやすいといった利点も挙げられる. GrADS を火星の観測データに応用するにあたって問題となったのは, 時刻系の違いである. 火星の時刻系としては, 地球と同様に火星上での 1 日を基準としたものの他に, 公転軌道上における太陽との位置関係を基準としたものがよく用いられている. GrADS はこのような時刻系は想定していないため, 新たに処理コマンドを作成するなどしてユーザが GrADS 上で MCS データを解析・可視化しやすくする工夫を行った.

2. 使用した元データ

本論文で用いるデータは, 米国航空宇宙局 (National Aeronautics and Space Administration : NASA) の Planetary Data System (PDS) の Web サイト^{*4)} において公開

^{*4)} http://pds-atmospheres.nmsu.edu/data_and_services/atmospheres_data/Mars/Mars.html (2016 年 8 月 19 日閲覧)

されているMRO MCS Derived Data Records (DDR) Version 4である。このデータのディレクトリ構造を図 1 に示す。地球時間で 1 ヶ月毎にディレクトリが分かれており、ディレクトリ名MROM_2xxxのxxxは 2006 年 9 月からの通算月を表す（つまり、2006 年 9 月は 001 である）。この 1 ヶ月毎のディレクトリはさらに日毎のディレクトリに分かれており、4 時間毎にひとまとまりのファイルが格納されている。データファイルには、データ本体が入っているファイル（YYYYMMDDHH.TAB）と、データ内容を記述するファイル（YYYYMMDDHH.LBL）の 2 種類が存在する。ただし、YYYYは西暦年、MMは月、DDは日、HHは時間を表す。ファイルは 4 時間毎であるが、欠測等で提供されていないプロファイルもあるため、1 つのTABファイルに含まれる高度プロファイル数は不定である。

ヘッダ行に続く観測物理量の高度分布としては、指定された各気圧面における気温、ダスト量、水蒸気量、水氷雲量、ドライアイス雲量の各量とそれらの誤差、及び周縁（limb）観測のタンジェントポイントの高度、緯度、経度が格納されている。なお、ドライアイス雲量に関しては、現時点（Ver.4）ではまだデータの検証が十分ではないとのことで欠損値（-9999）が与えられている。

3. データ整備

3.1 データの netCDF 化

ここでは、2 章で述べたMCSデータのnetCDF形式への変換について説明を行う。netCDFは、OSやプラットフォームに依存することなく多次元データの読み書きが可能な自己記述型のデータ形式で、地球の大気・海洋科学分野で広く使われている。netCDFを扱うために必要なライブラリやツールは、開発元である米国の University Corporation for Atmospheric Research (UCAR) の unidata プロジェクトの Web サイト^{*5}から取得できる。

あるデータを netCDF 化するには、最低限 1 つの次元を持たせる必要がある。MCS データには既に高度と言う次元が含まれているので、これをそのまま netCDF の 1 つ目の次元にすればよい。MCS データの気圧面はどの高度プロファイルでも同じであるため、気圧を軸に取って 1 つ目の次元とし、各物理量を変数として格納することとした。これらのデータを取得時刻の順に並べることで 2 つ目の次元とし、2 次元の netCDF データとして整備した。なお、netCDF 化するファイルの単位は、上述の MROM_2xxx ディレクトリ単位（つまり 1 ヶ月毎）とした。

このデータセットは、元のデータに平均等の操作を加えることなく単純に netCDF 化しているため、元データに含まれる情報を損なうことなく可読性・可搬性のみを向上させたものである。読み出し速度については、ファイルの単純な読み込みであれば 100 倍程度向上した。なお、圧縮性についてはそれほど劇的な向上はなく、テキストデータ（1 火星年当たり 12GB 弱）に対して netCDF データ（1 火星年当たり 9GB 弱）のサイズは 7 割程度で

^{*5} <http://www.unidata.ucar.edu/>（2016 年 8 月 19 日閲覧）

ある。しかし、昨今は市販の安価なハードディスクでも容量が数 TB 程度あり、圧縮の必要性は強くないと言える。

3.2 GrADS 利用のためのデータ格子化

前述したように、本論文では MCS データの解析や可視化を容易にするため、GrADS の利用を提案する。GrADS は 4 次元格子（経度、緯度、鉛直層、時間）上に配置された地球科学データに対するアクセス、操作、表示のための対話型ツールである。GrADS では netCDF を含む様々な格納形式のデータファイルを扱うことができる。GrADS のソフトウェア本体及び関連するドキュメント等は、開発元の Web サイト^{*6}から配布されている。

図 1 で示された MCS データは、高度プロファイルが取得順に格納されているだけの状態である。このデータを緯度・経度・時間方向にも格子化して 4 次元方向に自由に切り出せるようにすることが最終的な目標である。GrADS で利用するためには、データを 4 次元格子点に落とす作業（=格子化）が必要だからである。空間方向の次元のうち、鉛直方向はオリジナルのプロファイルデータが既に固定された気圧面で提供されているため、そのまま鉛直方向の次元として用いることとし、水平方向である経度と緯度に関してはそれぞれ 5 度ずつの間隔で平均を取ることで格子化した。

時間方向についての格子化は、少し工夫が必要である。火星での時刻の単位としては、Mars Year (MY), Solar Longitude (Ls), Local True Solar Longitude (LTST) がある。MY は火星における年を表し、公転周期が MY の 1 年に対応する（これは、地球時間で約 2 年に相当する）。MY1 は 1955 年 4 月に始まると定義されており、2015 年 6 月から 2017 年 5 月までは MY33 となる。Ls は火星の季節を表す。これは、火星の公転軌道上での位置を表すもので、北半球の春分点を 0 度（または 360 度）とし、夏至を 90 度、秋分を 180 度、冬至を 270 度とするものである。火星の公転軌道は楕円であるため、公転速度は一定ではない。それ故、1 度あたりの Ls に含まれる実際の時間の長さも一定ではない。LTST は、その場所での太陽の南中時刻を正午 12 時とする局所時刻系である。火星の 1 日は地球とほぼ同じ（約 24 時間 40 分）であるため、火星の 1 日を 24 等分した長さが火星での 1 時間とされる。ただし、MCS データにおいては、LTST を 0 から 1 までの間に正規化した値が格納されている。ここでは、Ls を時間方向の次元に取ることとし、5 度ごとに平均を取って格子化した。一方、MY については各火星年に対して、また、LTST については火星における昼（0.375～0.875）と夜（0.875～1.0 または 0.0～0.375）に対してファイルを分けることとした。このように格子化した MCS データは、第 3 章で述べたオリジナルの高度プロファイルデータの整理と同様に、netCDF 形式でファイルに格納した^{*7}。なお、今回は便宜上、上

^{*6} <http://cola.gmu.edu/grads/> (2016 年 8 月 19 日閲覧)

^{*7} MY28-MY33 の 6 火星年の各年については、昼と夜に対してそれぞれデータファイルを作成するものとして、計 12 個作成した（本論文執筆時点）。ただし、MY28 については観測開始以前の Ls に対して、MY33 についてはまだデータが提供されていない Ls に対して、ともに欠損値を挿入した。

述のような緯度・経度・Lsおよび時刻のグリッド幅を設定したが、ユーザが目的に応じて自由に変更できるようなルーチンを研究室内部にて整備済みである。

本来、GrADSは地球の大気・海洋科学分野のデータを扱うソフトウェアであるため、各次元については地球における空間・時間であることが想定されている。それ故、上述のように火星の空間・時間を扱うMCSデータを4次元格子化してnetCDF形式で格納しても、そのままではGrADSで扱うことができない。空間方向の次元（経度、緯度、気圧面）は地球における同等の次元として認識されるが、時間方向の次元として採用したLs（角度の単位で0～360度の範囲の値を取る）は地球における「時間」としては認識されないためである。そこで、データ記述ファイル（data-descriptor）^{*8}を導入し、Lsを強制的に地球時間へ対応させるようにした。具体的には、データ記述ファイルの中でLs=0度に対応する日時と格子間隔（Lsで5度）に相当する時間間隔（日数）を記述した。このように、データ記述ファイルを介してデータファイルを開く^{*9}ことで、上述の形で格子化したMCSデータをGrADSによって扱うことができるようになった。

なお、GrADSに元々備わっている機能だけではLsの値を指定したり設定されているLsの値を確認したりすることはできないため、これを実現するためのコマンド（GrADSスクリプト）を別途用意した。具体的には、「lsget」というコマンドによって現在設定されているLsの値を調べ、「lsset」というコマンドによってLsの値を指定する。ただし、Lsの間隔は地球における時間間隔と線形には対応しないので、Lsに対応する本来の地球時間とGrADSが表示する見かけ上の時間情報は一致しないという問題は残っている。

このように整備したMCSデータを実際にGrADSで可視化した例について、図2に示す。ここでは、まず、xdfopen コマンドによってデータ記述ファイルを開き、経度・緯度・気圧・Lsの範囲をそれぞれ設定し、display コマンドによって気温変数(t)をプロットしている。

^{*8} 通常は拡張子がctlで、CTLファイルとも呼ばれる。

^{*9} CFもしくはCOARDS規約に従った記述をされたnetCDFファイルをGrADSで開くにはsdfopenというコマンドを使うが、ここではデータ記述ファイルを介してnetCDFファイルを開くためxdfopenというコマンドを用いる。

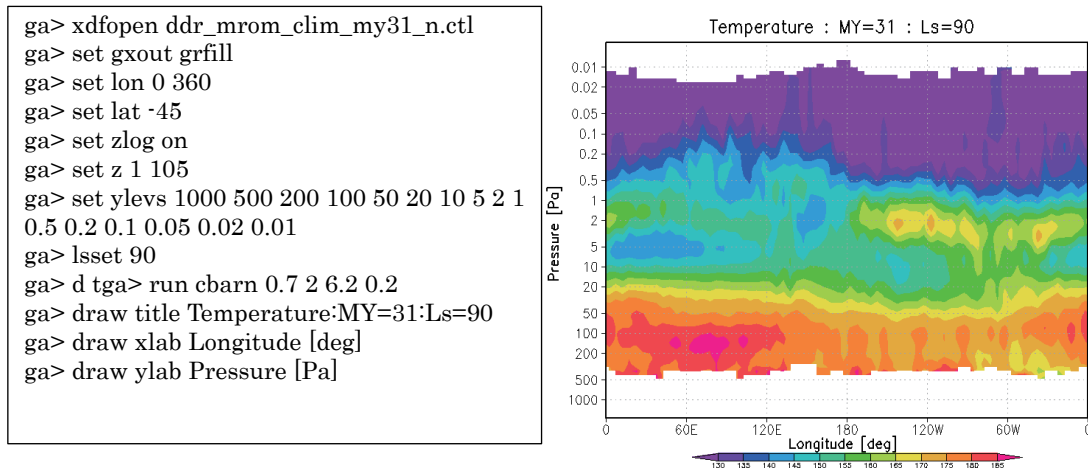


図 2 GrADS のコマンド例 (左) とそのコマンドで作成された図 (右). コマンド例において, 「ga>」は GrADS のプロンプトを表す. 図は, MY31, 南緯 45 度, Ls=90 度における気温の経度・気圧断面図である. 東経 50–100 度にかけて存在するヘラス盆地の上空 200–100 Pa 付近に気温の極大が現れている.

4. まとめ

米国 NASA の PDS で公開されている火星探査機 MRO 搭載の MCS 観測データの取り扱いを容易にするために, 二つの手法を提案した. まず, 配布されているテキスト形式の元データを netCDF 形式に変換して整理しなおすことで, データの可搬性と可読性を高めた. 次に, データを格子化して地球の大気・海洋科学分野で広く使われる GrADS で扱えるようにすることで, データの解析と可視化を容易にした.

本論文で解説をした MCS データ (netCDF 化及び格子化したデータ) は, 著者の所属する大学の Web サーバー^{*10}において一般に公開している. また, 学内における学生教育での利用も検討している. 今後, データ利用者からの意見も参考にして更なる改良を進めて行くとともに, 他の火星大気データについても同様な整備を実施することを目指したい. 本論文で紹介したような手法で整備したデータが他分野の研究者や学生にも広く使われることで, 火星大気研究の裾野が広がることを期待している.

なお, 本研究で行った第 1 段階の netCDF 化は情報量を失わない (可逆) ため, 本来はデータ配布元で実施されるべき内容かもしれない. 実際, PDS 内部の研究者からも将来的にはユーザの希望に応じたデータ形式の変換ツールを開発・配布する可能性が示されている.

^{*10} http://www.e.ics.nara-wu.ac.jp/~nogu/work/mars/database/mro_mcs/index.html (2016 年 8 月 19 日閲覧)

謝辞

MCS データは, NASA の PDS より配布されているものを使用させていただいた. また, MCS データの使用に際し, NASA Jet Propulsion Laboratory の Armin Kleinboehl 博士からは貴重なアドバイスをいただいた.

参考文献

- 1) Yamamoto, T. and K. Tsuruda, The PLANET-B mission, *Earth Planets Space*, 50, 175-181, 1998.
- 2) 佐藤毅彦, 久保田孝, 宮本英昭, 岡田達明, 松岡彩子, 今村剛, 尾川順子, 川口淳一郎, MELOS ワーキンググループ, MELOS 複合探査の概要と科学目標「火星はなぜ赤いのか?」, 日本惑星科学会誌「遊・星・人」18(2), 66-68, 2009.
- 3) 藤田和央, 佐藤毅彦, 火星着陸探査技術実証 WG, 火星着陸探査技術実証ミッション計画, 第 58 回宇宙科学技術連合講演会講演集, 1H09, 2014 年 11 月 12 日~14 日, 長崎ブリックホール.
- 4) 藤本正樹, MMX 計画, 日本地球惑星科学連合 2016 年大会, PCG10-16, 2016 年 5 月 22~26 日, 幕張メッセ.
- 5) McCleese, D. J., et al., Mars Climate Sounder: An investigation of thermal and water vapor structure, dust and condensate distributions in the atmosphere, and energy balance of the polar regions, *J. Geophys. Res.*, 112, E05S06, 2007.
- 6) Kleinboehl, A., et al., Mars Climate Sounder limb profile retrieval of atmospheric temperature, pressure, and dust and water ice opacity, *J. Geophys. Res.*, 114, E10006, 2009.
- 7) Rew, R. K., G. P. Davis, S. Emmerson, and H. Davies, *NetCDF User's Guide for C, An Interface for Data Access*, Version 3, 1997.
- 8) Doty, B. and J.L. Kinter III, The Grid Analysis and Display System (GrADS): A practical tool for earth science visualization. Eighth International Conference on Interactive Information and Procession Systems, Atlanta, Georgia, 5-10 January, 1992.

WebGL による「はやぶさ 2」「あかつき」のリアルタイム軌道可視化

柏井 勇魚^{*1}, 宮崎 剛^{*2}

Visualization of “Hayabusa2” and “Akatsuki” trajectories with WebGL in real-time

Isana kashiwai, Go Miyazaki

概略

筆者らは、2015 年 12 月 3 日の「はやぶさ 2」の地球スイングバイ、ならびに 2015 年 12 月 7 日の「あかつき」の金星軌道投入に際して、JAXA/ISAS などから一般向けに提供されている軌道データを利用して、Web ブラウザ上でのリアルタイム可視化を試みた。

Abstract

We have developed real-time visualization tools on web browsers for “Hayabusa 2” Earth Swing-by at 3, Dec, 2015 and “Akatsuki” Venus orbit insertion at 7, Dec 2015. These simulations used the data which is open to the public released by JAXA / ISAS or other sources.

Keywords

Data Visualization, Education, Outreach, Science Communication

キーワード

データ可視化, 教育, アウトリーチ, サイエンス・コミュニケーション

doi: 10.20637/JAXA-RR-16-007/0011

^{*} 平成 28 年 11 月 24 日受付 (Received November 24, 2016)

^{*1} isana.k@gmail.com (個人プロジェクトのため連絡先として E-mail アドレスを示す)

^{*2} wilhaftarkin@gmail.com (同上)

1. はじめに

本プロジェクトは、2015 年 12 月 3 日の「はやぶさ 2」の地球スイングバイ¹⁾、ならびに 2015 年 12 月 7 日の「あかつき」の金星軌道投入²⁾に際し、探査機の軌道、姿勢、位置などをユーザーが視点などを変更できるインタラクティブな 3D 表現によって、実際のミッションの進行と同時刻に Web ブラウザ上で再現するものである。

Web ブラウザ上での再現を行ったのは、ほぼすべてのパソコンにインストールされ、ごく一般的に利用されているソフトウェア上で探査機の位置や姿勢などをリアルタイムで提示することにより、誰でも手軽にこれらの情報にアクセスすることができ、またストリーミングによる動画中継、SNS などを通じた速報などと合わせ、コミュニケーションのツールとして利用されることで、イベントとしての一体感を産み、こうしたイベントを通じてユーザー同士のつながりが得られるであろうと考えたためである。

探査機の軌道投入などを可視化する一般向けのツールとしては、NASA が提供する NASA's Eyes³⁾、AGI 社の STK Viewer⁴⁾などがあるが、いずれもソフトウェアのインストールを必要とし、Web ブラウザ上では動作しない。また、動画中継などでこうした可視化ソフトウェアを利用した位置情報の提供なども行われているが、ユーザーによる視点の変更などのインタラクティブ性はない。本稿執筆時点では、ブラウザ上での探査機のリアルタイムシミュレーションは本プロジェクトがほぼ唯一の例であると思われる。

2. データの入手先

筆者らは当該プロジェクトとは直接の関わりのない外部の人間であるため、探査機の位置や姿勢などについては、Web サイトやアーカイブなどで一般に公開されているデータを利用した。

2.1 位置

一般向けに公開されている探査機などの位置データ入手先は、主に以下のようなものがあげられる。本プロジェクトではこれらのデータを併用する形で使用している。

- a) JPL HORIZONS⁵⁾
- b) JAXA DARTS (SPICE カーネル) ⁶⁾
- c) プロジェクト公式サイト等

a)JPL HORIZONS は NASA/JPL の軌道設計・ナビゲーション部門が運用している軌道データのデータベースであり、各惑星をはじめとする太陽系内天体の精密な軌道データならびに探査機などの人工天体の軌道データが入手できる。主に NASA の探査機のデータが公開されているが、「はやぶさ」，「はやぶさ 2」，「あかつき」など一部日本の探査機のデータも含まれている。

b)JAXA のサイエンスデータリポジトリである DARTS にも一部の探査機について軌道データが公開されている。「あかつき」は、DARTS において軌道データがテキストファイルおよび SPICE カーネルの形で公開されている。SPICE カーネルはバイナリデータであり、WebGL/JavaScript での利用には SPICE Toolkit に含まれるライブラリやユーティリティを利用してテキストファイルに変換する必要がある。

c)「はやぶさ 2」については、広報あるいはスイングバイの観測キャンペーンなどを目的としてリュウグウ到着までの軌道データならびに、スイングバイ時の詳細な軌道データがテキストファイルとして公開された⁷⁾。

本プロジェクトでは、「はやぶさ 2」については、公式サイトで公開されていたデータを中心に、再接近時など更に細かい刻み幅でのデータが必要な部分について JPL HORIZONS のデータを使用している。「あかつき」については、DARTS で公開された SPICE カーネルのデータを利用した。

「はやぶさ 2」「あかつき」のいずれの場合も光源の位置として太陽の位置の推算が必要であるが、これについては比較的精度の高い惑星略算式である VSOP87⁸⁾を利用した。また、背景となる恒星のデータについては Yale Bright Star Catalog⁹⁾を使用している。

2.2 姿勢

探査機の姿勢については、「はやぶさ」などで SPICE のデータとしてミッション終了後に公開されている例があるが、今回は「あかつき」，「はやぶさ 2」とともに事前の公開はなく、広報資料や探査機の仕様・形状などから推測する必要があった。

3. ブラウザ上での 3D 表現

現在、Web ブラウザ上での 3D 表現は WebGL¹⁰⁾が一般的になっている。現在の Web ページの標準規格である HTML5 に含まれ、最新版のブラウザであれば、スマートフォンやタブレットなどを含む、現行のほぼすべてのプラットフォームにおいて追加のプログラムやプラグインの導入無しで動作する。

本プロジェクトではこの WebGL を使って軌道データの可視化を試みた。実装においては、特殊な表現を多用する WebGL を JavaScript からアクセスしやすい形で表現し、また各プラッ

トフォーム間の差異などを吸収するオープンソースのラッパーライブラリ **Three.js**¹¹⁾を使用している。

4. 探査機の 3D データの作成

探査機のモデルを制作するにあたっては、「はやぶさ 2」「あかつき」ともに、図面などの公開はされていないため、一般に公開されている各種資料や機体公開時の写真などから細部を作成した。一部、それらの資料中に存在しない箇所については、他の探査機や人工衛星などから推定した部分もある。

Web ブラウザ上で利用するという制限から、使用する 3D オブジェクトはかなり軽量なものを制作する必要があった。正確で精細なデータは、美しく臨場感も増すが、一方でデータ量が大きくなり配信サーバに負荷をかけ、またユーザー側での計算量が増えて非力なマシンでは速度低下の恐れがある。

そのため、3D グラフィックソフト上で一旦精細なモデルを作成し、それを表面材質として簡易なモデル上に適用するという手法で一定のクオリティを保ったまま極めて軽量なモデルを実現した。いずれのモデルも探査機全体でファイルサイズは 1MB 以下となっている。

また、「あかつき」のモデルについては、太陽電池パネルの太陽指向を再現するため、本体と太陽電池を別のパーツとして作成して読み込んだ上で、プログラム上で配置し回転角度などを制御している。

図 1 に本システムで使用している「はやぶさ 2」および「あかつき」の 3D モデルを示す。

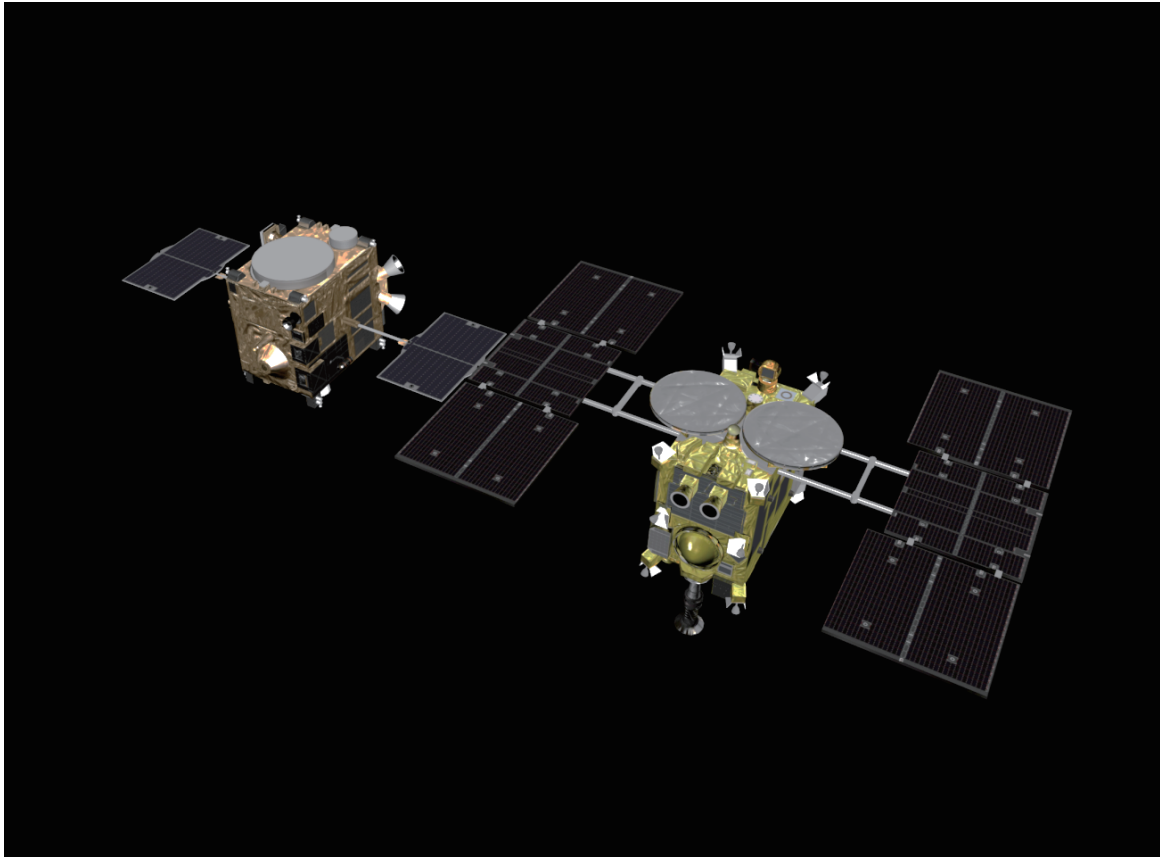


図 1: 同スケールで配置した「はやぶさ 2」および「あかつき」のモデル

5. 外部設計

本システムでは、対象となる天体の中心を 3D 空間の原点とし、座標系を各天体の赤道直交座標とした上で、探査機、惑星、恒星の各位置データを変換し各オブジェクトを配置した。カメラは探査機を中心としてユーザーのマウス操作によって視点を自由に変更できるようにした。

位置の更新頻度はいずれも 1 秒間隔としている。地球および金星への再接近時の動きが大きくなる限られた時間帯を除けば、位置情報の更新頻度を上げて画面には変化が判別できず、視覚上の効果はほとんどないためである。

本システムは、一般的なデスクトップパソコンやノート型パソコン、スマートフォン、タブレットなどで再生されることを意図している。そのため、位置や姿勢の推算においては、データポイント間を単純な線形補間によって位置を出力するなど、極力計算リソースの省力化を行った。また、ダウンロードされるデータのサイズをなるべく小さくしつつ、目標天体への接近時など位置変化が大きい部分についての密度を上げるために、中心天体との距離に応じて、データのポイントを 10 分毎、1 分毎、10 秒毎と変化させている。

さらに、探査機の 3D オブジェクトは実スケールではなく、中心天体との距離に応じてサイズを動的に変更するなどの操作をしている。これは数万キロないし天文単位スケールの位置データと、数メートルスケールの探査機の 3D オブジェクトを、ブラウザ上での精度の低い計算の中で同居させると、計算誤差にともなって画面上で探査機の位置が安定せず、大きくぶれて表示されてしまうためである。これを防止するために、それぞれの探査機は中心天体からの距離に応じて数十 km から数百 m までサイズを変化させて表示している。視覚的には判別することはできないため、ここではシミュレーションとしての正確さよりも画面表示上の安定を優先した。

画面設計においては、SNS などのオンライン上のコミュニケーションサービスなどで共有されることを意図し、インターフェイスは極力シンプルなものとし、また説明や設定項目などを極力排し、ブラウザで当該ページにアクセスするだけで、誰でもそれぞれの探査機の現在位置や状態、地球や目標天体との距離などが分かる、という状態を維持するよう務めた。

本システムの画面構成を図 2 に示す。図は「はやぶさ 2」地球スイングバイ時のものであるが、「あかつき」もほぼ同じ画面構成となっている（具体例は図 4ー図 8 を参照）。画面右下に現在時刻、打上からの経過時間、最接近までの時間、相対速度、天体表面までの距離を表示した。

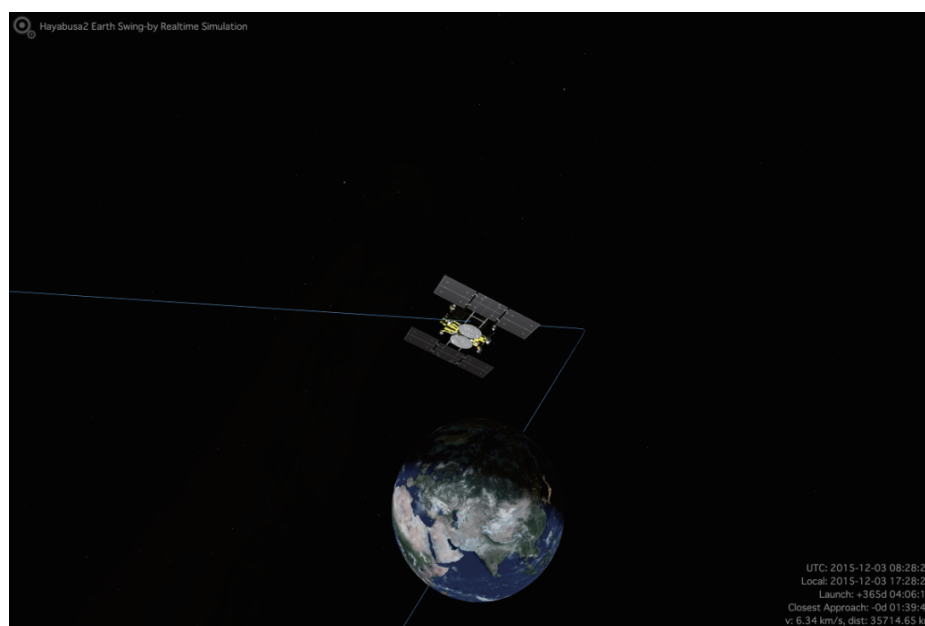


図 2: 本システムの画面構成。「はやぶさ 2」地球スイングバイ（2015-12-03 08:28 UTC）

5.1 「はやぶさ 2」の地球スイングバイ

「はやぶさ 2」の地球スイングバイ時の姿勢は、当初は、スイングバイ中の姿勢の変更がないと仮定した上で、事前の情報から、Z+面を太陽方向に向け、Y+面を天の北極方向に向けた姿勢とした。その後、はやぶさ 2 プロジェクトより STK Viewer を利用したシミュレーションが公開されたため¹²⁾、その姿勢を参考に天頂より 30 度傾けた姿勢とした。

スイングバイ後に「はやぶさ 2」プロジェクトから接近時および離脱時に撮影された画像が公開された。地球接近時の画像を本システムの表示と比較したものを図 3 に示す。

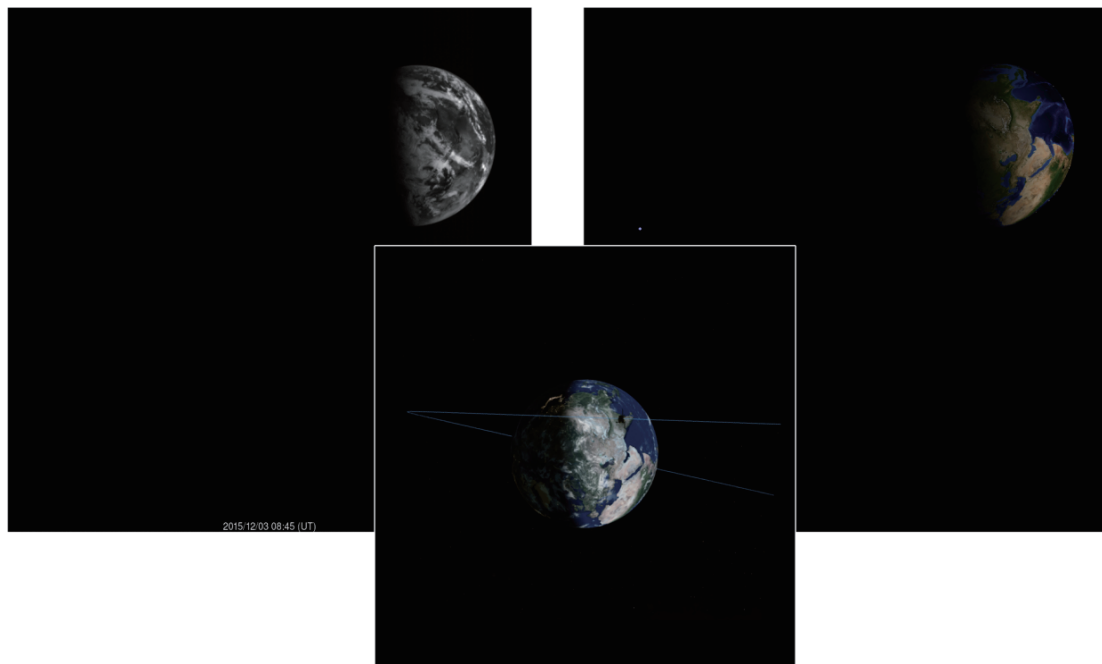


図 3: 右:「はやぶさ 2」の OCN W2 によって撮影された実画像¹³⁾、左:はやぶさ 2 プロジェクトチームによる視野再現¹³⁾、中央:同時刻のシステム上の表示。はやぶさ 2 は画面中央に位置する (2015-12-13 08:45 UTC)

本システムは「はやぶさ 2」からの視界そのものを再現するものではないため、この比較から予測した姿勢が正確であったかどうかを判断することは難しいが、同時刻の地球の姿勢や光のあたり方などが適切に表現されているのが確認できる。

5.2 「あかつき」の金星軌道投入

「あかつき」の金星軌道投入においては、スラスターの噴射やそれに伴う姿勢変更が発生する。これを本システムで再現するためには、各イベントのタイミングおよび姿勢を各種の資料から再構築する必要があった。金星軌道投入時に予定されていたマニューバを本システム上の表現と合わせて図 4 から 図 8 に示す。

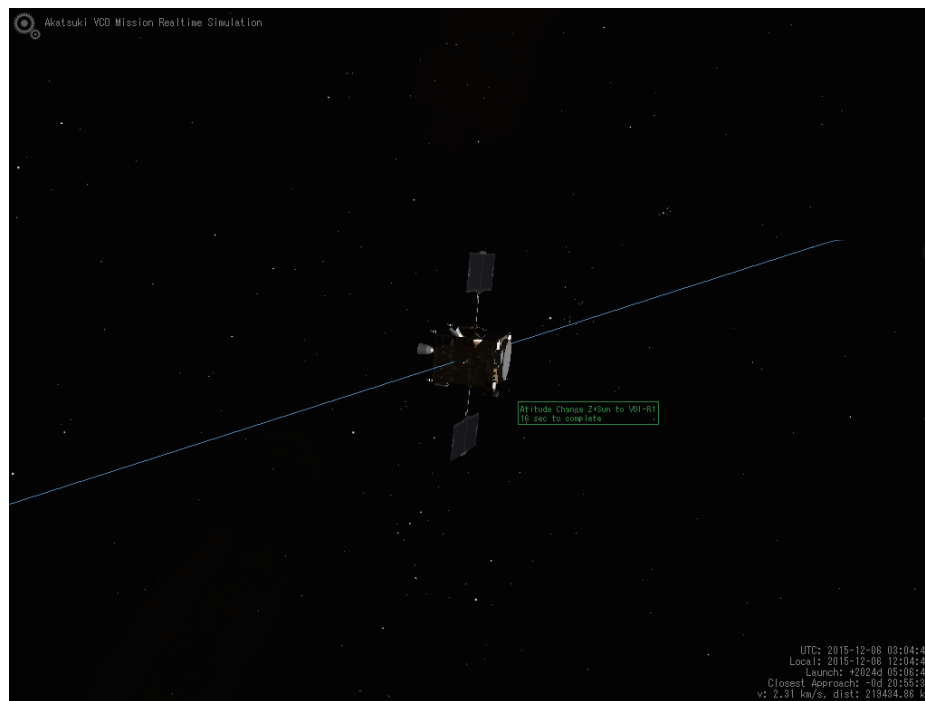


図 4: 太陽指向から軌道投入噴射姿勢(VOI-R1)へ(2015-12-06 03:00 UTC)

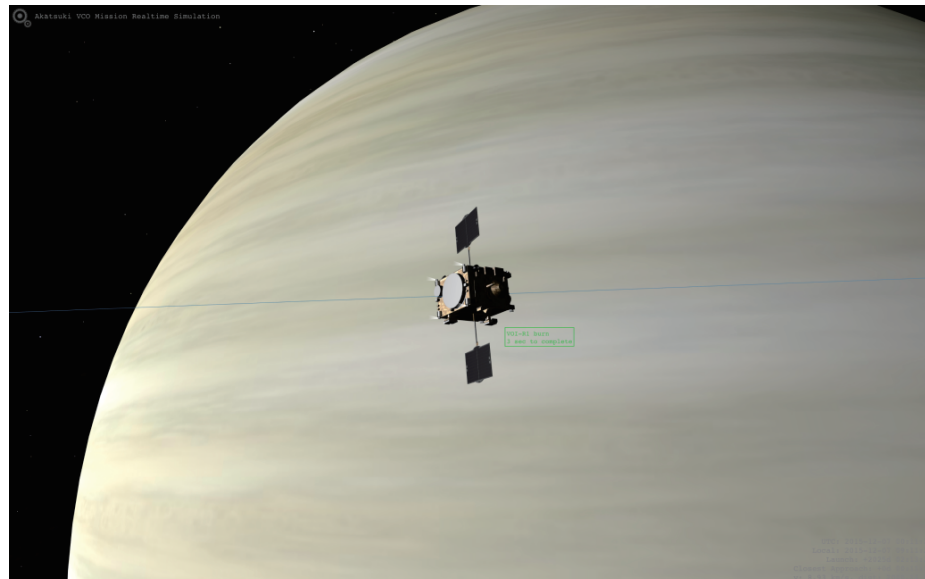


図 5: 軌道投入噴射(2015-12-07 00:05 UTC)

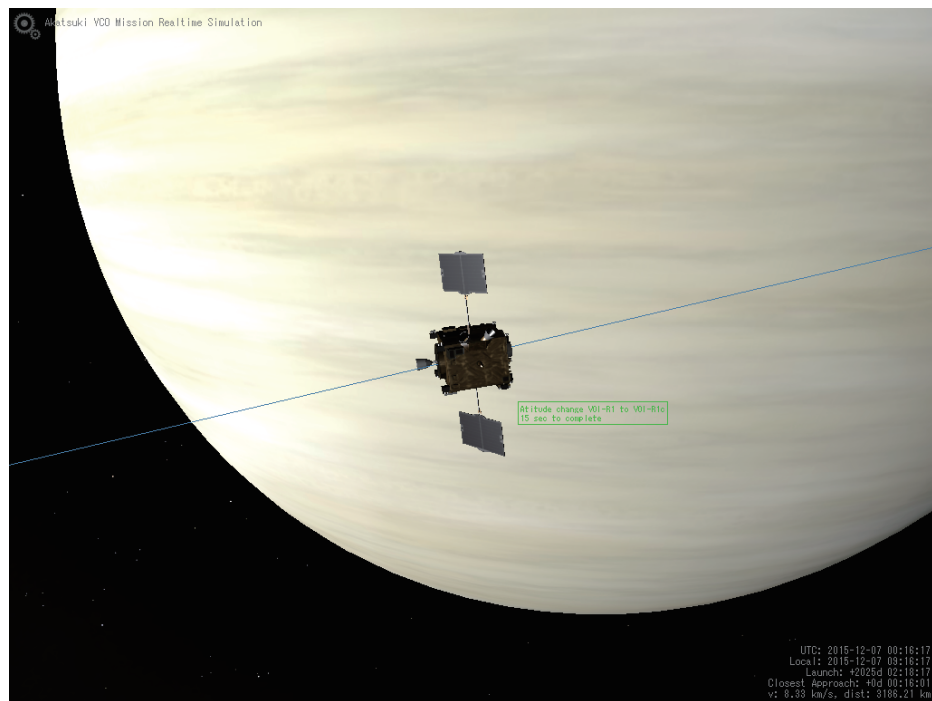


図 6: 軌道投入噴射姿勢(VOI-R1)から予備噴射姿勢(VOI-R1c)へ(2015-12-07 00:16 UTC)

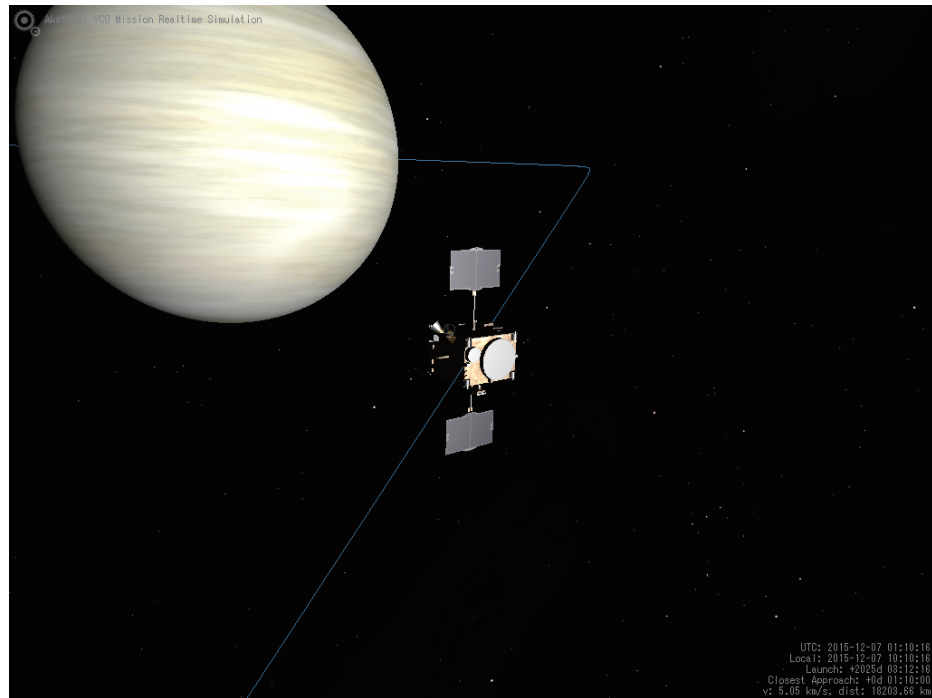


図 7: 予備噴射姿勢(VOI-R1c)から太陽指向姿勢へ(2015-12-07 01:10 UTC)

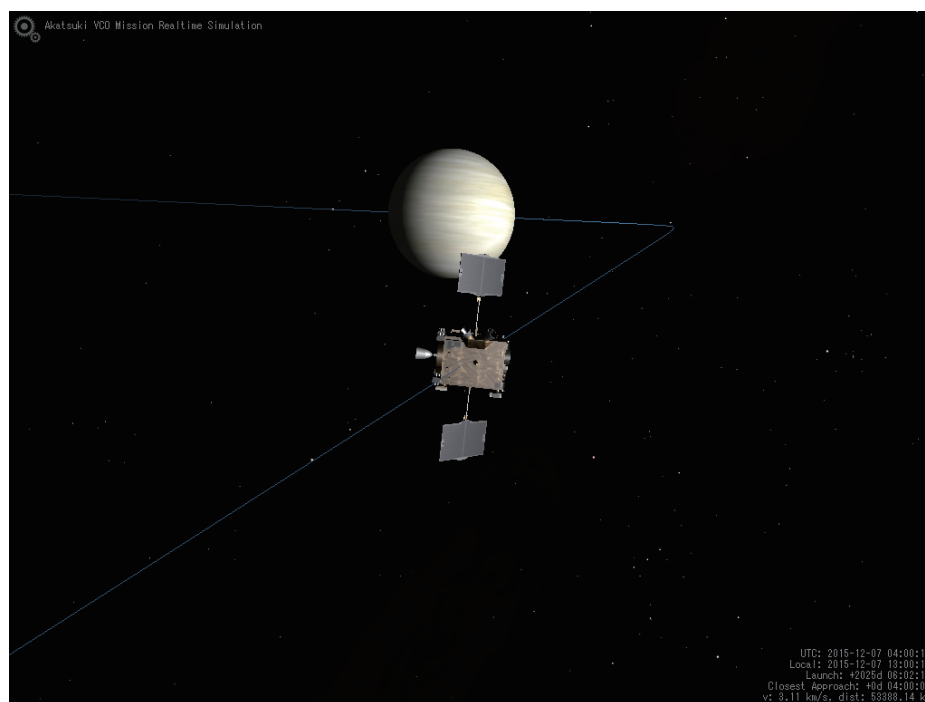


図 8:太陽指向姿勢から観測姿勢へ(2015-12-07 04:00 UTC)

これらの姿勢とタイミングについては、スラスタの方向などから姿勢を推定し、事前の記者会見で口頭並びに資料内での断片的な情報からシーケンスを再構築した。姿勢変更にかかる時間など、一部のタイミングについては言及がなかったが、他のマニューバの持続時間などから推測することとした。

なお、本システムの「あかつき」の金星軌道投入のシミュレーションでは、複数回の姿勢変更など、キーとなるイベントが多いため、イベントまでの時間をカウントダウンする小ウインドウを必要時応じて画面中央に表示するようにした(図9)。

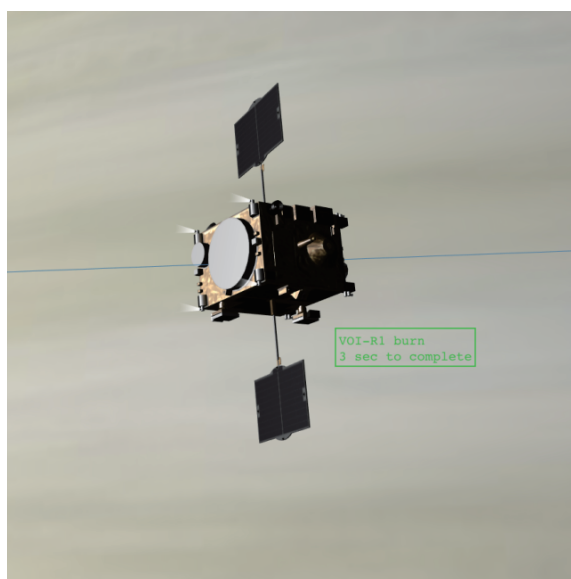


図 9:画面中央の拡大。探査機右下にカウントダウンが表示される

「あかつき」の軌道投入後に、軌道投入直後に撮影された金星の画像が公開された¹⁴⁾。同時刻の本システム上の表示と比較したものを図 10 に示す。

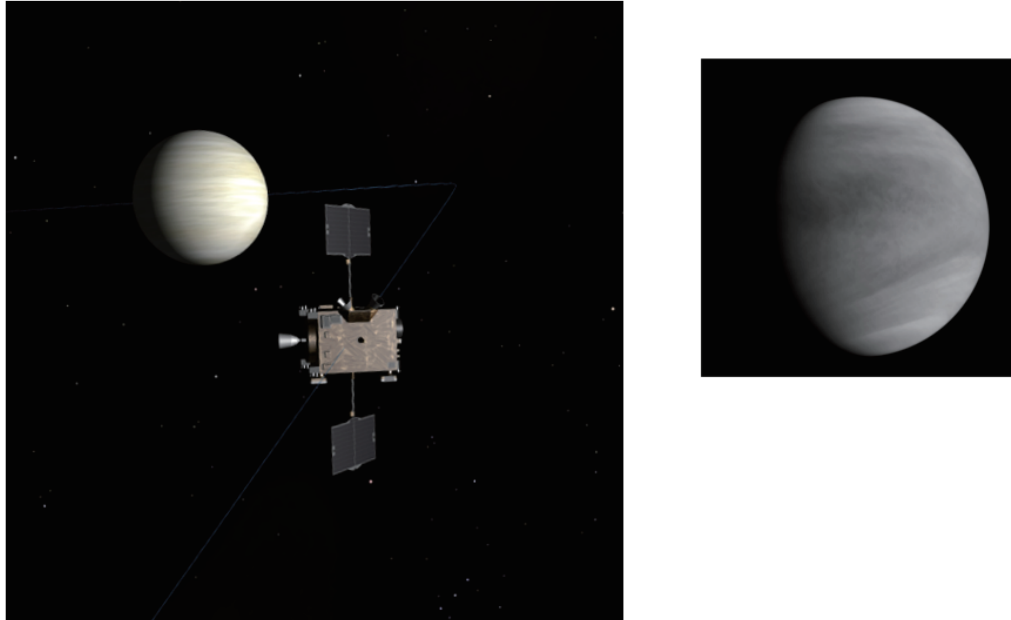


図 10: 右: 「あかつき」の軌道投入直後に撮影された金星, 左: 本システム上での再現(2015-12-07 05:19 UTC)

「はやぶさ 2」と同じく、本システムは「あかつき」に搭載されたカメラの視野を再現するものではないが、同時刻の金星への光の当たり方、観測面を金星に向けた探査機の姿勢などが適切に表現されていることが分かる。

6. まとめ

本システムの告知は、「はやぶさ 2」の地球スイングバイ、「あかつき」の金星軌道投入ともに、約 1 ヶ月前に公開し、SNS 上で定期的に告知をおこなったが、いずれも前後一週間で、100 カ国をこえる国と地域から 1 万~2 万 PV のアクセスがあった。また、殆どのアクセスが、イベント当日に集中していた。

個々のユーザーは、画面のスナップショットを SNS で共有しながら、当該探査機やイベントなどについて解説を加える、あるいは印象や感想を述べ合うといった使い方をしており、本システムを「体験を共有するためのツール」ないし「コミュニケーションのツール」として利用される場面が多く見られた。これは当初意図した通りの利用形態といえる。誰もが容易にアクセスできる環境での、リアルタイムでの位置情報の提供はアウトリーチにおいて非常に強力なツールとなる可能性があること示すものといえる。

課題としては、正確さの向上が挙げられる。本プログラムの主目的はあくまで啓蒙ないしエンターテインメントであり、サイエンスで要求されるような精度は必要とされない。とはいえ、事前の情報や資料、あるいは宇宙機についての常識などから逸脱した内容であっては、ユーザーの誤解や混乱に繋がるため、極力事前情報などと矛盾せず、事実 に即した内容となるよう努めた。結果的には、後日発表された撮影画像などとも大きな矛盾はなく、ユーザーにとって違和感のない内容になっていたといえよう。一方で、公開されていない姿勢やシーケンスなどについては推測に頼らざるを得ず、正確性という意味で課題を残した。

「はやぶさ 2」については、2018 年には本番ともいえる小惑星への接近、着陸が控えている。筆者らは、これらのイベントについても同様のシステムでの再現を計画している。小惑星への着陸に際しては、探査機の位置や姿勢について今回よりさらに精度の高いシミュレーションが要求される。さらに、計画が状況に応じて短期間に変更される可能性も高く、システムの汎用性や運用上の即応性も必要になることが予想される。こうした点が今後の開発課題となるだろう。

また、実現には探査機の位置や姿勢の計画値の公開、並びに継続的なアップデートが不可欠であり、これについてはぜひ JAXA/ISAS での公開を検討していただきたい。

参考文献

- 1) 柏井勇魚, 宮崎剛, Hayabusa2 Swingby Simulation <http://www.lizard-tail.com/isana/hayabusa2/misc/swingby/?t=201512030940> ※当日の再現, 2016. 06. 25 閲覧
- 2) 柏井勇魚, 宮崎剛, Akatsuki VOI Simulation, http://www.lizard-tail.com/isana/orbview/misc/akatsuki_orbit_insertion.html?t=20151207001100 ※当日の再現 2016. 06. 25 閲覧
- 3) NASA, “NASA’s Eyes”, <https://eyes.nasa.gov/>, 2016. 06. 25 閲覧
- 4) AGI “STK Viewer” <http://www.agi.com/products/stk/modules/default.aspx/id/stk-viewer>, 2016. 06. 25 閲覧
- 5) NASA/JPL, “HORIZONS System” <http://ssd.jpl.nasa.gov/?horizons>, 2016. 06. 25 閲覧
- 6) JAXA, “DARTS” <http://www.darts.isas.jaxa.jp/index.html>, 2016. 06. 25 閲覧
- 7) JAXA, はやぶさ 2 プロジェクト, はやぶさ 2 特設サイト; 小惑星遷移軌道計画情報 <公開用概略計算値>, <http://fanfun.jaxa.jp/countdown/hayabusa2/>, 2016. 11. 15 閲覧
- 8) Bretagnon, P.; Francou, G. “Planetary theories in rectangular and spherical variables - VSOP 87 solutions”, <http://adsabs.harvard.edu/abs/1988A&A...202..309B>, 2016. 06. 25 閲覧

- 9) Hoffleit, D. and Warren, Jr., W.H., 1991, "The Bright Star Catalog, 5th Revised Edition (Preliminary Version)". <http://tdc-www.harvard.edu/catalogs/bsc5.html>, 2016.06.25 閲覧
- 10) WebGL Working Group, WebGL - OpenGL ES 2.0 for the Web, <https://www.khronos.org/webgl/>, 2016.06.25 閲覧
- 11) three.js - Javascript 3D library, <http://threejs.org/>, 2016.06.25 閲覧
- 12) はやぶさ2プロジェクトチーム, "はやぶさ2 スイングバイ軌道3次元表示の公開について", <http://www.hayabusa2.jaxa.jp/topics/soft/>, 2016.06.25 閲覧
- 13) JAXA はやぶさ2プロジェクト, 「はやぶさ2」スイングバイ直前に撮影された地球の画像, <http://www.hayabusa2.jaxa.jp/topics/20151203/>, 2016.06.25 閲覧
- 14) JAXA/ISAS, "金星探査機「あかつき」の金星周回軌道投入結果について", <http://www.isas.jaxa.jp/j/topics/topics/2015/1209.shtml>, 2016.11.15 閲覧

多様な環境においてデジタル立体地球儀を実現するための ダジック・アースの開発

齊藤 昭則^{*1} 津川 卓也^{*2} 市川 浩樹^{*3} 島田 卓也^{*4}

Development of Dagik Earth to utilize digital globe
for various purposes

Akinori SAITO^{*1} Takuya TSUGAWA^{*2} Hiroki ICHIKAWA^{*3} Takuya SHIMADA^{*4}

Abstract

Three-dimensional digital globe that uses a spherical screen can display the time series of Earth and planetary science data, and space science data in 3D. The usage of it has been limited in large-scale science museums because the system is complicated, and costs many resources. A digital globe system, Dagik Earth, has been developed to utilize it for various purposes with low cost. Dagik Earth consists of 1) computer 2) PC projector 3) spherical screen, and 4) controller. Various sizes of spherical screens from 8cm to 16m, and various types of controllers, such as track-ball, touch-pad, and game-controllers can be used for Dagik Earth. Three types of digital globe software have been developed. “Chikyu Kuru Kuru (CKK)” is used on PCs with Windows and Mac. “Dagik earth On theWeb (DOW)” is executed on Web browsers on various devices. “Dagik Earth Appli” is for iPad tablet device. Using these software, Dagik Earth can be used with various type of computers including tablet PCs and smart phones. All of these three Dagik Earth software use a common set of data and image files. The set of data and image files consist of 1) surface map image files, 2) caption image files, and 3) configuration files. Users of Dagik Earth can create and present their own set of files. Dagik Earth enables to utilize digital globe in classrooms and science exhibition for which the conventional large-scale digital globe cannot be used.

Keywords: digital globe, virtual globe, 3D visualization, Earth and planetary science, public outreach

doi: 10.20637/JAXA-RR-16-007/0012

* 平成 28 年 11 月 24 日受付 (Received November 24, 2016)

^{*1} 京都大学 (Kyoto University)

^{*2} 情報通信研究機構 (National Institute of Information and Communications Technology)

^{*3} 愛媛大学 (Ehime University)

^{*4} CrossHat

概要

球形の立体ディスプレイに地球画像やデータを表示するデジタル立体地球儀は従来の地球儀に比べて表示画像を動的に変更することが可能であり、科学館での多様な地球惑星科学・宇宙科学データの展示に用いられている。このデジタル立体地球儀を多様な環境で容易に実現するシステム「ダジック・アース」の開発を行った。ダジック・アースの構成機器は、1) 表示デバイス、2) PC プロジェクター、3) 球形スクリーン、4) コントローラーの4点である。球形スクリーンとしては直径8cmのものから直径16mのものまで用いられており、コントローラーはトラックボール、タッチパッド、ワイヤレスマウス、ゲームコントローラーなど多様なものが用いられている。表示デバイス上で実行されるダジック・アース表示ソフトは、1) Windows 及び Mac 搭載のパソコンで実行ができるダジック・アース表示ソフト「Chikyu Kuru Kuru(CKK)」、2) 多くのデバイス上のウェブブラウザで実行ができる「Dagik earth On the Web (DOW)」、3) タブレット端末 iPad で実行ができる「ダジック・アース・アプリ」の3つが開発されており、多様な表示デバイスでの実行が可能である。いずれの表示ソフトも同一の「ダジック・アース・コンテンツ」ファイルを用いており、一つのコンテンツを開発することで多くの環境での表示が可能となる。「ダジック・アース・コンテンツ」ファイルは、1) 球面画像ファイル 2) 説明画像ファイル 3) 設定ファイル、の3種類のファイルからなり、利用者が独自に作成し表示することも可能である。従来は限られた科学館のみで使われていたデジタル立体地球儀がダジックアースによって学校での授業や科学イベントなど多様な環境で実施できるシステムが整備された。

1 はじめに

地球儀をソフトウェアで表現する Virtual Globe ソフトは 2004 年公開の NASA による World Wind¹⁾ によって地球惑星科学データの可視化に用いられるようになり、2005 年公開の Google による Google Earth²⁾ によってより広く利用されるようになった^{3,4)}。GoogleEarth は Keyhole Markup Language(KML) で記述したファイルを用いることでユーザーが画像やオブジェクトを Google Earth 上に表示させることが容易であり、地球科学データの可視化に多く用いられるようになった。Saito らは KML ファイルを用いた地球惑星科学データの可視化データベースである Dagik を開発し⁵⁾、様々なデータの Quick Look(簡易表示)を Google Earth 上で表示できる環境を構築した⁶⁾。

同時に Saito らは Virtual Globe ソフトの表示画面を PC プロジェクターを用いて球形のスクリーンに投影することで、デジタル立体地球儀表示ができることを着想し、ダジック・アースとして^{7,8)} その利用環境の整備を進めた。デジタル立体地球儀とは球形の立体ディスプレイに地球画像やデータを表示するもので、従来の地球儀に比べて表示画像を動的に変更することが可能で多様なデータの表示を行うことができる。また Virtual Globe ソフトによる平面ディスプレイでの表示では、球面である地表面の形状は広範囲を表示するときには大きくゆがんで表示されるのに対し、デジタル立体地球儀では地球全体を正しい形状で表示することができる点が特徴である。また、宇宙から地球を見るような視点で地球全体を眺めることが可能であるため、一般向けの展示機材として適している。ダジック・アースの最初の展示は 2007 年日本地球惑星科学連合同大会において実施された。

デジタル立体地球儀としては、日本科学未来館が大型のシステムである Geo Cosmos⁹⁾ を開発・運用しており、米国大気海洋局 (NOAA) による Science On a Sphere (SOS) が多くの科学館で

用いられている¹⁰⁾。これらのシステムが専用の機器を用いた大規模なものであるのに対し、ダジック・アースは通常のパソコンと PC プロジェクターを用いたものであり、従来の科学館での利用だけではないデジタル立体地球儀の幅広い利用が期待される。ダジック・アースはその特徴を活用するため、多様な環境での実施を目指して開発が進められた。また、多様なデータの表示も目的としており、ユーザーが独自にコンテンツを作成して表示できるようにすることでより多様なデータが表示できるようことを目指している。

本稿においては、このような多様性を実現するためのダジック・アースのソフトウェアについて記述する。まずダジック・アースのシステムについて概観し、そのシステム構成上の課題を述べ、その後、その課題を解決するためのシステムをコンテンツの構成、表示ソフトウェアに分けて述べる。

2 ダジック・アース・システムの構成機器と開発上の課題

ダジック・アースによるデジタル立体地球儀の様子を Fig. 1 に示した。左図は 2011 年 8 月 28 日に国立科学博物館で行われた展示の様子である。左側手前に設置された直径 2m の風船式の球形スクリーンに右側奥に設置された PC プロジェクターから投影している。アフリカからアラビア半島、インドにかけた領域の陸地と気象衛星による雲画像が投影されているが、球面上に正しい形状で投影されていることが分かる。中央の観覧者が、床に設置された白い四角形のゲームコントローラーに乗り、地球画像を操作しながら、その左側の説明者より解説を受けている状況である。PC プロジェクターとパソコンは机の上に設置されている。右図は、内部に PC プロジェクターを格納した箱に設置した半透明半球形スクリーンに内部から投影している例である。この場合は表示デバイス及びコントローラーとしてタブレット端末 iPad を使用している。



Figure 1: ダジック・アースによるデジタル立体地球儀展示例 (左) 直径 2m 球形スクリーン使用例 (右) 裏面投影半球形スクリーン使用例

ダジック・アースの機器の構成を Fig. 2 に示した。構成要素は、1) 表示デバイス、2) PC プロジェクター、3) 球形スクリーン、4) コントローラーの 4 点である。表示デバイスはパソコン、タブレット端末などであり、この機器で Virtual Globe ソフトであるダジック・アース表示ソフトが実行される。表示ソフトでは球形スクリーンに投影した時に正しい形状に表示される形で地表面画像を表示しており、その画像は接続された PC プロジェクターに出力され、球形スクリーンに投影される。コントローラーは表示デバイスに接続され、表示ソフトの操作に用い

られる．この操作とは，表示される地球画像の回転などの操作や，表示されるコンテンツの選択などである．Fig. 1 に示した例では，上に乗り体重移動で操作をする形態のゲームコントローラーが用いられている．

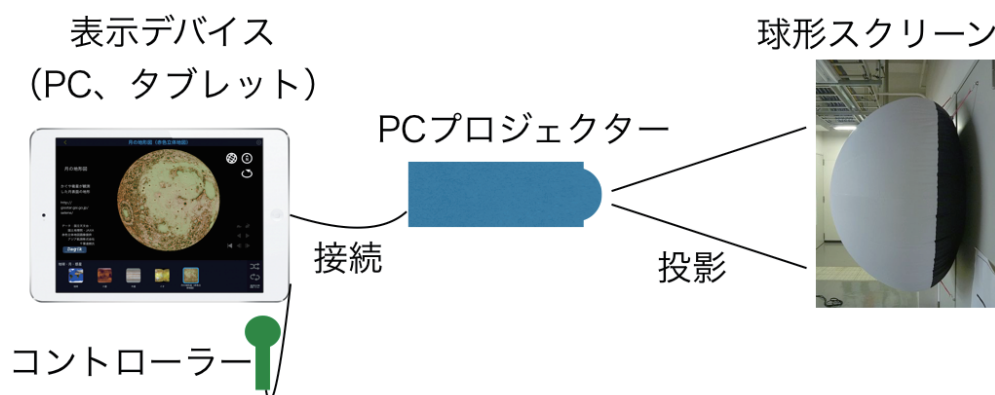


Figure 2: ダジック・アースの構成機器

多様な環境でデジタル立体地球儀を実現するために，多くの表示デバイスに対応することがシステムの開発において課題とされた．そのためのダジック・アース表示ソフトとしては，Windows 及び Mac 搭載のパソコンで実行できるダジック・アース表示ソフト「ChikyuKuru Kuru (CKK)」，多くのデバイス上のウェブブラウザで実行ができる「Dagik earthOn the Web (DOW)」，タブレット端末 iPad で実行ができる「ダジック・アース・アプリ」の 3 つが開発されている．これらにより，多くの表示デバイスを用いることが可能となっている．

ダジック・アースの投影に用いられる PC プロジェクターは通常のものであり，レンズなどに特殊な変更を加えられていない．球形スクリーン上の投影画像の各点から PC プロジェクターまでの距離差は，平面スクリーンへの投影で想定されている距離差とは異なるため，投影画像上で焦点が合わない部分が生じるが，多くの PC プロジェクターにおいてその焦点ずれによる画像のブレはさほど大きくはなく，投影に大きな影響を与えない．

PC プロジェクターから球形スクリーンへの投影方法は以下の 2 通りである，1) 球形スクリーンの球面に対して，球の中心と反対方向に PC プロジェクターを設置する場合 2) 球形スクリーンの球面に対して，球の中心と同一方向に PC プロジェクターを設置する場合．1) の場合は Fig. 1 (左図) や Fig. 2 に示した配置であり，球面の外側から投影する形態となる．この場合は，球形スクリーンは光を反射する素材のものとなる．1 台の PC プロジェクターによって半球への投影が可能であり，全球形ではなく半球形のスクリーンが用いられる場合も多い．PC プロジェクターの設置位置がスクリーンに近い場合は，1 台の PC プロジェクターから投影可能な領域は半球より狭くなることがある．複数の PC プロジェクターを用いることによって投影領域を半球よりも広げることが可能である．CKK では，2 つのウィンドウに異なる視点から見た地球画像を表示し両者を同期させて操作することができる．この 2 つのウィンドウを PC に接続した 2 つの PC プロジェクターに個々に出力し，球形スクリーンの両側から投影することによって広範囲な投影が可能である．この投影方法に用いられる球形スクリーンとしては，半球形，全球形という形状の違いだけではなく，大きさも直径 8cm から直径 16m までの多様なものが用いられており，その材質も風船式のビニールによるもの，発泡スチロール，FRP など様々なものが使用されている．

2) の場合は，球面の内側から投影する形態であり，Fig. 1 (右図) に示した配置である．この場合は，光を透過する半透明の素材で作られた球形スクリーンを用いることとなる．この配

置の投影では魚眼レンズなどを用いた超近接投影用のプロジェクターも使用可能である。半球形のスクリーンが用いられることが多いが、魚眼レンズを用いた PC プロジェクターの投影性能によって、半球形以上の広い領域の球形スクリーンへの投影も可能である。CKK では角度幅 270 度の投影まで対応しており、北極頂点として投影した場合、南緯 45 度以北の投影が可能となる。2) の投影方法における球形スクリーンとしては、直径 15cm から直径 1m までが用いられている。

立体地球儀投影のためには球形スクリーンを用いる必要があるが、平面スクリーン及びディスプレイでの投影であっても、地球画像を回転して異なる角度から見ることによって、表示されているものが球体であると認識することができるため、球形スクリーンが利用できない場合には簡易な表示として平面スクリーン及びディスプレイを用いることも可能である。

コントローラーは、地球画像の回転操作、時系列画像の再生・停止操作などや、メニューによってコンテンツの選択を行うために用いられる。デジタル立体地球儀の用途と利用形態によって多様なコントローラーが使用できるようにすることが開発の課題であった。そこで表示デバイスとしてパソコンを用いる CKK と DOW では、マウスとキーボードの両方による操作に対応させ、使用するコントローラーからの入力をマウスあるいはキーボードの入力信号に置き換えることで、多様なコントローラーの使用を可能にした。使用されているコントローラーはマウス、キーボード、トラックボール、タッチパッド、タッチディスプレイ、ワイヤレスマウス、ゲームコントローラー、ジョイスティック、ジェスチャーコントローラー、QR コードリーダー、NFC リーダーなど多岐にわたっている。

上記のように、4 つの構成機器に関していずれも多様な機器に対応することがダジック・アースの特徴であり、それにより幅広い用途での利用を実現している。そのような多様性への対応とコンテンツ開発の容易さの実現が開発における課題であった。

3 ダジック・アース・コンテンツの構成

表示デバイス上で実行されるダジック・アース表示ソフトは、Fig. 3 に示したように、一つのテーマによる画像などから構成された「コンテンツ」と、その複数のコンテンツの一覧を表示し、実行する「コンテンツ」を選ぶ「メニュー・プログラム」の 2 つから構成される。表示ソフトは上記のようにパソコン・ソフト「CKK」、ウェブブラウザ用「DOW」、タブレット用「ダジック・アース・アプリ」の 3 つが用いられているが、「コンテンツ」はこれら 3 つに共通で用いられる形式となっている。「コンテンツ」のテーマは、春の雲の動き、日本にきた台風、オゾン量の変化、台風・ハリケーン等の発生場所と海面温度、海面近くの海水の塩分、世界の地震分布、海洋底の拡大、地球と月と惑星、月の満ち欠け、金星の雲：「あかつき」の観測、星の配置、宇宙からみたオーロラ、日本最古の地球儀、クイズ：地球の昼と夜、などであり、約 150 コンテンツが作成されている。リアルタイム・コンテンツは実時間に近いデータを表示するものであり、サーバーから画像ファイルのダウンロードを行い、表示するものである。

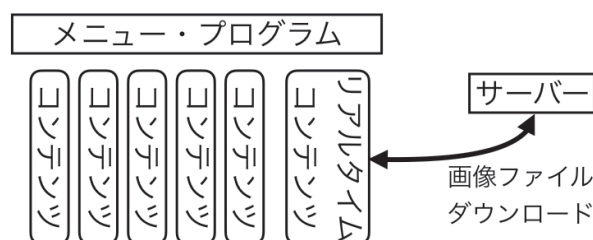


Figure 3: ダジック・アース表示ソフトの構成

ダジック・アースのコンテンツの構成は、Fig. 4 に示したように、すべてのコンテンツに共通する表示ソフトである CKK 及び DOW と、各コンテンツごとに異なる、1) 球面画像ファイル 2) 説明画像ファイル 3) 設定ファイル、の 3 種類のファイルからなる。これらのファイルのファイル名あるいは保存されているディレクトリは Table 1 にまとめられている。表示ソフトとしては、Windows 用の CKK の実行形式が Dagik Earth.exe, Dagik Earth 64bit.exe として含まれている。Dagik Earth 64bit.exe は 64bit Windows 用の実行形式である。Mac 用の CKK は Dagik Earth.app として含まれている。DOW はウェブブラウザで表示されるため HTML 形式のファイルとして作成されており、index.html, index-english.html, index-chinese.html として含まれている。index-english.html と index-chinese.html はそれぞれ英語用、中国語用の DOW である。



Figure 4: ダジック・アース・コンテンツの構成

Table 1: ダジック・アース・コンテンツの構成ディレクトリ名及びファイル名

	ファイル名・ディレクトリ名	ディレクトリ名 (第2レイヤー)
表示ソフト：CKK	Dagik_Earth.exeなど	
表示ソフト：DOW	index.htmlなど	
地表画像	data/images/map	data/images/map_second
説明画像	data/images/screen	data/images/screen_second
説明画像 (英語)	data/images/screen-english	data/images/screen_second-english
設定ファイル：コンテンツ毎	data/conf/conf.txt	
設定ファイル：共通	data/conf/init_conf.txt	
設定ファイル：DOW	data/conf/dow_conf.txt	

3.1 球面画像

球形スクリーン上に表示される球面画像は、正距円筒図法 (Equirectangular projection) で描画されたものを用いる。Fig. 5 に、球面画像ファイルの例を示した。経度幅 360°, 緯度幅 180° の球面を横と縦の比が 2 : 1 となる画像として表現しており、この例では緯度、経度それぞれ 10 度ごとに点線、30 度ごとに実線を引いている。画像ファイルの形式は JPEG が用いられている。解像度は、ダジック・アース表示ソフト「Chikyu Kuru Kuru(CKK)」で用いている OpenGL の古いバージョンにのみ対応しているパソコンでは画像の画素数が 2 の乗数以外では不具合が出ることがあるため 1024 ピクセル× 512 ピクセル、2048 ピクセル× 1024 ピクセル、4096 ピクセル× 2048 ピクセルなどが用いられているが、多くのコンテンツでは投影に用いられる PC プロジェクターの解像度と同程度の解像度になる 2048 ピクセル× 1024 ピクセルが多く用いられている。

球面画像ファイルが保存されているディレクトリは Table 1 に記述されているように “data/

images/map”であるが、これは設定ファイルにより変更可能である。CKK に用いられる第 2 レイヤーの球面画像ファイルは “data/images/map second” に保存されている。

3.2 説明画像

Fig. 6 に、説明画像ファイルの例を示した。この例ではコンテンツの題名とロゴのみが表示されているが、説明画像ファイルには、コンテンツの題名、データについての説明、データ提供元、コンテンツ作成者などの情報が記入される。またこれらの説明を多言語に表示するため、それぞれの言語に応じた説明画像ファイルが作成されている。Fig. 6 左図は日本語の説明画像ファイルであるが、同様のものを英語にしたものが Fig. 6 右図であり、このように、説明画像ファイルを日本語、英語、中国語にて作成し、表示ソフトによってどの言語の画像ファイルを用いるか選択することで、これらの 3 言語での表示を可能にしている。

説明画像ファイルが保存されているディレクトリは Table 1 に記述されているように “data/images/screen” であるが、これも設定ファイルにより変更可能である。CKK に用いられる第 2 レイヤーの球面画像ファイルは “data/images/screen second” に保存されている。

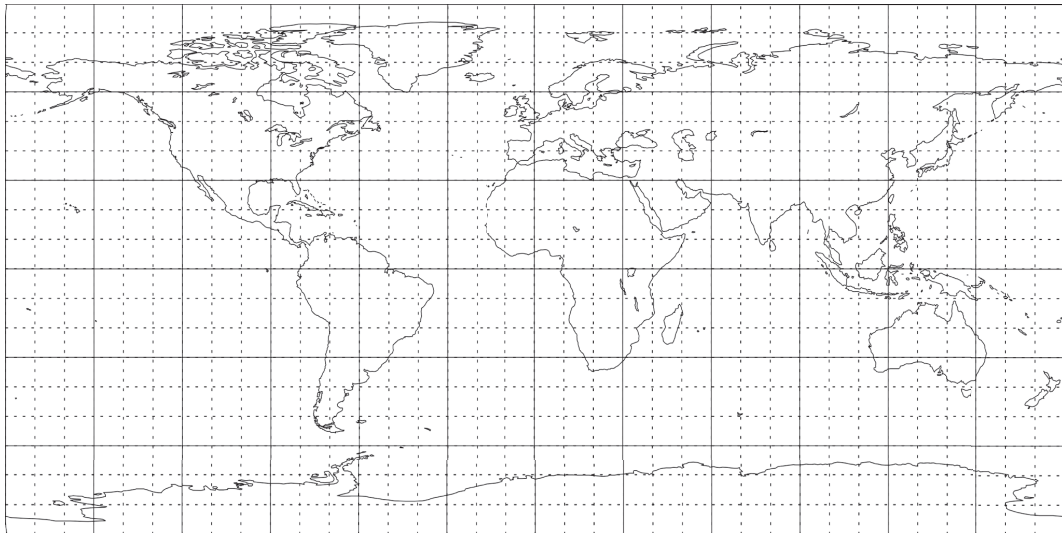


Figure 5: 球面画像ファイルの例. 正距円筒図法 (Equirectangular projection) で描画されている。

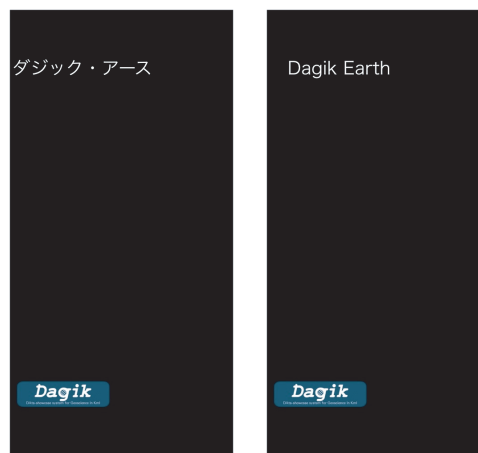


Figure 6: 説明画像ファイルの例. (左) 日本語用のファイル例, (右) 英語用のファイル例.

3.3 設定ファイル

球面画像ファイルと説明画像ファイルを用いたダジック・アース表示ソフト CKK による表示を Fig. 7 に示した。球面画像ファイルは球体上にマッピングされ、マウスとキーボードによって操作することができる。球面画像描画の投影法は、この例では垂直投影 (orthographic projection) が用いられているが、透視投影 (perspective projection) での表示も可能である。垂直投影は、無限遠に設置された PC プロジェクターから球形スクリーンに平行光で画像を投影した場合に正しい形状で投影されるものであるが、短焦点型ではない多くの PC プロジェクターにおいて十分正確な投影が可能であり、歪みも少ないので多くの場合に用いられている。この例では画面上の右側に寄せて表示されているが、大きさや位置は設定ファイルによって変更可能である。説明画像ファイルは、この場合は球体の左側に表示されている。白枠で囲まれた部分が Fig. 6 右図で示した説明画像の表示されている領域である。この説明画像ファイルも画面上の任意の場所・大きさに表示することができる。一方、画面右側には操作のためのアイコンが表示されている。このアイコンの機能は、左上より時計回りに、終了、縮小、拡大、北を上にする、自転の開始・終了、緯度経度線の表示・非表示、書き込みモードの開始、再生速度を遅く、再生速度を速く、逆向き再生・停止、再生・停止、初期画像へ戻る、逆向きコマ送り、コマ送り、アイコンの非表示・表示、説明画像の非表示・表示、である。これらのアイコンの表示・非表示、表示位置、アイコン画像ファイルも設定ファイルで指定可能である。

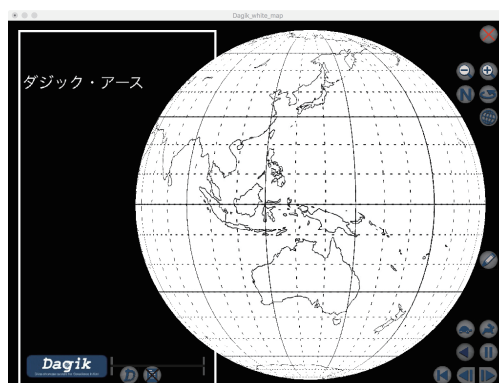


Figure 7: ダジック・アース表示ソフト CKK による表示例。

上記のように表示ソフトによる表示については、設定ファイルで設定ができる。Fig. 8 にその例を示した。ダジック・アースでは以下の 3 つの設定ファイルが使われている、1) 複数のコンテンツに共通する部分を書く `init conf.txt` 2) 各コンテンツに固有の設定を書く `conf.txt` 3) DOW 用の設定を書く `dow conf.txt`。設定はいずれのファイルでも共通の形式で行われ、同じ設定が可能である。表示ソフトにおいては 1), 2), 3) の順に読み込まれるため、後に読み込まれたファイルに書かれている設定が優先される。“#” で始まる行が有効な設定の部分である。これらの設定ファイルの保存場所は Fig. 1 の通りである。`init conf.txt` と `conf.txt` は、存在しなくてもデフォルト値が用いられるので問題ないが、`dow conf.txt` は DOW の実行には必ず存在しないといけない。

多言語表示のための設定ファイルは ‘data/conf/’ 以下に ‘lang-english.txt’ や ‘lang-chinese.txt’ として置かれ、最も最後に読み込まれる。この設定ファイルでは、それぞれの言語に対応した説明画像のファイル名の指定が行われている。

```

Initial Configuration file for Chikyu Kuru Kuru
Configuration line starts with single "#", e.g., #Latitude, #Spin.

```

```

#WindowSizeXY 1024 768
#WindowPositionXY 0 0
#FullScreen 0
#InvisibleCursor 0
#ClickLock 0
#NoMouseDragRotation 0
#PresentaionRemoteMode 0
#InertiaOfRotation 0.02 0.01 0.01 0.
#NumberOfEarth 1
#Scale 1.9
#EarthXY 0.3 0.0
#ScreenScaleXY 4.3 -2.6 1.9
#ScreenFront 0
#ScreenBGTransparency 255
#TimebarScaleXY .2 -1.6 -1.65
#Latitude 0.0
#Longitude 135.0
#Animation 1
#AnimationSpeed 8
#StopAt1stMap 1.0
#StopAtFinalMap 1.0
#Repeat 1
#Spin 0
#SpinSpeed 5 :[-200,200]

```

Figure 8: 設定ファイル

3.4 付属ファイル

各コンテンツには付属ファイルとして、情報 HTML ページ、解説ファイル、動画ファイル、手作り地球儀ファイル、が付属する。これらのファイルは必須ではなく、利用を支援するものである。情報 HTML ファイルは、コンテンツで表示されるデータの提供元、コンテンツ作成者、データの概要などが記述されており、コンテンツの利用者がより詳しい情報を得るのに利用されている。一部のコンテンツではコンテンツの説明画像にこの情報 HTML ページの URL(<http://dagik.org/L/1> など) が記載されている。解説ファイルは、コンテンツの内容を詳しく説明したものであり、授業や展示の解説の補助となるものである。動画ファイルは、コンテンツの表示状態を動画に収録したものであり、パソコンを用いずに動画ファイルのみによる投影に用いられる。

手作り地球儀ファイルは、それぞれのコンテンツについて通常の地球儀を作成するためのシートであり、ダジック・アースの表示ソフトと同じくダジック・アース・コンテンツを用いて作成される。Fig. 9 にその例を示した。これらは Fig. 5 に示した球面画像と Fig.6 に示した説明画像を用いて作成されている。上図は赤道が連続したもので 1 枚のシートになっている。下図は 6 枚の短冊状のシートからなっている。ともに球面画像部分を切り抜き、発泡スチロールやプラスチックなどの球体に貼り付けてそれぞれのコンテンツの地球儀を作成するものである。切り抜きと貼り付けが多少正確でなくてもズレの少ない地球儀を作成できるように、一つの地域を重ねて貼り付けるように図法が工夫されている。デジタル立体地球儀の実施環境がない家庭などでもデジタル立体地球儀で表示されるダジック・アース・コンテンツを再現することができるように、科学館などのワークショップでの作成に用いられている。作成された手作り地球儀を Fig. 10 にそれぞれの使用コンテンツ名とともに示した。この例では、直径 7.5cm のプラスチック球に貼り付けられている。

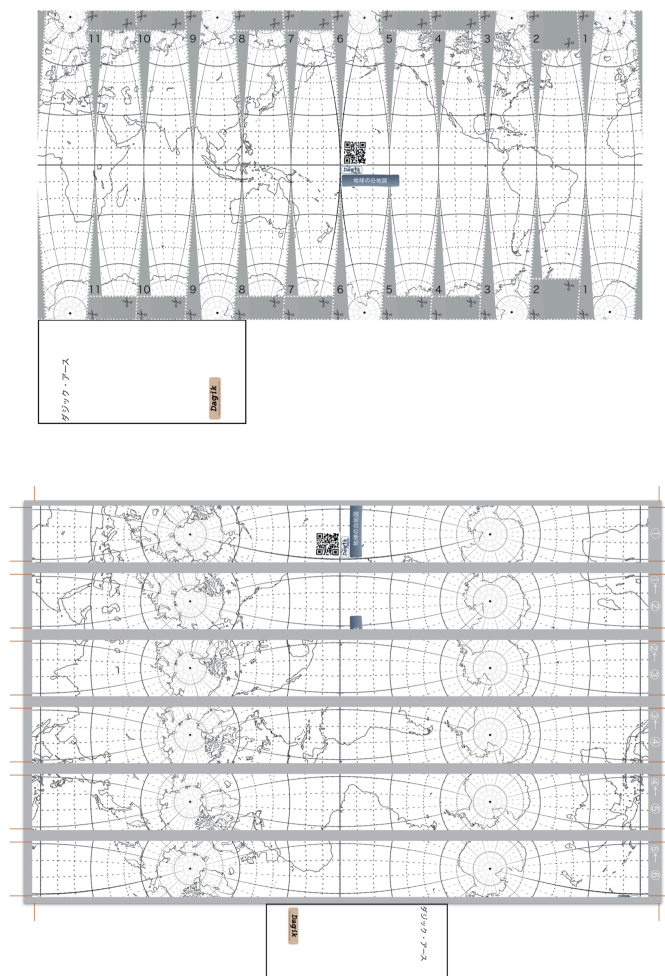


Figure 9: 手作り地球儀ファイルの例. (上) 赤道接続型シート, (下) 短冊型シート.

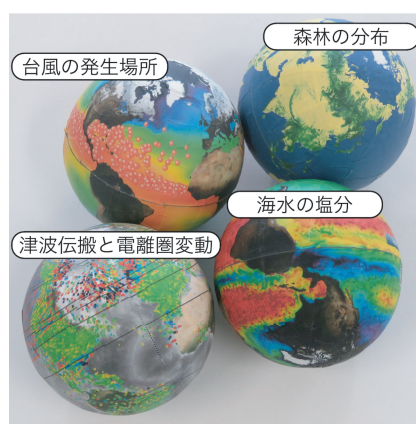


Figure 10: 作成された直径 7.5cm の手作り地球儀

4 表示ソフトウェア

ダジック・アースは多様なデバイスによって表示できることを目指しており、ダジック・アースの表示ソフトウェアとして 1) Windows 及び Mac 搭載のパソコンで実行ができるダジック・アース表示ソフト「Chikyu Kuru Kuru(CKK)」, 2) 多くのデバイス上のウェブブラウザで実行ができる「Dagik earth On the Web (DOW)」, 3) タブレット端末 iPad で実行ができる「ダジック・アース・アプリ」の 3 つが開発されている。上述のようにいずれも同一の「ダジック・アース・コンテンツ」ファイルを用いており、コンテンツの作成者はコンテンツを開発することで多くの環境での表示が可能となる。以下ではこれら 3 つのダジック・アース表示ソフトウェアについて述べる。

4.1 パソコンでの表示：CKK

パソコンでの表示ソフトウェアとして開発された CKK は C++ 言語で記述され、3D 表示のために Open GL¹¹⁾ を用いたライブラリ群である GLUT¹²⁾ を用いている。このため同一のソースから Windows と Mac の両方の実行形式が作成可能である。また、現在使用されている例はないが Linux での実行形式の作成も容易である。CKK による表示例は Fig.7 に示した。本稿の時点でのバージョンは 1.27 であり、その機能は Table 2 にまとめられている。「○」は対応している機能を示している。CKK は他のダジック・アース表示ソフトウェアに対して機能が多く、一つのコンテンツで二組の球面画像と説明画像が使用可能な「複数レイヤー」機能、マウスなどを使って自由に線での描画が可能な「書き込み」機能、CSV 形式のデータファイルを用いて任意の場所に線、印、矢印がプロット可能な「データプロット」機能などが利用でき、授業などの用途に用いられている。画像ファイルはサーバー上ではなくパソコンに保存されているため、それらのファイルを新たに作成することによりユーザーが独自のコンテンツを追加・使用することも可能である。「リアルタイム」機能はリアルタイムの球面画像、説明画像のダウンロードは別の補助プログラムを用いて行う必要があるが、表示可能である。引数で言語名を指定することで日本語以外に英語と中国語の説明画像の表示による多言語対応が可能であり、ダジック・アース・コンテンツの”zz_International”ディレクトリ以下にバッチファイルなどが置かれている。

CKK では個々のコンテンツが独立しており、それぞれのコンテンツのみで実行できるようになっている。そのため各コンテンツ毎に CKK の実行ファイルが含まれている。また、その反面、Fig. 3 で示したダジック・アース表示ソフトの構成のうち「メニュー」プログラムは CKK には含まれておらず、メニューから選択して利用する際には別のメニュープログラムを用いる必要がある。Windows では HTML Application (HTA) 形式で作成された「Dagik_Earth.hta」がメニュープログラムとして用意されており、Mac では他のソフトウェア・ランチャー・プログラムを利用することになる。

Table 2: ダジック・アース表示ソフトウェアの主な機能

	複数レイヤー	書き込み・ データプロット	ユーザーによる コンテンツ追加	リアルタイム	お気に入り登録	連続再生	多言語
パソコン・ソフト：CKK	○	○	○	△			○
ウェブ・ブラウザ：DOW			○	○			○
タブレット・アプリ				○	○	○	

4.2 ウェブブラウザでの表示：DOW

ウェブブラウザで実行ができる DOW は JavaScript 言語で記述され、3D 表示のために Web GL13) を用いたライブラリ群である three.js¹⁴⁾ を用いている。このため表示には WebGL が実行可能なウェブブラウザが必要である。DOW は表示用の index.html とそこから利用されている JavaScript ライブラリである dow.js からなる。dow.js などの JavaScript ライブラリはコンテンツ内の “data/js” ディレクトリに収納されている。DOW はオンラインとオフラインの 2 つの形態で利用可能である。サーバー上の <http://dagik.org/dow/> に置かれているダジック・アース・コンテンツのファイルを表示するオンライン型のもは Web 版ダジック・アースと呼ばれており、ローカルに保存したダジック・アース・コンテンツのファイルを表示するものは HTML 版ダジック・アースと呼ばれている。Web 版ダジック・アースは WebGL が実行可能なウェブブラウザが利用可能な多様なデバイスで利用可能であり、パソコンの他にスマートフォン、タブレットなどで利用できる。HTML 版はローカルに保存したファイルから JavaScript が実行可能な環境のみで利用可能である。多言語対応は多言語設定ファイルである ‘data/conf/’ 以下の ‘lang-english.txt’ や ‘lang-chinese.txt’ の読み込みを指定することで行われて、その設定をしたものを ‘index-english.html’ や ‘index-chinese.html’ としてダジック・アース・コンテンツに含んでいる。

ウェブブラウザによる DOW の表示例は Fig. 11 の左図に示した。ほぼ Fig. 7 と同じであるが、機能の違いにより操作アイコンの表示が異なっている。DOW の機能は、Table2 にまとめられているように限定されているが、サーバー上にファイルが置かれてる Web 版ではリアルタイムに更新されるコンテンツの表示が可能である。また、他の表示ソフトに比べて対応しているデバイスの種類が多いのも特徴である。

メニュー・プログラムとしては、index.html が用意されており、そこから各コンテンツの index.html が呼び出されて実行される。メニュー・プログラムも多言語用に index-english.html や index-chinese.html が用意されている。

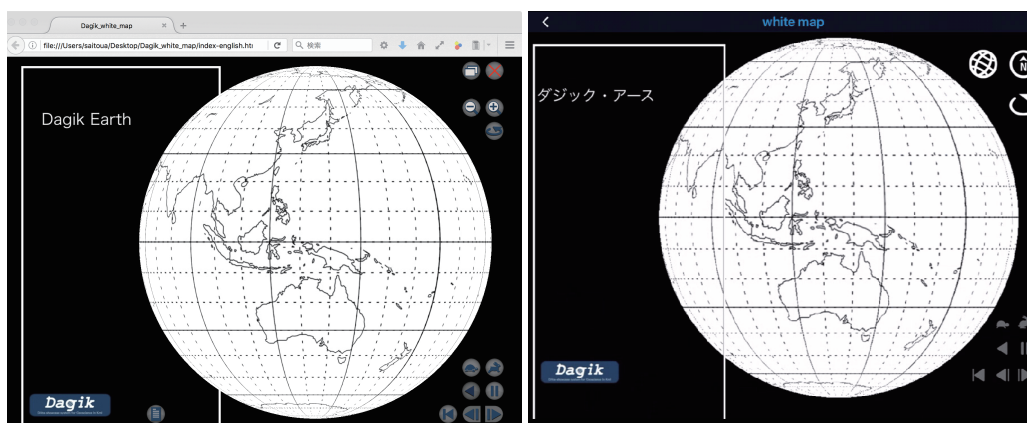


Figure 11: ダジック・アース表示ソフトによる表示例：(左) DOW による表示
(右) ダジック・アース・アプリによる表示

4.3 タブレットでの表示：ダジック・アース・アプリ

タブレット端末である iPad で実行ができるダジック・アース・アプリは Objective-C 言語で記述されている。2011 年発売の iPad2 以降の iPad で利用可能である。球面画像、説明画像、設定ファイルはサーバー上に置かれ利用に応じて各 iPad にダウンロードされ、保存されるが、

リアルタイム・コンテンツは iPad にはファイルは保存されず、表示毎にダウンロードされる。本稿作成時には日本語版のみが用意されており、多言語化はされていない。

他のダジック・アース表示ソフトと異なり、メニュー機能が付いており、アプリからコンテンツの選択と表示が可能であり、その機能を用いて、選択したコンテンツをグループにして登録する「お気に入り」機能や、選択したコンテンツを連続して表示する「連続再生」機能がある。その反面、コンテンツがサーバー上に登録されているため、利用者がコンテンツを作成して追加することはできない。

「ダジック・アース・アプリ」の表示例を Fig. 11 の右図に示した。ほぼ Fig. 7 と同じであるが、機能の違いにより操作アイコンの表示が異なっている。

5 まとめ

多様な環境においてデジタル立体地球儀を容易に実現するために「ダジック・アース」の開発を行った。システムを構成する表示デバイス、PC プロジェクター、コントローラーのいずれも広く使われている様々な機器が利用可能であり、球形スクリーンも直径 8cm のものから直径 16m のものまで多様なものが用いられている。表示デバイス上で実行される表示ソフトは多様なデバイスに対応するために 3 種類のもので開発されている。これらでは共通したコンテンツ・ファイルが使用されているためコンテンツの開発が容易であり、利用者が独自のコンテンツを作成・表示することも可能である。これによりコンテンツも多様なものが作成されている。従来は限られた科学館のみで使われていたデジタル立体地球儀が、ダジックアースによって学校での授業や科学イベントなど多様な環境で実施できるシステムが整備された。

謝辞

本研究の一部は、文部科学省宇宙利用促進調整委託費（平成 21-23 年度）及び宇宙科学技術推進調整委託費（平成 25-27 年度）による研究開発として実施された。

参考文献

- 1) NASA World Wind, <http://worldwind.arc.nasa.gov>, 参照日：2016 年 10 月 27 日
- 2) Google Earth, <https://www.google.com/earth>, 参照日：2016 年 10 月 27 日
- 3) Butler, D., Virtual globes: the web-wide world. *Nature* 439: 776-778, 2006.
- 4) Lisle, R.J., Google Earth: a new geological resource. *Geology Today* 22: 29-32, 2006.
- 5) Saito, A, and D. Yoshida, Dagik: A Data-Showcase System for the Geospace, *Data Science Journal*, 8, S92-S95, doi:10.2481/dsj.8.S92, 2009.
- 6) Dagik: DATA-showcase system for Geoscience In Kml, <http://dagik.org>, 参照日：2016 年 10 月 27 日
- 7) ダジック・アース, <http://earth.dagik.org>, 参照日：2016 年 10 月 27 日
- 8) Ebisawa, K., Y. Koyama, A. Saito, S. Sakamoto, M. Ishii, Y. Kumano, Y. Hazumi, Dagik Earth and IUGONET, *Astronomical Society of the Pacific Conference Series*, Vol. 495,3-5, 2015.
- 9) 地球を感じる Geo-Cosmos ジオ・コスモス[常設展示つながり]－日本科学未来館 (Miraikan), <http://www.miraikan.jst.go.jp/exhibition/tsunagari/geo-cosmos.html>, 参照日：2016 年 10 月 27 日

- 10) What is Science On a Sphere, <http://sos.noaa.gov> , 参照日 : 2016 年 10 月 27 日
- 11) OpenGL - The Industry Standard for High Performance Graphics, <https://www.opengl.org>, 参照日 : 2016 年 10 月 27 日
- 12) GLUT - The OpenGL Utility Toolkit, <https://www.opengl.org/resources/libraries/glut>, 参照日 : 2016 年 10 月 27 日
- 13) WebGL - OpenGL ES 2.0 for the Web, <https://www.khronos.org/webgl> , 参照日 : 2016 年 10 月 27 日
- 14) three.js - Javascript 3D library, <http://threejs.org> , 参照日 : 2016 年 10 月 27 日

科学データのドームを用いた映像化による研究成果の公表

田部一志^{*1}

Publication of scientific results by using of data visualization for dome

Isshi Tabe^{*2}

概 要

日本には公開されているドームシアターが 200 か所以上あるが、科学データの映像化、結果の公表という観点からは、十分に機能している状況にはない。ここでは、主に国内の状況を考察しつつ、ドームを用いた科学データの映像化と研究成果の公表についての可能性を述べる。併せて、国際科学映像祭ドームフェスタの有用性を考察する。

キーワード：科学データの可視化、ドームスクリーン

Abstract

There are more than 200 dome theaters that open for public dome in Japan. But it isn't in the situation functioning sufficiently from the point of view of visualization of scientific data and publication of results. In this paper, we discuss situation of dome theaters in Japan, also we guess a possibility of publication of scientific results by using dome screen and utilities of the Dome Fest of International Festival for Scientific Visualization.

Key word: dome screen, scientific visualization

1 はじめに

日本の映像用ドームスクリーンの多くはプラネタリウム用のものである。そのうち公開されている施設の 70%^{*3}には、何らかのデジタル全天映像投影装置が備えられている。それらは多種多様で、高価なメーカー製のシステムから安価なホームメイドのシステムまで様々なものが使われている。映像のみならずプラネタリウムの星像までがデジタル化され

doi: 10.20637/JAXA-RR-16-007/0013

^{*} 平成 28 年 11 月 24 日受付 (Received November 24, 2016)

^{*1} 株式会社リブラ代表取締役 tabe@libra-co.com

^{*2} Libra Corporation, director of board

^{*3} 日本プラネタリウム協議会データブック編集委員 (2010)「プラネタリウムデータブック 2010」掲載資料 2「日本のプラネタリウムリスト」に掲載された施設のうち恒常的に一般公開されている施設 231 館中 160 館にデジタル全天映像装置の設置が確認されている (2016 年 10 月調べ) ので 69%となる。

つつある。それによって、何が起きているかについては、人的な側面からは田部（2015）、ハード的な側面からは田部（2014）に詳しく述べた。

「プラネタリウム」という言葉については、多くの人々に漠然としたイメージがあるだけで、確固たる定義は存在しない。「プラネタリウム」という言葉は 1923 年に惑星の運行と背景の星野を、天球に見立てたドームスクリーン上に再現する、ドイツで発明された光学マシンを指していた。このマシンを中心として、ドームと椅子や音響装置などの付帯設備を含めて「プラネタリウム」、さらにそれを維持・運用する人的組織をも含めて「プラネタリウム」（図-1）と呼ぶのが一般的である。建物のことはプラネタリウム館と呼ばれる。

基本的なプラネタリウムの使い方は、来館者に今晚見られる星空を解説することと、小中学校からの来館者（多くは移動教室としてバスでやってくる）に対して行う授業のまとめ的な学習投影を行うことである。博物館や科学館に設置されたプラネタリウムでは、1 コマ（例えば 1 時間のプログラム）の中で、今晚の星空の解説（30 分）と「全天映像番組」（30 分）を見せることが一般化している。「全天映像番組」の内容はさまざまであるが、プラネタリウム館で星空解説とセットで上映される場合、宇宙や天文と何らかの関連のあるものが選ばれる傾向にある。以下、本論文ではこのようなパターンを「プラネタリウム」と呼ぶこととする。

プラネタリウム館のプログラムではあっても、1 コマまるまる「全天映像番組」だけで、内容も宇宙や天文とは関連なく、星空の解説もないといったパターンがある。このような投影は「プラネタリウム」ではなく「映画」ないしは、「ドーム映像」と呼ぶべきである。ここでは「ドーム映像」*4とよぶこととする。

本稿では 2 節で日本のドームスクリーンの現状、3 節で「ドーム映像」の現在から将来に向けての有用性、4 ではドーム映像コミュニティについて述べる。

2 日本における映像ドームの状況

2-1 プラネタリウムとドーム映像

プラネタリウムは科学の教育や普及を目的として主に地方自治体が設置したものであるが、教育用というより、博物館や科学館にある娯楽用の設備と思われるふしもある。事実、娯楽的色彩の強いドーム映像を好んで上映する施設もある。しかし、単なる娯楽の道具に留まらず、天文や宇宙科学に関する知識や興味関心喚起の場所として、地道に科学ファン*5を作り出している場所でもある。

プラネタリウムは 1920 年代に発明されてから 1970 年代までは機能的にはあまり変わらなかったが、1980 年代からドーム映像（初期は大型のフィルムを用いていた）を取り入れる

*4 ここでの定義ではプラネタリウム番組（1 コマ）の中で上映される「プラネタリウムのなドーム映像」を何と呼ぶべきか定義できない矛盾を含んでいる。しかし、「プラネタリウムのなもの」と、「プラネタリウムのでないもの」の 2 つに分けることは、実情を議論するうえで都合なのでこのようにした。

*5 「渋谷の空の向こうに」（五島プラネタリウム卒業文集制作委員会編、自費出版 2005 年、<http://miwa.way-nifty.com/bunsyu/>）には、多くの天文学研究者や天文愛好家がかつてプラネタリウムから強い影響や感銘を受けたことが記されている。

ようになり演出の幅が大きく広がった。

2000 年ころからプラネタリウムもドーム映像もデジタル化が本格化したが、ドーム映像がほぼ 100%デジタル化したのに対し、星像に関しては、まだ恒星原板をレンズで投影するアナログの恒星投影機の方がクオリティが高いので「アナログの恒星投影機」と「デジタルドーム映像装置」を併用している例が現時点では多い。一方、「アナログの恒星投影機」を排し、「デジタルプラネタリウム」のみで運用する施設もわずかずつであるが増えて^{*6}いる。

「デジタルプラネタリウム」と「デジタルドーム映像装置」はハードウェアは全く同じものである。「ムービープレーヤー」に相当するアプリケーションも同じである。星や宇宙や関連する現象を再現する「スペースエンジン^{*7}」の機能を持ったアプリケーションの有無で「デジタルプラネタリウム」であるか否かがきまる。もう 1 つは設置者（自治体）が施設の目的をどのように条例に規定するかによる。



図－1 典型的なプラネタリウムの例
ドームの中心には主に星像を投影するアナログ（光学）投影機（プラネタリウム本体）、周辺にはドーム映像を投影する機材が設置されている。この写真は、旭川市科学館サイパル（2005 年開館、ドーム直径 18m）の例。

2－2 プラネタリウムとドーム映像の分化

1980 年代半ば以降は傾斜型プラネタリウム^{*8}と 70mm フィルムの大型映像装置^{*9}を併用する施設が激増した時代があった。投影（上映）内容も従来のプラネタリウム投影（生解説）を模して自動化しただけのオート番組、海外の大型映像をドーム化^{*10}したものなど様々であったが、それらは多かれ少なかれ「プラネタリウム」的なものであった。しかし、主にデジタル化の進展によって、「内容にプラネタリウムの要素のないドーム映像」を主体にしたプ

^{*6} 小規模館を中心に、2016 年 10 月現在 20 館ほどが「デジタルプラネタリウム」で運用されている。

^{*7} デジタルプラネタリウムの機能を持ったアプリケーションのことを「スペースエンジン」と呼んでいる。

^{*8} ドームスクリーンを 10 - 30° ほど傾け、それにあわせて床にも段を設けた形態のプラネタリウムを傾斜型（または傾斜式）プラネタリウムと呼んでいる。

^{*9} 例えば IMAX 社の OMNIMAX（のちに IMAX-DOME）、World Odyssey 社の Ultra-70、五藤光学のアストロビジョン 70,35、Iwerks 社の Iwerksphere870、ウシオユーテック社の OMNI-U などがある。

^{*10} 70mm15p で作られた大型映像をドームに流用（ドーム化）した作品が 1980 代から現在にかけて相当数存在する。

プログラムやイベントの実施が可能になった。それらの一部は、制作に要する労力や経済的負担が、かつてとは比べ物にならないほど軽減されている。このことによって、プラネタリウム担当者の意識の中に、「プラネタリウム」と「プラネタリウムでないもの」（本論文で「ドーム映像」と呼んでいるもの）との切り分けが明確化してきているように思われる。ドームスクリーンに映されるものは全て「プラネタリウム」とその派生物という概念から脱却する傾向に転じた。これはデジタル化の一つの功績である。

2-3 歴史的な背景と規格化

ドームスクリーンに全天映像を簡単に投影したいという要請はプラネタリウムの世界から起こった。恒星と惑星しか映せない「アナログの恒星投影機」の補助投影システムとしてどうしても必要であった。それらは、主に経済的な理由により、投影する機材も、映す静止画も動画も既存（つまり市場に民生用として流通している）のデジタル機器と技術の流用で作られていた。これが、多くの人々の試行錯誤を経て、洗練された全天メディアプレーヤー（動画）と全天プレゼンテーションアプリケーション（スライド、ビデオ映像、ドーム映像をドームスクリーン上の任意の場所に配置できる）となり、投影機器（ビデオプロジェクター）も価格はあまり変わらないまま年々高性能化している。

一方、少し遅れてフィルムを使う大型映像の世界にデジタル化の要請が起こった。実現する過程で、フィルムの現像事業者の“生残り策”と、画像と音響も含めた総合的な質の確保、さらにはコピーガードやエンコードの仕組み作りの必要性とがあいまってデジタルシネマイニシアティブDCI（Digital Cinema Initiative）^{*11}という団体が推奨するハード・ソフトに対する総合規格（Digital Cinema Initiative Distribution Package：通称DCP規格^{*12}と呼んでいる）が作られた。ドーム映像にもDCP規格を適用しようとする動きもあったが、既にさまざまな投影方式がばらばらに用いられ、独自発展していたプラネタリウム内の投影システムとその上映作品を律することはできなかった。つまり、プラネタリウム及びドーム映像はDCIの恩恵も受けない代わりに、参入障壁もない自由を得たと言える。ドーム映像をやりとりするのに必要なのはドームマスター^{*13}という最低限の約束事だけである。

PCの性能の劇的な進化、投影装置（プロジェクターの解像度や明るさ、コントラスト比など）の進歩と低価格化にともなって、プラネタリウムでのドーム映像装置導入と映像作りに対する敷居はますます低くなっている。このことは、これまでとは違う新たな参入者を促す効果となってあらわれている。芸術家、写真家、映像作家、サブカルチャーの世界の人々....がドームを表現手段の新たな可能性として注目している。

2-4 もう1つの現実

ドーム映像を上映することのできるシアターは、そのほとんどがプラネタリウムである

^{*11} <http://dcimovies.com/>

^{*12} <http://www.digitalpreservation.gov/formats/fdd/fdd000200.shtml>

^{*13} http://fulldome.ryanwyatt.net/fulldome_domemasterSpec_v05.pdf

ことは既に述べた。上映のための予算は、プラネタリウム用の番組予算の流用である場合が多い。作品を販売しなければならない配給会社もコンテンツ制作者も、そのことは大いに意識しており、その結果、「プラネタリウム番組」と称する「ドーム映像」が出現することとなった。内容もプラネタリウム受けを狙ったものが多い。逆にそうしなければ、制作者や配給業者のビジネスは成立しないというのが現在の実情である。

最新の「プラネタリウムデータブック 2015」^{*14}は、2014 年 1 年間のプラネタリウムの観覧者数を 818 万人と見積もっている。この数字には、「プラネタリウム」もプラネタリウム館で上映している「ドーム映像」もドーム内での星空コンサートなどのイベント参加者数も含まれている。国内には年間入場者数 10 万人以上の館はほとんどなく、多くの館は数千から 2 万人程度である。季節による繁閑はあるが全体として座席占有率はそれほど高くない。地域のプラネタリウムの場合、学習投影で使われる他一般投影も行われるが、観客のニーズは「今晚見える星空に関する情報」である。さらに、観客の期待以上に科学興味（ここでいう興味とは今晚または晴れた晩に夜空を見てみようと思わせることである）を喚起するような内容が提供できれば、満足度も上がる。

ともすると一方通行の情報の提示に偏りがちになるプラネタリウムだが、能力のある解説者（科学インタープリター）が観客の客層や反応に応じて適切なコミュニケーションを保ちながら説明をすることができれば効果は倍増する。断片的であっても映像の力は非常に有効である。

プラネタリウム館は、内容が「プラネタリウム」であるか否かにかかわらず、科学の話を多くの人々に効率よく行うことのできる代表的な施設である。

3 科学映画としての映像

ある研究者が、一般の人に見てもらいたい研究結果や現象を画像にできるようなデータを持っているとしよう。その成果がプラネタリウムの番組、または生解説番組の一部として多くの人々の目に触れれば、研究者冥利に尽きるというものだろう。だが、数字の羅列であるデータから、一般の人に楽しんでもらいながら、なるほどと思ってもらえる作品にして、さらに上映されるまでには、さまざまなプロセスを経なければならない。それは、一般的には図-3 のようである。

^{*14} 日本プラネタリウム協議会 (JPA) 「データブック編集担当」2016 年刊
<http://planetarium.jp/public/databook/>

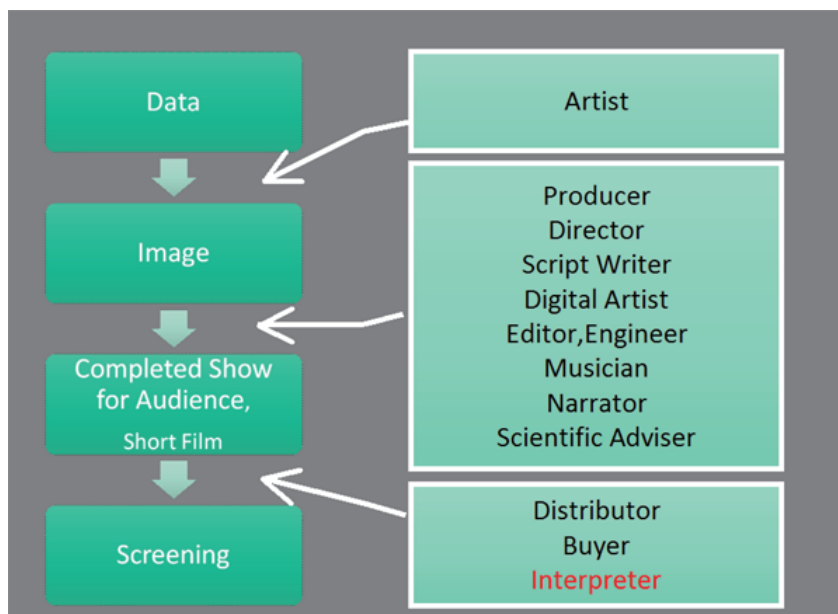


図-3 データから上映までのプロセス

3-1 アーティスト

NASA や ESO では組織内にデータを可視化する専門のアーティストがいて“きれいな”画像や CG を作っている。これに適切な科学的な解説（かなり詳細）が付加されてプレスリリースとして公表されている。見栄えのする動画なり静止画なりを作るのは科学とは別のデザインのスキルと芸術的センスが要求される。センスは才能まかせというわけではない。その分野で相応の訓練や経験を積んだプロのアーティストが必要である。

3-2 プロデュースと制作

多くの館へ配給できる作品を作るという観点で考えれば、最初の企画（プロデュース）は非常に重要である。観客のターゲットも然ることながら、購入して上映してくれる館の目的＝自治体が施設を設置する目的をふまえ、自治体が購入可能な価格設定にしなければならない。作風を統一感をもって規定できるか、監督や CG クリエーター、音楽担当者の選定も作品の出来を大きく左右する。大枠から子細と思われることまで考えつつ、販売経路や値段設定、供給会社間の利害調整も行わなければならない。

シナリオライター、音楽家、CG アーティスト、音響技術者、ポストプロエンジニアなどの人々の協働によって作品は完成する。各分野の担当者がドームとドーム映像に関する深い理解をもっている必要がある。そのような人材は限られている。

3-3 公表方法

このようなプロセスに乗って作品が完成し、広く上映され、一般の人々の目に触れるよ

うになる例は実際にはあまり多くない。かなり魅力的なテーマか、社会性のあるテーマでない限り、商業ベースに乗せることは難しい。もう1つの方法として、研究機関の予算で作品もしくは解説用ショートクリップを作って上映館へ配布することが考えられ、一部では行われている。国立天文台ではくつかのCGショートクリップをプラネタリウム館向に公開^{*15}している。ESAでもドーム用の番組やCGクリップを無償配布^{*16}している。ここで、どうしても必要なのは優秀な科学インタープリターである。科学の初歩（人々が直観できる事柄）から、最近のトピックスに結び付けて、相手に応じてその面白さを伝える役割を担う。

プラネタリウムが1980年代にオート化されて以来、長い間インタープリターや解説者はないがしろにされてきた。受難の時代であったとも言える。2010年代に入ると、ムービーのショートクリップに限らず、デジタルプラネタリウムが宇宙の現在の姿を任意の視点から任意の波長でリアルタイムで見せてくれるようになると、それを適確に解説できる能力を持った人材が求められるようになってきた。今後、解説者の地位が向上する方向に社会全体が進んでいくであろう。

4 コミュニティの重要性

プラネタリウムに映像の波が押し寄せた例は過去に2回ある。最初は1980年代フィルムによる全天周映画が流行った時代である。この頃はいわゆる"バブルの絶頂期"で、潤沢な資金力をバックに高価な70mmフィルムの映画が作られた。2回目が現在である。1980年代のように資金が潤沢にあるわけではないが、デジタル技術（主に投影装置とコンピューター）の発展に支えられ、映像制作のスキルさえあれば、誰でも手軽な設備でドーム映像制作に参入できるようになった。ことは前節でも述べた。

データ所有者とクリエイターとを結びつける試みは国立天文台と三鷹市が連携したプロジェクトとして2008年から2012年まで講習会^{*17}が開催され、多くのプロデューサーやクリエイターが巣立っていった。現在は終了したが、芸術分野も含めてドーム映像に興味を持つ人材を増やす役割を演じたように思われる。ある種の芸術家にとってまた自然系のフォトグラファにとってドームは格好の表現空間であるようだ。

国際科学映像祭実行委員会が主催し、科学技術映像全般を対象とする「国際科学映像祭」^{*18}は2009年にプレイベントを開催して以来6年間、毎年夏に開催されてきた。その中で、ドーム映像の上映や関連するワークショップを中心とした「国際科学映像祭ドームフェスタ」は最も集客力の高いイベントである。

4-1 ドームフェスタの有用性

「ドームフェスタ」と称するドーム映像作品の上映会はアメリカでもヨーロッパでも開

^{*15} <http://4d2u.nao.ac.jp/t/var/download/movie.html>

^{*16} <http://www.eso.org/public/videos/archive/category/fulldome/>

^{*17} 科学文化形成ユニット・科学映像クリエイター養成コース <http://prc.nao.ac.jp/ashub/about/create.html>

^{*18} <http://ifsv.org/>

催されている。Giant Screen Theater Association(GSTA) ^{*19}は毎年IMAX Filmの上映会とトレードショーを中心とする"GSTA International Conference and Trade Show"を開催している。また、Immersive Media Entertainments, Research, Science & Arts (Imersa^{*20}) Summitではもう少しプラネタリウムの要素(教育や科学の普及)そして技術開発に軸足を置いている。また、イエナ(ドイツ)で開催される Fulldome Festival^{*21}, エスピーニョ(ポルトガル)で開催される FestivalCinemaImmersivo^{*22}等では毎年上映会が行われる他、2年に一度の国際プラネタリウム協会(IPS^{*23})の年会に合わせて持ち回りで開催される上映会がある。

日本では、「国際科学映像祭ドームフェスタ」以外にも、日本ジャイアントスクリーン協会(JGSS) ^{*24}の「フィルムフェスティバル」や日本プラネタリウム協議会(JPA)の地域ブロック(支部)や、デジタルプラネタリウム研究ワークショップ(DPWS) ^{*25}で上映会が行われている。

一方、番組配給会社からすると、このようなフェスティバルへの出品より、各館へ試写用のファイルを持ち込み、上映作品の決定権者や直接の担当者に試写の機会を設けるのが営業的に早道である。

イエナで行われるドームフェスタは地元の大学などの協力もあって、若い芸術家や研究者の出品が多く見られる。日本でもショートフィルムのコンテストが行われるドームフェスタがあり、新規参入者への道が開かれている。あまり数が多いとは言えないが、毎年確実に新たな参入者があり、既成の枠にとらわれない新鮮な感覚の作品が上映され、参加者を楽しませている。

制作者、配給業者、上映者(プラネタリウム館を含めたドームシアターの担当者)が集うトレードショーの要素を持ったドームフェスタは、制作者が配給者や上映担当者の嗜好や動向について、直接的な情報を得ることのできる貴重な場である。しかし、商品として完成された作品主体の上映で、新たな試みや冒険性に富んだ、いわゆる奇抜な作品は上映されにくい傾向にある。作品の内容に関連したノベルティ等のサービス合戦も活発に行われ、施設(博物館や科学館)が本来持つ役割より集客力を重視した作品が受け入れられたりすることもある。

4-2 国際科学映像祭

研究者から最終アウトプットにかかわる人たちまでの距離感は遠く、コミュニティも様々で、放っておけば一同に会することは起こり得ない。新たなコミュニティの創出は簡単ではないが、異分野との接触間口の極めて広いプラネタリウム"業界"が中核となれば新

^{*19} <http://www.giantscreencinema.com>

^{*20} <http://www.imersa.org/>

^{*21} <http://fulldome-festival.de/>

^{*22} <http://iff.multimeios.pt/en/>

^{*23} <http://www.ips-planetarium.org/>

^{*24} <http://www.jgss.jp/>

^{*25} <http://digitalplanetarium.web.fc2.com/>

たな文化が生まれるのではないかと考えられる。

このような観点から、今後望まれることを列挙すると、

- ① 研究者からのデータや原理などのアピール
- ② プロデューサー目線での取捨選択
- ③ クリエーターや趣味人からのアイデアの提供やリクエスト
- ④ 科学映像を作る意欲を持った新規参入者の登場

ということになる。いろいろな立場の人にアイデアや作品発表の場所を提供するという観点から「国際科学映像祭ドームフェスタ」は、研究者（研究成果をドームで公表することに興味を持つ人）、それを具現するクリエイター、商売としての枠組みを整えるプロデューサー達が参加して交流が生まれることによって新しい何かが生まれること、ひいては業界が活性化することを目指している。単なる作品の品評会を超えて、広範な分野の人々が集い、情報交換ができるコミュニティの形成は社会的な使命であると考えられる。

5 まとめ

以上述べたことを摘要すると次のようになる。

- 1 プラネタリウムがデジタル化するにつれドームスクリーンの利用が多様化しており、これは世界的な傾向である。ドーム映像が扱う内容は自由である。
- 2 ドームは科学研究の成果を公表する場としては未開拓であるが、映像素材とスキルを持った解説者がいれば、教育や普及に大きな可能性を以ている。
- 3 研究者 - プロデューサー - クリエーター（既存、新進）のコミュニティ間の交流は密とは言えない。解決策として「国際科学映像祭ドームフェスタ」などの利用がある。

参考文献

- 1) 田部一志(2015)「デジタルプラネタリウムの発展」天界(東亜天文学会) vol.96 pp44-48
- 2) 田部一志(2014)「デジタルプラネタリウムの最新の技術動向」光技術コンタクト(日本オプトメカトロニクス協会) vol.52 pp.2-7

全天周映像作成のための、簡易ドームマスター生成手法について

三浦 昭^{*1}

A Simple Method to Create Dome Master for Full Dome Visualization

MIURA Akira^{*1}

概 要

全天周映像向けのフォーマットとして広く普及しているドームマスター形式の映像を様々なレイトレーシングソフトウェアでレンダリングするための、カメラオブジェクトと鏡面オブジェクトの簡易な組み合わせについて述べる。この手法は、ドームマスター形式に対応していないレイトレーシングソフトウェアにおいても適用可能なものである。

キーワード: ドームマスター, レイトレーシング, 鏡面, 回転楕円体

Abstract

The dome master format is widely used to film scenes for full dome theaters. This paper describes a simple combination of a camera object and a mirror object, which enables various ray tracing applications to render scenes suitable for the dome master format. This method is applicable for ray tracing applications that are not equipped with functions to render scenes for the dome master format.

Keywords: Dome Master, Ray Tracing, Mirror, Spheroid

1 まえがき

近年、全天周スクリーン用デジタル投影机 [1][2][3] の普及により、各地のプラネタリウム施設等で全天周映像番組が上映できるようになってきた。これらの全天周映像用のフォーマットとしては、ドームマスター形式 [4][5] が広く普及している。全天周映像としては、魚眼レンズなどを用いた実写映像の他に、レイトレーシング [6] ソフトウェア等で作られた CG 番組も多く提供されている。係る CG ソフトウェアも、ソフトウェア本体でドームマスター形式に対応したものや、ドームマスター形式に対応させるためのプラグイン等を用いるもの等が提供されている [7]。

一方で、ドームマスター制作の初心者、もしくは今後ドームマスターを習得するであろう予備軍を対象として考えた場合、彼等が使用していると思われる CG ソフトウェアは多様であり、必ずしも全ての CG ソフトウェアがドームマスター形式に対応しているとは限らない。そのような環境下で初心者向けのドームマスター制作講習

等を行おうとした場合、全ての受講者がドームマスター制作に対応した CG ソフトウェアを持ち寄れるかどうかは、保証の限りではない。さりとて、講習者側が画一的な CG ソフトウェアを用意して講習を行うことは、受講者としては、不慣れな CG ソフトウェアを用いた講習となる可能性が高く、そのようなケースではソフトウェアの操作自体の習熟に多くの時間を費やさねばならないというデメリットがある。

簡易のドームマスター制作手法としては、カメラの視線を一旦球体の鏡面に反射させることでドームマスターに近い映像をレンダリングする手法が提案されている [8]。また Bourke はポリゴンの法線方向を工夫することにより、ポリゴンの組み合わせでありながら滑らかなドームマスター映像を制作できる鏡面のモデルを提案している [9]。前者については、カメラと球体の組み合わせにより撮像系が構成できるため、講習会等の導入として取り扱える程度に単純である。しかしながら前者は真のドームマスターとの誤差は少なくない。また後者については、レンダリング精度の向上は期待できるが、構成

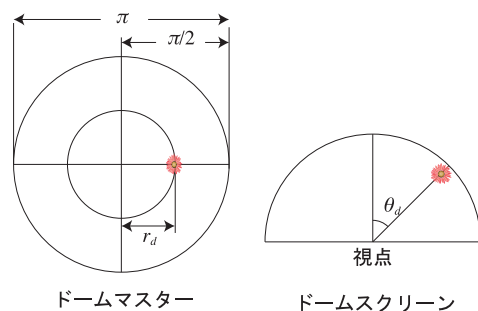


図 1: ドームマスターの模式図

左: ドームマスター, 右: ドームスクリーン上の投影

が複雑であり, 様々な CG ソフトウェアに対応させるためには, 相応の技術力が必要であると考えられる.

そこで本稿では, ドームマスター制作初心者向けの講習等においても, 受講者手持ちの CG ソフトウェア上で簡易にドームマスター制作を体験できるようにすることを目指して, カメラからの視線を一旦回転楕円体の鏡面に反射させることでドームマスターを模擬した画像をレンダリングする手法について述べる.

この手法は, 球体の鏡面に反射させる手法と同様に, 撮像系を構成するオブジェクトは単純かつ少数である. 同時に, オブジェクトの形状や配置を工夫することによりレンダリングの精度向上も期待されるものである.

2 検討する手法の概要

2.1 全天周映像におけるドームマスターの概略

全天周映像制作におけるドームマスターとは, 半球状のドームスクリーン面に描く映像を正距方位図法で円形に投影したものである [4]. ドームマスターの中心がドームスクリーン面の中心位置に相当し, ここを極点に例えると, 緯度に換算して赤道から極までの 90 度 ($\pi/2$) の範囲が描かれる. 模式図を図 1 に示す. ドームマスターの中心からの距離は, ドームスクリーン面上の対応する点と極点との緯度差に比例する. ここで, ドームスクリーンを構成する球の中心を視点とした場合に, 視点から見てレンダリングされるオブジェクトと極点とのなす角を θ_d とし, これに対応するドームマスター上の点の, 中心からの距離を r_d とする. ドームマスターの半径を $\pi/2$ で表すとすると, 式 1 の関係が成り立つ.

$$r_d = \theta_d \quad (1)$$

ドームマスターの例として, 球状の経緯線 (図 2) と, その中心に視点を置いてドームマスターで描いた結果

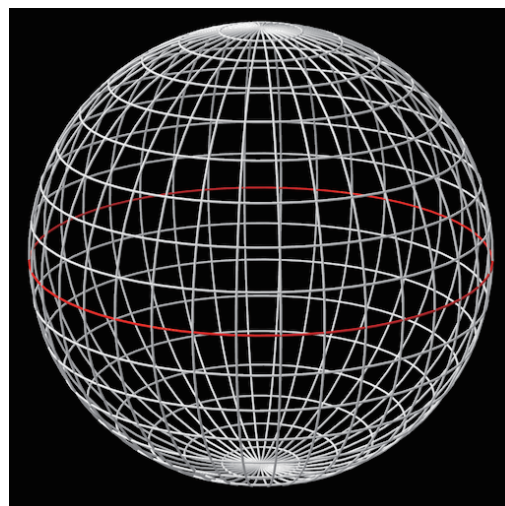


図 2: 経緯線の例 (経度, 緯度共に 10 度間隔)

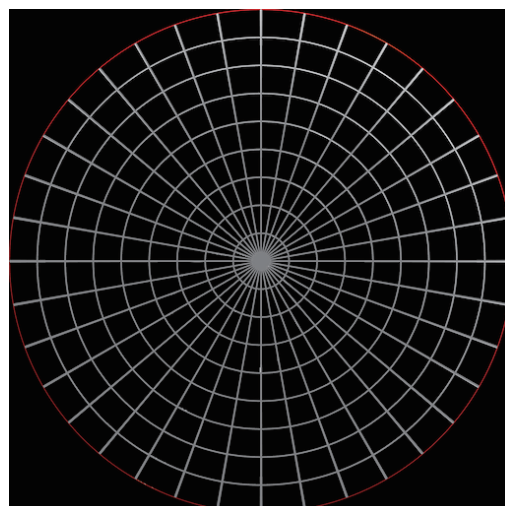


図 3: ドームマスターの例

(図 3) を示す. 両図中の赤線は, 赤道の位置を表している.

2.2 鏡面反射を用いたドームマスターの模擬

本稿においては, 全天周映像制作にあたって, ドームマスターのフォーマットを模擬するために, 回転楕円体の鏡面とカメラオブジェクトを組み合わせた撮像系を構成する. 概略を図 4 に示す. ここで, 描くべきオブジェクトは, 係る撮像系から十分遠くにあると仮定する. 図中の視線と仮想スクリーンが交わる点には, カメラの視線が鏡面に反射した先にあるオブジェクトが描かれると仮定する.

以後, 本稿においては, 鏡面反射を用いて生成するドームマスターについては, 「模擬ドームマスター」と称

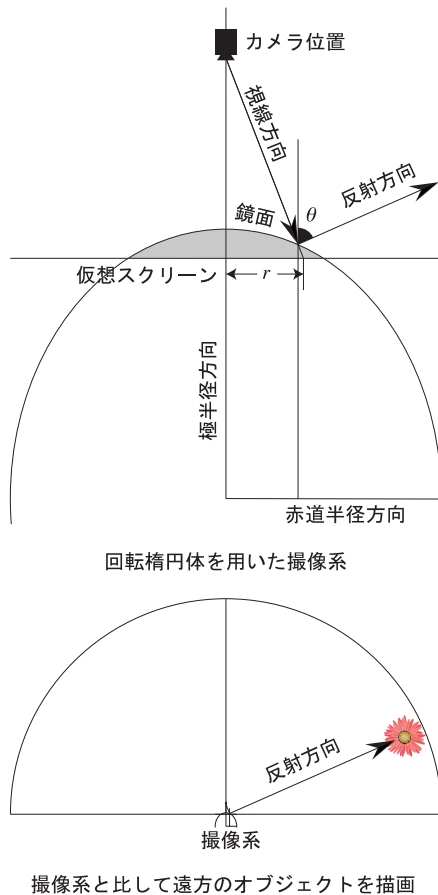


図 4: 鏡面反射を用いたドームマスターの模擬

することにより、歪の無い真のドームマスターと区別する。模擬ドームマスターの精度を上げるためには、図 4 中の θ と r が式 1 のような比例関係に近似できるような条件を求める必要がある。その計算の詳細は、3 節以降に示す。

2.3 本手法の適用範囲

本稿で検討する手法は、真球や回転楕円体を扱えるレイトレーシングソフトウェア一般に適用可能である。自由曲面等で球等を近似するソフトウェアへの応用も可能であるが、そのような場合、本手法におけるレンダリング結果の精度は、次節以降で求める誤差に加えて、係る CG ソフトウェアの曲面近似精度にも依存する。

2.4 本手法の制約

本稿で検討する手法は、鏡面反射を用いる結果として、幾つかの制約が生じる。

透明度 (α 値) 半透明オブジェクトや、オブジェクトが存在しない範囲等を識別するために、レンダリングされた画像に透明度 (α 値) を保存する場合があるが、本稿で述べた手法では常に鏡面オブジェクトの反射が描かれるため、適切な透明度を保存することができない。

鏡面オブジェクトの影 撮像系を構成する鏡面オブジェクトの影が、他のオブジェクトに現れることがある [9]。撮像系の鏡面オブジェクトは、影を落とさない設定にする必要がある。そのような設定の無い CG ソフトウェアにおいては、光源やオブジェクトと撮像系との位置関係に配慮する必要がある。

鏡面オブジェクトの写り込み 撮像系を構成する鏡面オブジェクトが他の鏡面反射に映り込む場合がある。撮像系の鏡面オブジェクトは、他のオブジェクトにおける反射に写り込まないように設定にする必要がある。そのような設定の無い CG ソフトウェアにおいては、他のオブジェクトと撮像系との位置関係に配慮する必要がある。

像の反転 模擬ドームマスターにおいては、鏡面における反射が 1 回あるため、得られる像は反転したものになる。そのため、オブジェクト配置もしくは撮像系のパラメータを予め反転させるか、後処理としての像反転が必要となる。

ただし、これらの課題の多くは、オブジェクト配置やパラメータ等の工夫によって解決可能であり、初心者向けの講習やドームマスター映像制作の習得用途としては、致命的な問題になることはないと考えられる。

2.5 図式等の簡略化

一般に円柱座標 (r, λ, z) と直交座標 (x, y, z) とは式 2、式 3 のような関係で表される。

$$x = r \cos \lambda \quad (2)$$

$$y = r \sin \lambda \quad (3)$$

ここで鏡面となる回転楕円体の中心を原点に置き、極を Z 軸上に置くとする。併せて視点、注視点を Z 軸上に置くとなると、鏡面反射に係る計算は、角度 λ の値によらず、円対称に扱うことが可能となる。以降、本稿における図・式等については、円対称に扱うことができる配置を用いて、円柱座標系における λ 方向の表記を省略し、X, Y 方向の 2 次元直交座標系を用いて撮像系を表現することとする。

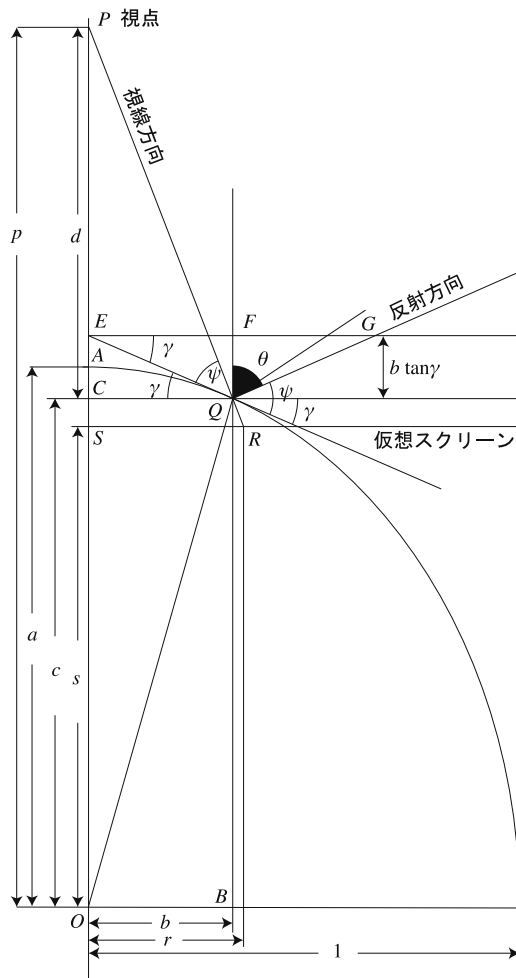


図 5: 模擬ドームマスター (透視投影)

3 透視投影を用いる場合

3.1 位置、角度等の定義

本稿においては原点 O に対して、図 5 に示すような各位置、角度等を、次のように定義する。

- 鏡面となる回転楕円体は、原点 O を中心として、水平方向 (X 軸方向) に赤道半径、鉛直方向 (Y 軸方向) に極半径を設定する。ここで赤道半径は 1、極半径は a とする。このとき、 Y 軸正方向の極点 A の座標は、 $(0, a)$ となる。
- 仮想スクリーンは、原点から s 離れた水平面上に置き、スクリーンの中心 S を $(0, s)$ とする。
- 視点 (カメラ位置) P は原点 O から鉛直正方向に置き、原点との距離を p とする。視点 P の座標は、 $(0, p)$ となる。
- 注視点は、仮想スクリーンの中心 S に置く。

- 視点 P から発する視線と鏡面との交差する点を Q とする。 Q から X 軸に垂直に引いた線と X 軸の交わる点を B とし、その座標を $(b, 0)$ とする。 Q から Y 軸に垂直に引いた線と Y 軸の交わる点を C とし、その座標を $(0, c)$ とする。これにより、点 Q の座標は (b, c) と表される。
- 線分 PC の長さは d と表す。
- 直線 PQ と仮想スクリーンが交差する点を R とし、その座標を (r, s) とする。
- 点 Q における鏡面との接線が Y 軸と交わる点を E とする。 E を通り X 軸に平行な直線が、直線 BQ と交わる点を F とする。
- 直線 PQ (視線) が点 Q において鏡面反射して生じる直線が、直線 EF と交わる点を G とする。
- 視線方向と直線 EQ がなす角を ψ と表す。 ψ は、鏡面上の点 Q における視線の入射角を表すと同時に、出射角とも等しくなる。
- 角 FEQ を γ と表す。図 5 中では、互いに平行な直線 EF と直線 CQ に係る、角 FEQ の同位角・錯角も、同様に γ と表している。
- 角 FQG を θ と表す。

以上の定義は、視点 P に関連する定義を除いて、次節以降も共通に使用する。

また、これらの定義より、 ψ , γ , θ の関係は、式 4 で表される。

$$\psi - \gamma = \frac{\pi}{2} - \theta \quad (4)$$

線分 CP の長さ d は、式 5 で表される。

$$d = b \tan(\psi + \gamma) \quad (5)$$

以上の諸条件から、 θ , r を求める式を導き出す。

3.2 球 (円) における関係式を用いた計算

ここで図 5 を Y 軸方向に $1/a$ 倍した図 6 を考える。図 6 において、鏡面は球 (図上では円) となる。図 5 の各点の位置や角に対応する点や角を表すために、添字 ' n ' を用いることとする。図 6 中で、添字の付いた各点、角等は、図 5 中の、添字 ' n ' の無い各点、角等に対応している。

また、三角形 BOQ_n について、角 BOQ_n を α と表し、角 OQ_nB を β と表す。 α , β は、互いに余角の関係であり、式 6 が成立する。

$$\beta = \frac{\pi}{2} - \alpha \quad (6)$$

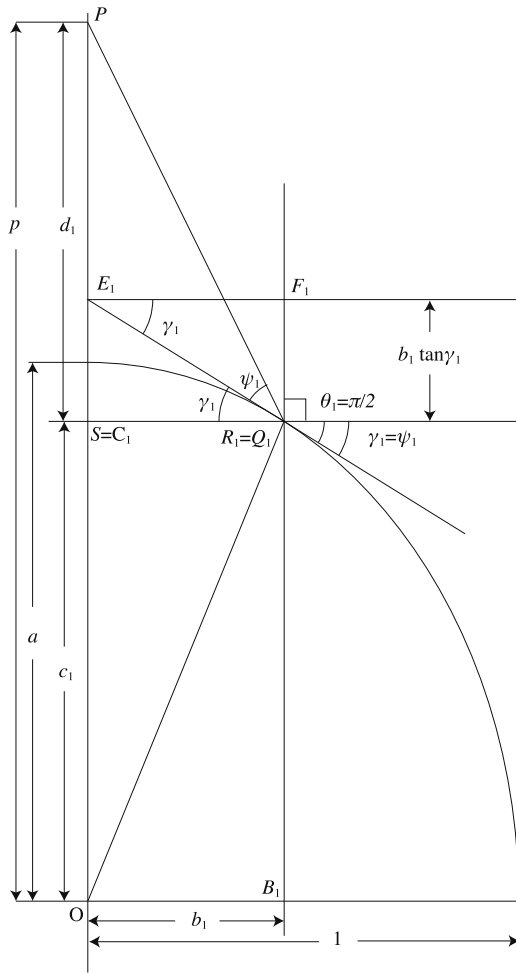


図 7: 模擬ドームマスター (透視投影, エッジ部分)

ここで $a \tan \beta_1$ が 1 となる場合は, 平行投影図法に相当する. 平行投影については, 4 節で述べることにし, 本節では透視投影が成り立つ範囲について述べる.

原点 O から視点位置 P までの長さ p は, 式 19, 式 22 を用いて, 式 23, 式 24 のように求まる.

$$p = c_1 + d_1 \quad (23)$$

$$= a \cos \beta_1 + b_1 \tan \{2 \tan^{-1}(a \tan \beta_1)\} \quad (24)$$

以上で, 角度 β_1 を定めた際の, 模擬ドームマスターのスクリーン位置と視点位置を求めることができた.

3.4 模擬ドームマスターのエッジ以外の計算

続いて, 図 5 において, 視線とスクリーンが交差する点 R の位置を求める. 点 Q の位置は, 式 10, 式 12 より, $(\sin \beta, a \cos \beta)$ と表されるので, これを用いて直線 PQ とスクリーン面が交差する位置 R の X 座標 r を求める.

線分 PC の長さ d は式 25 で与えられる. これと式 12 より式 26 が得られる.

$$d = p - c \quad (25)$$

$$= p - a \cos \beta \quad (26)$$

三角形 CPQ と三角形 SPR は相似形であるため, 線分 CQ の長さ b と線分 SR の長さ r の比は式 27 で表される. これより r の値は式 28 となり, 式 10, 式 26 と併せて式 29 が求まる.

$$\frac{r}{b} = \frac{d_1}{d} \quad (27)$$

$$r = \frac{bd_1}{d} \quad (28)$$

$$= \frac{d_1 \sin \beta}{p - a \cos \beta} \quad (29)$$

ここで, 仮想スクリーン上の模擬ドームマスターにおいても, 図 1 と同様に, 半径 $\pi/2$ となるように正規化することとする. 式 20 で得られた, スクリーンの中心からエッジまでの長さが $\pi/2$ となるように, 式 30 にて, r_d を定義する. これはスクリーンのエッジにおいては, 式 31 となる.

$$r_d = \frac{\pi}{2b_1} r \quad (30)$$

$$r_{d1} = \frac{\pi}{2b_1} b_1 = \frac{\pi}{2} \quad (31)$$

式 30 と式 20, 式 29 から, 式 32 が求まる.

$$r_d = \frac{\pi}{2 \sin \beta_1} \frac{d_1 \sin \beta}{p - a \cos \beta} \quad (32)$$

次に, θ を求める. 式 4 より, θ は式 33 で表される.

$$\theta = \frac{\pi}{2} - \psi + \gamma \quad (33)$$

また ψ については, 式 5 から式 34 が得られる.

$$\psi = \tan^{-1} \frac{d}{b} - \gamma \quad (34)$$

式 33, 式 34 から θ を求めると, 式 36 が得られる.

$$\theta = \frac{\pi}{2} - (\tan^{-1} \frac{d}{b} - \gamma) + \gamma \quad (35)$$

$$= \frac{\pi}{2} - \tan^{-1} \frac{d}{b} + 2\gamma \quad (36)$$

これに式 10, 式 16, 式 26 を適用すると, 式 37 が求まる.

$$\theta = \frac{\pi}{2} - \tan^{-1} \frac{p - a \cos \beta}{\sin \beta} + 2 \tan^{-1}(a \tan \beta) \quad (37)$$

このようにして求められた θ と r_d の差が小さいほど, 模擬ドームマスターのレンダリング結果は, 真のドームマスターに近くなると考えられる.

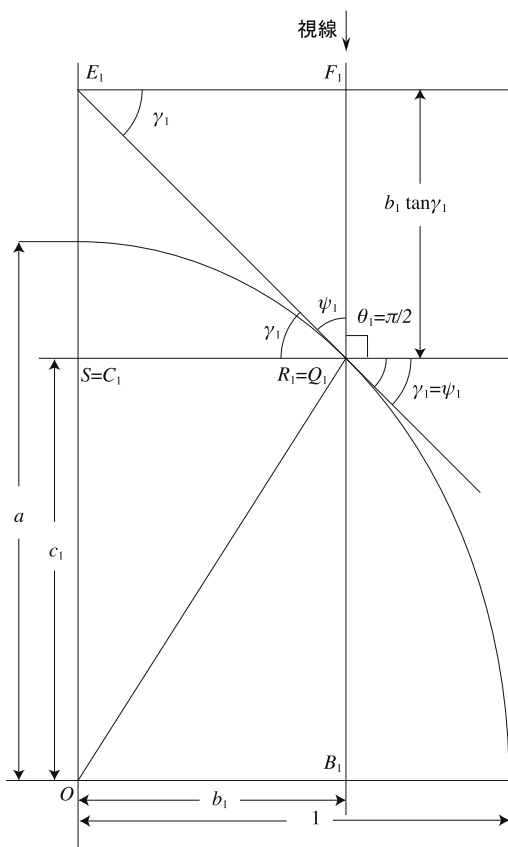


図 8: 模擬ドームマスター (平行投影, エッジ部分)

4 平行投影を用いる場合

平行投影図法を用いた場合, 模擬ドームマスターのエッジにおいて, 図 8 の関係が成り立つ. 視線は常に Y 軸に平行となる. ここで, 視点位置 P が関与しない式については, 本節と 3 節とで共通である.

θ_1 についても, 3 節と同様に, 式 17 を適用する. その結果, γ_1, ψ_1 は, 共に $\pi/4$ となり, 式 16 より, 式 38, 式 39 が求まる.

$$a \tan \beta_1 = 1 \quad (38)$$

$$\beta_1 = \tan^{-1} \frac{1}{a} \quad (39)$$

式 20 より, 式 40 が求まる.

$$b_1 = \sin \tan^{-1} \frac{1}{a} \quad (40)$$

エッジ以外においては, 図 9 の関係が成り立つ. ここで ψ と γ は, 式 41 の関係になる.

$$\psi = \frac{\pi}{2} - \gamma \quad (41)$$

よって, 式 33 は式 42 のように変形できる.

$$\begin{aligned} \theta &= \frac{\pi}{2} - \left(\frac{\pi}{2} - \gamma \right) + \gamma \\ &= 2\gamma \end{aligned} \quad (42)$$

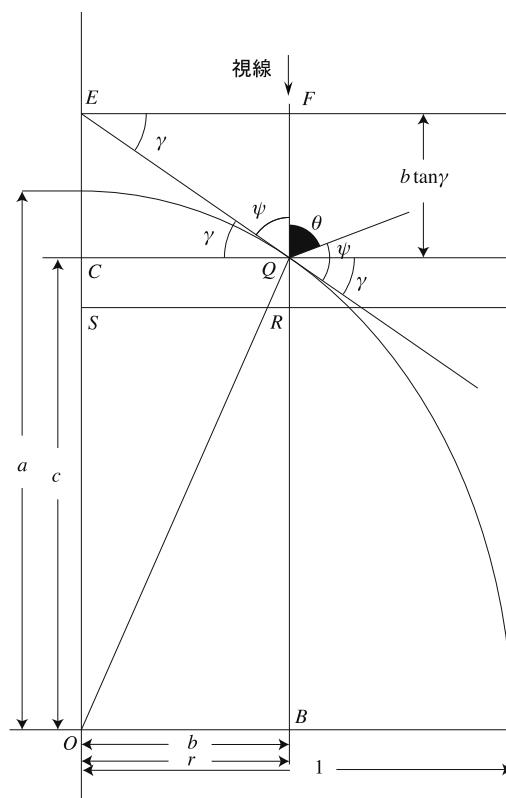


図 9: 模擬ドームマスター (平行投影)

式 16 より, 式 43 が求まる.

$$\theta = 2 \tan^{-1}(a \tan \beta) \quad (43)$$

また平行投影図法においては, r と b は等しいので, 式 10 を用いて, 式 44 が求まる.

$$r = \sin \beta \quad (44)$$

このとき, 式 30 の定義と式 40 より, r_d は式 45 のようになる.

$$r_d = \frac{\pi}{2 \sin \tan^{-1} \frac{1}{a}} \sin \beta \quad (45)$$

5 評価

3 節で述べた透視投影及び 4 節で述べた平行投影による各手法について, 模擬ドームマスターと真のドームマスターとの誤差を評価する.

5.1 透視投影

図 10 に透視投影を用いた場合の誤差の計算結果を示す. 横軸は, 鏡面の赤道半径に対する極半径の比 (a) を表す. 縦軸に表された誤差は, 以下の手法で求めたものである.

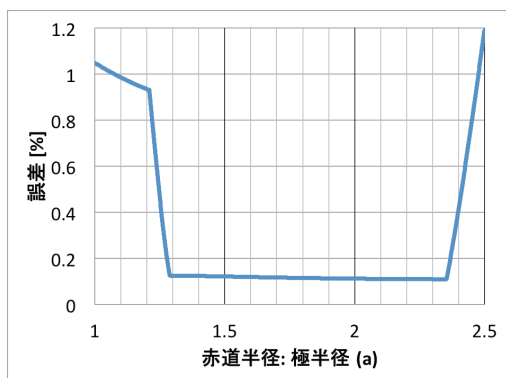


図 10: 透視投影における誤差 (全体)

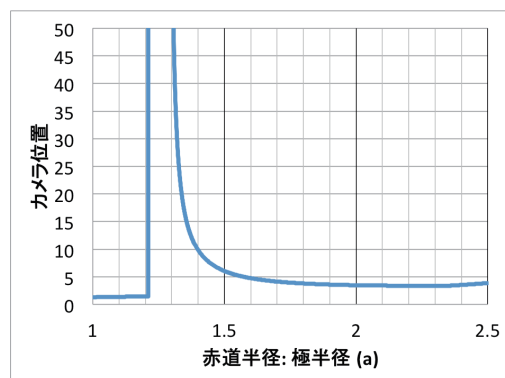


図 12: 透視投影におけるカメラ位置 (全体)

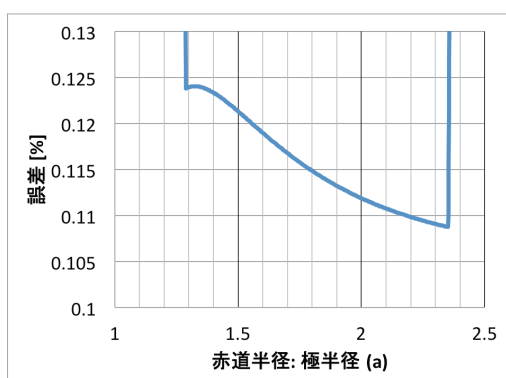


図 11: 透視投影における誤差 (拡大)

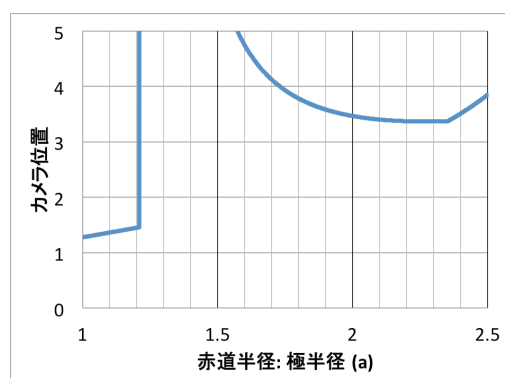


図 13: 透視投影におけるカメラ位置 (拡大)

- 与えられた a に対して、 $0 < a \tan \beta_1 < 1$ の範囲内にある β_1 について、次の計算をする。
- $0 < \beta < \beta_1$ の範囲内にある β について、 a と β から求まる θ と r_d の差の絶対値を計算する。
- θ と r_d の差の絶対値の内、最大となるものを、その a , β_1 の組み合わせにおける最大誤差とする。
- a , β_1 の組み合わせにおける最大誤差の内、 $0 < a \tan \beta_1 < 1$ の範囲内において最小となるものを、 a における誤差とする。

この図より、比が 1.3 付近から 2.35 付近までの誤差が一様に小さくなっていることがわかる。図 11 に、係る範囲の誤差の拡大図を示す。誤差が最小となるのは、比が 2.35 の付近となった。ここでの誤差は約 0.11% となった。これは、4K (4,096 ピクセル× 4,096 ピクセル) のドームマスターフォーマットに換算して、2 ピクセル程度となる。ドームスクリーンに投影された映像において 2 ピクセルの誤差を目視で判別することは事実上不可能であり、この範囲の誤差は十分実用レベルにあると考えられる。

図 10 における誤差を得たときのカメラ位置 p を図 12

に示す。同様に、その拡大図を図 13 に示す。 a が 1.2 付近から 1.4 付近の範囲では、カメラが鏡面から大きく離れており、1.2 付近から 1.3 付近の範囲では、事実上平行投影となっている。一方、図 11 で誤差最小を示していた $a = 2.35$ 付近ではカメラ位置も鏡面に近づいており、現実的な撮像系を構成できると期待される。

また、球 ($a = 1$) で鏡面を構成した場合のカメラ位置は 1 に近くなっている。球の半径 1 を差し引くと、カメラ位置と鏡面の距離は短い。図 10 によると、この時の誤差は 1% 余りあるが、球を鏡面に用いることは、撮像系をコンパクトに構成できるメリットがあると考えられる。

各 a について、誤差最大となる角度 (r_d) を図 14 に示す。図中では、角度をラジアンではなく度 ($^{\circ}$) に換算して表している。誤差最大となる角度は、概ね 30 度付近と 75 度付近に現れている。この 2 箇所誤差が最大となるケースでは、それぞれの角度付近における誤差は、ほぼ同一となった。

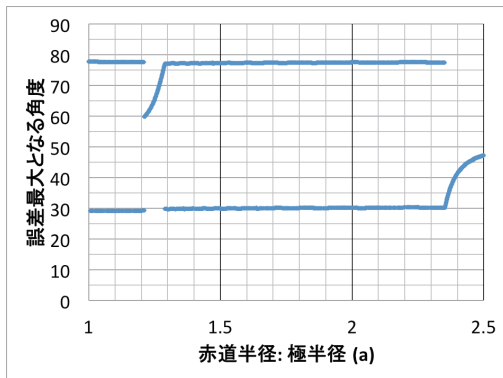


図 14: 透視投影における誤差最大となる角度

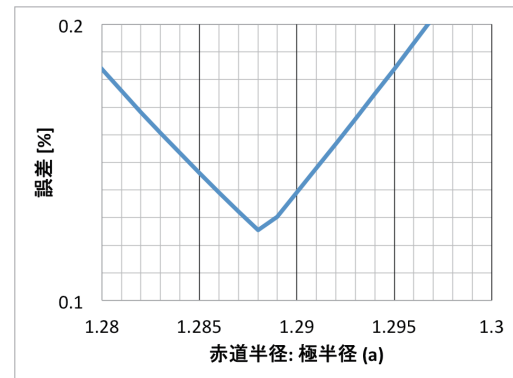


図 16: 平行投影における誤差 (拡大)

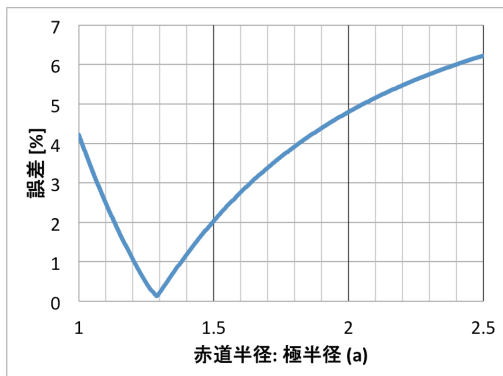


図 15: 平行投影における誤差 (全体)

5.2 平行投影

平行投影における誤差を、図 15 に示す。縦軸に表された誤差は、以下の手法で求めたものである。

- 与えられた a に対して、 $a \tan \beta_1 = 1$ となる β_1 について、次の計算をする。
- $0 < \beta < \beta_1$ の範囲内にある β について、 a と β から求まる θ と r_d の差の絶対値を計算する。
- θ と r_d の差の絶対値の内、最大となるものを、その a における誤差とする。

誤差が最小となる付近の拡大図を、図 16 に示す。平行投影においては、 $a = 1.288$ 付近が最も誤差が小さくなる。この付近における誤差は、約 0.13% となった。これは 4K のドームマスターフォーマットにおいて 3 ピクセル未満の誤差となる。平行投影の場合、カメラ位置を鏡面の直近におくことができるため、撮像系をコンパクトに構成することができる。平行投影に対応した CG ソフトウェアであれば、十分実用的な設定である。

以上の結果は、2.5 節で述べたように、円柱座標系 (r, λ, z) の角度 λ によらず、模擬ドームマスター形式の任意の角度方向に対して同様に適用できるものである。

表 1: 撮像系を構成するパラメータ

	投影	a	$s (c_1)$	p	誤差 [%]
(1)	平行	1.288	1.017	N/A	0.13
(2)	透視	1	0.930	1.277	1.04
(3)	透視	2.350	2.235	3.374	0.11

5.3 撮像系の構成例

以上の結果より、撮像系を構成するパラメータを選定する。候補を表 1 に示す。ここでは、(1) 平行投影にて良好な結果が得られる、 $a = 1.288$ のときの値、(2) 透視投影と球 ($a = 1$) の鏡面反射にて良好な結果が得られる値、(3) 透視投影と回転楕円体の鏡面反射にて良好な値が得られる $a = 2.350$ のときの値を例示した。いずれにおいても、撮像系を構成するカメラの画角調整はソフトウェア毎に設定方法が異なると考えられるが、それ以外に考慮すべき位置やサイズは、2 つ乃至 3 つのパラメータで表すことができる。(1), (2), (3) のいずれも、講習等で提示するパラメータ数としては、無理の無い範囲に収まっていると考えられる。

(1) は平行投影であるため、視点と鏡面の距離を短く設定でき、撮像系がコンパクトになる。(2) は球を鏡面として用いるため、扁平率を設定する必要がない。また視点と鏡面の距離も比較的短くなる。(3) は、計算上は最も高精度であると期待されるパラメータ設定である。前 2 者と比べて、考慮すべきパラメータが 1 つ増え、視点位置は鏡面から若干遠くなる。

なお実際に撮像系を構成する場合には、仮想スクリーン面下の不要な鏡面を視点から隠すための平面を、仮想スクリーンの位置に配置すると、ドームマスターとして描くべき範囲外の領域を隠すことができる。

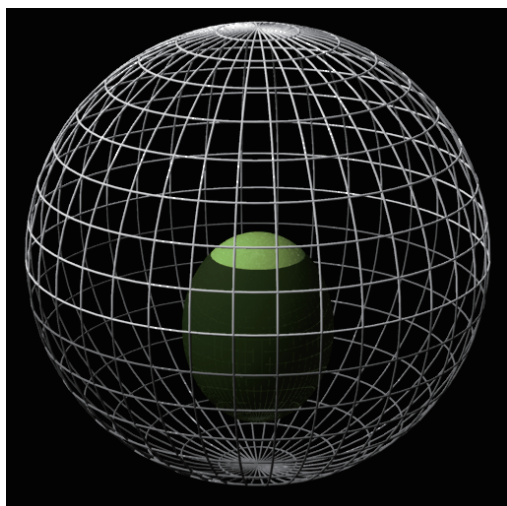


図 17: 経緯線レンダリングの模式図 (例示)

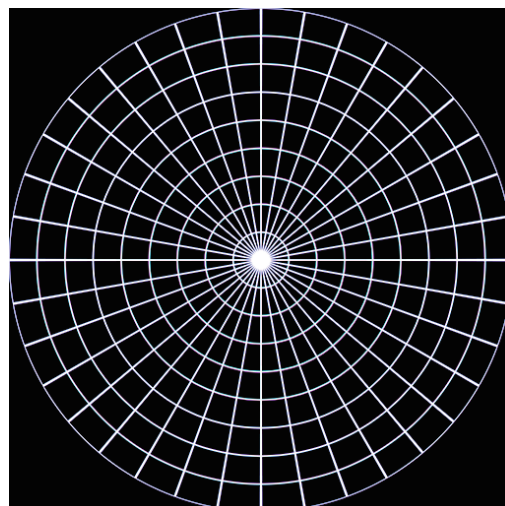
5.4 近景の歪

本稿においては、オブジェクトは撮像系に対して十分遠方にあると仮定したが、実際のシーンを構成するにあたっては、カメラ直近にオブジェクトが置かれる可能性もある。そのようなケースでは、歪の許容範囲を考慮しながら撮像系を構成する必要がある。

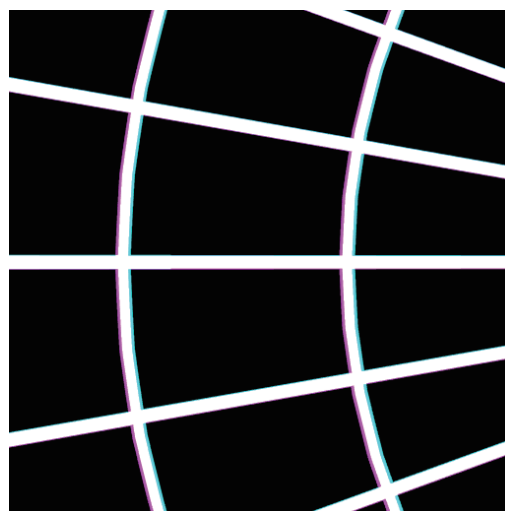
歪評価のためのレンダリングに用いたシーンの模式図を図 17 に示す。中央の黄緑に彩色した部分が鏡面を表し、周囲のグリッドが、10 度間隔の経緯線を表す。暗色の部分は、仮想スクリーン下にあり、ドームマスターとして表現される範囲の外にある部分である。なお、この図は例示であり、鏡面の形状が分かり易くなるように各オブジェクトの彩色やスケールを施している。後述のレンダリングに用いた鏡面は反射率 100% であり、経緯線と鏡面のサイズ比は、其々の図に示した通りである。

具体的な歪の例を図 18 から図 20 に示す。いずれも撮像系を表 1 における (1) の配置として描いたものである。これらの図においては、赤線が真のドームマスターにおける経緯線、水色の線が模擬ドームマスターで表現した経緯線となっている。白色の部分は、両者が重なっている部分である。

図 18 は撮像系の回転楕円体の赤道半径に対して 100 倍の半径の経緯線をレンダリングしたものである。真のドームマスターとの誤差は僅少であるため、誤差が大きくなる付近の拡大図を併せて示す。この誤差は 4K (4096 ピクセル) のドームマスターフォーマット換算で 3 ピクセル程度であり、オブジェクトが十分遠くにあると仮定した計算における誤差と大差ない。図 19 は同じく 10 倍の半径、図 20 は 1 倍の半径の経緯線をレンダリングしたものであり、それぞれ 4K のドームマスターフォーマット換算で 10 ピクセル、80 ピクセル程度の誤差となった。



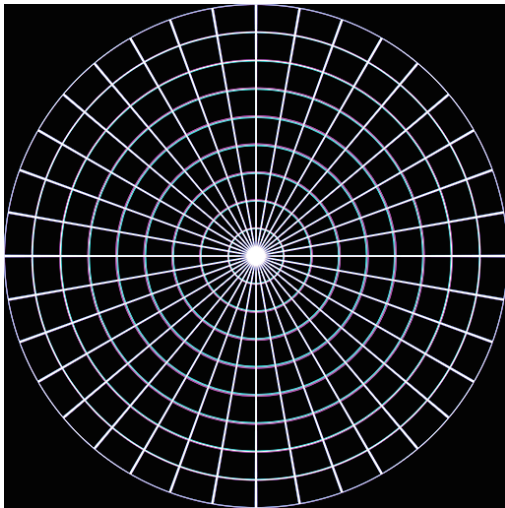
全体



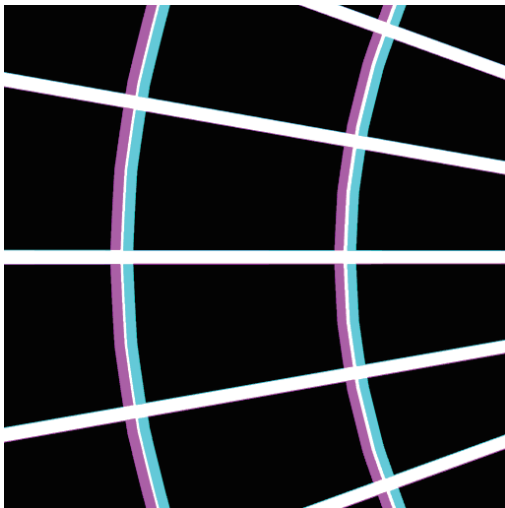
拡大 (8 倍)

図 18: 経緯線の遠近による誤差 (距離 100 倍)

このように、撮像系のサイズに対して、シーン内のオブジェクトを十分大きく、遠方に配置する (もしくは、シーン内のオブジェクトに対して撮像系を十分に小さく構成する) ことは、歪を抑える方法として有効である。一方で、レイトレーシングに用いられる浮動小数点演算の精度によっては、オブジェクトサイズの極端な不均衡はレンダリング結果に誤差を生じる可能性がある。そのような誤差が見受けられる場合は、撮像系に起因する誤差と、浮動小数点演算に起因する誤差とのトレードオフが必要となる。



全体



拡大 (8 倍)

図 19: 経緯線の遠近による誤差 (距離 10 倍)

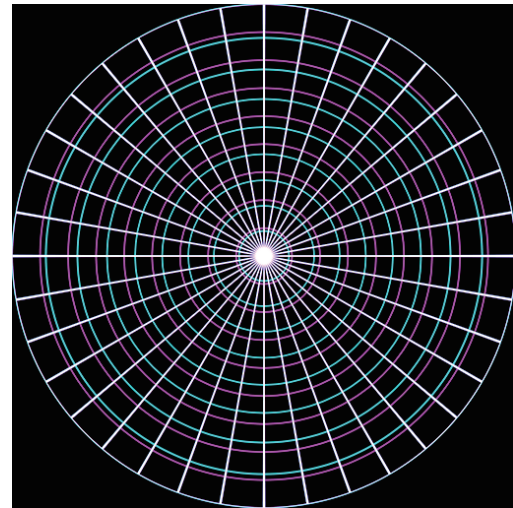


図 20: 経緯線の遠近による誤差 (距離 1 倍)

までには、若干の試行錯誤が必要となる。またシーンの中に撮像系と近接するオブジェクトがある場合のトレードオフについては、課題として残されている。

6 あとがき

全天周映像制作の初心者念頭に置いた、ドームマスター制作の模擬手法について述べた。本手法は、球体や回転楕円体の鏡面反射をサポートした CG ソフトウェアであれば共通に適用できる簡易的な手法である。また本手法は、ドームマスター制作において事実上問題の無い精度でのレンダリングが可能であることが明らかになった。

本稿においては、撮像系を配置するための設定を幾つか示したが、いずれも設定すべきパラメータ数は少なく、初心者向けの講習等での導入も容易であると考えられる。ただし撮像系を構成するカメラの画角の設定方法は CG ソフトウェアによって差があるため、適切な画角を得る

参考文献

- [1] コニカミノルタプラネタリウム株式会社, ‘プラネタリウム製品,’
<http://www.konicaminolta.jp/planetarium/>, (2016 年 9 月 参照)
- [2] 株式会社五藤光学研究所, ‘プラネタリウム,’
<http://www.goto.co.jp/planetarium/> (2016 年 9 月 参照)
- [3] 有限会社大平技研, ‘スーパープラネタリウム「MEGASTAR (メガスター)」シリーズのオフィシャルサイト,’
<http://www.megastar.jp/about/> (2016 年 9 月 参照)
- [4] Immersive Media Entertainment, Research, Science & Arts (IMERSA), ‘IMERSA | Standards,’
<http://imersa.org/standards/standards-forum> (2016 年 9 月 参照)
- [5] ORIHALCON Project, ‘全天周映像 (ドーム映像) の作り方,’
<http://orihalcon.jp/documents/domemaster-how-to.html> (2016 年 9 月 参照)
- [6] Wikipedia, ‘Ray tracing (graphics),’
[https://en.wikipedia.org/wiki/Ray_tracing_\(graphics\)](https://en.wikipedia.org/wiki/Ray_tracing_(graphics)) (2016 年 9 月 参照)
- [7] Jason Fletcher, ‘Fisheye & Spherical Collection: Supported Render Engines,’
<https://thefulldomeblog.com/2013/06/28/fisheye-lens-shader-options/> (2016 年 9 月 参照)
- [8] Paul Bourke, ‘Computer Generated Angular Fisheye Projections,’ May 2001,
<http://paulbourke.net/dome/fisheye/> (2016 年 9 月参照)
- [9] Paul Bourke, ‘Direct rendering technique for fisheye projection,’ Sep. 2000,
<http://paulbourke.net/dome/mirrorrender/> (2016 年 9 月参照)

宇宙データを使った世界同時開催ハッカソン「International Space Apps Challenge」の日本開催

湯村 翼^{*1,*2,*3}

概要

International Space Apps Challenge (SpaceApps) は、宇宙データを使って宇宙のアプリをつくる NASA 主催のハッカソンである。2012 年を皮切りに毎年 4 月に開催され、近年は世界中の 100 都市以上の会場にて総数 10,000 人以上が参加する世界最大級のハッカソンである。日本でも、東京をはじめとした会場にて SpaceApps を 2012 年より毎年開催してきた。当初は東京のみでの開催だったが、2016 年には国内の開催地は 7 箇所に至るまで増加した。

Organizing the space data hackathon “International Space Apps Challenge” in Japan

Tsubasa Yumura^{*1,*2,*3}

Abstract

International Space Apps Challenge (SpaceApps) is the hackathon to develop applications with using space data, which is organized by NASA. SpaceApps has been held every April since 2012. Recently SpaceApps has over 100 host cities and over 10,000 participants each time. It is one of the biggest hackathon events in the world. SpaceApps also has been held every year since 2012 in Japan. Although it was initially held only in Tokyo, domestic venue in 2016 was increased to seven.

1. はじめに

International Space Apps Challenge (以下 SpaceApps) は、宇宙データを使って宇宙のアプリをつくる NASA 主催のハッカソンである。2012 年を皮切りに毎年 4 月に開催され、近年は世界中の 100 都市以上の会場にて総数 10,000 人以上が参加する世界最大級のハッカソンである。

ハッカソン(hackathon)とは、より便利にするといった意味の「hack」と「marathon」を掛けあわせた造語で、定められた短い期間中にアプリケーションを開発するイベントの総称である。ハッカソンは近年急速に広まり、テクノロジーに限らずあらゆるテーマのハッカソンが開催されている。

SpaceApps は人類が抱える様々な課題解決を目的として宇宙データを使ったアプリケーションを作成

するハッカソンである。SpaceApps が誕生した 2012 年は、米国でオープンデータの動きが盛んとなった時期である。オープンデータの考え方は、政府や行政が有するデータを市民に公開し、アプリやサービスの開発への活用を推進し、市民の生活が向上するというものである。NASA が保有する宇宙データも、以前から知的財産権が発生しないパブリックドメインにて公開されていたが、宇宙データの活用をさらに推進するためにこの SpaceApps が立ち上げられた。

本イベントの名称は、2016 年より NASA Space Apps Challenge へと変更されている。略称は、当初は「SpaceApps」と International Space Apps Challenge の頭文字を取った「ISAC」が混在していたが、2016 年の正式名称変更により SpaceApps に一本化された。日本でも、2014 年までは ISAC を略称として使用してきたが、2015 年より、略称から内容を読み取れる

doi: 10.20637/JAXA-RR-16-007/0015

* 平成 28 年 11 月 24 日受付 (Received November 24, 2016)

*1 SpaceApps Tokyo

*2 国立研究開発法人 情報通信研究機構 (National Institute of Information and Communications Technology)

*3 北陸先端科学技術大学院大学 (Japan Advanced Institute of Science and Technology)

SpaceApps を公式に使用している。

本稿では、SpaceApps の日本での開催の報告とこれからの展望についての議論を行う。

2. SpaceApps の概要

SpaceApps は、実際に開発を行う 2 日間のハッカソンに加え、事前準備を行う BootCamp、ハッカソンでの審査で選出されたチームが進出するグローバル審査がある。また、当然ながら運営を行う事務局はハッカソンの何ヶ月も前から準備を行う。

SpaceApps では 20 個程度の「Challenge」と呼ばれる課題が設定されている。各チームはいずれかの Challenge を選択し、それに沿ったアプリケーション開発を行う。

ハッカソン開催前には Challenge の内容確認、チームビルディング、アイデアのブラッシュアップなどのハッカソンの前準備のための BootCamp を行うことが推奨されている。

ハッカソンは、週末の金曜夕方～日曜に設定され、最大 48 時間を使ってアプリケーションの開発を行う(図 1)。ハッカソンの最後に各会場で審査を行い、そこで選出されたチームがグローバル審査へ進出する。2016 年のレギュレーションでは、優秀 2 作品と、参加者投票による People's Choice の計 3 チームがグローバル審査へと進出する。また、グローバル審査の選考以外にも、開催地によってはスポンサー賞などの独自の賞も設けられる。

ハッカソン終了時に、各チームが成果物をプロジェクトページとしてまとめる(図 2)。プロジェクトページには、プロジェクトの概要や説明のほか、ソースコードやサービスサイトへのリンクを掲載する。このページは web のフォーム入力により簡単に作成できるようになっている。グローバル審査に選出さ

れたチームは、30 秒程度の動画を作成し、プロジェクトページに掲載する必要がある。

グローバル審査(図 3)はグローバルアワードを選出するためのもので、2016 年は「Best Use of Hardware」「Best Mission Concept」「Most Inspirational」「Galactic Impact」「People's Choice」「Best Use of Data」の 6 つのアワードが設けられた。アワードの選出は、People's Choice は Twitter を用いた投票で決定され、それ以外のアワードは審査員による審査にて決定される。審査はオンラインで行われ、プロジェクトページが審査対象となる。ハッカソン翌月の 5 月に、グローバル審査にて選ばれたグローバルアワード受賞者が発表される。

SpaceApps のまとめとして、参加者数などの詳細な開催情報が掲載されたミッションレポート[1]が公開される。

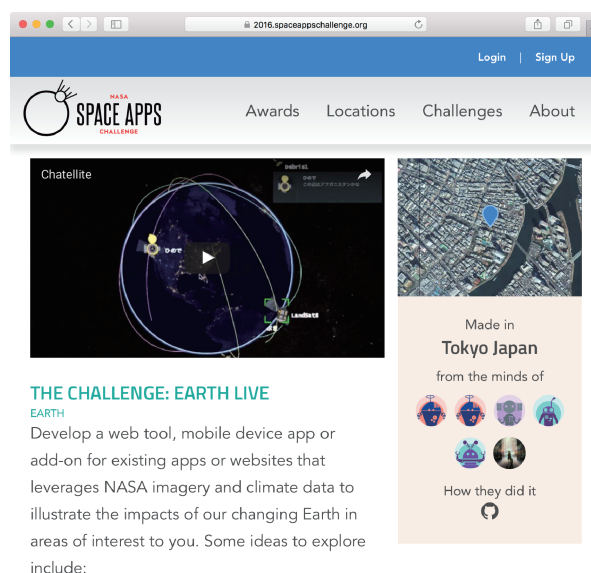


図 2 プロジェクトページ



図 1 開発の様子(SpaceApps Tokyo 2016 より)

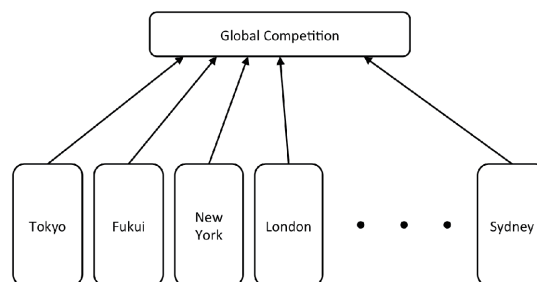


図 3 グローバル審査の概略図

3. 日本における SpaceApps

日本でも、東京をはじめとした会場にて SpaceApps を 2012 年より毎年開催してきた。当初は東京のみでの開催だったが、2016 年には国内の開催地は 7 箇所まで増加した(表 1)。参加者の属性は、ソフトウェアエンジニア、ハードウェアエンジニア、デザイナー、研究者、学生と幅広い(図 4)。

国内の参加者同士の交流用に SpaceApps Japan という Facebook グループ[2]も設置している。SpaceApps や宇宙データに関する情報共有を行い、ハッカソン期間にとどまらない交流を促進し、コミュニティを形成するのが目的である。また、日本語での情報発信用のブログ[3]も設置し、イベント前の情報告知や、開催報告に活用している。

表 1 日本での SpaceApps 開催地

年	開催都市
2012	東京
2013	東京
2014	東京
2015	東京, 会津(福島), 福井, 山口, 肝付(鹿児島)
2016	東京, 会津(福島), 福井, つくば(茨城), 相模原(神奈川), 宇部(山口), 熊本*

*震災のためハッカソンは未開催



図 4 SpaceApps の参加者(SpaceApps Tokyo 2016 より)

4. SpaceApps の運営体制

SpaceApps では、運営の全体取りまとめを行うグローバルオーガナイズチームと、各開催地の運営を行うローカルオーガナイズが存在する(図 5)。グローバルオーガナイズチームは、レギュレーションの策定や、運営ガイドラインの作成を行い、ローカルオーガナイズは、各開催地でのイベント運営の実働を担

う。グローバルオーガナイズチームとローカルオーガナイズを接続したオンラインミーティングが定期的に開催され、不明点などはそのミーティングにて議論や共有を行う。

国内の各開催地のローカルオーガナイズは、連絡体制強化のために、非公式の集まりとして Facebook グループを用いた情報共有を行っている。特に、本部からの連絡やミーティングはすべて英語で行われるため、準備に漏れが生じないように日本語での情報共有が重要となる。

国内の開催地では、スタッフはボランティアが主体となっている運営体制が多い。自治体と密に連携を図っている開催地もある。

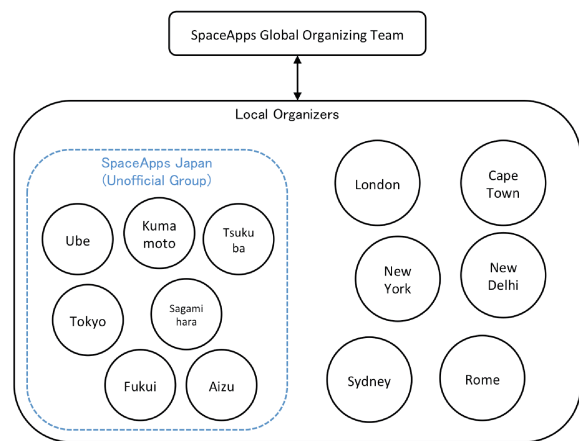


図 5 運営体制の概略図

5. 国内会場での開発事例

SpaceApps の国内会場で生まれた作品の代表例を紹介する。括弧内は SpaceApps の参加年を示す。

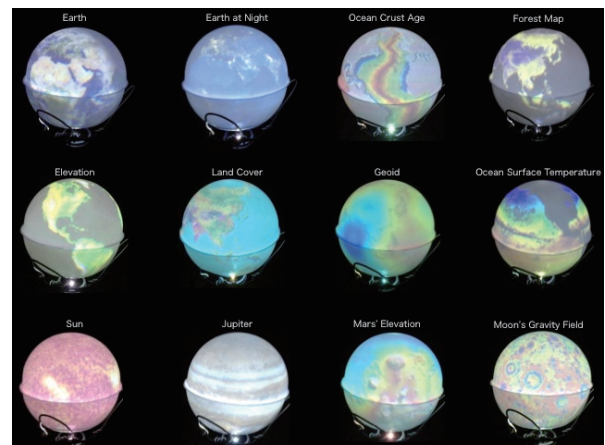


図 6 Personal Cosmos

Personal Cosmos(2013)[4]は、市販のプロジェクターなどの身近な材料を使ったデジタル地球儀を開発したプロジェクトである(図 6)。NASA や JAXA が公開する多数の地図データを画像変換し、地球儀として投影することができる。展示などで活用されている他、科学教育教材としての可能性も見出すことができる。

Marsface Project(2013-2016)[5]は、宇宙観測データの中から顔に見える特徴的な画像を、ディープラーニングなどの機械学習技術を利用して抽出する手法を提案した。これらの成果は国内学会や国際学会で多数発表している[6][7]。また、複数年にわたってプロジェクトを継続しており、3D プリンタを用いた月表面模様サンダーの作成や電子書籍の機械的生成など、年を追う毎に内容が発展している。



図 7 Marsface Project で発見した月面上の顔

Toward HyperLocation for Everyone(2015)[8]は、測量などで使われる高精度測位方式「RTK-GNSS」をスマートフォンや Raspberry Pi などの安価な端末を用いて実現した。安価な高精度測位の実現により、測位が必要なさまざまなサービスの発展が見込まれる。

Active Space Radar App(2015)[9]は、ガンマ線バーストが発生したらプッシュ通知を送る iPhone アプリケーションである(図 8)。開発には Swift 衛星[10]、Fermi 宇宙望遠鏡[11]、MAXI[12]のデータを用いている。

チャテライト(2016)[13]は、人工衛星の軌道を可視化するウェブアプリケーションである(図 9)。軌道可視化アプリケーションはこれまでも多数開発されてきたが、人工衛星にキャラづけをしてメッセージを発信することで飽きずに見続ける工夫が施されている。

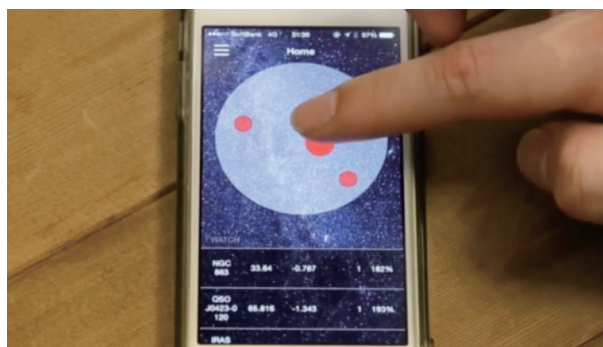


図 8 Active Space Radar App

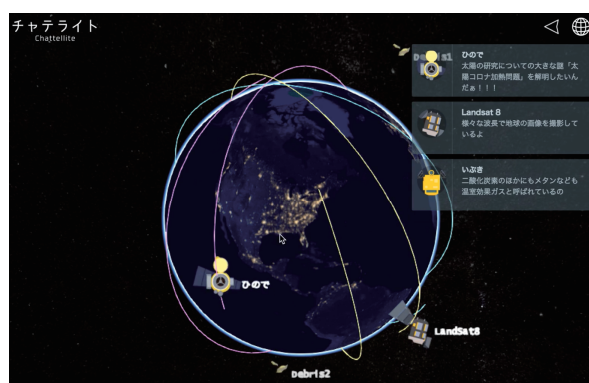


図 9 チャテライト

6. 課題

過去 5 年間にわたり SpaceApps を運営した過程で見えてきた課題を述べる。

運営に母体をおかず、ボランティア主体のため、安定した運営リソースが確保できない。そのため、運営メンバー個人の本業の忙しさなどに依存することとなる。

ハッカソンの参加者数は、東京では例年定員を超える参加申し込みがあるが、それ以外の日本の会場では定員が埋まるには至っていない。宇宙に興味のあるエンジニアは多数いると考えられるため、興味を持つ人達へ正しく情報を届けられるようにしたい。

参加者の多くは宇宙や人工衛星の専門知識がないため、これらについてアドバイスできるメンターが各会場に配置できればアプリケーション開発がより捗ると考えられる。大学や研究機関と連携して各会場に研究者を配置するなど、改善策を検討していきたい。研究者側にも、ハッカソンに参加するエンジニアと議論することによって良い相乗効果が生まれ、メリットを生み出せるのではないかと考える。

また、過去のプロジェクトのアーカイブ方法についても検討する余地があると考えられる。プロジェクト

ページを作ることによってプロジェクトのアーカイブは達成されているが、膨大なプロジェクト数になるため、知りたい情報にたどり着くことが困難となっている。検索しやすく、過去の知見が活用されるようなシステムが望まれる。

7. おわりに

2012年にはじまった SpaceApps を、2016 年までの 5 年間毎年継続して日本で開催してきた。今後もこれを継続し、宇宙データを活用した開発のためのコミュニティ形成をサポートしていきたい。

謝辞

SpaceApps は、非常に多くのボランティアスタッフによってその運営が支えられている。運営に携わった全てのスタッフへ感謝を表す。

参考文献

- 1) 2016 Space Apps Mission Report
<https://open.nasa.gov/documents/60/>
- 2) SpaceApps Japan Facebook グループ
<https://www.facebook.com/groups/spaceappstokyo/>
- 3) SpaceApps Japan ブログ <http://blog.spaceapps.jp/>
- 4) Personal Cosmos | International Space Apps Challenge
<https://2013.spaceappschallenge.org/project/personal-geo-cosmos/>
- 5) Marsface Project <http://marproject.org/>
- 6) 栗原一貴, 笹尾和宏, 山本光穂, 田中秀樹, 奈良部隆行, 國吉雅人, 会田寅次郎, 岡田裕子, 高須正和, 関治之, 飯田哲, 山本博之, 生島裕, Deep Learning に基づく画像認識を用いた月および火星表面の擬似不自然構造物探索, エンタテインメントコンピューティングシンポジウム 2014 論文集, pp.223-224 (2014).
- 7) Kazutaka Kurihara, Masakazu Takasu, Kazuhiro Sasao, Hal Seki, Takayuki Narabu, Mitsuo Yamamoto, Satoshi Iida, Hiroyuki Yamamoto: A Face-Like Structure Detection on Planet and Satellite Surfaces Using Image Processing, Advances in Computer Entertainment 2013 Proceedings, pp.564-567 (2013).
- 8) Toward HyperLocation for Everyone | 2015 SpaceApps Challenge
<https://2015.spaceappschallenge.org/project/toward-hyperlocation-for-everyone/>
- 9) Active Space Radar | 2015 SpaceApps Challenge
<https://2015.spaceappschallenge.org/project/active-space-radar/>
- 10) Swift/BAT Hard X-ray Transient Monitor
<http://swift.gsfc.nasa.gov/results/transients/>
- 11) Gamma Ray Astrophysics at the NSSTC
<http://gammaray.msfc.nasa.gov/gbm/science/pulsars.html>
- 12) 全天 X 線監視装置 MAXI
<http://maxi.riken.jp/top/index.php>
- 13) Chatellite [Space Apps 2016]
<https://2016.spaceappschallenge.org/challenges/earth/earth-live/projects/chatellite>

宇宙航空研究開発機構研究開発報告 JAXA-RR-16-007

JAXA Research and Development Report

宇宙科学情報解析論文誌 第六号

発 行 国立研究開発法人 宇宙航空研究開発機構 (JAXA)
〒182-8522 東京都調布市深大寺東町7-44-1
URL: <http://www.jaxa.jp/>

発 行 日 平成29年3月17日
電 子 出 版 制 作 松枝印刷株式会社

©2017 JAXA

※本書の一部または全部を無断複写・転載・電子媒体等加工することを禁じます。

Unauthorized copying, replication and storage digital media of the contents of this publication, text and images are strictly prohibited. All Rights Reserved.

