



ISSN 1349-1113  
JAXA-RR-13-010

# 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 第三号

2014年3月

宇宙航空研究開発機構

Japan Aerospace Exploration Agency

## 序 文

宇宙科学情報解析論文誌（Journal of Space Science Informatics Japan）第三号をお送りします。

JAXA 宇宙科学研究所・学際科学研究系では、その前身である宇宙科学情報解析研究系、宇宙科学情報解析センター（PLAIN センター）の時代より、主に科学衛星・探査機によって得られた宇宙観測データを収集し、広く国内外の研究者による利用を促進するためのサイエンスデータベースの開発、並びにデータ解析手法の研究を行ってきました。

近年、宇宙科学における様々な分野において、観測装置と計算機の大型化によって、以前に比べて飛躍的に大量のデータを取得することが可能になってきています。そのために、観測から優れた科学的成果を産出するためには、大規模データをいかに効率良く解析し、必要な情報を引き出すかが、重要な技術課題となりつつあります。また、デジタル技術の普及に伴い、デジタルプラネタリウムや様々なアプリケーションに代表されるように、宇宙観測データが科学研究以外の分野でも広く利用されるようになってきています。

そのような状況を鑑み、当研究系では、宇宙科学データ（シミュレーションデータを含む）に関する新しい処理、解析、利用の手法、データベース技術やそれを応用したシステムの構築・運用技術など、宇宙科学、情報科学、情報技術に関わる成果発表および情報交換の場として「宇宙科学情報解析シンポジウム」を毎年開催するとともに、それに関連した研究論文を集めた「宇宙科学情報解析論文誌」を発行して参ります。当論文誌が、周辺領域の研究者、技術者の間の情報交換や活発な議論につながることを期待しています。

2013 年 12 月 5 日

JAXA 宇宙科学研究所  
学際科学研究系

海老沢 研、篠原 育、高木 亮二、三浦 昭、山本 幸生  
plsinsympo@c-soda.isas.jaxa.jp

## 目 次

### アウトリーチ、デジタルアート

「科学におけるグラフィックデザインの役割～宇宙から細胞まで～」

佐藤 暁子 ..... 1

『はやぶさ』の小惑星イトカワ探査軌道・姿勢の可視化」

三浦 昭, 山本 幸生, 吉川 真 ..... 7

### 衛星試験・運用システム

「SIB2/GSTOS-1 における開発状況」

西村 佳代子, 松崎 恵一, 宮澤 秀幸, 高木 亮治, 山下 美和子,  
宮野 喜和, 福田 盛介, 馬場 肇, 永松 弘行, 山田 隆弘 ..... 17

『どこでも運用システム』の開発状況」

永松 弘行 ..... 27

### データセンター、クラウド

「月惑星探査研究拠点間連携によるデータセンター構想」

出村 裕英 ..... 33

「科学研究用クラウドシステム（NICT サイエンスクラウド）の提案」

村田 健史, 渡邊 英伸, 鶴川 健太郎, 村永 和哉, 鈴木 豊,  
磯田 総子, 山本 和憲, 久保田 康文, 長妻 努, 坂口 歌織,  
津川 卓也, 西岡 未知, 建部 修見, 田中 昌宏, 深沢 圭一郎,  
才田 聡子, 海老原 祐輔, 藤田 茂, 木村 映善, 黒澤 隆,  
村山 泰啓, 永井 亨, 水原 隆道 ..... 39

「NICT サイエンスクラウドによる大規模シミュレーションデータ分散可視化処理」

村田 健史, 磯田 総子, 渡邊 英伸, 深沢 圭一郎, 山本 和憲,  
建部 修見, 田中 昌宏, 木村 映善 ..... 57

### データ利用・配信システム

「地球惑星試料デポジトリ DREAM の構想と現状」

北川宙, 谷内勇介, 国広卓也, 中村栄三 ..... 71

「ALMA キューブデータ簡易解析 Web アプリケーションの開発」

江口 智士..... 79

「超高層物理学分野のメタデータ・データベースへの連想検索の適用」

小山 幸伸, 阿部 修司, 八木 学, 梅村 宜生, 堀 智昭,  
新堀 淳樹, 佐藤 由佳, 家森 俊彦, 田中 良昌, 橋口 典子,  
上野 悟, 谷田貝 亜紀代..... 89

**衛星データ処理、地上観測データ処理**

「CALET の運用及びデータ解析・管理システムの概要」

赤池 陽水, 浅岡 陽一、上野 史郎、田村 忠久、寺澤 敏夫、  
富田 洋、鳥居 祥二、中川 友進、仁井田 多絵..... 99

「SMILES Level 2 データ処理における逆問題の解析手法」

眞子 直弘, 鈴木 睦, 佐野 琢己, 今井 弘二, 光田 千紘,  
山田 道夫, 竹広 真一, 塩谷 雅人..... 109

「小型望遠鏡を用いた突発天体の自動広視野サーベイ」

前原 裕之.....119

**可視化技術**

「地球磁気圏 3 次元グローバル MHD シミュレーションの磁力線追尾可視化技術」

久保田 康文, 村田 健史, 山本 和憲, 深沢 圭一郎, 坪内 健..... 129

「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」

後藤 康路, 山田 竜平, 山本 幸生, 横山 昌平, 石川 博..... 137

**シミュレーション、データ同化**

「数値モデルを用いた惑星大気研究と地球流体電脳倶楽部 DCMODEL プロジェクト」

竹広 真一, 高橋 芳幸, 西澤 誠也, 林 祥介, 地球流体電脳倶楽部..... 147

“Effective system for simulating dust continuum observations on distributed computing resources”

Kazutaka MOTOYAMA, Yoshikazu TANAKA, Kento AIDA, Eisaku SAKANE,  
Kenichi MIURA..... 155



# 科学におけるグラフィックデザインの役割 ～宇宙から細胞まで～

佐藤 暁子 <sup>\*1,2</sup>

## Role of the Graphic Design in Science ～ From Space to Cells ～

Akiko Sato <sup>\*1,2</sup>

### Abstract

In order to progress research work and enhance research outcome, it is necessary to make research appealing to a broad audience. In order to achieve that, we need to highlight interesting aspects of each research topic. I propose that illustration gives a stronger impression than verbal description. Here, I demonstrate the importance of graphic design for research by presenting my recent work.

Keywords : Graphic Design, Visualize

### 概要

研究を発展させるためには、できるだけ多くの人々にその内容を理解してもらうこと、あるいは、それ以前に研究に興味を持ってもらうことが必要である。人に強い印象を与えられるものは、言葉ではなくまずビジュアルであろう。本論文では、科学研究をより発展させるために、研究内容を視覚で伝えていくグラフィックデザインの重要性を提起する。

## 1 はじめに

エンターテインメントや広告メディアなど、社会の多岐の分野に渡りグラフィックデザインが重要視されている一方で、科学分野においてグラフィックデザインはそれほど高く評価されていないように思われる。しかし、科学分野においても、研究を発展させるために、プレゼンテーションや論文に付加価値を付けるグラフィックデザインの力は欠かせない。当然、そのビジュアルも見難いものより、美しく分かりやすいものが求められる。どんなに立派な研究内容であっても、図解が拙劣であるとそれがマイナス要素となる可能性もある。しかし、研究者が美しいビジュアルを作成することは難しい。そこで、研究者自身ではなく、実際にデザインを学んできたものが、研究内容を理解した上で描いた図解や研究結果の総合的なイメージを、具体的な例を挙げて紹介する。

## 2 科学をビジュアル化する

ここでは、一つはブラックホールというマクロの世界のもの、一つはタンパク質というミクロ世界のもの、そしてアートにまで昇華させた科学分野の作例を紹介する。どれも我々人間の肉眼では見えないものであるが、人々の目に成り代わって研究内容を具現化し、ビジュアルとして提供するのがデザイナーの仕事である。その際に必要なことは、研究者が頭のなかでイメージしていることを噛み砕いて説明し、それをデザイナーが消化して自分の言葉で理解することである。何度も内容の確認や修正を繰り返すので、研究者とデザイナー双方のコミュニケーション力も重要である。また、デザイナーはプロとして、研究者の理想としているイメージ以上のものを描き上げるということを念頭に置く必要がある。

### 2.1 マクロの世界の表現 - ブラックホール

宇宙はどんな風に見えるのか？ 現在リアルな宇宙空間を一般人がこの目で確認することは難しい。そこで、宇宙研究者達が我々市民に、衛星から送られてくる情報やその後の解析、宇宙での新たな発見等を伝え、夢を与えてくれている。実際に撮影された美しい写真や迫力のある映像は驚きと感動を与えてくれ、このような魅力的なビジュアルは頭で考え

---

\*1 東京大学 生産技術研究所 (Institute of Industrial Science, the University of Tokyo)

\*2 JST ERATO 竹内バイオ融合プロジェクト (JST ERATO TAKEUCHI Biohybrid Innovation)

なくても目に入り興味をそそられる一方、素人には分からない難しい情報も沢山ある。その難しい情報を視覚化し理解の助けとなるのがグラフィックデザインである。今回はブラックホールに関する研究をビジュアル化した。

図 1-1 は、独立行政法人宇宙航空研究開発機構 (JAXA) 海老沢研教授らの発表した、ブラックホール周辺の X 線放射領域から放出される X 線の強度は時間変化していないが、その前をたくさんの物体が横切ることによって我々が観測する見かけの X 線強度が変化しているというモデルの図である<sup>2)</sup>。制作プロセスとして、まず研究者に描くべき図の説明を文字情報と簡単な絵で説明して貰う。今回は、絵を描いて欲しいと頼まれた段階ですでに研究者が描いた図があった (図 1-2)。これを元に描き始めるのであるが、宇宙研究の素人には、その業界の当たり前や言語が分からない。そのわからない言葉や理解できないことを質問し、ひとつひとつ丁寧に分かりやすいことばで答えて貰う。この初期の行程は大変重要なことである。なぜなら、研究を世に知らしめ、その後の研究の発展を期待するには、その業界の人達だけでなく、一般の人々にこそ知って貰う必要があるからである。ビジュアルを描くプロに絵を依頼する際、その者に対し、「科学知識を持っているかが問題」とする研究者の意見もあるが、その業界において知見のない素人が理解できれば一般の方達も理解できると考えて、一般人代表のパロメーターとして研究者は 制作者側に話をしてくれるとよいのではないだろうか。この言葉のやりとりを繰り返し、ビジュアルを作るのであるが、説明を聞いてすぐに絵が完成するわけではない。イラストという形にした段階で、研究者は伝えるべきことが正しく伝わっていない部分があることに気づき、再度要望や修正を伝える。それを繰り返すことによって研究者は内容をデザイナーに理解させ、デザイナーはより分かりやすい見せ方、美しい表現を研究者に提案しお互い納得の行く形にもっていく。ここでのやり取りはすべてメールで行った。この絵は描画ソフト Adobe Illustrator, 3D ソフトウェアの NewTek Light Wave 3D を使い、最終的に Photoshop で仕上げた。完成までのビジュアルの変化を図 1-3 に示す。

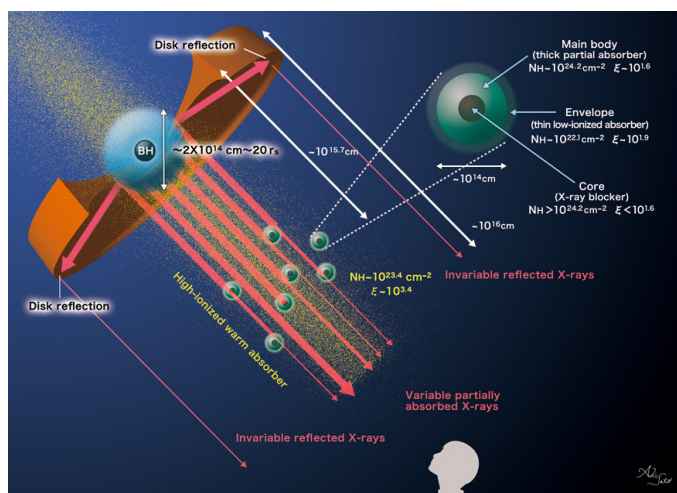


図 1-1 MCG\_6-30-15 のための部分吸収 (partial covering) モデル 2)

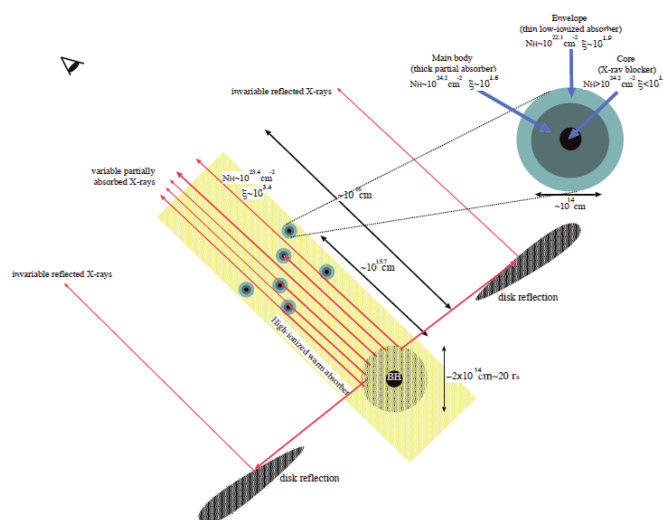


図 1-2 研究者が描いた研究概要を示す説明図 (JAXA 海老沢研教授 提供)

## 完成までの変化

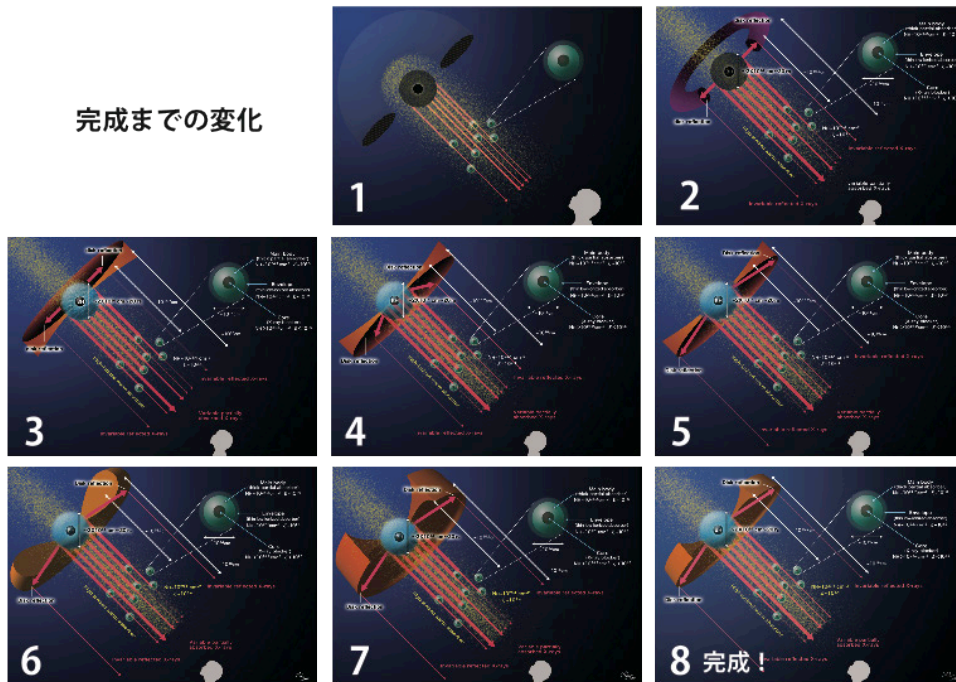
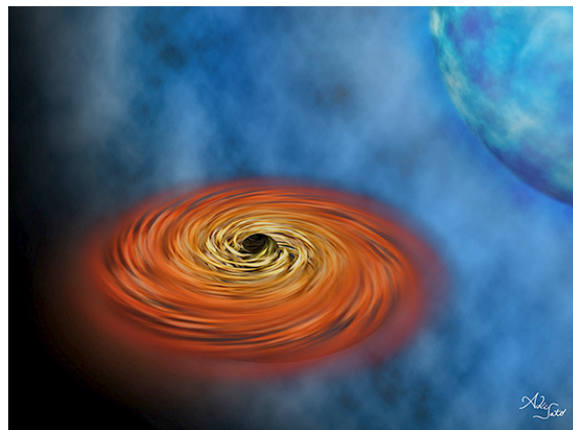


図 1-3 完成までのビジュアルの変化

1) 図 1-2: 研究者が描いた研究概要を示す説明図は、上から地面を観察している。天体を観察というと視線は下から上を見るイメージなのでブラックホール(BH)と視線の位置関係を反転させた。天体らしさを出すために背景は暗い色を選択。2) X線強度を示す矢印の強弱をつけるために矢印の太さ、細さで違いを出す。BHを取り囲む円盤の角度を変更。3) X線放射領域の色を中心のBHの色と区別させるために水色に変更。4) X線強度を示す矢印をX線放射領域の中心から出す。5) X線放射領域の中心からではなく放射領域の周りからX線が湧いて出て見えるように矢印の根元をぼかす。反射は円盤の端の方で起きているので円盤に向かう矢印を延ばす。6) 円盤を2Dで表現する限界を感じ3Dで制作。7) 円盤の形を3D作り直し。8) 円盤の形は問題なし。厚みを薄くした。

図2は、理研仁科加速器研究センター玉川高エネルギー宇宙物理研究室の山田真也 基礎科学特別研究員らを中心とした、10億度超の高温ガスを測定しブラックホール存在の証明に一步を刻んだ画期的な研究の成果で、「はくちょう座 X-1」ブラックホール連星を描いた想像図である<sup>3)</sup>。どのようにブラックホールにガスが引き込まれるのか、どんな色だと高温だということが伝わるか等、研究者とやりとりを続け、関係者も納得し、一般の人がその絵を見ることによってイメージを膨らませられるビジュアルを描いた。こちらは、宇宙空間を3Dソフトウェア NewTek Light Wave でモデリングし、そこに貼付けるテクスチャーを Adobe Photoshop で描き、最終的な合成も Adobe Photoshop を使用した。

図2 「はくちょう座 X-1」ブラックホール連星の想像図<sup>3)</sup>



## 2.2 ミクロの世界の表現 - タンパク質

身体の表面、皮膚や爪など目で見えるものもあるが、身体の内側は自分自身でも見えない。宇宙の表現と同様、体内や、顕微鏡でしか見えない微細なものを人に伝えるには、グラフィックデザインで見せることが有効である。

公益財団法人神奈川科学技術アカデミー (KAST) の「バイオマイクロシステム」プロジェクト (プロジェクトリーダー 竹内昌治) と (株) 森永生科学研究所、小山由利子研究員らは、簡易・高感度な免疫測定チップの開発に成功し、マイクロ化学、オンチップ化学に関する学会誌 “Lab on a Chip” 誌の表紙に掲載された<sup>4)</sup>。ここではその表紙デザインを紹介する。これも研究者から、研究内容、一番アピールしたい点などテレビ会議システムを使って話を聞き受け作り始めた。今回は研究の詳細な説明というより、雑誌の表紙ということで、人の目を引く迫力のあるデザインにしたかったので、チップ自体は画面に入れず、チップの内部でビーズに捉えられたタンパク質やガラスファイバーの隙間を流れていくタンパク質をダイナミックに構成した。図 3-1 は、必要なパーツを Light Wave 3D ソフトでモデリングし、角度を変えて画像をレンダリングしたものであり、これらを素材として使った。ガラスファイバーとタンパク質の置かれている奥行き間などを表現するために Photoshop で位置調整、手前にあるガラスファイバーの透き通った感じを表現するのに、不透明の設定で背景との透け具合を調整し、立体感をさらに出すために影を作った。ビーズに捉えられたタンパク質が複数あること、画面が切れた向こう側にもビーズに捉えられたタンパク質が存在しているということを表すために、素材を複製し、奥にあるビーズに捉えられたタンパク質は背景になじませると同時に、わざと個体全てが画面に収まらないようにレイアウトした。一方、手前にあるビーズに捉えられたタンパク質はきちんと見せるように全体を画面に収め、周りにあるものより目立つようにコントラストを強めた。流れるタンパク質の蛍光物質の光った感じを出すために鮮やかな黄緑色を追加し、流れに乗っているようにみせるため、ビーズに捉えられたタンパク質の方向にぼかしをかけた。以上のような行程を踏み完成させ、論文当該号表紙に採択されたものが図 3-2 である。

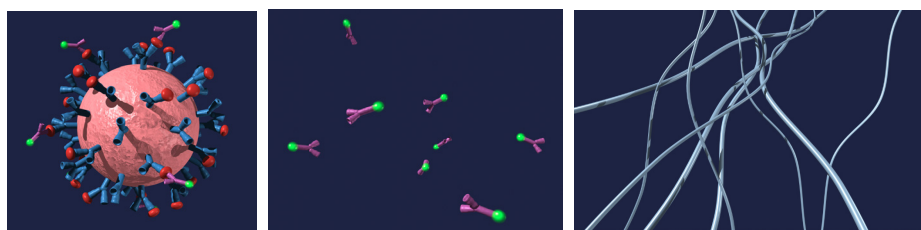


図 3-1 左から 3D で作った、ビーズに捉えられたタンパク質、流れるタンパク質、ガラスファイバー



図 3-2 “Lab on a Chip” 誌の表紙デザイン<sup>4)</sup>

## 2.3 アートへの展開

私が所属する ERATO 竹内バイオ融合プロジェクトでは、微細な加工・配置を得意とする MEMS 技術やマイクロ流体デバイス技術と組み合わせ、細胞をあたかもネジやバネ、歯車といった規格化された部品のように加工し、厚みを持った三次元組織を機械組み立てのように緻密かつ高速に構築することを目指している。将来は再生医療にも役立てたいと

いうこのプロジェクトの内容から、細胞でできた「部品」を組み合わせ、人工的な臓器、心臓を組み立てて作った。まず、細胞でできた「部品」ということで、テクスチャー素材として細胞のイラストを Photoshop で描く。これは色を変えたりサイズを変更したりすることで、沢山のパターンが用意できる。次に Light Wave 3D ソフトで、歯車やネジ、バネなど数種類をモデリングし、先ほどのテクスチャー素材をモデルに張込み、3D 上でのライティング、カメラの角度などの調整した後レンダリングする。これらをまさにパーツとして使い Photoshop で心臓の形に作り上げる。パーツを複製、サイズを変更し、色味を変え、様々な種類の部品があるようにし、心臓の形になるようにレイアウトした。心臓の鼓動のように、歯車とリンク機構で動き出すようなデザインにし、最終的に出来た形のバランスを見つつ全体の色味を調整した。完成した作品は、バイオテクノロジーと工学が融合したハイブリッドなシステムを表現し、プロジェクトが推進する新しい分野のモノ作りをビジュアルでアピールすることに成功した。この作品は、アジアのデジタルコンテンツを紹介する ASIAGRAPH2011 の CG アートギャラリー、2011Asia Digital Art Awards で共に入選した（図 4-1）。また、研究をわかりやすく紹介できると判断され、独立行政法人 科学技術振興機構（JST）発行の 2013 年 7 月号の『JSTnews』の表紙<sup>5)</sup>にも採用された（図 4-2）。



図 4-1 『機械仕掛けの生命体』  
ASIA GRAPH 2011 CG アートギャラリー 静止画部門 入選  
2011Asia Digital Art Awards 入選



図 4-2 『JSTnews』表紙<sup>5)</sup>

## まとめ

科学をビジュアル化する場合、制作の流れとして、1、研究内容を聞く。2、分からない部分を研究者に投げかける。3、返答を貰う。4、研究者の要望を聞く。5、どのような表現方法が適切か考え提案する。6、ラフなデザインを作り見せる。7、フィードバックをもらう。8、作り込み。9、完成という手順となる。これら科学分野のビジュアルを作るのに大事なことは、研究者とデザイナーのコミュニケーションであると考えている。お互い辛抱強く、「伝える」「理解する」を繰り返すことによって、科学として正しく、デザインとして美しい、そして、それを見る人が研究を理解できる助けとなるような、それぞれが納得する最終的な形が見えてくるのである。

科学分野におけるグラフィックデザイナーの活躍の場はまだ少ない。最先端科学技術の研究室に身を置き、研究員メンバーの研究成果をビジュアルとして作成している立場から、後回しに考えられがちなビジュアルの大切さを伝え、科学研究と人々を繋げ、研究に付加価値を与えるグラフィックデザインの重要性を世に訴え広めていきたい。

## 謝辞

今回発表する機会を与えてくださった宇宙科学研究所の海老沢研教授、国立大学になかったデザイナーの職を認めてくださった東京大学生産技術研究所の竹内昌治准教授、初の論文作成にあたりご指導くださった東京大学生産技術研究所の大崎寿久特任助教、英語訳のアドバイスをくださった東京大学生産技術研究所の Daniela Serien 氏、大事な研究にビジュアル制作という面で関わらせていただいた、独立行政法人理化学研究所 山田真也研究員、森永生科学研究所 小山由利子研究員に感謝の意を表する。

## 参考文献

- 1) 田中佐代子, 小林麻己人, 三輪佳宏, 科学者によるサイエンスイラストレーション作成の実体, 芸術研究報 32, 2011, pp.59-70.
- 2) Takehiro MIYAKAWA, Ken EBISAWA, Hajime INOUE, A Variable Partial Covering Model for the Seyfert 1 Galaxy MCG\_6-30-15, PASJ: Publ. Astron. Soc. Japan, vol.64, 2012, pp. 140-1 - 140-17.
- 3) [http://www.riken.jp/pr/press/2013/20130404\\_2/](http://www.riken.jp/pr/press/2013/20130404_2/)
- 4) Yuriko OYAMA, et al., A Glass fiber sheet-based electroosmotic lateral flow immunoassay for point-of-care testing, Lab on a Chip, vol.12, 2012, pp.5155-5159.
- 5) 独立行政法人 科学技術振興機構 (JST) 発行, JSTnews, 2013 年 7 月号。

# 「はやぶさ」の小惑星イトカワ探査軌道・姿勢の可視化

三浦 昭<sup>\*1</sup> 山本 幸生<sup>\*1</sup> 吉川 真<sup>\*1</sup>

## Visualization of Trajectories and attitudes of “Hayabusa” probing Itokawa

Akira MIURA<sup>1</sup>, Yukio YAMAMOTO<sup>1</sup>, Makoto YOSHIKAWA<sup>1</sup>

### Abstract

We have developed methods to visualize trajectories and attitudes of “Hayabusa” spacecraft probing around the asteroid Itokawa. We visualized trajectories and attitudes of “Hayabusa” using public data of those parameters. Then, around the touchdown phase, we corrected the positions of “Hayabusa” using the public data and engineering data, until we succeeded to enhance the accuracy of the trajectories in the visual of the rendezvous of “Hayabusa” and Itokawa.

### 概 要

筆者らは、探査機「はやぶさ」のイトカワ周辺における探査軌道及び姿勢の可視化について幾つかの手法を開発し、映像化への応用を試みた。筆者らはまず公開データに基づいて「はやぶさ」の位置・姿勢を再現した。続いてイトカワへの接近フェーズについては、公開データに加えて工学データを用いた位置補正を行った。その結果「はやぶさ」とイトカワのランデブー映像化において、位置精度の向上を実現した。

## 1. はじめに

### 1.1. 背景

小惑星探査機「はやぶさ」が小惑星イトカワの周辺でどのような軌跡を辿ったかについては、様々な文献やデータとして公開されており、プレスリリース等でも逐次広報されてきた。また「はやぶさ」の映画や映像作品も数多く制作され、リアリティのある「はやぶさ」の映像が生み出されてきた。しかしながら、体系的に「はやぶさ」の軌跡を実データに基づいた3次元CGで再現することは行われて来なかった。

研究者は数値データや2次元のグラフを読み解くことで「はやぶさ」の位置・姿勢を理解することができるため、3D表現は必ずしも研究の本質ではなかった。「はやぶさ」の位置・姿勢データが公開されているとは言え、それを活用してリアリティのあるCGを合成するためには、ある程度の専門知識が必要となり、広く一般に「はやぶさ」の軌跡を容易にCG化できる状態にもなっていない。

### 1.2. 目的

このような状況に鑑み、本研究においては可視化の目的を以下のように定める。

#### 1.2.1. 記録映像としての可視化

「はやぶさ」の位置・姿勢データ等は、後述のSPICEカーネル等の形式で、だれでも利用可能な状態で公開されている。過去のデータの保存・公開という観点ではこれで十分とも考えられるが、これらのデータを参照して、実際に「はやぶさ」の軌跡を理解できるのは、その分野の研究者や軌道計算に詳しい者等に限定されてしまう。この枠を越えて「はやぶさ」の軌跡についての理解を深めるためには、可視化は強力な手段であり、後続の類似ミッション等の検討材料として活用するためにも、映像として「はやぶさ」の軌跡を記録することは意義の大きいものである。したがって、可視化の一つの目的は記録映像として保存し、後世の役に立てることとする。

<sup>\*1</sup> 宇宙航空研究開発機構 (Japan Aerospace Exploration Agency)

### 1.2.2. 広報・アウトリーチとしての可視化

「はやぶさ」帰還前後から、「はやぶさ」をモチーフとした映画等の映像作品が幾つか制作されており、2013 年は「はやぶさ」10 周年ということもあり、また新たな映像も制作され、イトカワの微粒子も公開される等、「はやぶさ」に関連するアウトリーチ活動は継続している。これまで広く支持されてきた「はやぶさ」の軌跡を、単に記録するのみでなく、公にフィードバックして行くこともまた重要なアウトリーチ活動であると考えられる。広く一般に訴求するにあたっては、「はやぶさ」やイトカワがある程度のリアリティを以て描かれることが重要であると考えられる。したがって、もう一つの目的は広報・アウトリーチ活動のための、リアリティを伴った映像化とする。

### 1.3. 可視化の概要

前述のような観点から、可視化の概要を述べる。

本稿において主要な構成要素となるのは、「はやぶさ」と小惑星イトカワである。「はやぶさ」とイトカワの相対位置・姿勢が時々刻々と変化する様子を一目で分かるように可視化することが基本となる。可視化の際に配置するオブジェクトも、現存する「はやぶさ」とイトカワの形状データに基づいてモデリングするものとする。またイトカワや「はやぶさ」の陰影を再現するために、当時の位置データに基づいて計算された太陽方向に光源を置くものとする。

「はやぶさ」の軌跡を直感的に把握するために、実際に「はやぶさ」がとった軌跡を、実際の時間に比例する再生速度で可視化することとする。現実には動きの多いシーンや少ないシーン等で再生速度の倍率は調整する必要があるが、同一シーン内での倍率は同一とする。

可視化にあたっての各オブジェクトの配置は、原則として現存する公開データ（位置、姿勢等）に基づくものとする。ただし公開データに補正をかけることで精度を向上できると判断される場合は、適切な補正を検討する。

可視化にあたっては、視点（仮想的なカメラの位置）や注視点（仮想的なカメラの視野中心に位置する箇所）を工夫し、適宜補助的な可視化要素を配置することによって、距離感覚や方向感覚を補うものとする。

## 2. 使用するデータ

以下に、可視化に使用するデータについて述べる。

### 2.1. オブジェクトの形状データ等

可視化にあたって使用するオブジェクトは以下の通りである。主要なオブジェクトはイトカワと「はやぶさ」

であるが、両者の相対的な位置関係が視覚的に把握できるように、背景に恒星を配置することとする。

#### 2.1.1. イトカワの形状データ

イトカワの形状は、宇宙航空研究開発機構（JAXA）宇宙科学研究所（ISAS）が運営する Data ARchive and Transmission System（DARTS）で提供されている Gaskell 形状モデルのうち、49,152 面モデル（itokawa\_f0049152.tri）と 3,145,728 面モデル（itokawa\_f3145728.tri）を用いる。前者は遠方にあるイトカワの描画や補助的な可視化要素の描画に用いる。後者は近傍にあるイトカワの描画に用いる。具体的なモデルの使い分けにあたっては、両者の描画結果を視覚的に比較し、後者の高精細なモデルを用いた効果が識別できないと判断されるケースでは処理速度を優先して前者のモデルを用いる。

#### 2.1.2. 「はやぶさ」の形状データ

「はやぶさ」の形状は、「M-V-5/MUSES-C 飛翔実験計画書（衛星編）」<sup>[1]</sup>の図面、太陽パネルのデザイン<sup>[2]</sup>、公開されているフライトモデルの写真等に基づいてモデリングする。各パーツの色や材質は、公開されているフライトモデルの写真を参考としながらも、描画結果が見やすくなるよう、適宜調整を加える。「はやぶさ」の形状データのモデリングにあたっては、3D ソフトである Shade<sup>2</sup>を用いる。

#### 2.1.3. 恒星のデータ

「はやぶさ」とイトカワのみ描画したのでは、相対的な位置関係がどのように変化しているのか不明確となる。今回の可視化においては、注視点をイトカワとの相対位置で決定する。視点を「はやぶさ」との相対位置で決定した場合、何らかの視覚的な補助が無いと、「はやぶさ」とイトカワの相対位置の変化を把握するのが困難となる。そこで本研究では、距離以外の相対的な位置関係を把握できるようにするために、背景に恒星を描画することとする。恒星の位置情報や色情報は、HIP-<sup>3</sup>PARCOS 星表から、7 等級より明るい天体（整数部分が 6 等級より明るい天体）を採用する。位置は J2000 分点で記述する。実際に描画する際の恒星の明るさや大きさは科学的な考察に基づいたものではなく、描画結果を見比べて、「はやぶさ」やイトカワに対して十分暗く小さく、かつ、「はやぶさ」とイトカワの相対位置の変化が背景の恒星の動きで認識できる程度に明るくなるように

1 [http://darts.jaxa.jp/planet/project/hayabusa/shape\\_ja.pl](http://darts.jaxa.jp/planet/project/hayabusa/shape_ja.pl)

2 <http://shade.e-frontier.co.jp/>

3 <http://www.rssd.esa.int/index.php?project=HIPPARCOS>



設定する。ヒッパルコス星表において、色は B (Blue)、V (Visual)、I (Infrared) で表現されており、一般に CG で用いられる R (Red)、G (Green) の比率を導き出すことはできない。今回の可視化では  $R = (2V+I)/3$ 、 $G=V$  と変換し、尤もらしい描画結果になるよう調整する。これはあくまでも恒星が位置関係把握の補助であるための措置であり、科学的な考察に基づいて色を決定するものではない。

## 2.2. 位置・姿勢データ等

各オブジェクトの位置・姿勢等は、SPICE カーネル形式で公開されている「はやぶさ」やイトカワの時系列データを用いて再現する。

SPICE カーネルは、様々な宇宙科学に関連する補助データを取り扱うための枠組みであり、NASA の Navigation Ancillary Information Facility (NAIF)<sup>4</sup> によって維持管理されている。NAIF は SPICE カーネルと共に、それらにアクセスするための、SPICE ツールキット<sup>5</sup> と呼ばれるライブラリを提供している。SPICE カーネルは、データの種類毎に幾つかのフォーマットが存在する。以下に、使用した主要なデータについて述べる。下記以外にも、座標系等を定義する各種カーネルを使用している。

### 2.2.1. イトカワ

小惑星イトカワの位置データは、itokawa\_1989\_2010.bsp を用いた。これにはイトカワの太陽に対する相対位置が記録されている。実際の可視化にあたっては、イトカワを座標中心に置いて各オブジェクトを配置するため、イトカワの位置データは、イトカワに対する太陽の相対位置を計算するのに用いる。

イトカワの自転に関するデータは、作成されたイトカワのモデルの種類毎に複数作成されている。今回は、前述の Gaskell 形状モデルとの整合をとるために、itokawa\_gaskell\_n3.tpc を用いる。

### 2.2.2. 「はやぶさ」

「はやぶさ」の位置データは、原データである hayabusa\_itokawarendezvous\_v01.bsp と、これに AMICA や LIDAR 等、他の理学データに基づいた補正が施された hay\_osbj\_050911\_051118\_v1n.bsp とを用いる。いずれも、イトカワに対する「はやぶさ」の相対位置が記録されているが、後者は前者に比べて精度が高い代わりにデータが格納されている期間が限られている。

「はやぶさ」の姿勢データは、hayabusa\_itokawarendez-

vous\_v02n.bc を用いた。

## 2.2.3. 可視分光撮像カメラ

可視分光撮像カメラ (Asteroid Multiband Imaging Camera; AMICA) とは、「はやぶさ」搭載の望遠カメラである。同カメラは航法で用いる際には、望遠型光学航法カメラ (Optical Navigation Camera-Tele; ONC-T) と呼ばれている。AMICA で撮像された画像は数多く公開されており、本研究では可視化結果の妥当性確認と、「はやぶさ」軌跡映像化の際の補助映像として、AMICA の視野に基づいた可視化を行う。

「はやぶさ」実機における AMICA の取り付け位置・角度を正確に再現するために、「はやぶさ」に対する AMICA のアラインメントを記したファイル amica31.ti を用いる。

## 3. 可視化手法

以下に、前節の各種データを用いて可視化する手順を示す。

### 3.1. 可視化環境

可視化に用いるレンダリングソフトは POV-Ray<sup>6</sup> である。POV-Ray は、光源、カメラ、各オブジェクト等の配置やテクスチャ等のパラメータを全てスクリプトで定義できる 3D レンダリングソフトウェアである。POV-Ray 自体にモデリングの機能は無いが、数値計算やループ処理等のスクリプト機能で形状を合成することが容易であり、科学データ等の数値群で表された形状をバッチ処理でレンダリングするには適したツールである。今回使用するバージョンは、Linux 版の 3.7.0.RC7 である。OS 環境は Cent OS や Yellow Dog Linux, CPU は Intel Core i7 や Xeon、Opteron 等の混在であるが POV-Ray のバージョンさえ一致していれば、環境によらない均質な描画を得ることができる。

### 3.2. 各形状データの変換

2.1 節で用意した各形状データは、可視化に先立って POV-Ray 形式のオブジェクトに変換する。

イトカワの形状データ (itokawa\_f3145728.tri 等) は三角形の集合で表現されており、簡単なバッチ処理で POV-Ray のメッシュオブジェクトに変換できる。

Shade で作成した「はやぶさ」のモデルは、一旦 LightWave 3D 形式に変換した上で、PoseRay<sup>7</sup> を用いて

4 <http://naif.jpl.nasa.gov/naif/>

5 <http://naif.jpl.nasa.gov/naif/toolkit.html>

6 <http://www.povray.org/>

7 <https://sites.google.com/site/poseray/>

POV-Ray のオブジェクトに変換する。各パーツのテクスチャは Shade と POV-Ray とで表現が異なるため、尤もらしい表面材質となるように、手作業で微調整を施す。

恒星のデータは、HIPPARCOS 星表から抽出した座標値、等級、色情報を POV-Ray 形式の配列に変換した上で、POV-Ray のスクリプト機能を用いて、個々の天体に対応した球体オブジェクトを生成する。

### 3.3. 時系列データの生成（位置・姿勢）

2.1 で述べた SPICE カーネルの内、時間変動のある位置・姿勢データについては、時系列データを POV-Ray 形式の配列に格納する。用途によっては配列長が 1 の場合もありうる。

この作業にあたっては、SPICE ツールキットを用いる。SPICE ツールキットは、対象となる各 SPICE カーネル群に対してデータを取得する時刻を指定することで、対応する時刻における位置データや姿勢データを返す関数が提供されている。POV-Ray 形式の配列に格納するにあたって、位置はイトカワ中心を原点として J2000 分点で表現する。姿勢は POV-Ray の座標変換機能に則って、オイラー角で表現する。

### 3.4. オブジェクトの配置と描画

POV-Ray のスクリプトで各オブジェクトの配置を記述する。配置する基本的なオブジェクトは以下の通りである。配置する空間の座標はイトカワ中心を原点とした J2000 分点で表す。

イトカワのオブジェクトは中心を原点に置いて、指定時刻の姿勢に基づいた回転を与える。

「はやぶさ」のオブジェクトは指定時刻の姿勢に基づいた回転を与えた上で、指定時刻の位置に基づいた平行移動を行う。

恒星のオブジェクト群は天球上に固定する。

太陽の位置に光源を置き、太陽からイトカワ中心への平行光を定義する。光源の明るさは、「はやぶさ」やイトカワが適切な色調となるように調整する。

可視化にあたっての視点、注視点、視野角等（カメラのパラメータ）は、用途毎に定める。

この他、各種描画条件等を定義した後、バッチ処理で POV-Ray を起動し、配列に格納された各時刻毎の静止画を描画する。実際の描画にあたっては、マルチ CPU 環境下のパラメータ並列で実効速度の向上を図る。

連続した静止画から動画像を生成するにあたっては商用の映像編集ソフトを用いる。

## 4. 妥当性の検証

今回用いる各種データや手法が正しく可視化に供されるか確認するために、AMICA が実際に観測したイトカワの画像や既存の視野角ビジュアライザと、本手法（2 節のデータ群と 3 節の可視化手法の組み合わせ）での描画結果とを比較した。

視野角ビジュアライザ (FieLd Of View Visualizer<sup>8</sup>) とは、SPICE で提供される各種宇宙機の位置・姿勢情報等に基づいて、観測機器の視野を再現するツールである。視野角ビジュアライザの源泉となる SPICE カーネルは、本手法と同一のファイルを用いることとした。

描画にあたっては、3.3 節の手法で AMICA が実際にイトカワを観測した時刻の位置・姿勢データを取得し、2.2.3 節のアラインメントに基づいて AMICA を模擬したカメラのパラメータ（視点、視線方向、視野角）を定めた。「はやぶさ」はカメラの視野外になるが、「はやぶさ」の影がイトカワに落ちる場合も想定して、「はやぶさ」のオブジェクトも配置する。「はやぶさ」の位置データは、2.2.2 節の hay\_osbj\_050911\_051118\_v1n.bsp を用いた。

3 者の比較結果を図 1 に示す。3 色合成の内、本手法による描画結果（赤）と視野角ビジュアライザの描画結果（青）とは、イトカワの輪郭・背景の恒星共に、画像サイズに対して 0.1% 以下の精度で一致している。画像上端で青が強く出ているのは視野角ビジュアライザの描画において影（shadow）判定がされていないためであり、形状に関する本質的な相違ではない。したがって本手法における SPICE カーネルデータの取り扱いや、各オ

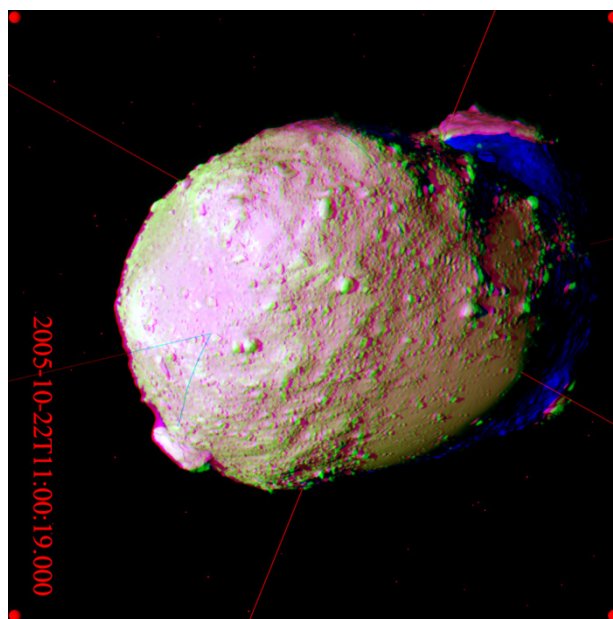


図 1 描画結果比較

8 <http://darts.jaxa.jp/planet/tools/flow/>

プロジェクトの座標変換は適切になされていると考えられる。

本手法による描画結果（赤）と AIMCA の実際の撮影画像（緑）とを比較すると、画像サイズの 0.7% 程度の誤差で一致している。画像左端で赤が強く出ているのは、AMICA のフィルタの特性によるものであり、形状に関する本質的な相違ではない。

他の撮影画像においても幾つか比較を行ったが、hay\_osbj\_050911\_051118\_v1n.bsp に格納されたデータは 1% 程度の精度で 3 者が一致しており、可視化に用いるにあたっては十分な精度であると考えられる。

## 5. ONC-W を用いた位置補正

hay\_osbj\_050911\_051118\_v1n.bsp で提供されている期間以外の位置データについては、広角光学航行カメラ（Optical Navigation Camera-Wide; ONC-W）を用いた補正を試みた。「はやぶさ」がイトカワに接近した状態では、AMICA の画像はイトカワの一部しか撮像できず、位置情報の補正に用いるためにはイトカワ上の特徴点を特定する等の高度な画像処理が必要となる。

この度は可視化の目的の範囲内で十分な位置精度を得ることが目的であるため、接近時にもイトカワを広く撮像しており照合が容易と思われる ONC-W のデータを用いて、2005 年 11 月 19 日のイトカワ降下からタッチダウン 30 分前（20:33 UT）までの位置補正を試みた。この時刻以降は「はやぶさ」の位置データが大きく変化しており、信頼性を判断するのが困難であったため、位置補正の対象外とした。

この時期の ONC-W の画像には、イトカワと共にイトカワ表面に落ちた「はやぶさ」の影が捉えられている。この時期は「はやぶさ」の位置データを補正するにあたって、イトカワの輪郭とイトカワ上の「はやぶさ」の影を用いることができる。高度な画像処理を行えば他の特徴点も位置補正に利用可能であると考えられるが、この度の可視化にあたっては割愛した。位置補正にあたって用いた手法は以下の通りである。各手順の中で生成する ONC-W の模擬画像は、4 節と同様の手法で ONC-W を模擬したカメラのパラメータを定め、3 節の手法を用いて生成する。なお「はやぶさ」の姿勢データはスタートラッカにより生成されており信頼性が高いと考えられるため、以下の手順の中で「はやぶさ」の姿勢データは正しいものとして扱う。

- (a) ONC-W の実画像と模擬画像がおおまかに一致するように位置データを修正する。
- (b) 両画像においてイトカワに対する「はやぶさ」の影

の相対位置が一致するように、「はやぶさ」の位置を修正する。

- (c) 両画像においてイトカワの輪郭が一致するように、「はやぶさ」の位置及び ONC-W のアラインメントを修正する。なお ONC-W のアラインメントは、幾つかの画像で照合した結果を元に固定するものとする。
- (d) 必要に応じて (b) から (c) を繰り返す。

以上の手順を用いた位置補正前後の画像を図 2 に示す。AMICA の実画像と模擬画像とは並行移動のみで歪無く一致させることができたが、ONC-W の実画像は、模擬画像と比較すると、光学系が原因と思われる歪が見受けられる。さらに高精度に位置補正するためには、光

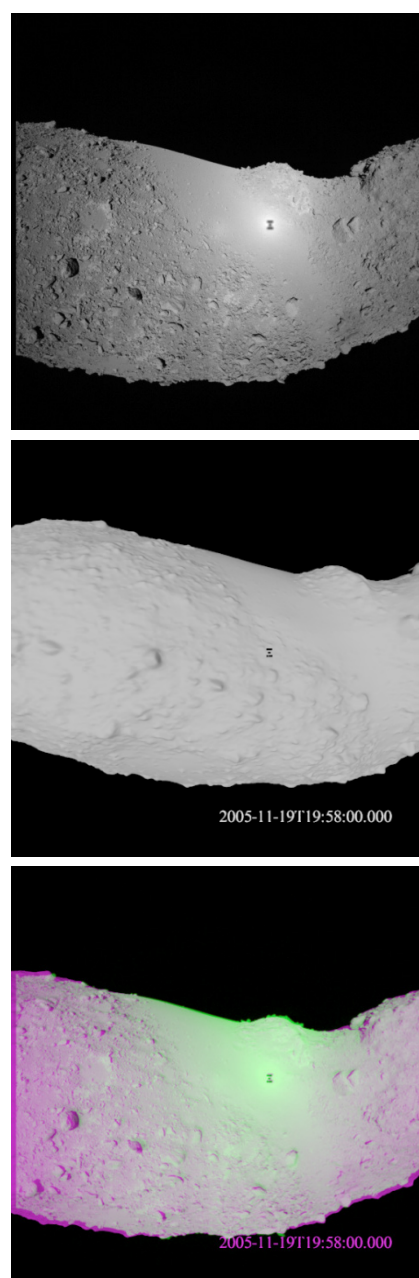


図 2 ONC-W 画像を用いた位置補正（画像）  
上：実画像、中：補正前、  
下：補正後（緑＝実画像、赤＝模擬画像）



学系の特性を考慮した照合が必要と考えられる。しかしながらこの相違は、視覚的には両者を重ね合わせてようやく視認できる程度の相違である。

係る期間の ONC-W 実画像を用いた補正の比較を図 3 に示す。黄色の実線が補正前の位置データに基づいた軌道であり、その脇の黄色の点々が、ONC-W の実画像が存在する各時点における、補正後の位置である。補正前の位置データによると、「はやぶさ」はイトカワに衝突している状態であったが、補正後のデータはイトカワ近傍への接近を破綻なく再現している。それぞれの時点における模擬画像も図 2 に示すのと同程度の誤差で実画像と一致している。補正データの科学的利用には課題が残るところであるが、軌跡の可視化にあたっては十分なレベルの補正がなされていると考えられる。

この度の照合作業は手作業で実施したが、今後は照合作業を自動化し、ONC-W の光学系をも考慮した補正手法を確立すると共に、LIDAR 等の他の観測機器のデータも考慮した精度向上が課題である。また ONC-W の特性が分かれば、「はやぶさ」の影がイトカワに落ちていない期間においても「はやぶさ」とイトカワの相対位置を推定する有効な手段となる。

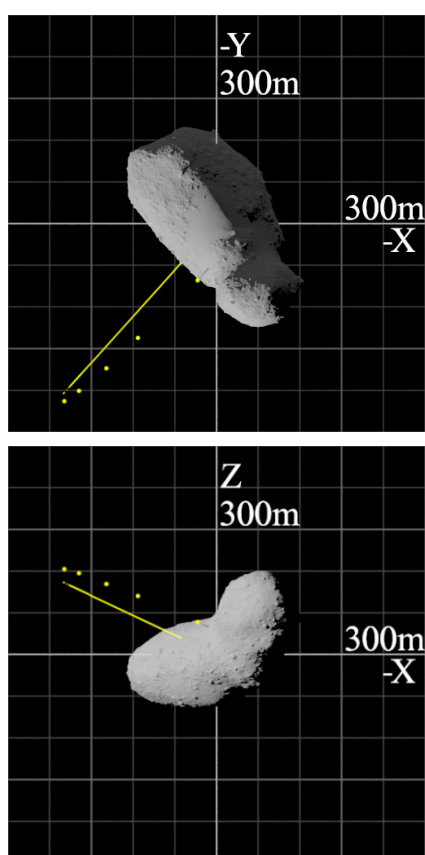


図 3 ONC-W 画像を用いた位置補正（軌跡）  
黄色の実線：SPICE の位置データに基づいた軌跡、  
黄色の点：補正された位置

## 6. 軌跡の可視化

### 6.1. 主画面の描画と課題

まず「はやぶさ」の軌跡を可視化する際の主画面について検討する。3 節の手法で「はやぶさ」とイトカワを可視化した結果を図 4 に示す。この描画に用いたデータは 2005 年 9 月 30 日のものであり、水平方向の画角は 90 度に設定している。35mm フィルムカメラに換算すると 18mm 程度の超広角となるが、これは「はやぶさ」がイトカワに接近したシーンにおいてもイトカワを画角内に納めるためである。超広角とは言え、視野内に「はやぶさ」とイトカワを、両者が同時に視認可能な状態で描画できるカメラワークは限られてくる。イトカワ周辺において「はやぶさ」は、太陽電池パネルの向きや、当時の「はやぶさ」から見て太陽の反対側にあった地球との交信等を考慮して、概ね太陽—「はやぶさ」—イトカワの並びになるように軌道を描いている。したがって仮想カメラの配置としては、「はやぶさ」の太陽側から、「はやぶさ」とイトカワが並ぶように見る位置が基本となる。

図 4 は、上記のような条件で「はやぶさ」とイトカワの位置関係や姿勢を当時のデータに忠実に再現したものであるが、直感的に「はやぶさ」とイトカワの位置関係を把握することは困難となっている。原因の一つには、「はやぶさ」がイトカワに対して 8km 近く離れているにもかかわらず、「はやぶさ」とイトカワの間に大きさや距離を推測する補助になるような物体が存在しないことが考えられる。背景に描いた恒星も、動画像で相対位置の変化を認識する補助にはなるが、星図を記憶している人でない限りは、静止画像における相対位置を認識することは困難である。

またこの距離ではイトカワは小さく映っており、詳細は視認できない。描画時に視野が狭くなるように調整すれば、この時期の位置関係でもイトカワを大きく描くことは可能であるが、その視野を維持しようとすると、「はやぶさ」がイトカワに接近した際に描画されるイトカワが大きくなりすぎるため、時期によって視野を調整する等の操作が必要となる。その結果として、描画結果全体

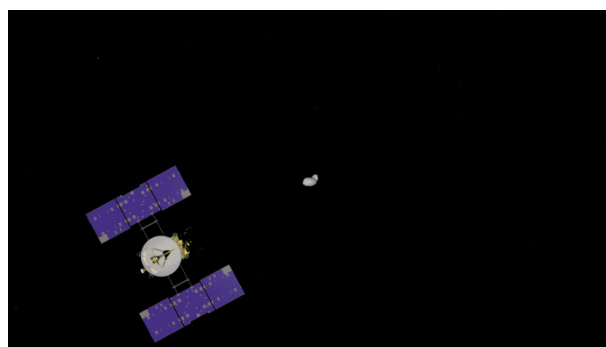


図 4 主画面の可視化結果

を通して見た際に、イトカワの見た目の大きさは距離を把握する判断材料とはならなくなる。

## 6.2. 補助的な可視化

このような問題を解決するために、距離や位置関係を把握する補助となる視覚要素を別途描画し、映像化の段階で合成することとする。またイトカワの詳細を把握できるようにするために、4節で用いた AMICA の再現画像を適宜挿入することとする。補助的な要素を含めた合成映像を図 5 に示す。この図においては以下に記す視覚要素を合成しているが、実際の映像制作においては、各時点において必要な視覚要素を取捨する。

### 6.2.1. AMICA の再現画像

4 節に示したような AMICA の再現画像を補助画像として用意する。AMICA の実画像は動画を生成できる程には存在しないため、今回は全ての AMICA 画像を、CG で描画する。図 5 において AMICA の画像が 90° 回転し

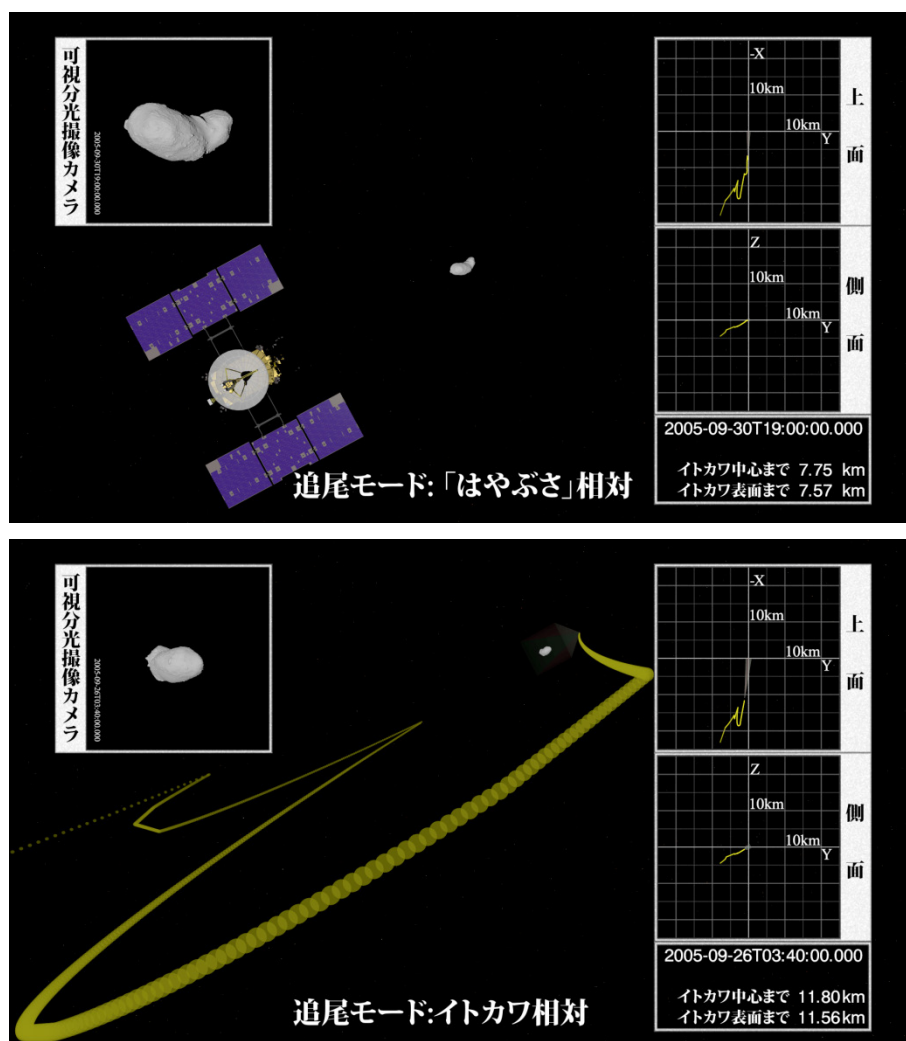
ているのは、主画面で描画されているイトカワと向きを揃えるためである。他のシーンにおいても、「はやぶさ」の姿勢によって AMICA の画像の向きは適宜調整する。

### 6.2.2. AMICA の視野の可視化

描画対象となっている時刻の AMICA の視野を四角錐で描く。AMICA は「はやぶさ」の底面に固定されており、常に底面方向を観測している。AMICA の視野を描くことにより、その時の「はやぶさ」の姿勢を視認することが可能となると共に、AMICA の再現画像が描いている範囲を把握することも容易となる。

### 6.2.3. 軌跡（履歴）の可視化

「はやぶさ」の位置を時系列で重ね描きすることで、辿った経路を可視化する。その際、「はやぶさ」のオブジェクトを重ねるのではなく、「はやぶさ」の大きさに比例した球体を一定時間間隔で描き、位置・速度の履歴を把握しやすいよう考慮する。



#### 6.2.4. 位置関係を表す画面

位置関係を表す画面は、上面図（Z 軸方向から原点を見る図）と側面図（X 軸もしくは Y 軸から原点を見る図）を用意する。夫々、イトカワを中心とした J2000 分点の座標上に「はやぶさ」の軌跡（履歴）とイトカワを描く。各図の軸は、主画面に描画された位置関係に準じて、「はやぶさ」が画面手前（上面図においては、画面下）に位置するように選択し、各軸には目安となる距離を記す。

#### 6.2.5. 時刻・距離を表す画面

時刻や距離の情報を数値として描画する。距離は位置関係を表す画面からも推測可能ではあるが、表示できる距離情報としては、レーザ高度計で計ったイトカワまでの距離も公開されており、位置情報から計算した距離と共にレーザ高度計の値も併記することとする。

### 6.3. 天周映像としての可視化

全天周映像は 6.1 において制約となっていた画角とカメラワークの問題を解決する有効な手段である。平面映像においては画角を無理に広くすると画像周囲の歪が大きくなり、移動するオブジェクトを描く際にはフレームアウトしないように注意を払う必要もあるが、全天周映像においては、少なくともドーム中心から観る限りは描画位置による歪を気にする必要は無く、フレームアウトの制限も緩やかなものとなる。

また広報・アウトリーチの手段としても、全国に展開されているプラネタリウムもしくは全天周シアターに提供できる素材を制作できるというメリットがある。

全天周映像を制作するにあたっては、幾つかのフォーマットがあるが、本稿ではドームマスター形式の描画について述べる。ドームマスターのフォーマットでは、正方形に内接する円に半球を描画する。円の中心を原点とする円座標系を用いて、ドーム中心（仰角  $90^\circ$ ）を原点に据える。ドーム上の方位、仰角は円座標系の偏角、動径にマップする。動径は  $90^\circ - \text{仰角}$  に比例し、円の端が仰角  $0^\circ$  になる。

CG ソフト用にドームマスター形式の描画をするプラグインも存在するが、筆者らは回転楕円体の鏡面と平行投影を併用することにより、視覚上問題のない精度でドームマスター形式の映像を描画した。この手法は反射回数が増えるためレンダリングの負荷としては若干不利であり、描画結果が鏡面反転した映像となるといった癖もあるが、ドームマスター用のプラグインが存在しない CG ソフトにおいても容易にドームマスター形式の描画が可能となるメリットがある。

主画面の描画にあたっては、3 節の手法におけるカメ

ラの位置に、平行投影のカメラと回転楕円体の鏡面を配置することで事足りる。回転楕円体のサイズによっては、画像中に同物体の影等が映りこむことがあるので、描画結果を見て適宜調整する。

補助的な視覚要素については注意が必要である。平面映像では別途描画した視覚要素を市販の映像編集ソフト等で 2 次元面上に適宜配置することが可能であったが、ドームマスター形式の描画においては、補助的な視覚要素もドームマスター形式で想定される 3 次元空間上にオブジェクトとして配置して描画する必要がある。

全天周映像の描画例を図 6 に示す。平面映像ではどこに配置しても同じ形状であった正方形の視覚要素が、配置する位置によって異なる形状に変換されている。

## 7. まとめ

「はやぶさ」の軌跡の可視化手法について述べた。SPICE カーネルで公開されている「はやぶさ」の位置・姿勢データの内、AMICA や LIDAR 等の観測機器のデータ等を用いて補正された位置データについては、そのまま可視化するに足る精度であり、画角を考慮し視覚の補助となる画像を併せて合成することで、実用的な軌跡を映像化することが可能である。「はやぶさ」がイトカワに接近したフェーズでは、ONC-W の実画像と照合することで、「はやぶさ」の位置データを映像化に適した精度まで補正することができた。映像化にあたっては、平面映像のみならず全天周映像の描画も試み、自由度の高い映像を合成できることを確認した。

今後の課題は、以下の通りである。

- (a) 位置データの補正：本稿では ONC-W を用いた位置補正は、「はやぶさ」がイトカワに接近したフェーズでの手動補正のみ試みたが、補正の自動化や、他のフェーズでの補正も考慮する必要がある。ONC-W の光学系の歪を考慮した補正や、他の観測機器のデータを併用した補正も今後の課題である。
- (b) リアルタイムの描画：本稿においては Shade や POV-Ray を用いてリアリティの高いモデリングや描画を試みたが、一方で位置データの補正等に用いる模擬画像はリアルタイム性の高い描画手法が適している。GPU を用いたハードウェアレンダリングでは 300 万ポリゴン程度の描画はリアルタイムに実現可能であり、十分検討の余地がある。また GUI ベースでイトカワ周辺の「はやぶさ」の軌跡を可視化できれば、今後の類似の探査ミッションの計画立案や探査中のデータ可視化等への応用も可能となる。

## 参考文献

Proceedings of the Sixth European Conference.

1. M-V-5/MUSES-C 飛翔実験計画書（衛星編）：宇宙科学研究所，2003.
2. Takahashi, K., Okada, Y., Arai, H., Tajima, M., and Higuchi, K. MUSES-C Solar Array Electrical and Mechanical Design.: European Space Agency, 2002. Space Power,

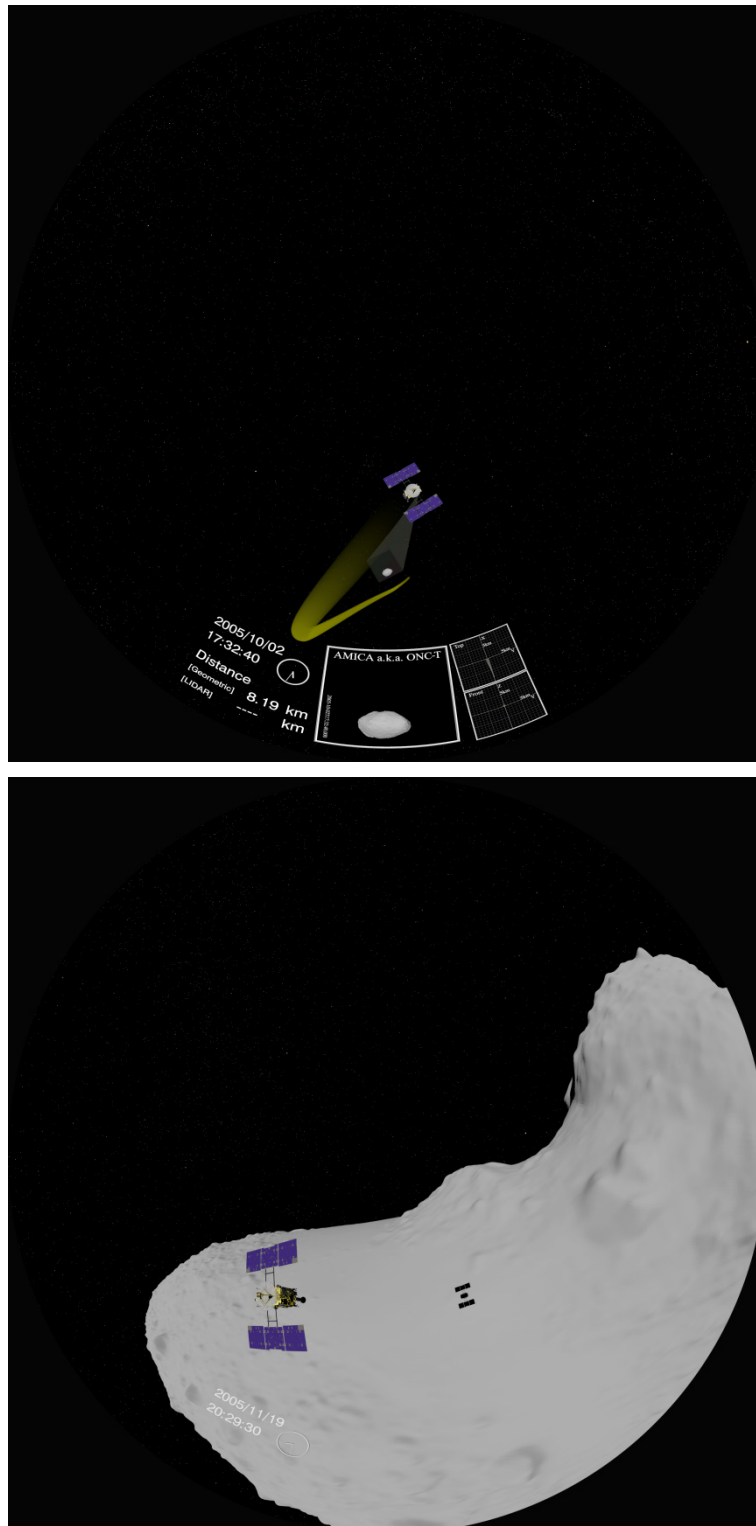


図 6 ドームマスター形式の描画例



# SIB2/GSTOS-1 における開発状況

西村佳代子<sup>\*1</sup> 松崎恵一<sup>\*2</sup> 宮澤秀幸<sup>\*1</sup> 高木亮治<sup>\*2</sup> 山下美和子<sup>\*3</sup>  
宮野喜和<sup>\*1</sup> 福田盛介<sup>\*4</sup> 馬場肇<sup>\*5</sup> 永松弘行<sup>\*1</sup> 山田隆弘<sup>\*4</sup>

## Development status of SIB2/GSTOS-1

Kayoko NISHIMURA<sup>\*1</sup>, Keiichi MATSUZAKI<sup>\*2</sup>, Hideyuki MIYAZAWA<sup>\*1</sup>, Ryoji TAKAKI<sup>\*2</sup>,  
Miwako YAMASHITA<sup>\*3</sup>, Yoshikazu MIYANO<sup>\*1</sup>, Seisuke FUKUDA<sup>\*4</sup>, Hajime BABA<sup>\*5</sup>, Hiroyuki NAGAMATSU<sup>\*1</sup>,  
Takahiro YAMADA<sup>\*4</sup>

### Abstract

GSTOS is Generic Spacecraft Test and Operations Software applied for test and operation system of future ISAS (Institute of Space and Aeronautical Science) spacecraft based on database SIB2 (Spacecraft Information Base version 2). SIB2/GSTOS-1 is a project to provide SIB2/GSTOS for SPRINT-A, ASTRO-H, Bepi/MMO. This paper describes goal, achievement relative to typical test and operation system for former ISAS spacecraft, development status and tasks for the future of SIB2/GSTOS-1.

### 概要

GSTOS (Generic Spacecraft Test and Operations Software; 汎用衛星試験運用ソフトウェア) は、ISAS (Institute of Space and Aeronautical Science) の今後の衛星の試験と運用に使用される汎用のソフトウェアであり、SIB2 (Spacecraft Information Base version 2) に基づき動作する。SIB2/GSTOS-1 プロジェクトは、SPRINT-A, ASTRO-H, Bepi/MMO 向けの SIB2/GSTOS を開発するプロジェクトである。本論文では、SIB2/GSTOS-1 プロジェクトが目指すゴール、従来の ISAS における典型的な衛星試験・運用システムに対する改善点を述べるとともに、現在までの開発状況、今後の課題について述べる。

Keywords: functional test, spacecraft operation, SIB2, GSTOS

### 1. はじめに

ISAS の衛星開発・衛星運用で使用されている現行の衛星試験・運用システムは<sup>1)2)3)4)</sup>、のぞみ衛星と合わせて開発されたものであり、同一のワークステーション上で複数の衛星の運用をサポートするという特長を持つ。これまで、いくつかの問題点・リスクをかかえながらも、衛星ごとに開発・改良され、維持・運用されている。この衛星試験・運用システムを改善するため、山田は SIB2/GSTOS の構想を提案し、開発要求を設定した<sup>5)</sup>。これを具現化し、また、衛星試験・運用システムのその他の問題点を改善するため、ISAS は、SPRINT-A, ASTRO-H, Bepi/MMO を対象衛星とした、SIB2/GSTOS-1 プロジェクトを立ち上げた<sup>6)</sup>。SIB2/GSTOS-1 プロジェクトは、SIB2/GSTOS 構想を基に、現行の衛星試験・運用システムの特長を生かしつつも、問題点を解消し、より効率的なシステム構築のため、システムを構成する汎用のソフトウェアを開発するプロジェクトである。

本論文では、SIB2/GSTOS-1 プロジェクトの概要と開発状況及び今後の課題について述べる。まず、2 章では、SIB2/GSTOS-1 プロジェクトにいたるまでの背景を、3 章では、開発方針、設定目標について示す。4 章で、現在までの開発状況と今後の課題について説明し、5 章にまとめる。

---

\*1 宇宙科学研究所 科学衛星運用・データ利用センター (Center for Science Satellite Operation and Data Archive, ISAS)  
\*2 宇宙科学研究所 学際科学研究系 (Department of Interdisciplinary Space Science, ISAS)  
\*3 宇宙科学研究所 Bepi Colombo プロジェクトチーム (Bepi Colombo Project Team, ISAS)  
\*4 宇宙科学研究所 宇宙機応用工学研究系 (Department of Spacecraft Engineering, ISAS)  
\*5 宇宙科学研究所 科学推進部 (Management and Integration Department, ISAS)



## 2. 背景

### 2.1. 衛星運用

衛星運用とは、衛星に与えられたミッションを遂行するために、地上から、人工衛星の状態を監視し、人工衛星の制御を行うために必要な業務全般である。衛星運用を行うにあたって、必要となるのは、大きくわけて軌道力学系システム、地上局システム、衛星運用システムである。軌道力学系システムは、軌道予測・決定を行う機器で構成され、地上局システムは、地上にあるアンテナ及びアンテナで送受信する電波の変復調を行う機器で構成される。衛星運用システムは、衛星運用の計画作成、人工衛星の状態監視、人工衛星への制御を行う端末を含む機器で構成される。以下では、衛星運用システムを用いて、地上局と人工衛星が通信できる時間帯に、人工衛星の状態監視、人工衛星への制御を行うことを可視運用という。

可視運用中及び可視運用前後の各システムを用いた流れは以下のようになる(図1)。

#### [可視運用前]

- ・ 軌道力学系システムが衛星の軌道予測値を計算する。
- ・ 衛星運用システムで、衛星の軌道予測に従って、どの地上局で運用を行うかを決定し、可視運用のみならず非可視の運用を含めた、衛星運用の計画を作成する。この中には、ミッションや姿勢制御等の計画に従って、衛星に発行する制御コマンドなどが含まれる。

#### [可視運用中]

- ・ 地上局システムを経由して、衛星運用システムで衛星の状態監視や、衛星へのコマンド発行を行う。
- ・ 地上局システムでは、軌道決定に必要な測距データを取得する。

#### [可視運用後]

- ・ 可視運用で得られたテレメトリデータに対して、異常がないか診断を行う。
- ・ 可視運用で得られた測距データをもとに軌道力学系システムが衛星軌道を決定・予測し、可視運用前の流れに戻る。

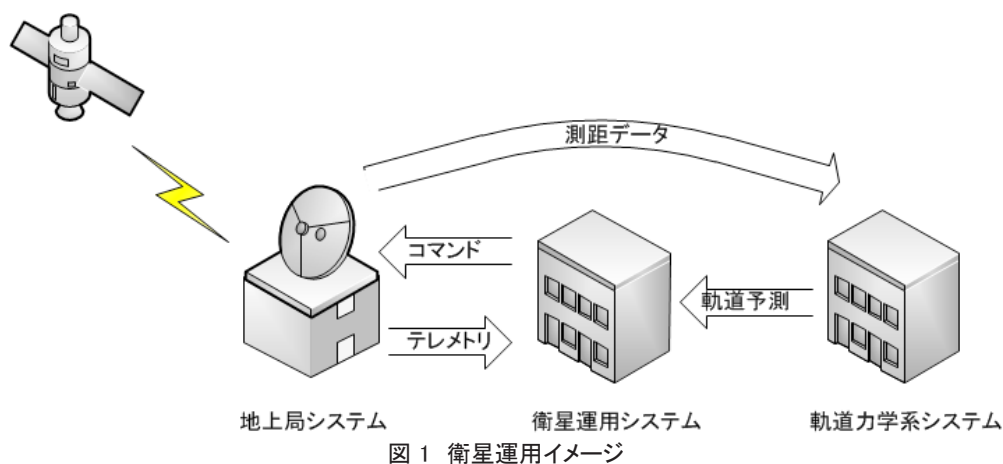


図1 衛星運用イメージ

### 2.2. ISASにおける衛星運用システム

ISASの現行の衛星運用システムは、典型的に、衛星の状態監視・制御、衛星運用の計画作成、衛星の診断を行うサブシステムで構成されている(図2、表1)。本著では、衛星の状態監視・制御を行うサブシステムを、衛星試験及び衛星運用で使用されることから試験・運用系と呼び、また、計画作成や診断を行うサブシステムは、主に定常運用時で使用されることから定常運用系と呼ぶ。個々の衛星のテレメトリ・コマンドの定義情報はSIBというデータベースに登録され、テキストファイル形式で蓄積される。SIBを作成・管理するためのシステムをSIBシステムといい、SIBシステムも衛星運用システムに含まれる。試験・運用系、定常運用系それぞれのサブシステムが、SIBシステムから最新のSIBを取得し、衛星運用システム全体で共通のSIBに基づき動作する。試験・運用系とSIBシステムの部分は、可視運用向けのみならず、衛星試験時向けにも構築され、このばあい、衛星試験システムと呼ばれる。以下では、衛星運用システム、衛星試験システムを総称して、衛星試験・運用システムと呼ぶ。

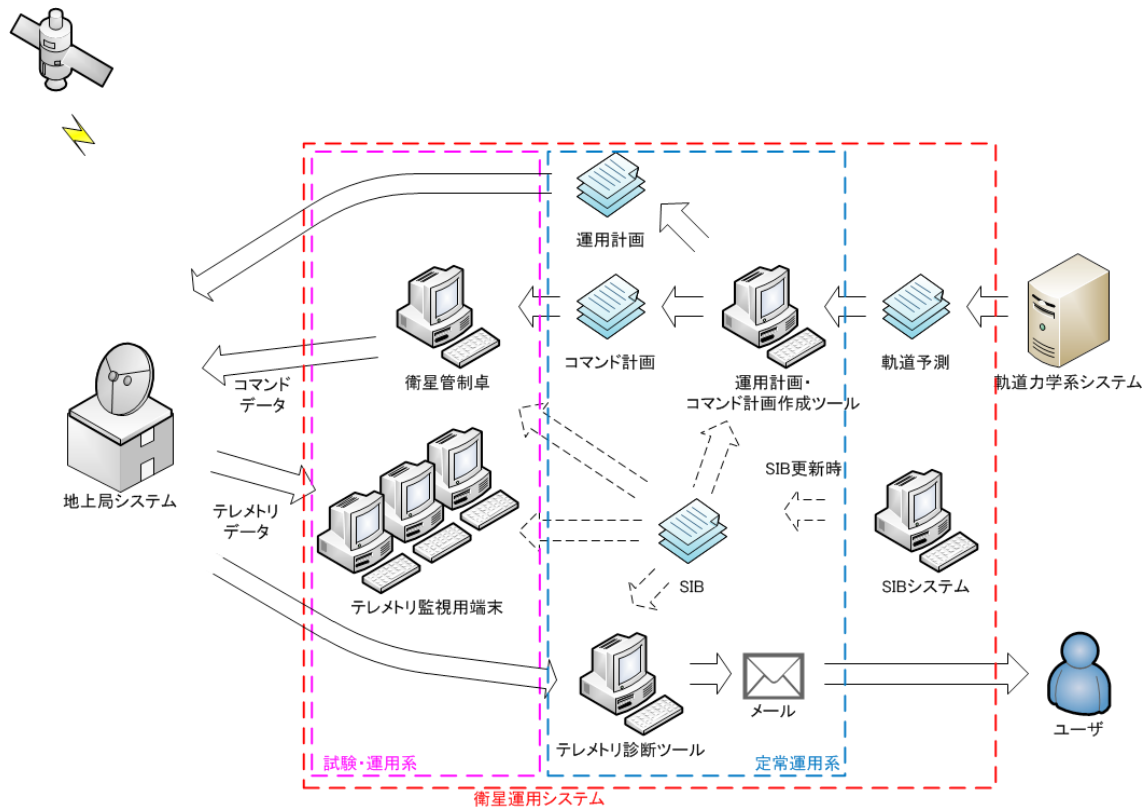


図 2 衛星運用システム

表 1 衛星運用システム構成要素

構成要素	機能概要
SIB 関連	-
SIB システム	SIB 作成, 構成管理
試験・運用系	-
衛星管制卓	衛星に対するコマンド発行, コマンド発行結果の照合
テレメトリ監視用端末 ( 状態監視卓, 共通 QL, PI QL )	現在や過去に衛星から受信したテレメトリの文字やグラフによる表示
定常運用系	-
運用計画・コマンド計画作成ツール	衛星運用の計画作成 コマンド計画ファイルの作成 計画内容の検証
テレメトリ診断ツール	テレメトリの内容のチェック, ユーザ通知

以下, ISAS の衛星運用システムを構成する個々のサブシステム・ツールについて各項で説明する。

### 2.2.1. SIB システム

SIB システムは, SIB の入力, 構成管理を行うためのツールで構成される。搭載機器の開発段階で, 機器設計者が入力ツールを通じて, 設計結果を SIB の各項目として定義を行う。また, 衛星システム担当者が機器の設計を取りまとめるとともに, 衛星システム全体の設計を, 入力ツールを通じて定義する。定義された情報はテキストファイル形式で出力される。衛星システム試験や衛星運用において, 衛星システム担当者が取りまとめた情報は, 衛星運用システムの他ツールやサブシステムに配布される。

### 2.2.2. 試験・運用系：衛星管制卓

衛星管制卓は、ユーザの画面操作により、衛星へコマンドを発行するために用いる。また、発行したコマンドの送信結果、実行結果などの照合を行う。衛星管制卓は、衛星打ち上げ後の可視運用時のみでなく、衛星試験時にも用いられる。衛星に対して発行するコマンドの内容は、テキストファイル形式のコマンド計画ファイルに記述しておく。コマンド計画ファイルは、運用計画・コマンド計画作成ツールなどの外部ツールで作成するか、衛星管制卓の内部機能で作成する。

### 2.2.3. 試験・運用系：テレメトリ監視用端末

テレメトリ監視用端末は、いずれの機器も、現在や過去に受信したテレメトリを文字やグラフで表示することにより、ユーザが衛星や搭載機器の状態を監視するために用いる。衛星管制卓と同様に、可視運用時のみでなく、衛星試験時にも用いられる。従来、対象ユーザや端末上の位置づけにより、テレメトリ監視用端末として、状態監視卓、共通 QL (Quick Look), PI (Payload Instrument; 搭載機器) QL が存在し、主に、状態監視卓は衛星システム担当者、共通 QL はバス機器担当者、PI QL は各搭載機器担当者が利用する。ここで、端末名称に使用している QL とは、リアルタイムでテレメトリを画面表示し、衛星の状態を監視する業務を意味している。

### 2.2.4. 定常運用系：運用計画・コマンド計画作成ツール

運用計画・コマンド計画作成ツールは、衛星の定常運用におけるコマンド計画ファイルの作成を省力化するものであり、2つの部分、コマンド計画作成部、計画検証部から構成される。ここで、計画検証部には、運用計画作成部が含まれる。従来は、コマンド計画作成部にあたるものとして ISACS-PLN というシステムが開発・運用されてきた。コマンド計画作成部では、定常運用時に、衛星管制卓でコマンド発行時に使用するコマンド計画ファイルを作成する。「日照・日陰など、軌道上環境が切り替わる際にこのコマンドをこの順番で実行する」など、定常運用時のパターン化した運用要求を運用要求記述言語 (ORL; Operation Request Language) で記述し、日々変わる可視運用を行う局や時刻の情報を含む地上局情報や軌道上環境と組み合わせ、コマンド計画ファイルを作成する。

また、計画検証部では、地上局情報や軌道上環境、運用内容から、衛星や搭載機器の状態を簡易的に予測し、電力、データ容量、熱、姿勢などの制約条件があれば、それぞれ条件に対する検証を行う。運用計画作成部では、衛星の軌道予測をもとに、各地上局で可視運用を行える時間帯を計算し、運用計画ファイルを作成する。実際に、可視運用を行う時間は、他衛星等との調整を経て決められ、その結果が地上局情報としてコマンド計画作成部、計画検証部への入力となる。

### 2.2.5. 定常運用系：テレメトリ診断ツール

テレメトリ診断ツールは、衛星・搭載機器の異常状態をタイムリーに、機器担当者に知らせるためのものである。従来は ISACS-DOC というシステムが開発・運用されてきた。診断ルールに基づいて、可視運用中の衛星のリアルタイムテレメトリや、衛星に蓄積されたデータを再生したテレメトリ (リブロードデータ) の値をチェックし、異常があれば機器担当者にメールで通知する。診断ルールには複数のテレメトリを参照する条件を設定できるなど、テレメトリ監視用端末上でチェックできない条件を設定可能である。

## 2.3. 従来の衛星運用システムの問題点

ISAS では、前節で説明したような衛星運用システムが開発・運用されてきたが、従来の衛星運用システムでは、以下の4つの問題点があり、衛星開発のコストやスケジュールに大きく影響を与えていた。

#### (1) 搭載機器の設計と SIB の関係についての問題点：

搭載機器の設計の記述レベルが設計者毎に異なる。SIB システムによる電子的管理が不十分で非効率的である。

#### (2) 運用計画・コマンド計画作成ツールの問題点：

運用計画・コマンド計画作成ツールのうち計画検証部は、衛星の打上前後に、若手研究者が作成していた。これは、本来研究に充てるべき時間をツールの作成やメンテナンスで消費することになり、科学的な生産性を落とすと共に、ツールの検証不足による不具合発生などのリスク要因となることが多い。

#### (3) 衛星プロジェクト毎の開発・運用体制の問題点：

衛星毎の新規性がほとんどない箇所でもプロジェクト毎に開発・運用を行い、それぞれのプロジェクトでメーカーと対応を行っており非効率的である。また、各サブシステムがソフトウェアを含むブラックボックスの装置として整備され、ソフトウェア部分の開発規模が不明瞭となり、開発・運用のコストや品質の管理が困難となるとともに装置毎にコストが発生することとなっている。

(4) 維持管理の問題点：

メーカーや JAXA 内の特定の個人のみが詳細な設計内容を把握しているという属人的な開発が行われてきたため維持管理に必要な設計情報が不足している。

これらの問題点を解消し、より効率的な衛星開発・衛星運用を行うことが可能なシステムを構築することをめざし、ISAS は、SPRINT-A、ASTRO-H、Bepi/MMO を対象衛星とした、SIB2/GSTOS-1 プロジェクトを立ち上げた。なお、現在では、SIB2/GSTOS-1 プロジェクトで開発したソフトウェアは、はやぶさ 2、ERG においても使用されている。

### 3. SIB2/GSTOS-1 プロジェクトの目標・開発方針

2.3 節で挙げた従来の ISAS の衛星運用システムの問題点に対して、SIB2/GSTOS-1 プロジェクトでは、それぞれ以下に示す解決策を策定し、またこれらの解決策に対して表 2 に示す目標レベルを設定した。

- (1-1) 機能モデルによる衛星設計を実現
- (1-2) 従来を凌ぐ利便性の実現
- (2-1) 計画作成・検証系の汎用ツール化
- (3-1) 衛星毎の差分開発の極小化
- (3-2) ソフトウェア化し、開発元とは別の運用部隊による効率的な管理の実現
- (4-1) 利用者が理解可能な資料を整備
- (4-2) JAXA 側が差分開発を継続できるよう資料を整備

表 2 目標レベル

レベル	内容
ミニマムサクセス	各衛星プロジェクトのスケジュールに合わせてまずは必要最低限の機能を提供すること
	従来システムで実現していた基本的な機能を実現すること
フルサクセス	属人性の排除・文書化
	従来システムで実現していたすべての機能を実現すること
エクストラサクセス	その他の新機能の実装

また、より具体的な開発方針として、以下を設定した。SIB2/GSTOS-1 プロジェクトでは、これらの目標・開発方針に基づき、開発を実施した。

- (A) 試験・運用系については、従来のシステムと同レベルの機能を保持する。
- (B) 定常運用系については、既存のソフトウェアの機能の取捨選択を行う。
- (C) 衛星試験・運用で最低限必要な SIB2 と試験・運用系の検討・開発から着手
- (D) 検討フェーズでは、各衛星プロジェクトと協力し、従来システムに対する改善要求の洗い出しや、衛星設計との矛盾がないよう設計検討を行う。
- (E) 上記の改善要求などについては、共通部と衛星固有部に区別し、共通部の開発を SIB2/GSTOS チームで担当し、衛星固有部については、衛星プロジェクト側で担当することとする。

#### 4. SIB2/GSTOS-1 プロジェクトにおける開発状況と今後の課題

SIB2/GSTOS-1 プロジェクトでは、従来の ISAS の衛星運用システムと山田の示す GSTOS への要求仕様<sup>4)</sup>を踏まえ、各サブシステムの機能定義を行い、ソフトウェアの開発を行った。試験・運用系、定常運用系ソフトウェアは、近年の衛星運用システムの計算機採用状況に合わせ RHEL(Red Hat Enterprise Linux) 系 OS を想定している。2013 年度上期現在、開発は一段落し、維持管理フェーズに入っている。SIB2 及び試験・運用系については、すでに複数の衛星プロジェクトの衛星試験で並行して使用されており、SIB2/GSTOS-1 プロジェクトとして、不具合や改善内容の水平展開を効率よく行うことができている。定常運用系についても、衛星プロジェクトへの提供が始まり、軌道上運用での利用にむけた準備が進んでいる。ここで示した開発したツールや関連文書に関しては、専用 web ページを作成し、公開している。ただし、ダウンロードできるユーザについては各衛星プロジェクトに対して、アカウントを発行し、制限をかけている。

以下では、サブシステム・ツール毎に、現在までに開発した内容及び今後の課題について述べる。

##### 4.1. SIB2

SIB2/GSTOS-1 プロジェクトでは、目標 (1-1) に示すように、搭載機器の機能設計方法を統一するために、機能設計に対する統一的な枠組みとして山田が提唱する衛星の機能モデル<sup>7)</sup>の概念を導入することを各衛星プロジェクトへ提案した。衛星の機能モデルとは、オブジェクト指向プログラミングにおけるオブジェクトの概念に基づく。衛星や搭載機器がもつ様々な機能のまとまりをクラスに似た機能オブジェクトとして定義したり、ステートマシンとして定義したりするのに加え、衛星や搭載機器設計固有な情報を定義する。従来の SIB にかわるデータベースとして、搭載機器が機能モデルに基づいて設計されていることを前提として、SIB2 のデータ構造設計を行った<sup>8)</sup>。これにより、入力ファイルにおいて、従来の SIB との互換性が無くなるとともに、扱う情報も多く、複雑になったため、従来よりも利便性の高い入力ツールの開発<sup>9)</sup>、履歴・構成管理方法の確立、また、従来検討不足であった SIB システムについて、ユーザが扱いやすくなるように検討・開発を行った (目標 1-2 に対応)。従来のシステムとの相違点の詳細は、現時点で以下の通りである。

- 機能モデルに基づくことで、SIB2 から衛星搭載ソフトウェアのテレメトリ・コマンドを処理する機能部分を自動で生成することが可能となった。また、これを実現するツール (SIB2Generator) を開発した<sup>10)</sup>。
- 従来の SIB では、衛星システム担当者が、機器の設計を手作業で取りまとめ、機器の設計を転記することで衛星全体の SIB を構築する必要があった。SIB2 では、衛星システム担当者の負担や作業ミスを軽減するよう、機器設計の定義内容をツールが自動でマージすることとした。
- SIB2 のデータベースの源泉 (機能情報個別定義) の表記には XML を採用することとし、機能情報個別定義はサブシステムや搭載機器など、任意の取りまとめ単位毎に複数の XML ファイルで構成した。XML ファイルを源泉とするため、SIB2 を入力するツール (SIB2 入力ツール; SIB2UI) には、XML ファイルを出力する機能のみならず、XML ファイルを取り込む機能も持たせた<sup>11)</sup>。なお、SIB2 入力ツールなど Excel 上から XML を入出力するツールを効率よく開発するため、汎用なライブラリ xmlExcelHyper を開発し、各種のツールの構築に適用している。
- SIB2 の利用形態として、試験・運用系など従来の SIB の形式が利用しやすいサブシステム向けに、従来から利用していた SIB ファイルの形式 (配布用 SIB) を、新規の衛星に対応させたものを SIB2 テレコマビューファイル形式と定義した。また、機能情報個別定義 XML ファイルをテレコマビューファイル形式に変換するツール (テレコマビュー変換ツール) を開発した。
- 従来のシステムでは、SIB を管理するため、RDB (Relational Data Base) に SIB を登録するサーバを配置したシステム構成となっていた。このサーバは有意義な機能を持っていなかったため、SIB2 では、このサーバと RDB を廃する設計とした。
- SIB2 の定義範囲を拡大するため、衛星運用システム開発メーカーが独自で持っていたツールについても洗い出し、JAXA 側で内容を把握・維持管理できるようにするなど、ツールの拡充を図った (テレメトリ表示画面定義ファイル作成ツール、衛星メモリ管理用ツール、搭載機器冗長系定義ツール)。また、これらのツールや SIB2UI など、SIB2 関連の全てのツールを一括でインストール可能なよう取りまとめたパッケージ (SIB2Suite) を開発した。
- 設計変更の履歴を残し、また、過去の設計に戻せるよう、汎用な履歴管理ソフトウェア (mercurial) を用い、履歴・構成管理を行うようにし、SIB2Suite にこの機能を持たせた。
- コマンド計画ファイルを誤りなく効率よく編集するため、SIB2 に基づく専用のエディタ (SIB2PlanEditor) を開発した<sup>9)</sup>。



- ・従来の SIB の定義項目のうち、テレメトリの値によりリミット範囲を変更する定義やグラフ表示画面定義など、一部複雑な定義については未対応。

各ツールの開発は完了し、各衛星プロジェクトにて使用されている。なお、今後の課題として、文書レベルが一部まだ不十分であることと、システム構成が検討不足なことから、ユーザが使いにくい状態になっている箇所があるため、ユーザの意見を収集し、SIB2/GSTOS チームにて、システム構成の見直しを行い、改善作業を進めていく。また、これまで未対応である搭載機器の経年変化による定義内容の切替等への対応についても検討する。

#### 4.2. 試験・運用系

従来、ISAS は衛星管制卓、テレメトリ監視端末というように、衛星運用システム開発メーカーに対しての要求として、サービスを提供するシステムの開発・セットアップを依頼していた。SIB2/GSTOS-1 プロジェクトでは、それぞれ、ソフトウェアとして開発し、任意の数のハードウェアへ導入可能なようにすることとした。また、ソフトウェア化することにより、衛星毎の差分が局在化するようにした（目標 3-1 に対応）。衛星管制卓・状態監視卓に対応するものとして、コマンド発行・状態監視ソフトウェアを開発し、共通 QL に対応するものとして、テレメトリ監視ソフトウェアの開発を行った。各ソフトウェアは従来と同等の機能を保持し、従来と同じシステム構成とすることが可能である。なお、コマンド発行・状態監視ソフトウェアのうち、状態監視ソフトウェアの機能は、SIB2/GSTOS-1 プロジェクトの範囲には無かったが、はやぶさ 2 プロジェクトの協力により開発されたものである。このほか、従来システムとの相違点は、以下の通りである。

##### コマンド発行・状態監視ソフトウェア

- ・従来から有する機能や信頼性を維持するため、従来の SIB ファイルの形式（配布用 SIB）を、新規の衛星に対応させたものである SIB2 テレコマビューファイルに基づき動作する設計とし、従来から衛星運用システム開発メーカーが有していたソフトウェアに対する微改修に留めた。なお、本ソフトウェアは、PC-Solaris 系と PC-Linux 系にマルチプラットフォーム対応した SELENE 衛星管制システム<sup>4)</sup>の流れを汲むあかつきにも対応したシステムを引き継いだものである。
- ・コマンド発行時に、テレメトリの値をチェックする機能、条件成立まで一時停止する機能を追加した（各機能に対応し、コマンド計画ファイルの文法も追加した；CHECK 文、WAIT\_UNTIL 文）。
- ・コマンドのパラメータは、エンコード済みの 16 進数表記ではなく、パラメータ毎の工学値を指定できるようになった。

##### テレメトリ監視ソフトウェア

- ・従来に比べ文書の拡充を目指し、新規のソフトウェア開発を実施した。SIB2 に基づいて動作する設計とした。
- ・従来の共通 QL は、テレメトリ表示画面を作成・変更するには、開発メーカーへプログラムの改修の依頼が必要だった。テレメトリ監視ソフトウェアでは、機器担当者が自ら画面定義のツールを用いて定義を行えるようにした。
- ・ユーザ要求を踏まえ検索性など機能向上した。
- ・テレメトリの工学値変換の方法として、状態監視ソフトウェアと同一の C 言語の関数による変換、任意の関数式による変換を扱えるようにした。

##### 両ソフトウェア共通

- ・従来、ソフトウェアのセットアップ手順は、メーカー内のみで保持されていた。SIB2/GSTOS-1 プロジェクトの試験・運用系ソフトウェアは、いずれも RPM (RedHat Package Manager) パッケージ化することで、インストール手順の簡略化を図ると共に、ユーザに対し、セットアップ手順を明らかとした。

各ソフトウェアの開発は完了し、SIB2 システムと合わせて実際に各衛星プロジェクトの衛星試験で使用されている。今後の課題としては、文書レベルが一部まだ不十分なので、ユーザの意見を収集し、改善作業を進めていく。

---

1 ただし、あかつきで使用されているシステムは PC-Solaris 系のみである。

### 4.3. 定常運用系

#### 4.3.1. コマンド計画検証ソフトウェア

SIB2/GSTOS-1 プロジェクトでは、運用計画・コマンド計画作成ツールに対応するものとして、コマンド計画検証ソフトウェアの開発を行った。これは、従来のシステムでもある程度共通化されていたコマンド計画作成部、衛星プロジェクトそれぞれで衛星打ち上げ前後にインハウスでツールを作成するなどして対応していた運用計画作成部、計画検証部から構成される。このうち、計画検証部は、衛星毎の差が大きいことと、また、Bepi/MMO では既に一部ツールを作成していたこともあり、まずは近地球を対象を絞ることとして共通部の開発を行った。特に、SPRINT-A に対して全モジュールの構築を実施することで、モジュール構造にて目的を達成できることを示した。このほか、従来システムとの相違点は以下の通りである。

##### 計画作成部

- ・従来のツールで仕様が不明確であった各コマンドの優先度や時刻の割付方法について明確にするため、新規に開発を実施した。
- ・運用要求記述言語の見直しを行った。見直すにあたっては、独自の言語設計がミニマムになるよう、プログラミング言語 Ruby の構文に沿うよう ORLG (Operation Request Language for GSTOS) を定義した。
- ・条件判断や衛星のモード遷移など高度な機能については現段階では未対応

##### 計画検証部

- ・一つのソフトウェアではなく、複数のツール群として開発し、ツールは衛星毎に入れ替えられるようなモジュール構造とした。
- ・各種の計算モジュールは、座標系や精度が明らかになるよう仕様書を整備した。
- ・共通部を作成したことで、従来に比べ衛星プロジェクトの負担は少なくなった (目標 2-1, 3-1 に対応)。

##### 計画検証部の運用計画作成部

- ・ひので衛星で開発されたプログラムを微改修し、SPRINT-A で使用する地上局に対応するモジュールを作成した。

今後は、ユーザの意見を収集し、改善作業を進めていく。

#### 4.3.2. テレメトリ診断ソフトウェア

SIB2/GSTOS-1 プロジェクトでは、テレメトリ診断ツールに対応するものとして、テレメトリ診断ソフトウェアの開発を行った。従来システム (ISACS-DOC) と同様に、事前に設定した診断ルールに基づき、リアルタイムテレメトリおよびリプロデータの診断を行い、その結果をユーザに通知する。従来システムとの相違点は、以下の通りである。

- ・他のソフトと重複していた機能の整理を行った。特に、工学値変換の機能については、コマンド発行・状態監視ソフトウェアと同一のプログラムを利用する設計とした。特に、この整理により、従来は診断ルールとして記述されていた、複数のテレメトリを加工し新たな値を得る計算を、工学値変換エンジンの機能により実現した。
- ・SIB2 に定義されたリミットチェックはデフォルトで実施し、それ以外の診断ルールを別途設定することとした。
- ・現時点ではテレメトリを診断した結果をメールで通知する機能に限定している。グラフ表示機能、診断ルール入力インタフェース、データの時間方向の加工 (フィルター機能など) については未対応。

今後は、ユーザの意見を収集し、改善作業を進めていく。

### 4.4. SIB2/GSTOS-1 プロジェクト全体としての目標達成状況

各ソフトウェアの開発状況を踏まえ、SIB2/GSTOS-1 プロジェクト全体としての目標 (問題点に対する解決策) 毎の達成状況を表 3 に、目標レベルの達成状況を表 4 に示す (○は達成、△は今後達成予定)。これを踏まえ、今後は各ソフトウェアの維持管理を行いながら、フルサクセス、エクストラサクセスを目指していく。

表 3 目標達成状況

従来システムの問題点		目標 (問題点に対する解決策)		状況
(1)	搭載機器の設計と SIB の関係についての問題点	(1-1)	機能モデルによる衛星設計を実現	○
		(1-2)	従来を凌ぐ利便性の実現	○
(2)	運用計画・コマンド計画作成ツールの問題点	(2-1)	計画作成・検証系の汎用ツール化	○
(3)	衛星プロジェクト毎の開発・運用体制の問題点	(3-1)	衛星毎の差分開発の極小化	○
		(3-2)	ソフトウェア化し、開発元とは別の運用部隊による効率的な管理の実現	△
(4)	維持管理の問題点	(4-1)	利用者が理解可能な資料を整備	△
		(4-2)	JAXA 側が差分開発を継続できるよう資料を整備	△

表 4 目標レベル達成状況

レベル	内容	状況
ミニマムサクセス	各衛星プロジェクトにのスケジュールに合わせてまずは必要最低限の機能を提供すること	○
	従来システムで実現していた基本的な機能を実現すること	○
フルサクセス	属人性の排除・文書化	△
	従来システムで実現していたすべての機能を実現すること	△
エクストラサクセス	その他の新機能の実装	△

## 5. まとめ

ISAS における衛星運用システムが抱えていた問題点を解消し、より効率的な衛星開発・衛星運用を行えるよう SIB2/GSTOS-1 プロジェクトとして SIB2 定義、各ソフトウェアの開発を行った。開発は完了し、各衛星プロジェクトでの衛星試験等で使用され、ミニマムサクセスは達成した。今後は、文書整備やユーザの意見収集による改善作業などを行うことにより、フルサクセス、エクストラサクセスの達成を目指している。

## 参考文献

- 1) 山田隆弘: 宇宙情報システム講義第1部 衛星データシステムをこう作ってきた (第1回 はじめに), PLAIN センターニュース第 160 号 (2007), pp. 1
- 2) 山田隆弘: 宇宙情報システム講義第1部 衛星データシステムをこう作ってきた (最終回 衛星運用), PLAIN センターニュース第 166 号 (2007), pp. 1
- 3) 栗山祐一, 久保雅嗣, 島村敬之, 江川宏, 春木和寿: X 線天文衛星 ASTRO-E 対応衛星管制システムの開発, NAS 技法 No. 12(2001), pp. 40-45
- 4) 矢野昌邦, 島村敬之, 久保雅嗣, 大橋清一: 科学衛星向け衛星管制システム, NAS 技法 No. 17(2006), pp. 45-50
- 5) 山田隆弘: 汎用衛星試験運用ソフトウェア (GSTOS) 要求仕様 (案) 第 0.4 版, 2009 年 1 月 30 日
- 6) 馬場肇, 松崎恵一, 福田盛介, 山田隆弘, 飯塚祐介, 山地尋之, 大石克己: SIB2/GSTOS1 の開発, 第 11 回宇宙科学シンポジウム, 2011 年 1 月
- 7) Takahiro Yamada: Functional Model of Spacecraft (FMS) DRAFT, Issue 0.9, March 2011
- 8) 山田隆弘, 松崎恵一: 衛星情報ベース 2 定義 (案) 第 0.9 版, 2009 年 11 月 11 日
- 9) Keiichi Matsuzaki, Takashi Kominato, Shinya Okunishi, Yoshikazu Miyano: SIB2/GSTOS Tools for Preparing Test Procedure – PlanEditor and SIB2Viwer, 29<sup>th</sup> International Symposium on Space Technology and Science, 2013
- 10) Matsuzaki, K., Saito, T., Okunishi, S. and Kominato, T.: Automatic generation of on-board software from the model – Spacecraft Information Base Version 2 (SIB2), Trans. JSASS Aerospace Tech. Japan, 10, ists28 (2012), pp. Tf\_11-17.
- 11) 松崎恵一, 山下美和子, 馬場肇, 西村佳代子, 宮野喜和: xmlExcelHyper を用いた SIB2UI の開発, 宇宙科学情報解析論文誌 第二号 (2013), pp. 113-130



### 謝辞

SIB2/GSTOS-1 プロジェクトは、日本電気株式会社、NEC 航空宇宙システム、NEC システムテクノロジー、富士通株式会社、富士通エフ・アイ・ピー・システムズの協力の下、各ソフトウェアの開発を実施しました。各社のエンジニアをはじめ、構想の具現化に協力を頂いた方々に、感謝の意を表します。

# 「どこでも運用システム」の開発状況

永松 弘行<sup>\*1</sup>

## Development of Anywhere Satellite Operation System

Hiroyuki Nagamatsu<sup>\*1</sup>

### Abstract

We are in progress to develop a system for automatic operation of a satellite in order to reduce human load at satellite steady operation phase. The ground station for small satellite REIMEI is used as a test bench for verification of the proposed method. The automatic operation system is nearly completed for downlink operations of the data recorder that account for 80% of REIMEI steady operation. This approach is very effective to reduce psychological and physical load of operators. To extend or complement functions of automatic operation system, we are also in progress to develop a system for remote operation of satellite using mobile gear, for example, tablet or smart phone, as well as personal computers. With this remote operation system, operators can monitor status of satellite and ground station anywhere. In addition, on receiving emergency call, operator can send appropriate commands or operation plans by controlling automatic operation system.

### 概要

衛星の定常運用における運用者の負担軽減を主目的とする衛星自動運用システムを開発中である。「れいめい」衛星の運用局(相模原局)を実験ケースとして試験運用を継続しており、実用化の目処は立った。また、自動運用システムを補完、あるいは拡張するための遠隔運用システムをあわせて開発中であり、「どこからでも衛星が運用・監視可能なシステム(どこでも運用システム)」の構築を目指している。本稿では、「どこでも運用システム」の開発状況を紹介する。

### 1. はじめに

宇宙機を打ち上げる目的は様々だが、ミッション期間中に高品質の観測データを安定して提供し続ける運用システムの整備が、運用人員の確保も含めて肝心である。しかしながら、運用システムを維持するためのこれら要素は、衛星の開発において、どうしても後回しにされがちであり、運用システムの整備維持が不十分、運用人員の確保が不十分、運用人員の負担増、運用コストは削減傾向にある、などの問題がある。定常運用の負担軽減、効率化、コスト削減のための現実的な方策が求められると考える。運用システムを可能な限り自動化してオペレータの負荷を軽減する方法がいくつか提案され、また、実施している例、機関もある<sup>1), 2), 3)</sup>。

筆者らはこれまで、特に定常運用における運用者の負担軽減を主目的に、ISAS/JAXAが開発した小型科学衛星「れいめい」の地上局(相模原局)<sup>4)</sup>をテストベッドとした衛星自動運用システム<sup>5)</sup>の開発を行ってきた。このシステムは衛星管制系と独立に整備可能なシステムで、ユーザの要求に応じて機能拡張可能なように設計されている。また、「れいめい」衛星による、観測データ/HKデータのダウンリンク運用(定常運用の約8割)の実績を積んでおり、自動運用システム導入の効果は大きい。また、自動運用アルゴリズムの検証を容易にするため、衛星実機の挙動に近い(擬似)信号パターンを容易に生成できるシステム(「簡易シミュレータ」と呼ぶ)を導入し、自動運用システムの完成度向上に役立てている<sup>6)</sup>。

一方、緊急時の対策や遠隔地からの衛星状態のモニタなど、自動運用システムの機能を補完あるいは拡張するためのシステム開発については、これまで課題として紹介するにとどまっていた。昨年後半より、実運用適用へ向けての開発を本格化したところである<sup>7)</sup>。本稿では、開発中の遠隔運用システムについて紹介する。遠隔運用システムの目的は、1) 衛星運用の操作一式(運用時間調整、運用、取得データ管理、安全管理)を、携帯端末により実現し、2) 自動運用システムと組み合わせることで、「どこからでも衛星が運用・監視可能なシステム」(以下、「どこでも運用システム」と呼ぶ)

<sup>\*1</sup> 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)

を実現することである。

本稿の構成は、以下の通りである。第2章で、開発・相模原局にて試験運用中の自動運用システムの基本的な考え方を述べる。第3章で、自動運用システムを補完・拡張し、「どこでも運用システム」の核となる遠隔運用システムの開発経緯と基本的な考え方を述べ、システムの開発状況(一部、実運用での使用状況を含む)を紹介する。第4章で今後の課題を示し、本稿をまとめる。

## 2. 衛星自動運用システム

衛星自動運用システム(ASURA: Adaptable Scheduler for Ubiquitous and Rational Application)は、「スケジューラ」が、運用者の代役として運用シーケンス全体を管理・統括・制御する<sup>4)</sup>。スケジューラは、運用系PCに対する個々の操作(キーボード入力やマウスクリック、メニュー選択など)を利用するための通信窓口・自動化API(Application Program Interface)を通じて、可視時間中のアンテナ制御やコマンド送信、衛星や地上局の状態監視などの機能を実現する。図1に、ASURAの動作シーケンスを示す<sup>4)</sup>。自動運用システム(ASURA)の機能確認のため、「れいめい」実機および相模原局を用いての自動運用試験を、2010年末より継続している(実験回数は2013年5月末現在、500回以上)。「れいめい」運用においては、観測データ/HKデータのダウンリンク運用が、定常運用の約8割を占める。相模原局では、このすべてを自動運用システムを用いて実施している。

人間のオペレータが運用する場合、例えばコマンド送信は、コマンド選択、必要ならパラメータ入力、送信、結果確認、の4ステップで構成されるが、これらすべてをキーボード・マウス操作で実施する。自動運用システムでは、この4ステップすべてをソフトウェアが制御しており、オペレータは運用中にキーボード・マウスを操作する必要はない。「自動運用の様子を見ているだけ」である。また、オペレータは運用中に衛星・地上局の状態監視・緊急時対応を行うが、これは自動運用システムの衛星・地上局状態監視機能が代行し、例えば信号ロックオフ検知・修復は自動化される。軽度障害の場合は、あらかじめ準備してある(試験検証済みの)緊急スケジュールを自動送信する。

一方で、地上局機器トラブル、セーフホールド時対応など、人間が介入せざるを得ない場合は、オペレータに通知・運用を引き継ぐ。自動運用による定常運用負担軽減の効果は大きいが、人間が介入せざるを得ない(緊急時対応)場面は存在する。人間が介入しやすい仕組みを取り入れ、自動運用システムの特長を損なうことなく、その機能を補完あるいは拡張する仕組みの導入が望まれる。一つの方法として、第3章にて、開発中の携帯端末を用いた遠隔運用システムの概略をまとめる。

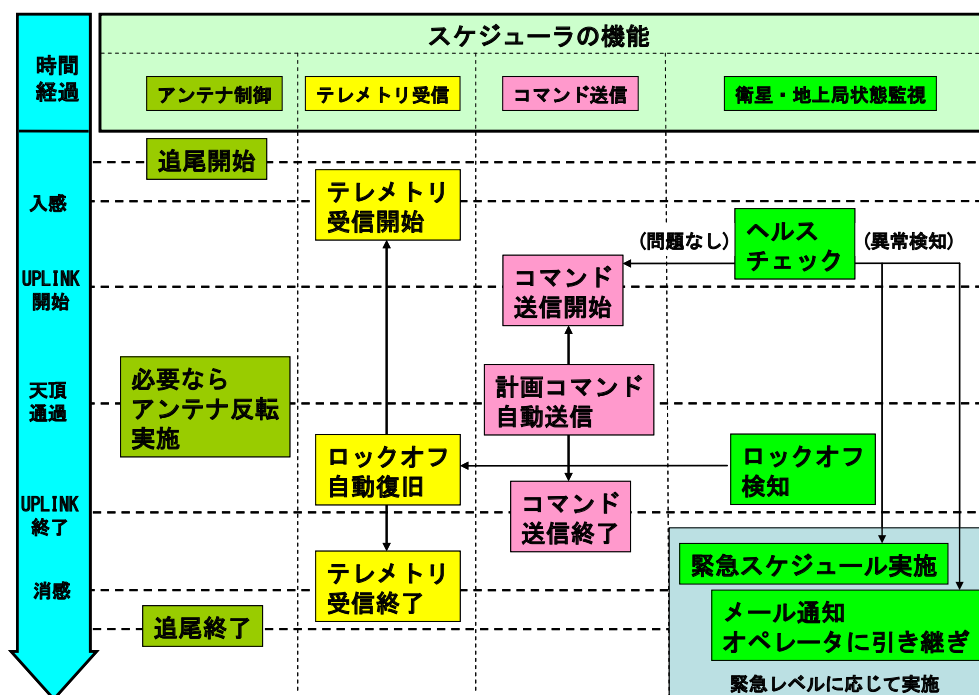


図1 自動運用システム ASURA による自動運用シーケンス概略

### 3. 遠隔運用システムを用いた「どこでも運用システム」

自動運用システムは、「観測者(ユーザ)が運用計画を立て、それをサーバに登録すれば、以後の検証、運用、運用データのアップロード、運用ログ配信などの作業を自動的にしてくれる」(ほぼ無人の)運用システムの実現を目的とする<sup>5)</sup>。しかし、ユーザが衛星や地上局の状態を監視したり、場合によっては緊急時の対応をせざるを得なかったり(セーフホルド状態からの復帰、緊急コマンド送信、災害対策など)、というケースも考えられる。そのため、人間と自動運用システムとの間に何らかの通信手段が欲しいところである。

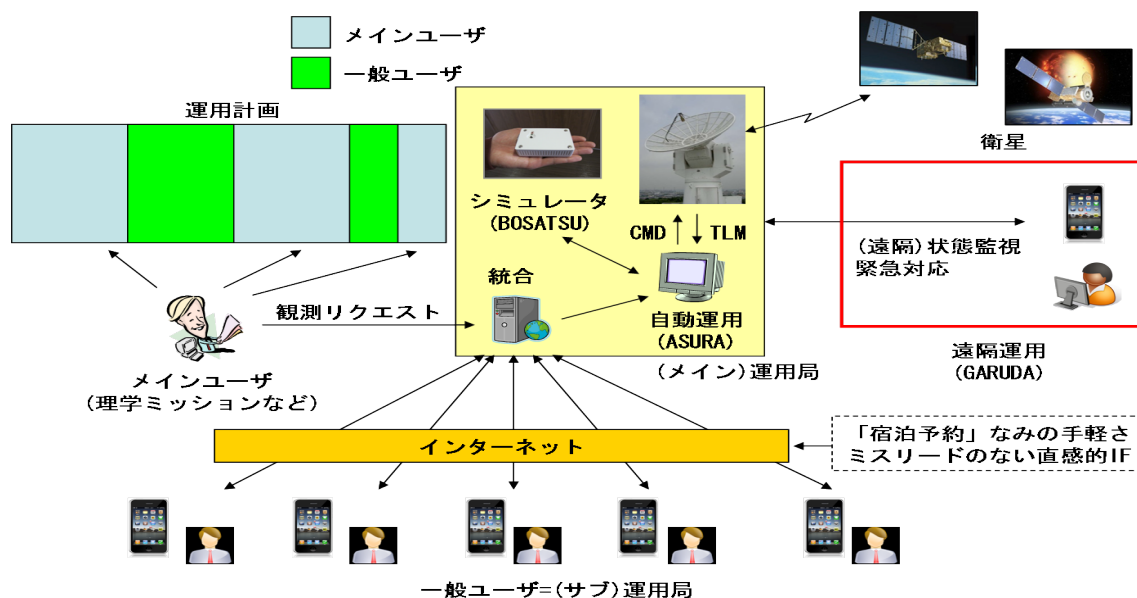
人間が極力運用システムに(時間的にも空間的にも)拘束されないための一つの手段として、ここでは遠隔運用システム<sup>6)</sup>を現在の自動運用システムと組み合わせることによって、

- 1) 衛星運用に必要な操作一式(運用時間調整、運用、取得データ管理、安全管理)を携帯端末により実現する
- 2) 自動運用システムとの組み合わせにより、「どこからでも衛星が運用・監視可能なシステム」すなわち「どこでも運用システム」を実現する

ことを目指す。開発のゴール「どこでも運用システム」をイメージとしてまとめたものを、図2に示す。携帯端末を自由に持ち運び、タッチパネルからコマンドや運用計画を選んで送信したり、ネット経由で運用結果や設備情報を確認したり、が遠隔運用システムの操作イメージである。本章では、図2の「(遠隔)状態監視、緊急対応」の部分について記述する。

遠隔運用システムの開発にあたり、1)システムを実装する端末自体の入手性・操作性のよさ、2)開発環境の充実、3)導入および運用コストに留意する。近年、iPadに代表されるタブレット端末やスマートフォンが広く普及している。これらは安価でありながら、場所を選ばずに(無線環境があればどこでも)情報のやりとりが可能である。必要なときに即座に取り出して使え、不要になったらしまえるなど、使い勝手がよい。開発環境もタブレット自体のシミュレータ(エミュレータ)がある。煩雑になりがちな配線も大幅に削減可能で、また、消費電力も低く抑えられる。用途を絞れば、低コストで使い勝手のよい運用システムの構築が期待できそうである。

なお、遠隔運用システムは、自動運用システムと連携可能であり、どこからでも使えるフットワークの軽さをふまえ、GARUDA(Generic Attachment for Reviewers' Utility and Data Archives)と命名した。図2の他の要素に関しては、別の機会にあらためて発表させていただく予定である。



- 携帯端末のタッチパネルから、観測日時、観測時間、解像度などを選択してリクエスト
- 結果をネット経由で確認、(将来的には)コマンド運用も視野に

図2「どこでも運用システム」の概略

遠隔運用システムは、図3に示すサーバ・クライアントモデルを元に開発し、開発手順としては、通信回線の問題(必要な帯域の確保)、セキュリティの問題などから、まず、衛星や地上局の状態監視を行うシステム(衛星・地上局の情報を受信するだけのパッシブなシステム)を開発した。現時点では「れいめい」の衛星状態監視機能の一部を実装し、実

運用にて自動運用システムとあわせて試験中である。スクリーンショットを、図4に示す。

この衛星状態監視システムは、運用システムのLANに設置した無線アクセスポイント(Wi-Fi)経由で衛星・地上局の情報を取得する(相模原局の環境では、有効距離は10m弱だった)。ただし、取得した情報は、自動運用システムにおいて既に工学値変換されたものであり、iPad側は通信回線・処理速度の関係から表示に特化している。「れいめい」衛星状態監視システムは簡易的なものではあるが、通信インフラが整備された環境下では、衛星管制室・運用室以外の場所においても衛星や地上局の状態を(ほぼ)リアルタイムで把握できるものである。また、衛星管制室・運用室内で使用する場合でも、人の移動の自由度が増す(人の行き来のための動線が確保しやすくなる)メリットは大きいと考える。いろいろな制約条件から、衛星の管制・運用を専用に行うための設備を収納するための部屋や建屋を確保・維持管理するのは一般に困難であると考えられるが、衛星管制・運用のための(簡易的ではあるが)システム自体が時間と場所を選ばず構築・利用でき、また、維持管理が容易であれば、導入検討の余地があるのではないかと考える。

iPad版衛星状態監視システムは、「れいめい」実運用に適用して、2013年5月末の時点で、約9ヶ月が経過した。これまでのところ、特に大きなトラブルなく運用できている。衛星・地上局に対して情報(コマンド)を送信するアクティブなシステムの開発も検討中で、原理的に可能なことは実運用試験にて確認済みである。自動運用システムの開発に用いた簡易シミュレータ(BOSATSU: Basis Of Simulator Architecture for Total Satellite/Spacecraft Utility)<sup>7)</sup>を利用すれば、一見敷居が高く見える遠隔地からのコマンド送信や携帯端末による遠隔操作などのある程度思い切った試験でも比較的手軽に試せるという利点がある。

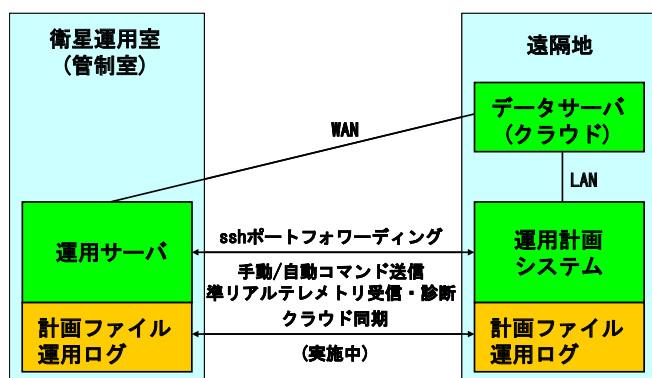


図3 GARUDAのサーバ・クライアント構造

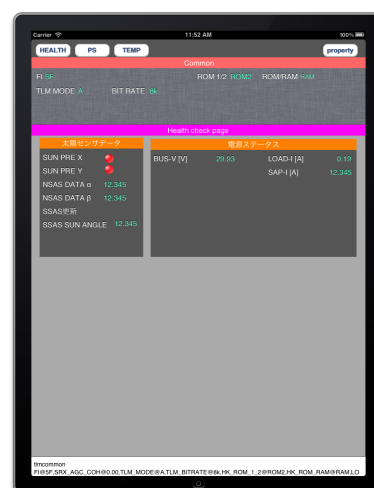


図4 iPad版衛星状態監視システム

iPad版衛星状態監視システムは、現時点では衛星・地上局の全情報を表示できていないが、今後徐々に実装範囲を拡大する。また、緊急時対応のメール受信、緊急内容に応じたコマンド(緊急スケジュール)発行の機能なども、徐々に実装する予定である。前述のとおり、遠隔地からコマンド送信可能であることは、実運用試験にて確認しており、遠隔コマンド運用の目処は立っている。

#### 4. おわりに

時と場所を選ばず「どこからでも衛星が運用・監視可能なシステム」すなわち「どこでも運用システム」の開発状況について述べた。どこでも運用システムの核の一つとなるのが、衛星自動運用システムの機能を補完あるいは拡張するために開発中の遠隔運用システムである。遠隔運用システムと自動運用システムを核とした「どこでも運用システム」の実現が、当面の最終ゴールである。

自動運用システムと遠隔運用システムの開発に用いている簡易シミュレータもあわせて、運用を意識した衛星開発(逆に衛星開発を意識した運用システム開発)が可能なシステムを実現することも、今後の大きな課題の一つとしたい。このシステムは、現状では基本的には独立して行われている衛星と地上システムの開発を、比較的早い段階から統合的に検討可能なシステムの構築を目指す。また、運用訓練や衛星のコマンド計画立案・検証のための支援システムとしての用途にも期待している。

今回は「どこでも運用システム」に始まるシステム開発の一環として、iPad 版衛星状態監視システムの現状を紹介した。今後も、システム機能の充実を図りつつ、「れいめい」衛星ならびに相模原局以外の衛星および地上局へと適用範囲を広げてゆきたいと考えている。相模原局自体も、引き続き自動運用・遠隔運用システムのテストベッドとして有効活用してゆく。また、自動運用・遠隔運用を前提とする運用システムとしてのパッケージングを視野に入れた検討を進めてゆきたい。例えば大学衛星をはじめとするビギンバッグ衛星向けのシステムも含めた検討を進めてゆきたい。

### 謝辞

日頃からご指導ご鞭撻頂いております、「れいめい」開発メンバ、運用メンバの皆様、自動運用システムのソフトウェア開発でお世話になっております秋丸忠隆氏に、この場を借りてお礼申し上げます。

### 参考文献

- 1) J. Jean-Marc Soula: “Automation of CNES Ground Station Networks” , 8th International Workshop on RCGSGO, Automation No.3, Tsukuba, Japan, 2009.
- 2) D.Chung et.al.: “Concept Design of the Generic Mission Operation System” , 8th International Workshop on RCGSGO, Automation No.12, Tsukuba, Japan, 2009.
- 3) J.Jackson et.al.: “NASA’ s Ground Network’ s Right Sizing and Transition to Outsourcing” , 8th International Workshop on RCGSGO, Outsourcing No.5, Tsukuba, Japan, 2009.
- 4) H. Saito et.al. : “An overview and initial in-orbit status of “INDEX” satellite” , 56th International Astronautical Congress, IAC-05-B5.6.B.05, Fukuoka, Japan, Oct. 2005.
- 5) H. Nagamatsu and H. Saito : “Development of Automatic Operation System-Using REIMEI Ground Station as a Test Bench” , 62nd International Astronautical Congress, IAC-11-B4.3.4, Cape Town, South Africa, Oct. 2011.
- 6) 永松：衛星の遠隔運用システムの開発，第 56 回宇科連，JSASS-2012-4109，2012.
- 7) 永松：簡易衛星シミュレータの開発，SANE，SANE2011-163，2012，pp.19-21.



# 月惑星探査研究拠点間連携によるデータセンター構想

出村 裕英<sup>\*1</sup>

## Concept of Data Center supported by a consortium of Lunar and Planetary Exploration Bases

Hirohide DEMURA<sup>\*1</sup>

### Abstract

This report emphasizes significance of data centers as outside of JAXA, which is cooperated with JAXA and universities in research community. This also summarizes activities and histories for invitation of the data center to University of Aizu in Fukushima and establishment of a consortium of Lunar and Planetary Exploration Bases.

### 概要

JAXA と大学等外部機関とが連携して、大学等研究者コミュニティ側にもデータセンターを設置する意義を述べる。本稿では、特に月惑星探査データセンターに関する会津大学の取り組みと、コンソーシアムを形成しつつある月惑星分野における連携の動きを併せて紹介する。

### 1. 本構想の背景

深宇宙探査を手がける宇宙機関は様々な媒体で探査成果を世界に公開・アピールしている。引き続き行われる探査や調査のための基礎資料としてだけでなく、広く宇宙科学の発展のために、探査データの再利用可能な形で公開が重視されていて、データ共有の枠組は例えば国際惑星データ連合 [0] (IPDA: International Planetary Data Alliance) において議論や調整がなされている。日本の宇宙機関 ISAS/JAXA には科学衛星運用・データ利用センター (C-SODA: Center for Science-satellite Operation and Data Archive) が設置され、JAXA の宇宙科学ミッションが取得した高次データプロダクトおよび関連する国内外データをアーカイブし、DARTS (Data ARchives and Transmission System) を通じて広く公開している。JAXA の宇宙科学ミッションの包含する分野は、天文学 [X 線、電波、赤外線]、太陽物理学、太陽地球系物理学、月惑星科学に大別され、更に国際宇宙ステーションで生成した一部のデータプロダクトも収集対象として整理されている。データ・アーカイブを効率よく管理する必要性から、上述の分野間で共通認識を持つための機会として、この C-SODA 運営委員会において、科学データの公開・利用に関するあるべき姿について数年にわたる議論がなされた。委員の構成メンバーは上述の分野を横断した大学・天文台・JAXA 教職員から成り、国内外の現状をレビューしつつ、分野間で異なる用語・概念の違いを確認しながら再定義して、科学データ利用のあるべき理念と基本概念をまとめた。それを宇宙科学研究所の理学委員会、工学委員会、月惑星探査プログラムグループの探査委員会、そして国際宇宙ステーションを管轄する有人宇宙ミッション本部の宇宙環境利用委員会に報告した。その成果は 2011 年 9 月に『科学データの公開・利用について』としてウェブで公開 [1] された。そこに記されている、JAXA の科学データ利用に関する三原則をここに引用する。

---

<sup>\*1</sup> 会津大学 コンピュータ理工学部 (兼) 先端情報科学研究センター E-Mail: demura[at]u-aizu.ac.jp

### 1) データプロセッシングの原則

すべての科学データについて、機器校正やデータ処理アルゴリズムを適用し、公知の知識だけでそこから科学的成果を引き出せるような段階に至るまでの処理(プロセッシング)を行う。

### 2) データ保存の原則

取得したすべての科学データは、使用できる状態で永久に保存する。

### 3) データサービスの原則

データセンターは、データプロバイダを明らかにした上でそのデータが長期にわたってできるだけ広い範囲のユーザーに使われるようにするための基盤サービスを無償で提供する。

ISAS/JAXA は大学共同利用機関としての性格を持ち、上述のオーサライズを経たことで、研究者コミュニティの総意として上記三原則が支持されている。更に、科学衛星の観測データに留まらず、探査機によるサンプルや宇宙環境での実験結果など多様な科学的価値の高いアウトプットもアーカイブ化されるべきであると、当時の C-SODA 運営委員会ではみなされていた。しかしながら、こうした原則を実行する体制・予算が組まれているかという点、そうではない現実がある。そのため、米国 Planetary Data Science System[2] (以下 PDS) の大学等機関委託制度をモデルとし、大学等外部機関との連携を強化することで利用性の向上を図る方向を模索している。米国 PDS が 2013 年現在各機関に委託しているノードには、ワシントン大学に Geoscience Node[3]、米国調査所に Imaging Node[4]、ジェット推進研究所に Navigational & Ancillary Information (以下 NAIF) Node[5]、等がある。分野ごとに分けて拠点を置いたことで、各分野固有の事情に合わせて小回りのきく運営がなされ、研究者コミュニティだけでなくアウトリーチ活動までも活発に行われている。こうした事情は、中枢にだけ重点的に予算配分する日本の傾向と比較して、注目に値する。先にあげた 3 つの原則を米国 PDS の体制は満たしており、日本にデータセンターが今後作られる場合は、既に C-SODA がそうであるように、共同利用研究施設に類した運営体制が望ましいことは言をまたない。後発参入大学・機関の可能性を排除しない拠点分散・連携型の体制とすることで、大学コミュニティのマンパワー等リソースを有効に取り入れられることが期待される。理学・情報工学等との異分野連携・協調や、分野内競争ができる環境こそが、『より良いもの』を生み出す母体となり、宇宙科学の促進に資するだろう。

また、三原則第 2 項データ保存の原則に関係して、データセンター ISAS 開所以来のデータが関東にのみ置かれている現状が憂慮されている。関東以外の比較的自然災害の起こりにくい複数の場所に、ミラーリングしたデータセンターや部分的なアーカイブを分散して設置することは、結果として遠隔バックアップとして機能し、リスク管理上のメリットがある。分散させるデータセンターの単位はアーカイブする容量に応じて、大学、部局、あるいは研究室と様々なレベルがあり得るが、少なくとも地理的に近接した場所に集中している現状は、可能な限り速やかに改めた方が良好だろう。

## 2. 福島県および会津大学の背景

日本でも、データセンター設置に手を挙げうる機関は複数あると思われるが、ここでは福島県および会津大学の背景について紹介する。

福島県は東日本大震災前からデータセンターの誘致に積極的である。それは、福島県が火力だけでなく水力・風力・地熱といった再生可能エネルギーも含めて多大な電力を関東へと送っており、農林水産業や観光業の振興に加えて電気を用いた産業の育成と誘致も重視してきたためである。その一環として、情報工学に特化した福島県立会津大学を 1993 年に開学し、情報産業の誘致と育成を図っている。

会津大学の特徴のひとつとして、実践を重視した課題解決型学習 (PBL: Project Based Learning) を数多く含むカリキュラムが挙げられる。専任教員の約半数が外国人であり、国境の無い情報産業に特有の国際的環境を形作っている。1000 人余の学生数という小規模校ながら、公立大学の中では突出した大学発ベンチャー企業数で首位にあり (2013 年 4 月現在 25 社)、震災後は復興支援プロジェクトの産学連携を通じて情報系企業の大手が数多く進出している。

宇宙理学と情報工学とが連携する場として、2009 年 4 月には、先端情報科学研究センター宇宙情報科学クラスター (CAIST / ARC-Space) [6][7] が発足しており、はやぶさ・かぐやのデータ解析を支援する可視化ツールの開発や形状モデリングの実績を積みつつ、はやぶさ 2 など宇宙ミッションへの参画や提案も積極的に行っている。開発成果例として、小惑星など不規則形状天体を可視化し 3D-GIS として統合する取組である、平田ほか (2011) 『ポリゴン形状モデルを基盤



とした不規則形状小天体観測データ検索・解析システム』[8]、かぐや月データを念頭に置いた統合解析プラットフォームの例である、寺菌ほか(2012)『月・惑星探査データ表示・解析システム「WISE-CAPS」』[9]、そして産総研と共同開発中のかぐや月面分光データを可視化・解析するツール『月光』[10]などがある。また、国立天文台および宇宙航空研究開発機構とは航空宇宙分野の人材育成を目的とした大学院教育連携協定を締結しており[11]、学生や産業界も巻き込む形で、宇宙データ利用の新規開拓を目的とした宇宙ハッカソンへの参画や運営も行っている[12][13][14]。

2013年には、国と福島県とが出資して復興を後押しするための産学官連携の場として復興支援センター[15]が発足した。福島県民健康データを30年スパンで管理する体制を兼ねつつ、「会津産学官連携クラウド」として企業等も利用可能なIaaSサービスが提供される予定である。この情報基盤は、建物・設備に関する基本要件がJ-Tier3レベルに準拠している。すなわち、建物や火災に対して一般建物より高いレベルでの安全性が耐震設計等で確保され、8時間以上の専任による常駐監視ならびに常駐時間以降は外部委託によるリモート監視をおこなう管理体制をとる。民間のデータセンターと同等のサービス品質を、高いセキュリティのもとで公的機関が中長期に渡って運営するという特徴がある。福島県の復興のため、雇用を生むだけでなく文化的で若者を惹き付けられるデータセンター・プロジェクト候補のひとつとして宇宙開発・宇宙科学は位置づけられており、その誘致に向け福島県をあげた支援が得られている。後述のデータセンターとその連携に関する事務的な体制は既に整っている。共同利用を意識したデータセンター構想の提唱に至る経緯、会津大学の組織対応、直接関連する研究プロジェクトの採択実績と本構想の準備や誘致・提案活動を年表にまとめた(表1)。但し、年表中の参考文献番号は本稿末尾の参考文献と対応している。

表1 データセンター構想に関わる提案諸活動・研究費採択実績

年月日	事案
2007年10月	福島県提供ラジオ番組 知事室からこんにちは(対談:佐藤知事-出村) USGSフラッグスタッフ宇宙地質学チーム(PDS Imaging Node)を手本にして、世界から人が集まる、情報と宇宙を組み合わせた拠点を会津に作る構想の提示。[16]
2007-9年度	科研費基盤A採択 出村ほか 「深宇宙探査機データのGISへの統合と遠隔共同解析技術の研究」、情報工学と地球惑星科学の融合プロジェクト。会津大学ほか3機関。
2008年度	日本惑星科学会、月惑星探査育英会[17]が発足、協力 月惑星探査データ解析ユーザーの裾野を広げる活動として、データ解析実習会の第1回から協力し、現在も継続。
2009年4月	理学・情報工学連携、会津大学CAIST/ARC-Space発足[6][7]
2010年度	日本学術会議大型研究計画マスタープラン改訂公募に「月惑星データセンター」提案 C-SODA、国立天文台RISE、会津大学の3機関連携提案。当時のヒアリングでは、アーカイブ管理基盤が弱い日本の学術分野の状況を踏まえて平座長(当時)から高く評価された。しかし、拠点分散型連携機能と月以外のアーカイブが省かれ、次期月探査プロジェクトにマージされた(学術の大型施設計画・大規模研究計画 マスタープラン 2011[18])。
2011年度	宇宙科学研究所(ISAS) 科学衛星運用・データ利用センター(C-SODA)運営委員会から 『科学データの公開・利用について(2011)』[1]公開
2012年1月	会津大学復興支援センター・クラウドデータセンター構想発表
2012年7月	「惑星科学の今後を考える会」にて研究者コミュニティの議論 ARC-Spaceと他機関の取組が報告・議論され、連携の合意が得られた[19]。
2013年3月	国と県の出資で会津大学復興支援センター発足
2013年3月	日本学術会議大型研究計画マスタープラン再改訂[20]の公募に連携提案 惑星科学会長提案「太陽系生命前駆環境の実証的解明のための統合研究プログラム」、月惑星探査を支えるコミュニティ側から提案されたコンソーシアム構想。会津大学はデータ解析プロジェクト・データセンタ部門として参画。
2013-15年度	科研費基盤B採択 平田ほか 「高精度形状モデルを基盤とした小惑星地質活動の解析」、情報工学(コンピュータビジョン)と惑星科学の融合プロジェクトで、会津大学ほか3機関。
2013年11月	月惑星探査コンソーシアム(仮称)設立準備会合

### 3. 本構想の骨格

本構想は、第1節の科学データ利用の理念[1]に基づく、アーカイブサイエンス促進およびサービスを提供する基盤を、拠点連携で実現しようというものである。具体的には、付加価値をつけたデータプロダクトおよびオンライン地図配信(GIS, Web Mapping Service)、各種解析ツールの開発と配信、解析ノウハウの蓄積と公開、それを通じた人材育成を行うというものである。共同利用研究施設に準じて学外委員を含む運営体制で行うことにより、1拠点に閉じず、大学研究者コミュニティと一体となった基盤のひとつとして機能することが前提である。

現在、日本の研究者コミュニティが利用する JAXA 科学衛星のデータセンターは、C-SODA が運営する DARTS が唯一のものである。JAXA の科学データ利用に関する三原則のうち1) データプロセッシングの原則、2) データ保存の原則、は比較的満足されているが、3) データサービスの原則、については「できるだけ広い範囲のユーザーに使われるようにする」という観点で、海外と比べて課題が残る。現行のプロジェクト縦割りとなっている研究者向けインターフェースやデータベース構成は、新規参入者にとって使いにくく分かりにくい。それらを改善し、データマイニングや解析のための可視化を進める必要性が高まっている。こうしたことの実現には、情報工学分野との連携が必須である。例えば、米国 PDS のトップページ[2]に見られるようなファセット・ナビゲーションが有効であることが、会津大学で行われたハッカソンで示されている[12]。

JAXA の科学データ利用に関する三原則第3項のデータサービスを充実させるには、ニーズが存在する大学コミュニティ、ユーザ側に近く、可塑性の大きい環境である方が有利である。また、宇宙科学分野の裾野を広げていくための人材育成の舞台として、大学は有力な要素である。データ解析ノウハウの蓄積、解析ツールの開発と配布、そして理学・情報工学の研究を通じて月惑星データ利用を促す人材を育成するためには、手堅い DARTS だけでなく、現行の DARTS ラボのような研究段階の試作サービスが極めて重要である。また、データセンターのハードウェア・リソースは集約した方が効率的だが、データプロダクトの質の保証やソフトウェア開発については、それを手がけるチーム単位で行って成果を蓄積・公開する場を用意した方が、ユーザー側に近い分だけ効率的である。自然と様々なサイズのチームが組合わさる形で機能することになるが、相互作用やデータ保管等の冗長性を維持する手間を費やすことで、より科学的成果の豊穡な環境に成長することが期待できる。新しくデータを取ってくることやそのための研究開発も大切だが、米国 PDS を参考にして、既に得られたデータを他の知見と結びつけつつ研究成果を挙げたり、新規ミッションを検討したりすることが大切である。それらは「アーカイブサイエンス」と総称されるもので、日本発のデータでそれを根付かせていかねばならない段階に来ている。

会津大学に DARTS ないしその一部である月惑星ミッションを扱うデータセンターを誘致する形で実現した場合は、第2節で示した復興支援センター内クラウドサービスにミラーリングする形で DARTS ないしその一部の遠隔バックアップを実現する。本構想におけるデータセンターは、枯れた技術で手堅く管理するバックエンド部分と、ユーザビリティを追求し情報工学研究要素のあるフロントエンド部分に分けられる。後者を中心に科学データの一望性・一覧性を向上させるためのインターフェース部分の研究開発を行い、コミュニティへ提供する。データ解析のノウハウを蓄積・公開すると共に、必要なソフトウェア・ツールを PBL で開発し供給する。理学と情報工学の境界領域として、以下に示す形で人材育成を図る。学部・大学院では情報工学教育カリキュラムに宇宙科学と宇宙利用に関するオープンデータの取扱に関する知見を補い、知識基盤社会の中核となる専門人材および産業を支える職業人材を育成する。PD・スタッフ級の人事交流では、理学側には最新の情報工学の知見と開発技術を学ぶ機会を提供し、情報工学側にはオープンデータとしての宇宙科学データを素材とした新規技術の開発機会を提供し、それぞれ成果を挙げる形で、相互触発する環境を形成する。

### 4. おわりに

本構想の趣旨は、JAXA の科学データ利用に関する三原則のうち、海外と比較して手薄である第3項データサービスの原則を拡充することが最大の目的であり、次いで第2項のデータ保存の原則を全うするため現状で関東にしかない DARTS 月惑星データを分散させ大規模災害等でデータが失われないようにしたい、というものである。例えば、日本発の月惑星データを会津大学にデータセンターとして整え、会津大学で開発されつつあるデータサービスと組み合わせる。それにより、日本発のデータの利用価値を高めて月惑星科学アーカイブサイエンスを促進することを意図している。会津大学復興支援センターに整えられる情報基盤を生かすかたちで本データセンター構想を実現できる準備がなされてい

るので、興味関心のある研究者・機関は著者まで連絡を頂ければ幸いである。

日本の月惑星科学分野では、本構想と並行して神戸大学惑星科学研究センター（CPS: Center for Planetary Science）を事務局としたコンソーシアムを形成し、月惑星探査を支えるコミュニティ基盤を整える動きがある。会津大学がこれまで提唱してきたデータセンター構想は、単なるデータの箱物ではなく、上述の通りコミュニティに対するデータサービスを重視した取組である。そのため、そうしたコンソーシアムの一翼を担う形で密接に連携していきたい。

## 参考文献

- [0] 国際惑星データ連合 IPDA: International Planetary Data Alliance  
<http://planetarydata.org>
- [1] 宇宙航空研究開発機構 (JAXA) 宇宙科学研究所 (ISAS) 科学衛星運用・データ利用センター (C-SODA) 運営委員会 (2011) 『科学データの公開・利用について』  
<http://c-soda.isas.jaxa.jp/bibliography/KagakuData.pdf>
- [2] Planetary Data System  
<https://pds.nasa.gov>
- [3] Geoscience Node in Wasington University  
<http://geo.pds.nasa.gov/default.htm>
- [4] Imaging Node in U.S. Geological Survey  
<http://img.pds.nasa.gov>
- [5] NAIF node in JPL  
<http://naif.jpl.nasa.gov/naif/>
- [6] 出村 (2009) PLAIN News no.189 『日本の月惑星探査と科学データアーカイブ第2回 理・情報工学連携拠点：会津大学 ARC-Space の紹介』  
[http://www.isas.ac.jp/docs/PLAINnews/189\\_contents/189\\_1.html](http://www.isas.ac.jp/docs/PLAINnews/189_contents/189_1.html)
- [7] 出村 (2009) 日本惑星科学会誌遊星人 18 巻 3 号 pp 157-158 『会津大学 CAIST/ARC-Space 紹介』  
<https://www.wakusei.jp/book/pp/2009/2009-3/157.pdf>
- [8] 平田ほか (2011) 宇宙科学情報解析論文誌 vol.1 pp.133  
『ポリゴン形状モデルを基盤とした不規則形状小天体観測データ検索・解析システム』  
<http://repository.tksc.jaxa.jp/pl/dr/AA0065236017>
- [9] 寺藺ほか (2012) 宇宙科学情報解析論文誌 vol.2 pp.89  
『月・惑星探査データ表示・解析システム「WISE-CAPS」』  
<http://repository.tksc.jaxa.jp/pl/dr/AA0061914010>
- [10] 林ほか (2013) 第46回月・惑星シンポジウム集録 (in press)  
『かぐや搭載スペクトルプロファイラデータの検索・表示ツールの開発』
- [11] 航空宇宙分野の人材育成で JAXA と連携  
<http://www.u-aizu.ac.jp/events/agreement-jaxa.html>
- [12] 出村ほか (2012) 日本惑星科学会誌遊星人 21 巻 4 号 pp386-388 『宇宙ハッカソン報告』  
<https://www.wakusei.jp/book/pp/2012/2012-4/2012-4-386.pdf>
- [13] 宇宙科学ミッションのデータを活かそう ―会津大学で「宇宙ハッカソン」― 2012 年 8 月 28-31 日  
<http://www.u-aizu.ac.jp/research/caist/arc-space/uoa-spacehackathon.html>
- [14] ARC-Space 学生チーム開発「月旅行アプリ」に審査員特別賞！ 2013 年 4 月 20-21 日  
<http://www.u-aizu.ac.jp/events/arc-spaceisac.html>
- [15] 会津大学 復興支援センター  
<http://www.u-aizu.ac.jp/research/uarc/uarc-mission.html>

- [16] 2007 年 10 月 福島県 知事室からこんにちは（対談：佐藤知事-出村）  
[http://wwwcms.pref.fukushima.jp/pcp\\_portal/PortalServlet;jsessionid=79D4F95AC018E1702B41A618229F2093?DISPLAY\\_ID=DIRECT&NEXT\\_DISPLAY\\_ID=U0000004&CONTENTS\\_ID=11168](http://wwwcms.pref.fukushima.jp/pcp_portal/PortalServlet;jsessionid=79D4F95AC018E1702B41A618229F2093?DISPLAY_ID=DIRECT&NEXT_DISPLAY_ID=U0000004&CONTENTS_ID=11168)
- [17] 月惑星探査育英会 データ解析実習会  
<https://www.cps-jp.org/~tansaku/wiki/top/>
- [18] 学術の大型施設計画・大規模研究計画 マスタープラン 2011  
<http://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-21-h135-1.pdf>
- [19] 小林ほか (2012) 21 巻 3 号 pp324-329 『「惑星科学の今後を考える会」の報告』  
<https://www.wakusei.jp/book/pp/2012/2012-3/2012-3-324.pdf>
- [20] 第 22 期学術の大型施設計画・大規模研究計画に関するマスタープラン「学術大型研究計画」の公募  
<http://www.scj.go.jp/ja/member/iinkai/ogata/22-koubo.html>

# 科学研究用クラウドシステム（NICTサイエンスクラウド）の提案

村田 健史<sup>\*1</sup> 渡邊 英伸<sup>\*1</sup> 鶴川 健太郎<sup>\*2</sup> 村永 和哉<sup>\*2</sup> 鈴木 豊<sup>\*2</sup> 磯田 総子<sup>\*3</sup>  
山本 和憲<sup>\*1</sup> 久保田 康文<sup>\*1</sup> 長妻 努<sup>\*1</sup> 坂口 歌織<sup>\*1</sup> 津川 卓也<sup>\*1</sup> 西岡 未知<sup>\*1</sup>  
建部 修見<sup>\*4</sup> 田中 昌宏<sup>\*4</sup> 深沢 圭一郎<sup>\*5</sup> 才田 聡子<sup>\*6</sup> 海老原 祐輔<sup>\*7</sup> 藤田 茂<sup>\*8</sup>  
木村 映善<sup>\*9</sup> 黒澤 隆<sup>\*10</sup> 村山 泰啓<sup>\*1</sup> 永井 亨<sup>\*11</sup> 水原 隆道<sup>\*12</sup>

## The NICT Science Cloud — A Proposal of Cloud System for Scientific Researches —

Ken T. MURATA<sup>\*1</sup> Hidenobu WATANABE<sup>\*1</sup> Kentaro UKAWA<sup>\*2</sup> Kazuya MURANAGA<sup>\*2</sup> Yutaka SUZUKI<sup>\*2</sup>  
Fusako ISODA<sup>\*3</sup> Kazunori YAMAMOTO<sup>\*1</sup> Yasufumi KUBOTA<sup>\*1</sup> Tsutomu NAGATSUMA<sup>\*1</sup> Kaori SAKAGUCHI<sup>\*1</sup>  
Takuya TSUGAWA<sup>\*1</sup> Mlchi NISHIOKA<sup>\*1</sup> Mamoru ISHII<sup>\*1</sup> Shinichi WATARI<sup>\*1</sup> Osamu TATEBE<sup>\*4</sup>  
Masahiro TANAKA<sup>\*4</sup> Keiichiro FUKAZAWA<sup>\*5</sup> Satoko SAITA<sup>\*6</sup> Yusuke EBIHARA<sup>\*7</sup> Shigeru FUJITA<sup>\*8</sup>  
Eizen KIMURA<sup>\*9</sup> Takashi KUROSAWA<sup>\*10</sup> Yasuhiro MURAYAMA<sup>\*1</sup> Masao OGINO<sup>\*11</sup> Takamichi MIZUNAGA<sup>\*12</sup>

### Abstract

This paper is to propose a cloud system for science, which has been developed at NICT (National Institute of Information and Communications Technology), Japan. The NICT science cloud is an open cloud system for scientists who are going to carry out their informatics studies for their own science. The NICT science cloud is not for simple uses. Many functions are expected to the science cloud; such as data standardization, data collection and crawling, large and distributed data storage system, security and reliability, database and meta-database, data stewardship, long-term data preservation, data rescue and preservation, data mining, parallel processing, data publication and provision, semantic web, 3D and 4D visualization, out-reach and in-reach, and capacity buildings.

Keyword Science Cloud, Data-intensive Science, Informatics, Virtual Laboratory

### 概要

データ指向型科学は、実験科学、理論科学、数値シミュレーション科学に続く第4の科学研究パラダイムと言われている。サイエンスクラウドは、データ指向型科学の基盤環境として期待されているが、現在、サイエンスクラウドとは何か、サイエンスクラウドには何ができるかなどの議論が始まったばかりである。NICTサイエンスクラウドは、情報通信研究機構が2010年より構築を進めている科学研究専用のクラウドシステムである。本稿では、NICTサイエンスクラウドについて宇宙科学研究をはじめとした科学研究の利活用面から議論することで、サイエンスクラウドの定義を行うとともに、今後、サイエンスクラウドを活用した科学研究手法・環境構築についての筋道を示す。

キーワード サイエンスクラウド、データ指向型科学、インフォマティクス、仮想研究室

- 
- \*1 情報通信研究機構 (National Institute of Information and Communications Technology)
  - \*2 株式会社セック (Systems Engineering Consultants Co., Ltd.)
  - \*3 株式会社サイエンス・サービス (Science Service Co., Ltd.)
  - \*4 筑波大学計算科学研究センター (Center for Computational Sciences, University of Tsukuba)
  - \*5 九州大学情報基盤研究開発センター (Research Institute for Information Technology, Kyushu University)
  - \*6 統計数理研究所 (Institute of Statistical Mathematics)
  - \*7 京都市立大学生存圏研究所 (Research Institute for Sustainable Humanosphere, Kyoto University)
  - \*8 気象大学校 (Meteorological College)
  - \*9 愛媛大学医学部 (Department of Medical Informatics Ehime University)
  - \*10 株式会社日立ソリューションズ東日本 (Hitachi Solutions East Japan, Ltd.)
  - \*11 名古屋大学情報基盤センター (Information Technology Center, Nagoya University)
  - \*12 株式会社クレアリンクテクノロジー (CLEALINK TECHNOLOGY Co., Ltd.)



## 1. まえがき

これまでに発展してきた実験科学（第1の柱）、理論科学（第2の柱）、計算科学（第3の柱）に対して、データ指向型科学は科学的発見のための第4の柱であると Jim Gray は提唱した<sup>1)</sup>。筆者（村田）は、実験科学を支えるインフラストラクチャは実験装置や観測装置（宇宙科学研究では科学衛星による観測やロケット実験など）であり、計算科学を行う基盤となるのがスーパーコンピュータであるのに対して、データ指向型科学を支えるインフラとして提案するのが、科学研究専用のクラウドシステム（サイエンスクラウド）であると考えている<sup>2,3)</sup>。

サイエンスクラウドは、2008 年ごろにイリノイ大学によって提唱された<sup>4)</sup>。2010 年に第1回サイエンスクラウドワークショップ (ScienceCloud Workshop) が米国・シカゴにおいて開催された<sup>5)</sup>。その後、毎年1回の同ワークショップが開催され、サイエンスクラウドについての議論が進められてきた<sup>6,7)</sup>。これらのワークショップにおいてなされた議論は、クラウド基盤上での科学研究事例検討、クラウド環境および技術の有効性評価、クラウドシステムの耐障害性（fault tolerance）および信頼性、クラウド上でのデータ指向型研究の工程やツール、Map Reduce などのプログラミングモデルの利用、クラウドストレージ設計、クラウド内での I/O やデータ管理、クラウド内のワークフローやリソース管理、NoSQL などのクラウド技術の科学研究への利用、データストリーミングや動的アプリケーション、HPC 環境におけるクラウドコンセプトの応用、高機能並列分散ファイルシステムと仮想環境接続、クラウドセキュリティーに関する研究および実験などと多岐にわたる（表1）。

表1 第1回 ScienceCloud ワークショップでの議題一覧 (ScienceCloudWorkshop2010<sup>5)</sup>)

Scientific application cases studies on cloud infrastructure
Performance evaluation of cloud environments and technologies
Fault tolerance and reliability in cloud system
Data-intensive workloads and tools on clouds
Use of programming models such as Map-Reduce and its implementations
Storage cloud architectures
I/O and Data management in the cloud
Workflow and resource management in the cloud
Use of cloud technologies (e.g., NoSQL databases, etc) for scientific applications
Data streaming and dynamic applications on clouds
Application of cloud concepts in HPC environments
High performance parallel file systems and interconnects in virtual environments
Research and best practices in cloud security

これらの議論からわかることは、データ指向型科学研究のインフラストラクチャとしてサイエンスクラウドは重要であるという共通認識はあるものの、多くのクラウド技術は商用クラウドに共通したものであり、それらの技術を科学研究に流用しているにすぎないという点である。すなわち、技術的議論を通じての多くの参加者の興味や目的は、サイエンスクラウドとは何か、いわばサイエンスクラウドの定義を行う事である。サイエンスクラウドが、ビッグデータ科学のインフラストラクチャとして位置付けられることには誰も異論がないが、具体的に何をどのように問題解決するものであるかという事が明確ではない。

本稿ではサイエンスクラウドの個別の技術について議論するのではなく、利用面からのサイエンスクラウドを提案する。サイエンスクラウドを使った研究手法（研究スタイル）を規定することで、サイエンスクラウドにより期待される科学研究成果や、サイエンスクラウドに求められる個別の技術が明らかになるからである。

## 2. サイエンスクラウドとビッグデータ

### 2.1. サイエンスクラウドの事例

現在、欧米諸国を中心に様々なサイエンスクラウド（ScienceCloud）またはそれに準ずるシステムが提案または運用さ

れている（表2）．サイエンスクラウドは、前述のとおりその定義が議論されている状況であり、これらのシステムを通じてその有効性や必要とされる技術が検討されている．また、表2を含む多くのシステムの中で、特定の研究テーマについてテストケースとしての成果が挙げられているものはあるが、複数の研究テーマで大きな成果を挙げた汎用性の高いサイエンスクラウドはまだ存在しない．

表2 世界のサイエンスクラウドの動向

国名または組織名	クラウド名またはグループ名	URL	概要
CERN, CNR-IREA	HELIX NEBULA	<a href="http://helix-nebula.eu/">http://helix-nebula.eu/</a>	To create a multi-tenant Open Market Place for Science
INRIA Rennes (France)	Centre de recherche commn	<a href="http://www.inria.fr/centre/rennes">http://www.inria.fr/centre/rennes</a>	Scalable Storage for Sharing Application Data on Azure Clouds: TomusBlobs
	The Venus-C project	<a href="http://venusfeedback.codeplex.com/">http://venusfeedback.codeplex.com/</a>	
	SCIDP-ES	<a href="http://www.scidip-es.eu/">http://www.scidip-es.eu/</a>	Long-term data preservation for Earth Science
(Lavanya Ramakrishnan)	FRIEDA	<a href="http://frieda.lbl.gov/">http://frieda.lbl.gov/</a>	Flexible Robust Intelligent Elastic Data Management in Cloud Environments
	The Open Science Data Cloud (OSDC)		A petabyte-scale science cloud managed and operated by the Open Cloud Consortium (OCC)
	“Science Clouds”	<a href="http://scienceclouds.org/">http://scienceclouds.org/</a>	Science Clouds Blog

## 2.2. ビッグデータ処理の基盤技術

IT ビジネス分野の中心となりつつあるビッグデータという概念は、科学研究分野でも適用され始めている．1 節で議論した第3の研究手法である数値シミュレーションなどにおいて大規模計算の中核をなす技術の一つが HPC（High Performance Computing）である．HPC は計算指向型（compute-intensive）と言われており、個別のアプリケーション（プログラムや数値計算コード）の処理速度を最大化することが主目的となる．しかし、データ指向型（data-intensive）科学では、数値シミュレーションを含むあらゆる科学データを処理することが目的である．

一方、多様な計算機リソースを融合し、データ分散、並列データ処理やコンピュータとデータファイルをローカライズする工夫など、総合的なデータ処理環境をめざすのが MTC（Many-Task Computing）という考え方である．クラウド計算機リソースによる並列分散処理などがこれに該当する．

また、計算処理ではなくデータに視点を置いた考え方で提案されているのが DIC（Data-intensive Computing）である．Jim Gray が提言した第4のパラダイムであるデータ指向型科学では、数値シミュレーションを含むあらゆる科学データを処理することが目的である．クラウド内に配置された大規模科学データを解析するビッグデータ科学のためには、DIC 技術の向上が必須である．

特定パターンの商用サービスを実現することが多い民間クラウドと比較すると、複雑で特殊な目的実現型のサイエンスクラウドでは HPC はその一部であり、MTC 環境の実現が重要となる．

## 2.3. NICT サイエンスクラウド

1 節や 2.1 節で述べたように、現在欧米を中心にサイエンスクラウドに関する取り組みが行われているが、特定の研究目的に設計されたものや、民間のクラウドを模したシステムも多い．その結果、汎用性が高く多数の大規模プロジェクトで成果を達成したサイエンスクラウドはこれまでに存在しない．

本論文で議論する NICT サイエンスクラウドは、あらゆる科学分野でほとんどのデータがデジタル化されている現在、科学研究の第1の柱、第2の柱、第3の柱<sup>1)</sup>によるあらゆる科学データを収集・保存・解析（可視化）・公開データ処理するデータ指向型科学研究環境として設計された（図1）．サイエンスクラウド構築においては、利用者（すなわち研究者）の視点が欠かせない．NICT サイエンスクラウドはユーザ指向型の設計を行うため、2010 年ごろからプロトタイプを構築してユーザ利用を進めると並行して、ユーザからのヒアリングを積極的に行った．

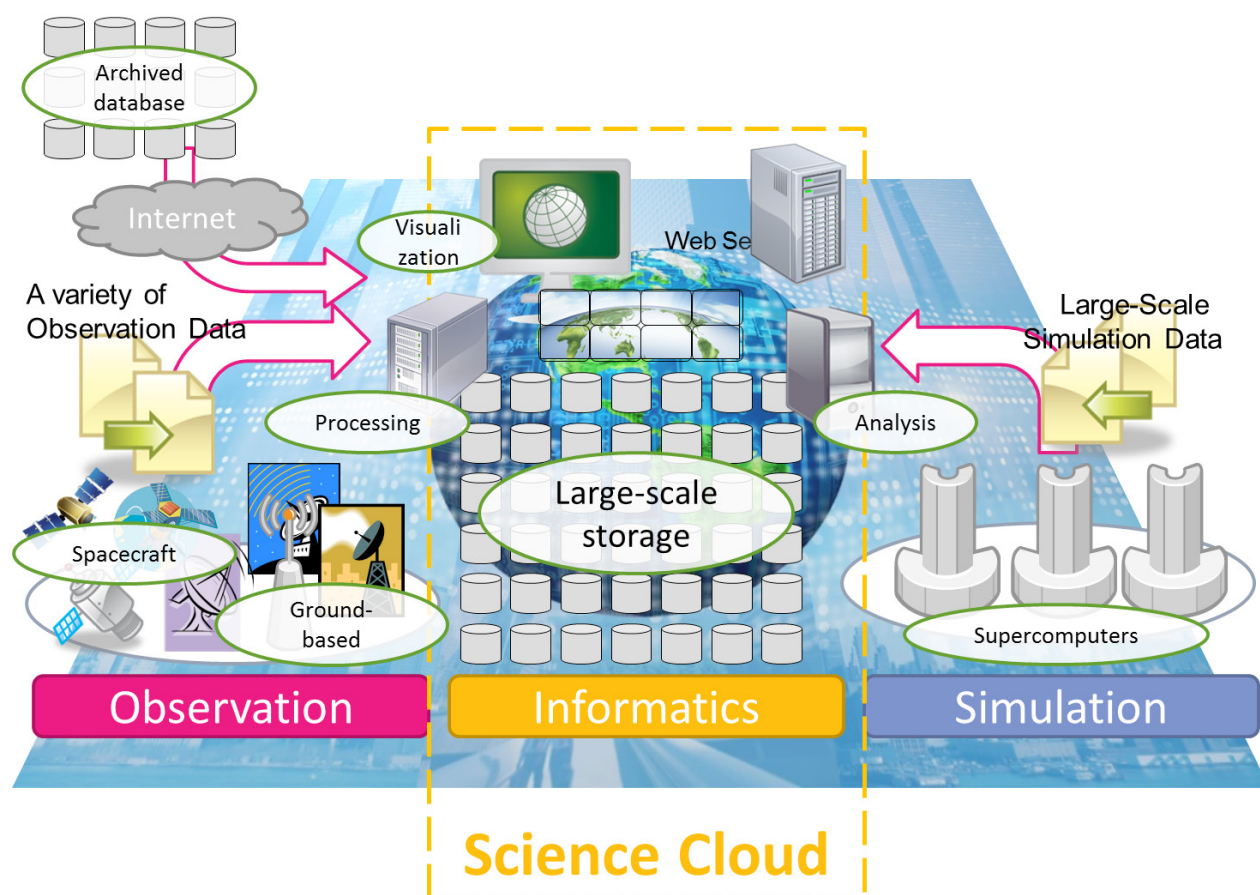


図 1 NICT サイエンスクラウド基本コンセプト

世界科学データシステム（WDS: World Data System）は、科学界の国連と呼ばれる国際科学会議（ICSU: International Council for Science）が実施しており、科学データに関する国際的取組の高度化を目指すプログラムである。情報通信研究機構（NICT）と ICSU は、本プログラムの実施主体となる国際プログラムオフィス（IPO: International Programme Office）に関する協力文書を 2011 年に締結した<sup>8)</sup>。本研究ではこの活動を背景として、広く科学研究者が要求または期待する要件を収集した。これまでに調査したサイエンスクラウドに求められる機能またはサービス要件を分析すると、クラウドサービスとしての機能とデータ指向型科学研究のインフラストラクチャの 2 つの側面があることが分かった。

クラウドサービス機能は、サイエンスクラウドの利便性に関わる機能である。民間クラウドと共通する機能であり、研究者がサーバ、ネットワーク、ユーザ管理等の管理から解放されることで研究の利便性が向上する側面である。民間クラウドと異なるのは、研究用のサービス（アプリケーション、ツール、システム等）が有効である点である。民間クラウドは、一般には計算機リソースや商用アプリケーションサービスの提供が主体となるが、サイエンスクラウドは研究に特化したサービスや研究用にカスタマイズできるツールを提供することで有効性が高まり、また商用クラウドとの差別化も行われると考えられる。NICT サイエンスクラウドの研究支援サービスについては、4 節において議論する。

データ指向型科学研究のインフラストラクチャ機能は、サイエンスクラウドの高度利用に関わる機能である。その中でも、特に大規模データ処理（ビッグデータ科学）と複合型データ処理が要求される機能の中心である。大規模データ処理では、単に大規模なストレージにデータを保存するだけでなく、それらのデータの収集、管理、公開と処理が期待される。また、複合型データ処理では、データ観測や数値計算、可視化やアウトリーチなど、サイエンスクラウド上においてすべてのプロセスを実現することが期待される。データ指向型科学研究事例については、5 節において議論する。

大型科学研究プロジェクト推進では、協調的研究環境としてのサイエンスクラウド利活用も重要である。近年、ほとんどの大型研究計画は国際プロジェクト化されており、高い成果を挙げるためには国際的協調研究環境が必須である。NICT サイエンスクラウドの重要な機能の一つが、国際的協調研究を行うための仮想研究室（バーチャルラボラトリー）機能（図 2）である。国際的協調研究環境を NICT サイエンスクラウド上に構築することにより、これまでよりもはるか



に短い期間で、多くの研究成果を生み出すことが期待される。バーチャルラボラトリーの実例については、5 節において議論する。

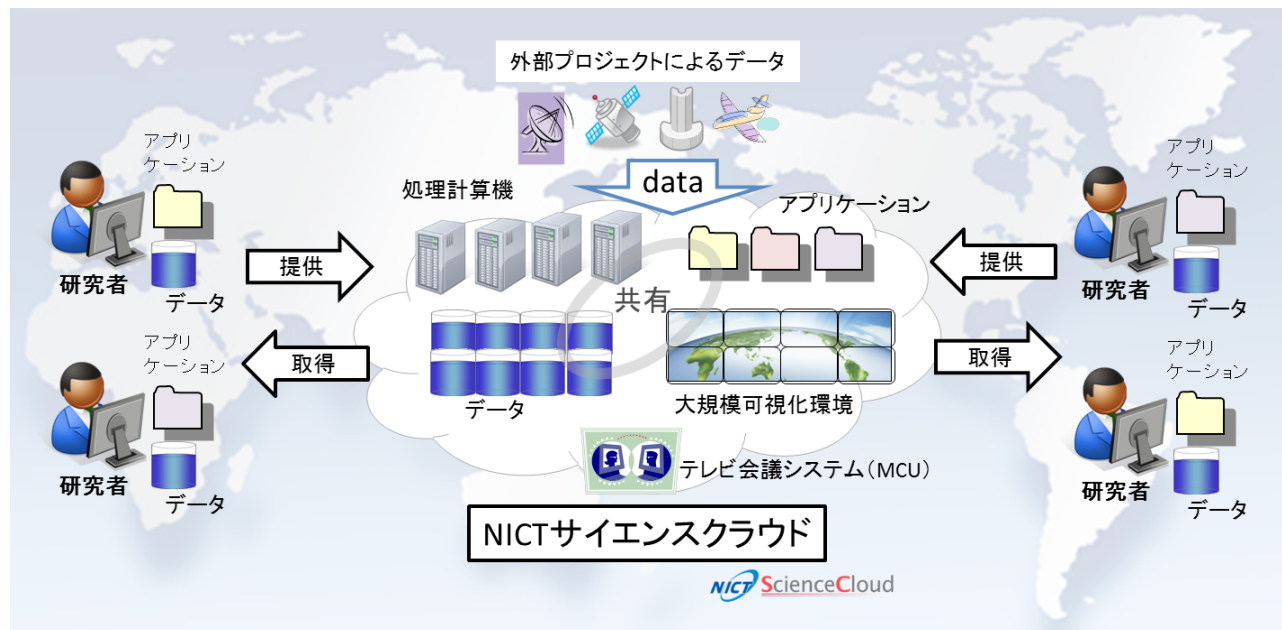


図2 仮想ラボラトリーによる協調的研究開発

### 3. NICTサイエンスクラウドのシステム概要

本節では、NICTサイエンスクラウドのシステム概要について議論する。NICTサイエンスクラウドは、広域分散型のクラウドシステムであり<sup>9)</sup>、2013年現在で国内5か所のDC（データセンター）をL2接続することで構成されている（図3）。インターコネクトには、情報通信機構の広域テストベッドネットワークであるJGN-X<sup>10)</sup>を用いている。

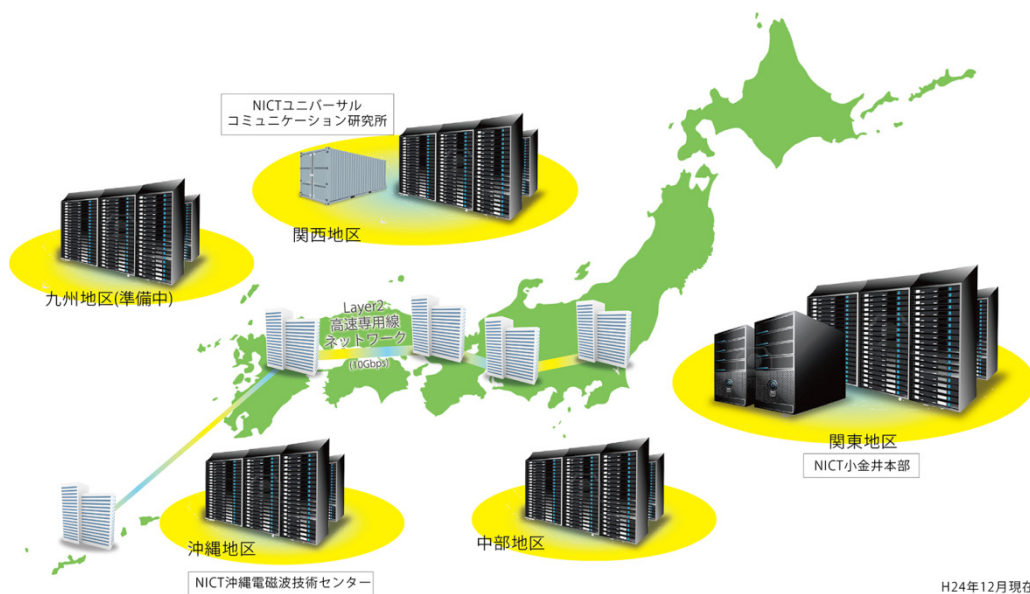


図3 NICTサイエンスクラウドデータセンター

図4に、NICTサイエンスクラウドのシステム構成図を示す。図4はクラウドリソース全体を表しており、すべてのクラウドユーザが利用できる安定で冗長化されたクラウド共通リソースと、ビッグデータ処理または定常データ処理を対象とするクラウド拡張リソースから構成される。NICTサイエンスクラウドは、図4に加えて、4節で述べる各種研究支

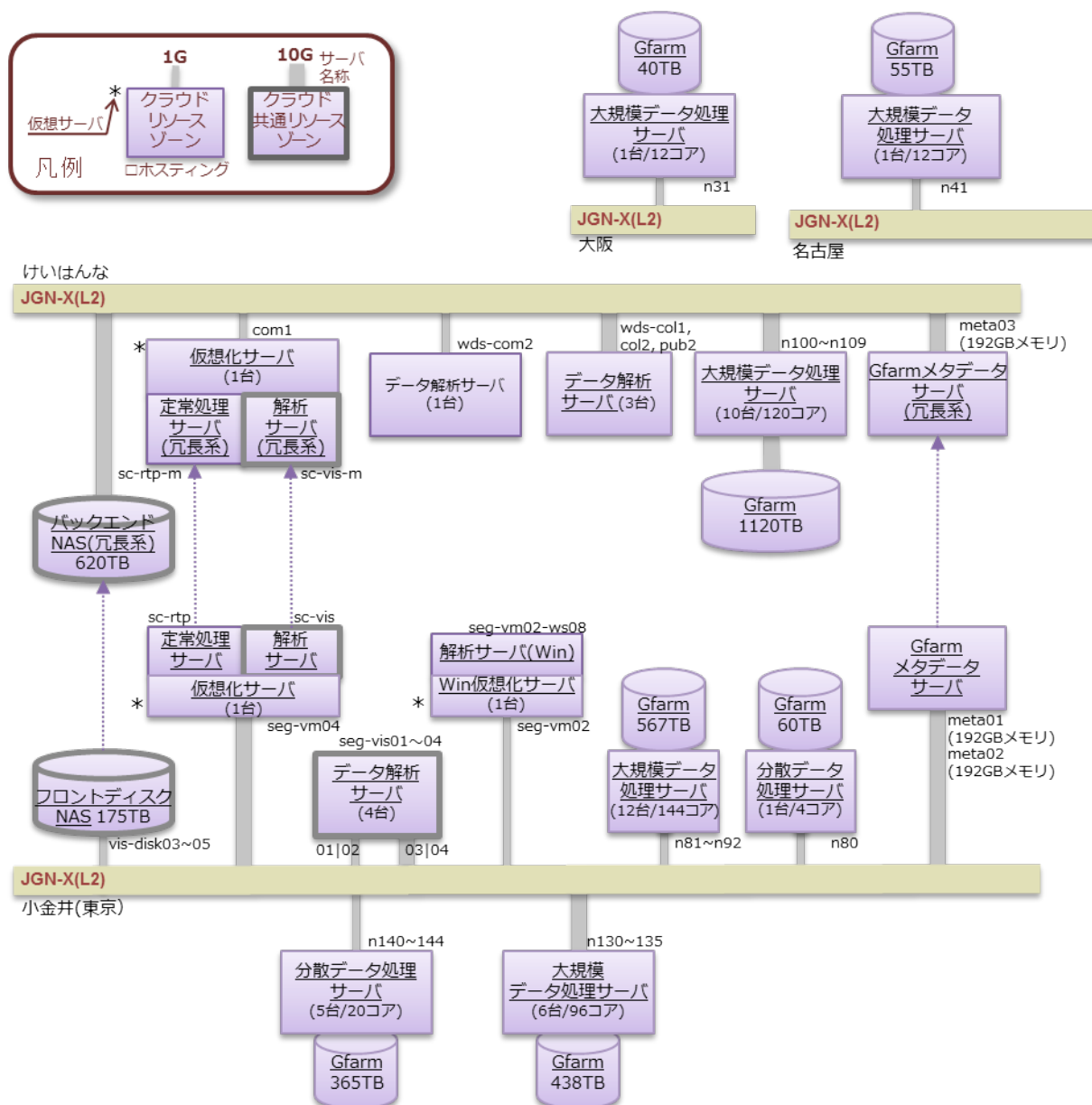


図4 NICTサイエンスクラウドシステム構成図

援サービスを行うサーバ群と4.6節に示す独自研究環境（ハウジング・ホスティングサーバ群）から構成されているが、本節ではクラウドリソースについてのみ議論する。

図5は、NICTサイエンスクラウドリソース利用を、クラウドリサービス利用と研究開発システム利用に分けて示している。多くの研究者は、クラウド利用を小規模な研究からスタートする。図6に示すクラウド的利用であり、商用クラウド利用と比較的似た利用である。クラウド的利用の場合には、安定性・冗長性に優れた共通リソースの利用が必須である。データの保存が多重化されているだけでなく、システム停止の頻度も小さいため、継続的な研究環境が保障されるからである。また、NICTサイエンスクラウド側がファイアウォール（IPSを含む）、ユーザ管理（LDAP）、データファイルのウィルススキャンなどを行うことで、研究者（ユーザ）の利便性は独自の研究環境と比較して高くなる。また、NICTサイエンスクラウドにライセンスされた民生品の研究用アプリケーションが利用できることも、サイエンスクラウドの利便性の一つである。これまでは、多くの研究者が個別にこのような研究環境を構築・運用してきたが、NICTサイエンスクラウドを用いることで構築・運用から解放され、研究に集中できることが期待される。

研究が進展すると、NICTサイエンスクラウド利用目的は、単純な利便の利用から、処理データの大規模化や特殊化、多目的化などに発展することが多い。また、4.6節で述べるように独自研究環境をサイエンスクラウドに移植し、クラウ



ドリソースを活用して目的を達成することもある。NICTサイエンスクラウドを活用した研究開発システム利用（図5）については、利活用する研究支援サービスを4節で、利活用事例を5節で述べる。

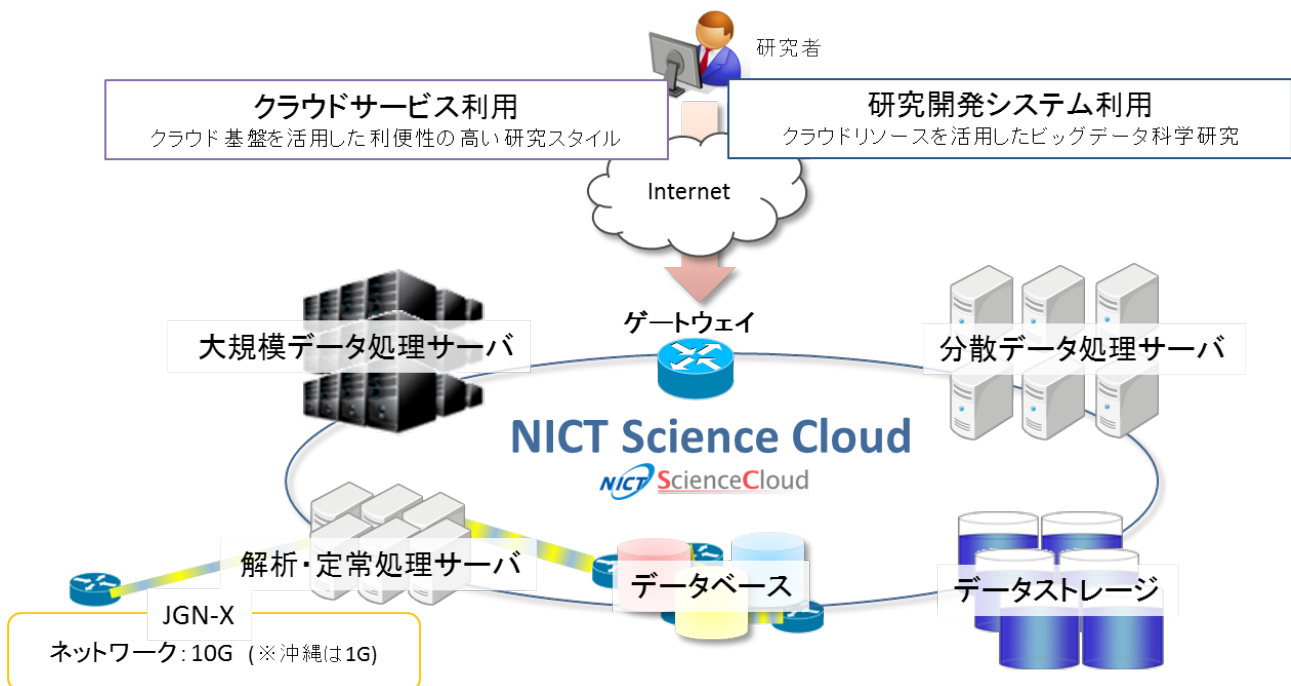


図5 NICTサイエンスクラウド利活用概要

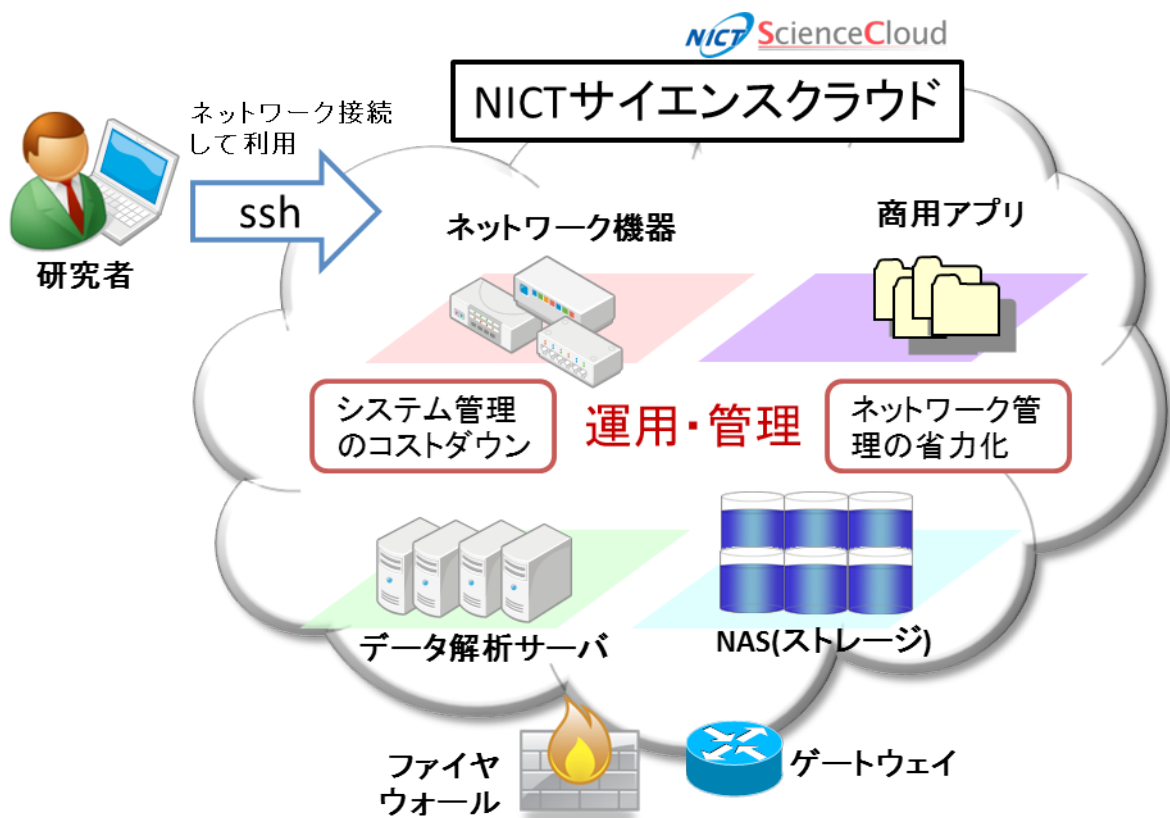


図6 クラウド研究環境利活用（クラウド的利用）

## 4. NICTサイエンスクラウドの研究支援サービス

### 4.1. クラウドエコシステム

商用クラウドでは、多くのソリューションが特定ベンダの独自技術と仕様に基づいており、その結果として囲い込み（ベンダロックイン）がユーザの利便性を損なうことが多かった。この状況を改善するため、ソフトウェア開発者、SIer、クラウド事業者、さらにクラウドユーザから構成されるステークホルダーによる、オープンソースのクラウド基盤ソフトウェアの利用と協業によるクラウドサービスやクラウドソリューションの提供という流れがある。オープンクラウドで市場を活性化し、共存共栄で成長するクラウドの生態系をつくり出そうという取り組みはオープンクラウドエコシステムと呼ばれている。

### 4.2. 科学研究用エコシステム

エコシステムは商用クラウドで提案された概念であるが、サイエンスクラウドでもエコシステムに取り組む事例もある<sup>11)</sup>。Helix Nebulaのエコシステムは、科学データから価値のある情報を抽出することを目的としており、これにより科学研究者と民間産業を結びつける試みである。

しかし、NICTサイエンスクラウドユーザからのヒアリングによると、クラウド基盤を使った新しいサイエンス（特にビッグデータ科学）を行う場合には研究テーマごとの特殊性が高く、エコシステムの実現は容易ではない。ただし、商用クラウドと比較すると専門性は高いが、図7に示すような科学研究分野で汎用性があるアプリケーション、ツールおよびシステムは可能である。NICTサイエンスクラウドではこのような汎用的な科学研究用アプリケーション・ツール・システム（科学研究用エコシステム）を開発し、公開してきた（表3）。本節では、それらの一部を紹介し、その有効性や汎用性について議論する。

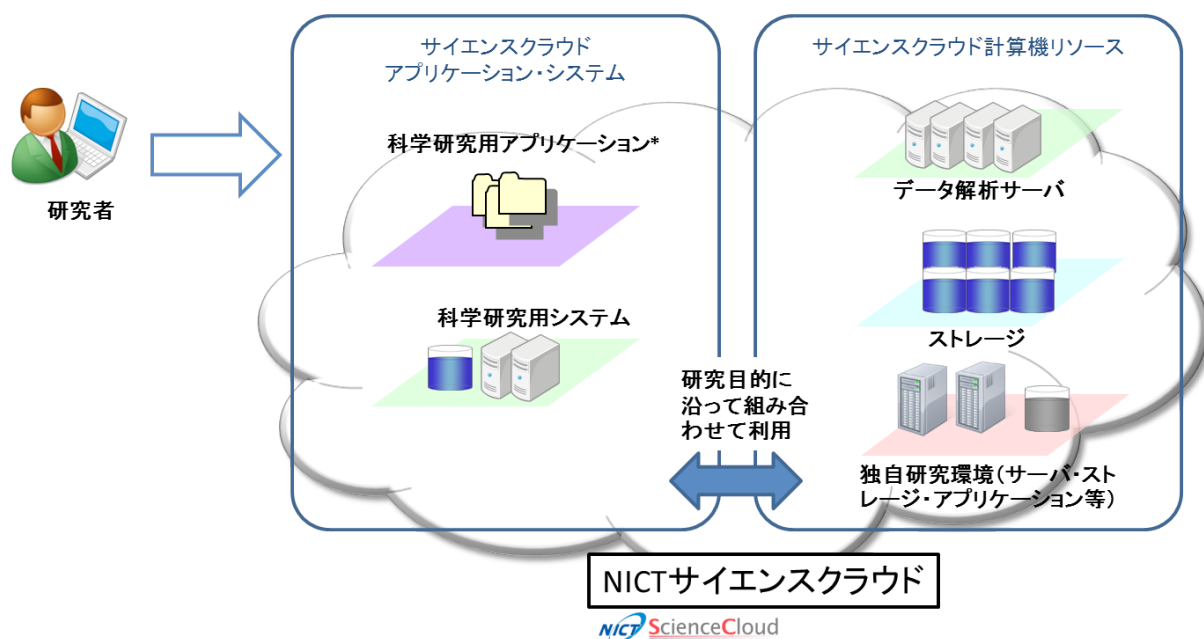


図7 科学研究用アプリケーション・システム利活用

### 4.3. 科学データファイルクローラ（自動収集）サービス（NICTY/DLA）

サイエンスクラウドにおいてデータ処理を行うために重要な機能の一つが、データ自動収集である。現在、科学研究データの多くはインターネット上で公開されているが、公開されるデータ量は多くの分野で増加している。図8は、NICTが収集している国内外の公開GPS（GNSS）受信機データ数の経年変化である。現在、5000を超えるGPS受信機の観測データが公開されており、公開サイトも20を超えている。このような公開データの多くは、各組織が独自のポリシーに従って公開している。データ更新の頻度、公開プロトコルはもちろんのこと、場合によってはデータ較正の変更のために公開サイト上の過去データファイルが連絡なく更新されることもある。また、サイト上で公開されるデータファイル形式、ファイル名命名方式、データ公開範囲やアクセス制限なども、公開サイトによって異なる場合がある。

表 3 NICT サイエンスクラウド科学研究用エコシステム利用例<sup>21)</sup>

科学研究用エコシステム名	概要
NICTY/DLA	インターネット上で公開されている科学データ（特に時系列観測データ）のメタ情報および実データファイルをクロールするツール。
バーチャルオーロラツール	地球および惑星磁気圏・電離圏の MHD シミュレーション（多流体系を含む）を 3 次元時系列可視化するためのツール。AVS Express（Linux および Windows）で動作する。
広域観測ネットワーク監視システム（WONM システム）	地球広域の観測網において観測所・観測装置からのデータ伝送を継続的に行うための監視システム。データ伝送サーバの監視、データ伝送状況の監視と再送、ネットワーク状態のモニタリングなどを行う。
WSDBank アプリケーション	サイエンスクラウド上にデータを保存または公開するための Web アプリケーション。（データファイルは NICT サイエンスクラウドのフロントエンドストレージに保存され、定期的にバックエンドストレージにバックアップされる。）
STARS（Solar-Terrestrial data Analysis and Reference System）	各種衛星観測・地上観測データを統合的に解析するツール。全対象データファイルをカプセル化したクラスライブラリ（SEDOC）のみを利用することもできる。

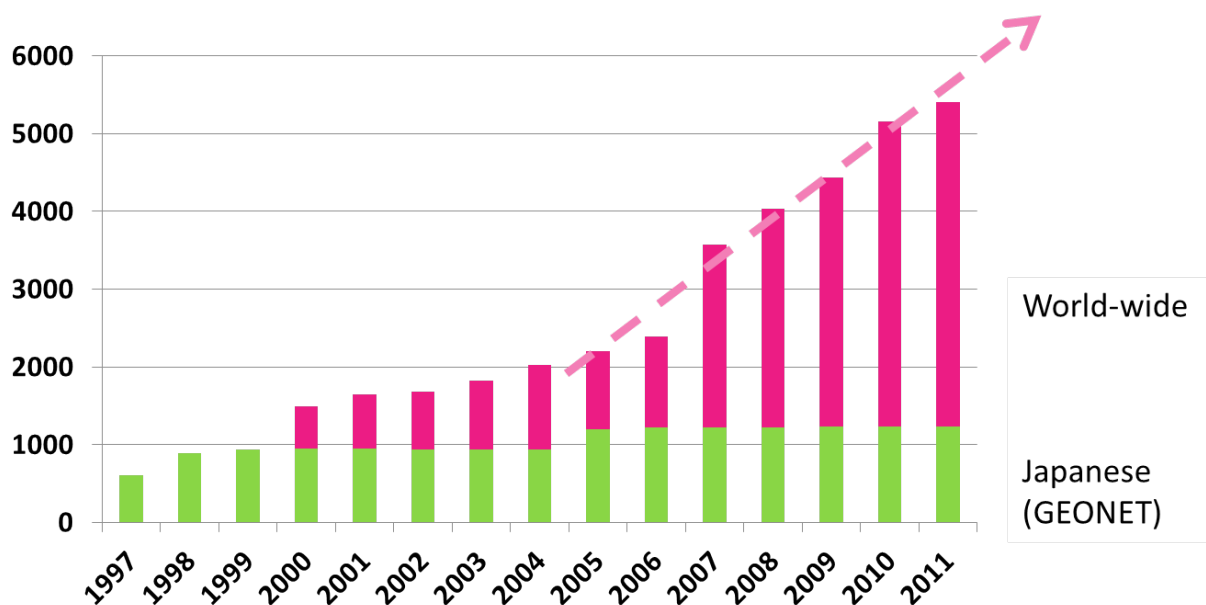


図 8 NICT が収集している国内外の公開 GPS 受信機データ数のトレンド

NICTY/DLA は、RSS によりメタ情報を記述し、FTP/HTTP などの標準的なプロトコルに対応することで、公開データの独自性や流動性を吸収できる<sup>12)</sup>。図 9 に NICTY/DLA（Download Agent）の機能概要を示す。NICTY では、あらかじめ対象となるデータの情報（ホスト名、ディレクトリ名、ファイル名フォーマットや公開ポリシーなど）を登録することで公開データのメタ情報をクロール・データベース (DB) 化し、これをもとにして各種プロトコルで公開されている実データファイルのダウンロード・DB 化を行う<sup>12)</sup>。

NICTY/DLA により独立して収集されたデータファイルは多様であり、多数のデータを解析するユーザがそれらのデータファイルから容易に情報を取り出すことができる環境が重要である。しかし、太陽地球系物理分野を含む関連する分野の科学データ記述形式を統一化することは容易ではない。NICT サイエンスクラウドでは、収集するデータのうち 400 種類のデータをカプセル化したクラスライブラリ (SeDOC クラス) を準備した<sup>12)</sup>。これによりユーザは、データ ID とデータ期間 (StartTime と EndTime) を指定するだけで、データファイル形式やデータの場所を意識することなく get 関数によりデータを容易に取得できる。SeDOC クラスは C++ 言語により記述されており、Java 言語からの呼び出しも可能である。

図 10 は、NICTY/DLA によるデータファイルの自動収集実績である。NICT サイエンスクラウドでは、NICTY/DLA をシステムとして実装し、定常運用を行っている。2013 年 4 月現在で、47 ホスト (428 データ) に対応した。これらのデータ対応において RSS に新たな定義を追加することで、順次データ対応を進めた。これまでに NICTY/DLA の収集対象と

なった 428 データにおいては、データポリシを RSS で吸収できなかった例はない。図 10 には、NICT サイエンスクラウドが NICTY/DLA により自動収集しデータベース化した科学研究データファイルのファイル数とデータサイズも示した。1200 万を超える観測データファイルと 7TB を超えるデータファイルが NICT サイエンスクラウド上に蓄積され、現在も最新の公開データが 6000 ～ 10000 ファイル/日で収集されている。

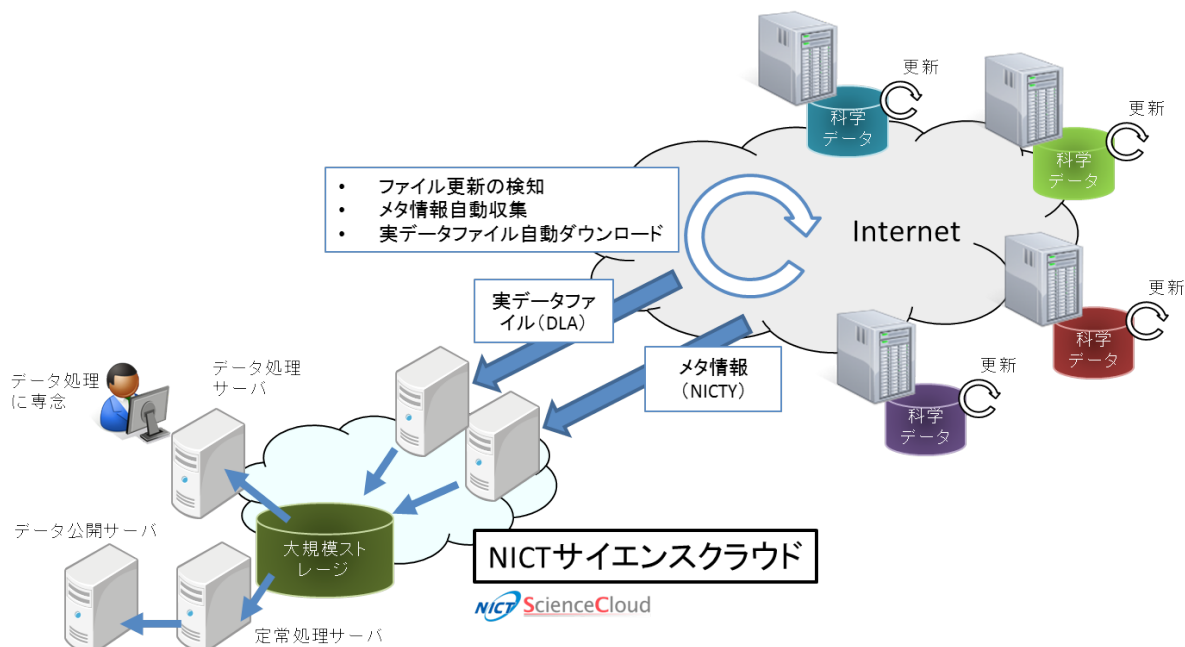
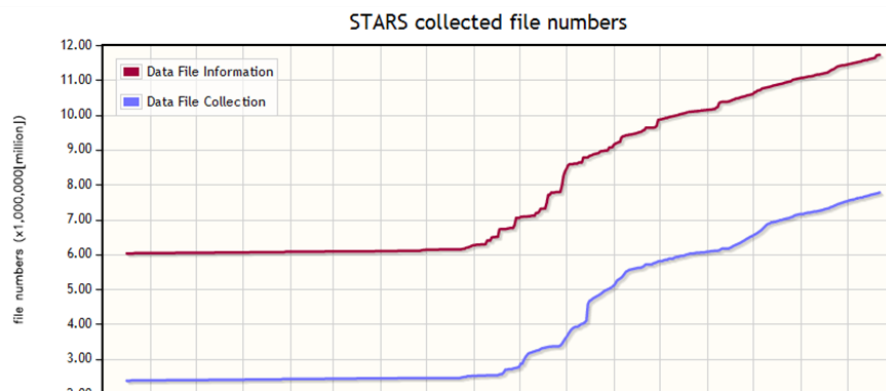
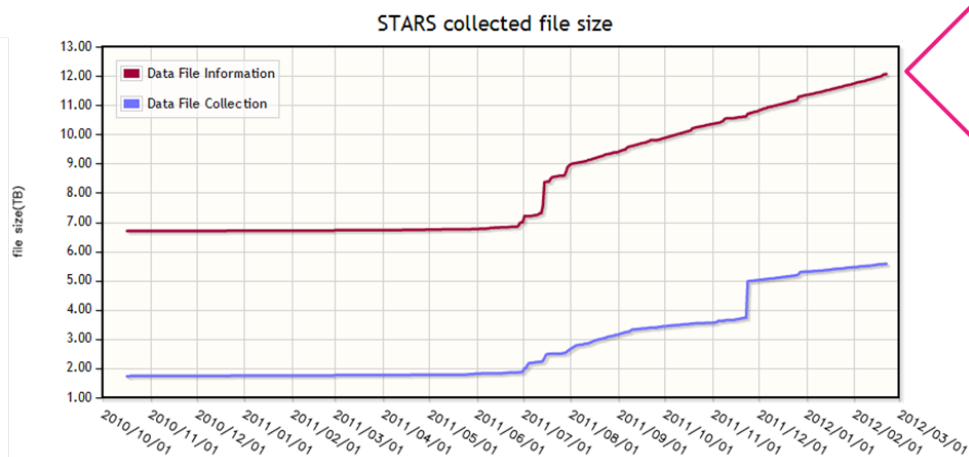


図 9 科学データファイルクローラ（自動収集）サービス（NICTY/DLA）の仕組み<sup>12)</sup>



More than 11 million files



More than 12 tera bytes

図 10 NICTY/DLA<sup>12)</sup> によるデータ自動収集実績（赤線が NICTY、青線が DLA）

#### 4.4. 広域観測ネットワーク監視システム

NICT サイエンスクラウドに示す通り、NICT サイエンスクラウドの基本コンセプトは、さまざまな形で公開されている科学データを収集し、格納し、解析することである（図 1）。地球・宇宙観測データについては、NICTY が収集対象とするようなインターネット上での公開データが対象となるが、研究機関が観測するデータを観測施設（Observatory）から直接伝送し、DB 化することもある。

現在、地球規模での地上観測ネットワーク網の試みが、様々な分野で進められている。これらの観測拠点数は、少ないものでは数か所であるが、多いものでは国内外で数千か所を超える大規模ネットワークもある。業務系・現業系の組織は別として、研究組織においてはこのような多数の観測拠点または観測装置を管理・運用することは容易ではない。一方で、地球規模での地球・宇宙環境・現象の理解には、このような大規模観測網が不可欠である。

特別な専用システムを除くと、一般研究機関での地球規模の観測システム運用やデータ収集には、専用ネットワーク回線ではなくインターネットが利用される。しかし、インターネットの物理層は 10Gbps を超える超高速なものから kbps オーダーで不安定な場合、また有線、無線などが混在している。また、インターネットでは QoS（Quality of Service）を確保することが難しく、ベストエフォートでのデータ伝送しか期待できない。これらの事情が、一般的な地球規模観測を困難にしている。たとえば、情報通信研究機構（NICT）の地上観測網も例外ではない。NICT 宇宙環境インフォマティクス研究室では国内外を含めて 30 以上の拠点に観測所を直接的または間接的に運用しており、40 種類以上のデータが日々伝送されている。これらの観測拠点の観測機器およびデータ伝送の維持運営が大きな負担となりつつある。そこで、NICT サイエンスクラウドでは、広域観測拠点と NICT サイエンスクラウドを結ぶネットワークを仮想化し、運用を一元化するためのシステムとこれを管理運営する Web アプリケーション（広域観測ネットワーク監視システム）を開発した（図 11）。ここで使われるネットワーク技術は新しいものではないが、L3/L2（VPN）ネットワークを混在させ、データ再送制御とデータファイル整合性の確保、機器の遠隔監視、ルーティング制御、分散データベースとの連携などを組み合わせることで、悪条件化でも観測とデータ伝送を継続的に行う OCM（Observation Continuity Model）に基づいた設計となっている。さらに、これを小型サーバに組み込むことで、1 台 10 万円程度の安価なデータ伝送システムパッケージも実装した。

現在、NICT 宇宙環境インフォマティクス研究室では国内外 8 拠点にこのシステムを配置しており、拠点数はさらに増える予定である。室内実験として数 10 台のサーバでのこのシステムを試験運用し、悪条件ネットワーク環境でデータが

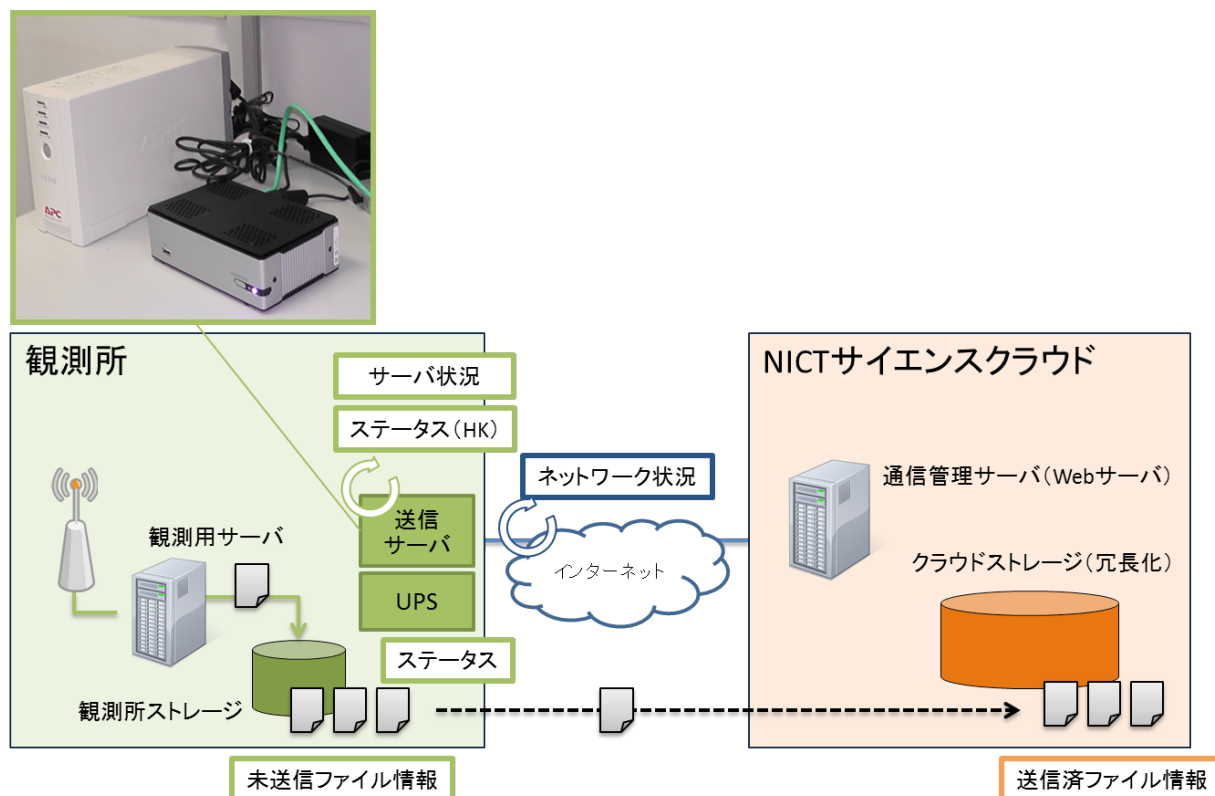


図 11 広域観測ネットワーク監視システム（写真はパッケージ化されたデータ送信サーバと UPS）



定常的に収集できることを実証している<sup>13)</sup>。なお、このシステムは、TCP/IP ベースであればネットワーク環境を選ばず、観測種類も選ばない汎用性の高いシステムである。

#### 4.5. 地球磁気圏 3 次元可視化ツール：バーチャルオーロラツール

筆者らの研究グループでは、宇宙・地球科学シミュレーションの可視化を積極的に進めており<sup>14,15,16,17,18,19)</sup>、ツールの標準化が有効であることを経験してきた。そこで、2008 年より地球近傍の宇宙空間（電離圏・磁気圏など）を対象とした 3 次元流体シミュレーションを時系列可視化するツールであるバーチャルオーロラツール（図 12）を開発している。バーチャルオーロラツールはサイバネット社の AVS Express<sup>20)</sup> により実装された、地球磁気圏物理学で必要となる標準的な可視化機能を網羅したツールである。バーチャルオーロラツールでは、対象となる時系列電磁流体（MHD）データを HDF5 形式とし、データファイル情報（メタ情報）を XML 形式記述された SPL（StarsProjectList）ファイル<sup>21)</sup>として表記することで、上記の可視化事例を含む様々な MHD シミュレーションコード（多流体型を含む）に対応しており、2013 年 4 月時点で、4 種類の Global MHD シミュレーションコードに対応している<sup>22)</sup>。

バーチャルオーロラツールは、AVS Express が対応している Windows と Linux の主要なディストリビューションで動作し、NICT サイエンスクラウド内のサーバ上であればライセンスを利用できる。

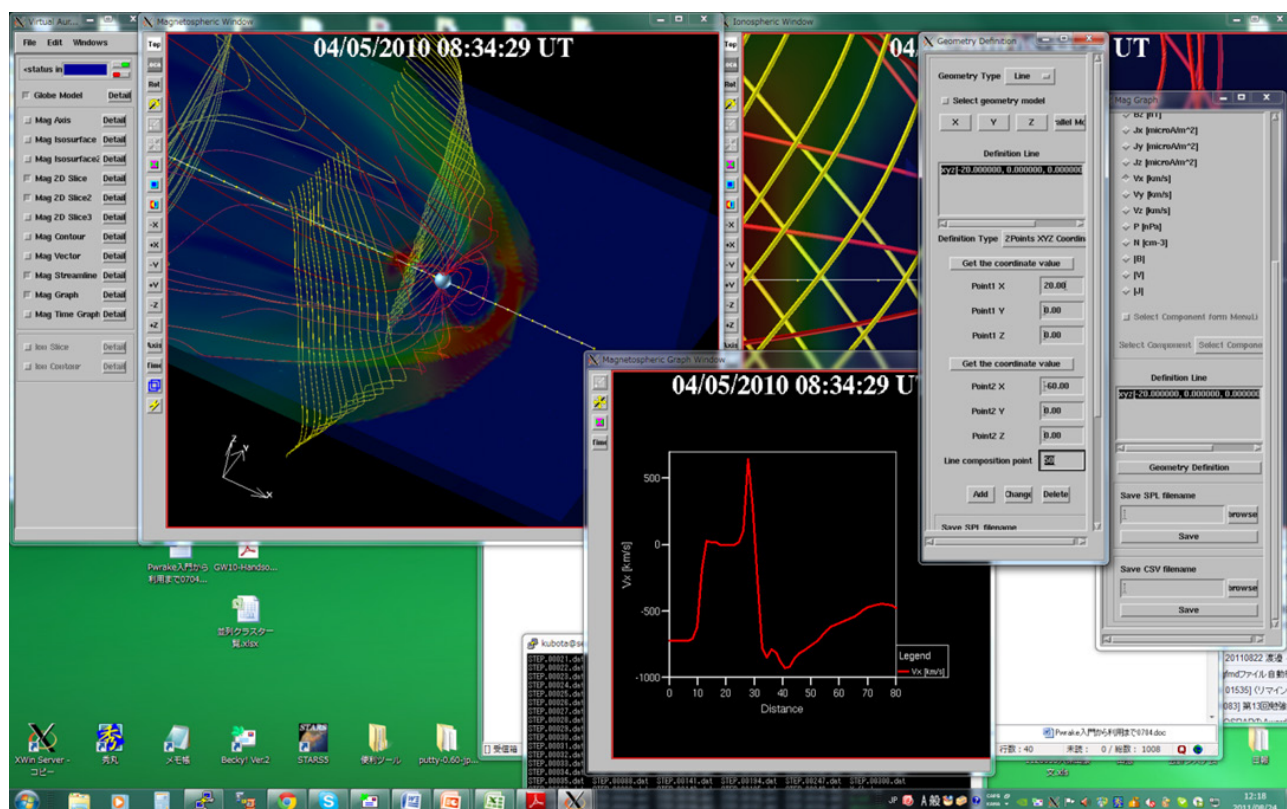


図 12 地球磁気圏 3 次元可視化ツール：バーチャルオーロラツール（図は Linux 版）

#### 4.6. 独自研究環境利用サービス（ハウジング・ホスティング）

アプリケーションやシステムといった科学研究用エコシステムとは別に、独自の研究環境を NICT サイエンスクラウド上で実現するためのハウジングやホスティングサービスも、サイエンスクラウドでは最も重要な機能の一つである。NICT サイエンスクラウドでは、特定の研究プロジェクトのためではなく汎用性の高いコンピュータリソースの提供を目指している。したがって、個別の研究テーマにおいて、NICT サイエンスクラウドのリソースで目的が達成できるとは限らない。そこで、NICT サイエンスクラウドでは、独自の研究環境を導入し、個別の研究目的に有効な機能のみを NICT サイエンスクラウドから利用する独自研究環境利用サービスを提供している（図 13）。

なお、独自研究環境利用サービスは商用のハウジングサービスとは (1) サーバ管理はユーザが行う、(2) NICT サイエンスクラウドのリソース（ファイアウォール／IPS やユーザ管理サービスを含む）を積極的に利用できるなどの点で異な

るため、NICTサイエンスクラウドではハウジングサービスと呼んでいない。ホスティングサービスは、商用クラウドとほぼ同じであるが、上記(2)のNICTサイエンスクラウドのリソースを積極的に利用する点が目的達成型のサイエンスクラウドの特徴である。

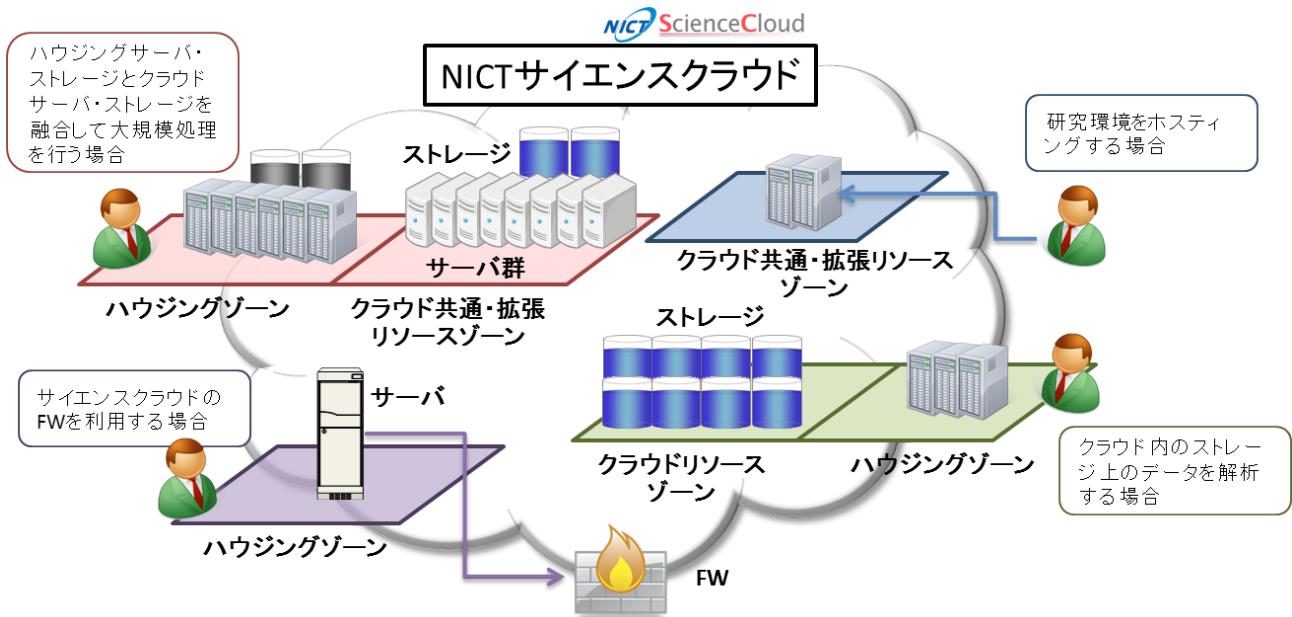


図 13 独自研究環境利用サービス（ハウジング・ホスティハウジング）

## 5. NICT サイエンスクラウド活用事例

前節までに、商用クラウドと比較したサイエンスクラウドの特徴や現状（2節）とNICTサイエンスクラウドのシステム（3節）および研究支援サービス（4節）について議論した。NICTサイエンスクラウドは、表2に示す各国のサイエンスクラウドと比較して、特に各研究テーマのためのカスタマイズ機能を重視している。本節では、宇宙科学研究の2つのテーマについて、NICTサイエンスクラウドの活用事例を示す。特に、前節までに議論したNICTサイエンスクラウドの各サービスをどのように導入・システム化するかについて重点的に述べる。

### 5.1. 放射線帯高エネルギー粒子予測システム

情報通信研究機構における放射線帯高エネルギー粒子予測システムをNICTサイエンスクラウドにおいて構築した例が図14である。放射線帯高エネルギー粒子予測システムでは、インターネット上から太陽風および静止軌道衛星データをNICTY/DLA（4.3節）により定常的に収集する（表4）。収集されたデータは、データ解析サーバ上で開発された放射線帯粒子予測アルゴリズム<sup>23)</sup>を用いて定常処理サーバ上で予測に用いられる。放射線帯粒子予測結果は、予測の精度確認（Verification）のための事後結果とともにWeb上で公開される。

今後、放射線帯高エネルギー粒子予測を高度化するためには、シミュレーションとの連動・融合が不可避である。本システムでは、NICTサイエンスクラウドが提供する大規模データ処理サーバを用いた放射線帯粒子計算と、NICTスーパーコンピュータ（Hitachi SR16000）によるGlobal MHDシミュレーション<sup>24)</sup>を融合させる計画である。これら3種類（衛星観測データ、粒子計算データ、MHDシミュレーションデータ）はすべてフロントエンドストレージに保存され、データ処理サーバで解析される。これは、図1に示すデータ指向型科学研究スタイルに一致している。

### 5.2. GPS全電子数（TEC）データ処理システム

図15は、NICTサイエンスクラウド上に構築されたGPS全電子数（TEC）データ処理システムの概要である。NICTでは、電離圏変動予測研究の一環として、世界中で公開されているGPSデータを収集し、独自のアルゴリズム<sup>25)</sup>によるGPS-TECデータの定常作成を行っている。GPSデータ自動収集については4.3節において述べたが、図15では収集したGPS

データ（表 5）をフロントエンドストレージに保存し、一部のデータ（日本上空の GEONET データ）については大規模処理サーバにおいて自動で TEC データに変換している。国内のデータについては、準リアルタイム（数時間程度の遅れ）で GPS-TEC 処理を行っている。生成された GPS-TEC データはマップ化され、様々な電離圏研究<sup>26,27,28)</sup>で活用されるほかに、NICT サイエンスクラウドの Web サーバ上で公開される。また、生成した GPS-TEC データは、地球科学・宇宙天気データを球面スクリーン上に立体的に表示する Dagik Earth<sup>29)</sup>でも利用され、展示イベントや出前授業などのアウトリーチに活用されている。NICT サイエンスクラウドの一部である可視化展示室でも閲覧可能である。

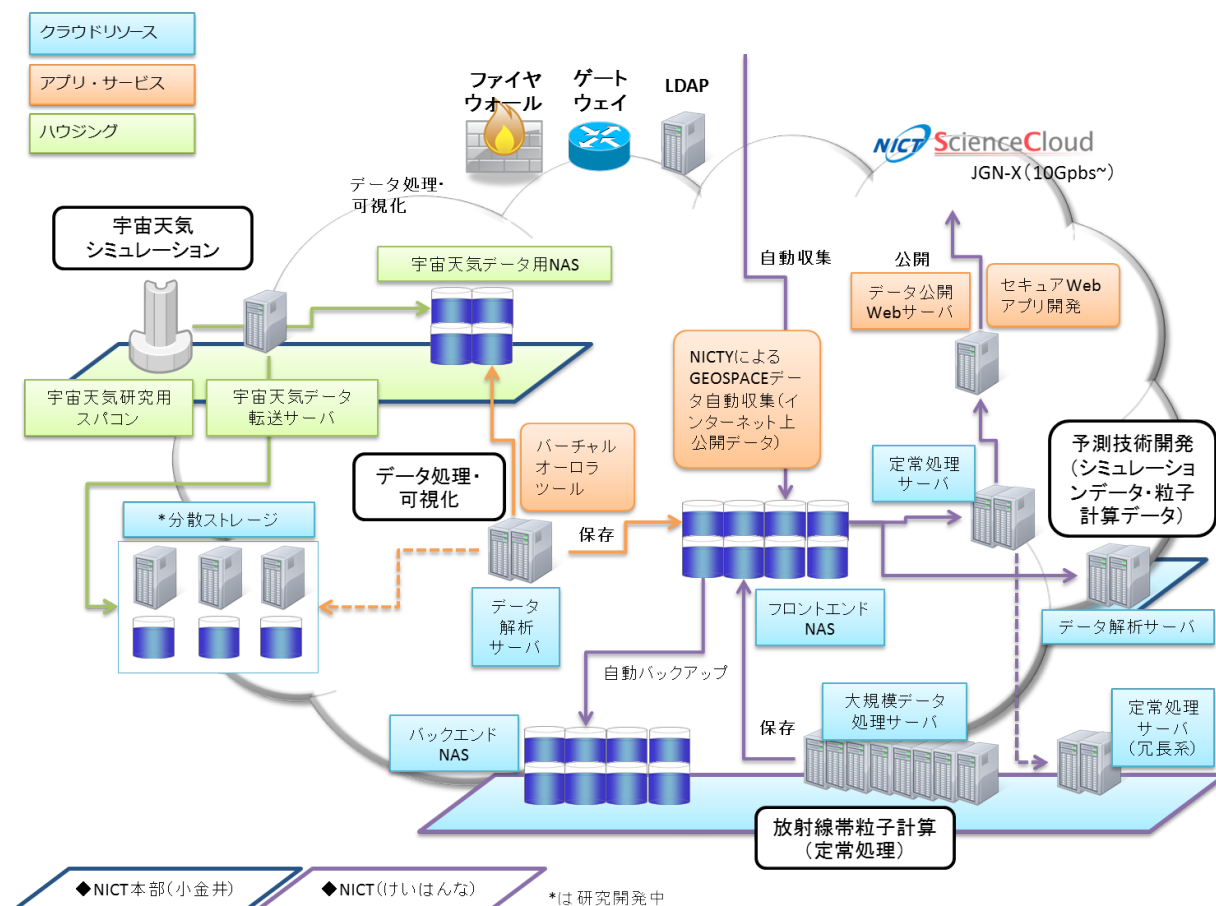


図 14 放射線帯高エネルギー粒子予測システムにおける NICT サイエンスクラウドシステム図  
(計画中や研究開発中の項目を含む)

表 4 NICTY/DLA による放射線帯高エネルギー粒子予測対象データ

データ名 (STARS メタデータベース)	データ説明	収集データ数
PARTICLE_5M/Realtime/GOES	静止軌道粒子データ / 5 分値 (GOES 衛星 / Secondary) 静止軌道粒子データ / 5 分値 (GOES 衛星 / Primary)	2 (ファイル/日)
FLUX/DOM/DRTS	静止軌道粒子データ (こだま衛星)	1 (ファイル/日)
SWEPAM_1M/Realtime/ACE	太陽風速度データ / 1 分値 (ACE 衛星) 太陽風密度データ / 1 分値 (ACE 衛星)	1 (ファイル/日)
SWEPAM_1H/Realtime/ACE	太陽風速度データ / 1 時間値 (ACE 衛星) 太陽風密度データ / 1 時間値 (ACE 衛星)	1 (ファイル/月)
MAG_1M/Realtime/ACE	太陽風磁場データ / 1 分値 (ACE 衛星)	1 (ファイル/日)
MAG_1H/Realtime/ACE	太陽風磁場データ / 1 時間値 (ACE 衛星)	1 (ファイル/月)
EPAM_5M/Realtime/ACE	太陽風粒子データ / 5 分値 (ACE 衛星)	1 (ファイル/日)
AE_Index_QL/GeomagneticIndex	地磁気データ (AE 指数)	5 (ファイル/月)
Dst_Index_QL/GeoMagneticIndex	地磁気データ (DST 指数)	1 (ファイル/月)

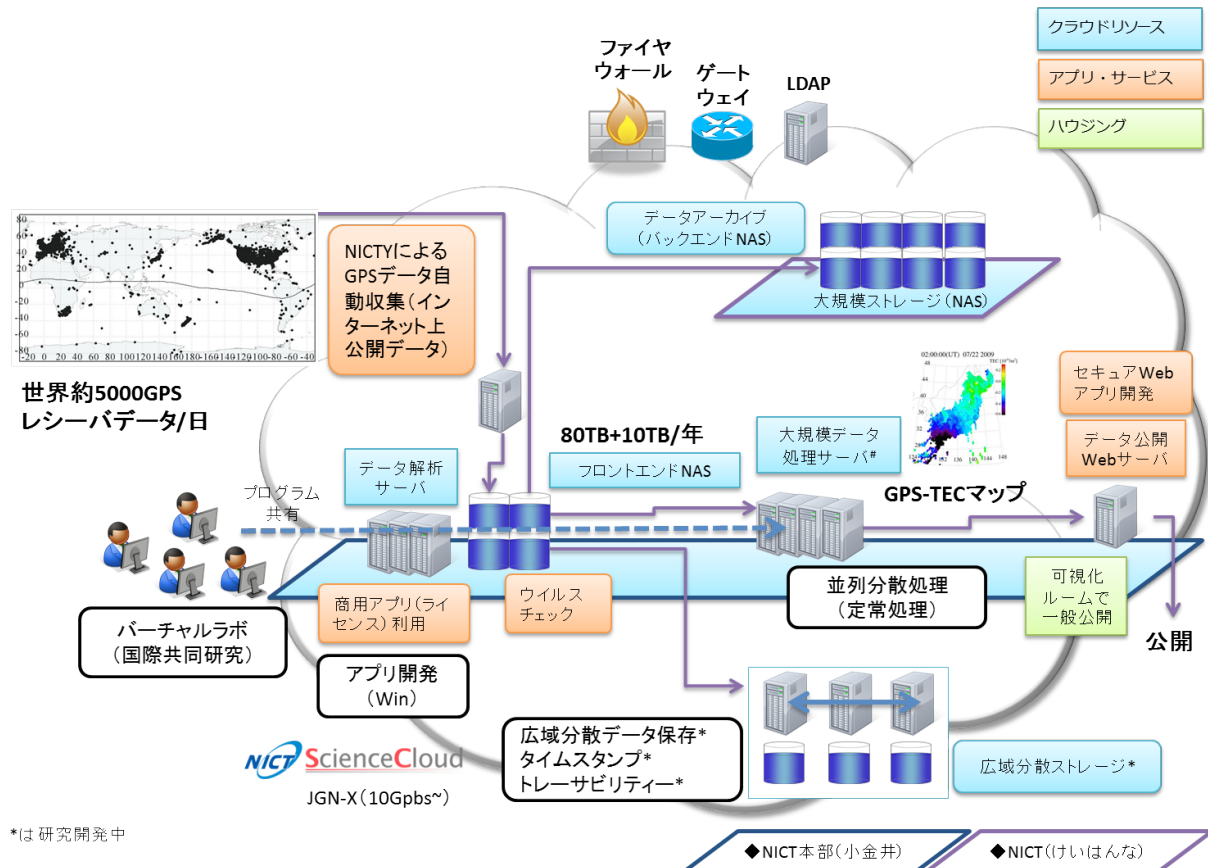


図 15 GPS 全電子数 (TEC) データ処理システムにおける NICT サイエンスクラウドシステム図  
(計画中や研究開発中の項目を含む)

表 5 NICTY/DLA による GPS-TEC 作成のための収集対象データ

データ名 (STARS メタデータベース)	データ説明	収集データ数
GPS/RINEX/UNAVCO	GPS/Compacted RINEX データ	2000 (ファイル/日)
GPS/RINEX/EU_DUT	GPS/Compacted RINEX データ	15 (ファイル/日)
GPS/RINEX/CDDIS0	GPS/Compacted RINEX データ	400 (ファイル/日)
GPS/RINEX/USCORS	GPS/Compacted RINEX データ	1800 (ファイル/日)
GPS/RINEX/ESEAS0	GPS/Compacted RINEX データ	1 (ファイル/日)
GPS/RINEX/EU_ASI	GPS/Compacted RINEX データ	40 (ファイル/日)
GPS/RINEX/EU_BKG	GPS/Compacted RINEX データ	160 (ファイル/日)
GPS/RINEX/EU_IGN	GPS/Compacted RINEX データ	250 (ファイル/日)
GPS/RINEX/EUOLG2	GPS/Compacted RINEX データ	4 (ファイル/日)
GPS/RINEX/GEOSCI	GPS/Compacted RINEX データ	320 (ファイル/日)
GPS/RINEX/GE_BKG	GPS/Compacted RINEX データ	16 (ファイル/日)
GPS/RINEX/IGSIGN	GPS/Compacted RINEX データ	410 (ファイル/日)
GPS/RINEX/AUSIPS	GPS/Compacted RINEX データ	4 (ファイル/日)
GPS/RINEX/IGSKAS	GPS/Compacted RINEX データ	400 (ファイル/日)
GPS/RINEX/PANGA1	GPS/Compacted RINEX データ	600 (ファイル/日)
GPS/RINEX/PANGA2	GPS/Compacted RINEX データ	10 (ファイル/日)
GPS/RINEX/BRIBGE	GPS/Compacted RINEX データ	80 (ファイル/日)
GPS/RINEX/ITACYL	GPS/Compacted RINEX データ	36 (ファイル/日)
GPS/RINEX/TRIGNT	GPS/Compacted RINEX データ	50 (ファイル/日)
GPS/RINEX/SOPAC0	GPS/Compacted RINEX データ	2100 (ファイル/日)
GPS/RINEX/NRCAN0	GPS/Compacted RINEX データ	22 (ファイル/日)
GPS/RINEX/GEONET	GPS/Compacted RINEX データ	1230 (ファイル/日)



### 5.3. 考察

本節では、NICT サイエンスクラウドを活用した研究システム事例を2例紹介した。サイエンスクラウドを活用した科学研究テーマの多くは単一の研究機関では実行が困難であり、2.3節で述べた仮想研究室（バーチャルラボラトリー）による国際的な共同研究体制が必要となる（図2）。図15に、5.2節のGPS全電子数(TEC)データ処理システムでのバーチャルラボラトリーとしてのNICTサイエンスクラウド利用について示した。バーチャルラボラトリーを構築するにあたり、ファイアウォール、ゲートウェイ、ユーザ認証（LDAP）などはNICTサイエンスクラウドが用意するシステムやサーバを利用できることも、サイエンスクラウド上にシステム構築する利点の一つである。

図14や図15では、定常処理結果をWebで公開している。Webアプリケーションによる情報提供は、NICTサイエンスクラウドの運用面では重要である。近年、Webアプリケーションのセキュリティ対策は重要度が増す一方で専門性が高く、一般の科学研究者には対応が容易ではない。NICTサイエンスクラウドでは、セキュアWebアプリケーション開発手順書を作成し、NICTサイエンスクラウドWebで公開している<sup>22)</sup>。本手順書に沿ってWebアプリケーションを開発することで、セキュリティ対策を立てることができる。

2.2節においてHPC/HTC/MTCの概念を紹介した。放射線帯高エネルギー粒子予測システムでは、HPC環境としてのスーパーコンピュータを、独自研究環境利用（ハウジング）としてNICTサイエンスクラウドにネットワーク接続している。図14や図15の大規模データ処理サーバも、小型のHPCシステムである。一方で、図14や図15のシステム全体は、多種多様な計算機リソースを組み合わせることで目的となる速度でデータ処理を行うためMTC環境と言える。NICTサイエンスクラウドのリソースと独自研究環境（ハウジング）を組み合わせることにより、衛星・地上観測データ（Grayによる第1の研究手法）および数値シミュレーションデータ（Grayによる第3の研究手法<sup>1)</sup>）を融合したMTC型の研究環境を実現していることが分かる。

なお、各事例では研究内容の進展と並行してシステム利用の形態は変化する。本事例紹介は、2013年4月時点での事例であり、今後、発展・修正、新規追加、機能削除などが常時おこなわれることになる。

## 6. まとめ

宇宙科学研究分野は様々な科学研究分野の中でも研究規模が大きな分野である。宇宙科学の主要な研究手法である衛星観測および広域地上観測に対して、数値シミュレーションは観測を補完する手法であり、宇宙科学研究の二本柱と言ってもよい<sup>30)</sup>。これらの手法は、近年、予算の大型化に伴うメガサイエンスとなっている。それに伴い、出力されるデータも大規模化されている。宇宙科学では、他の科学研究分野同様に、観測とシミュレーションの融合が研究のブレークスルーとなることが期待されているが、二つのメガサイエンスの融合は20世紀後半に期待されたほどには進んでいない。これは、保存された大量のデータファイルから有益な情報を抽出する環境と手法が不十分であることを示唆している。

このような背景のもと提唱されているのが、第4のパラダイムであるデータ指向型科学研究である<sup>1)</sup>。一般にデータ指向型科学研究の基盤となるのがサイエンスクラウドであると考えられているが、サイエンスクラウドについては、技術的側面および利活用面からの議論が始まったばかりである。言い換えると、どのような方法でサイエンスクラウドがデータ指向型科学研究を実現するのかは明確ではない。本稿では、情報通信研究機構（NICT）で構築した科学研究用クラウドであるNICTサイエンスクラウドについて、「サイエンスクラウドとは何か」「サイエンスクラウドには何ができるか」という視点で議論を行った。

NICTサイエンスクラウドがデータ指向型科学研究に対して貢献できるのは、バーチャルラボラトリー構築を含むクラウドの利便性とビッグデータ処理を含む大規模データ科学研究環境の提供という2本柱の機能提供である。本稿では、特に前者に重点をおいて議論を行った。サイエンスクラウドにより真に大規模な研究プロジェクトを進めるためには、商用クラウドの様に計算機リソースを仮想化して提供するだけでは不十分である。セキュリティ対策やシステム冗長化はもちろんのこと、バーチャルラボラトリーを構築するためのユーザ管理、商用アプリケーションライセンスの提供、研究目的に特化したアプリケーションやシステムの提供、独自研究環境とクラウドリソースの融合などが総合的に機能して、初めてサイエンスクラウドとしての成果が期待できる。セキュリティについても、システムセキュリティの担保だけではなく、研究成果をWeb等で公開するためのセキュリティ対策のガイドラインなども必要である。

ビッグデータ科学の重要性が認識されつつある現在、今後も研究リソースを一か所に集約することでサイエンスクラウドを構築する研究機関が増えるであろう。しかし、単に大規模な計算環境を作っても、サイエンスクラウドとしては



機能しない。本論文は、真に実用的なサイエンスクラウド構築の助けになるだろう。

## 謝辞

本論文の研究は情報通信研究機構の NICT サイエンスクラウドおよび JGN-X を用いて行われました。バーチャルオーロラツールは情報・システム研究機構 新領域融合研究センターの育成融合研究課題として 2006 ～ 2008 年度に開発されたものを基盤としています。

## 参考文献

- 1) Hey, T., S. Tansley, and K. Tolle, The Fourth Paradigm: Data-Intensive Scientific Discovery, Microsoft Research, Washington, 2009.
- 2) 村田 健史, サイエンスクラウドは第四の研究基盤となるか?, 学術の動向, vol.17, no.6, 2012, pp.42-47.
- 3) 村田 健史, NICT サイエンスクラウドによる新しい科学研究の姿, 情報管理, Vol.55, No.8, 2012, pp.552-561.
- 4) “Science Clouds” Web site, <http://scienceclouds.org/>.
- 5) 2010 年第 1 回 ScienceCloud ワークショップ Web Site, <http://datasys.cs.iit.edu/events/ScienceCloud2010/>.
- 6) 2011 年第 2 回 ScienceCloud ワークショップ Web Site, <http://datasys.cs.iit.edu/events/ScienceCloud2011/>.
- 7) 2012 年第 3 回 ScienceCloud ワークショップ Web Site, <http://ceng.usc.edu/~simmhan/ScienceCloud2012/>.
- 8) ICSU/WDS Web サイト, <http://www.icsu-wds.org/organization/structure/ipo>.
- 9) Murata, K., T. Watari, S., Nagatsuma, T., Kunitake, M., Watanabe, H., Yamamoto, K., Kubota, Y., Kato, H., Tsugawa, T., Ukawa, K., Muranaga, K., Kimura, E., Tatebe, O., Fukazawa, K. and Murayama, Y., A Science Cloud for Data Intensive Sciences, Data Science Journal, Vol. 12, 2013, pp.WDS139-WDS146.
- 10) 尾家祐二, 超高速研究開発ネットワーク JGN II, 情報通信研究機構季報, Vol.51, Nos.3-4, 2005, pp.3-11.
- 11) Helix Nebula Web site, <http://helix-nebula.eu/>.
- 12) 石倉 諭, 村田 健史, 久保 卓也, 木村 映善, 山本 和憲, 篠原 育, RSS1.0 を利用した科学衛星・地上観測データのメタデータ自動収集, 電子情報通信学会論文誌 (B), Vol.J91-B (4), 2008, pp.499-509.
- 13) Watanabe, H., Yamamoto, K., Tsugawa, T., Nagatsuma, T., Watari, S., Murayama, Y. and Murata, K., T., An Integrated Management System of Multipoint Space Weather Observation, Data Science Journal, Vol.12, 2013, WDS2011-CS607.
- 14) 松岡大祐, 村田健史, 藤田茂, 田中高史, 山本和憲, 木村映善, Global MHD シミュレーションによる磁気フラックスロープの 3 次元構造解析, 可視化情報学会論文集, Vol.28, No.6, 2008, pp.38-46.
- 15) 松岡大祐, 村田健史, 藤田茂, 田中高史, 山本和憲, 大野暢亮, 3 次元可視化とビジュアルデータマイニング, 情報通信研究機構季報, Vol.55, Nos.1-4, 2009, pp.459-469.
- 16) Murata, K., T., Yamamoto, K., Matsuoka, D., Matsumoto, H., Okada, M., Mukai, T., Sigwarth, J., B., Fujita, S., Tanaka, T., Yumoto, K., Ogino, T., Shiokawa, K., Tsyganenko, N., A., Green, J., L. and Nagai, T., Development of the Virtual Earth’ s Magnetosphere System (VEMS), Advances in Polar Upper Atmosphere Research, No.19, pp.135-151, 2005.
- 17) Saita, S., Kadokura, A., Sato, N., Fujita, S., Tanaka, T., Ebihara, Y., Ohtani, S., Ueno, G., Murata, T., K., Matsuoka, D., Kitamoto, A., Higuchi, T., Displacement of conjugate points during a substorm in a global magnetohydrodynamic simulation, Journal of Geophysical Research-Space Physics, Vol.116 (A6), 2011, A06213.
- 18) Kubota, Y., Yamamoto, K., Fukazawa, K., and Murata, K., T., Visualization of the Flux Rope Generation Process Using Large Quantities of MHD Simulation Data, Data Science Journal, Vol.12, 2013, pp.WDS134-WDS138.
- 19) Fujita, S., Yamagishi, H., Murata, K., T., Den, M. and Tanaka, T., A numerical simulation of a negative solar wind impulse: Revisited, Journal of Geophysical Research, vol.117, 2012, A09219.
- 20) Murata, K., T., Yamamoto, K., Matsuoka, D., Matsumoto, H., Okada, M., Mukai, T., Sigwarth, J., B., Fujita, S., Tanaka, T., Yumoto, K., Ogino, T., Shiokawa, K., Tsyganenko, N., A., Green, J., L. and Nagai, T., Development of the Virtual Earth’ s Magnetosphere System (VEMS), Advances in Polar Upper Atmosphere Research, No.19, 2005, pp.135-151.

- 21) Murata, K., T., Yahara, H. and Toyota, K., Software design via object-oriented methodology and network database for solar-terrestrial observation data, Database system, 123-5, database, 2001, pp.31-36.
- 22) NICT サイエンスクラウド Web site, <http://sc-web.nict.go.jp>.
- 23) Sakaguchi, K., Miyoshi, Y., Saito, S., Nagatsuma, T., Seki, K. and Murata, K., T., Relativistic electron flux forecast at geostationary orbit using Kalman filter based on multivariate autoregressive model, Space Weather, Vol.11 (2), 2013, pp.79-89.
- 24) Yasubumi Kubota, Ryuho Kataoka, Mitsue Den, Takashi Tanaka, and Tsutomu Nagatsuma, GlobalMHDsimulation of the magnetospheric response to large and sudden enhancement of the solar wind dynamic pressure, PEM05-07, JpGU2013.
- 25) Tsugawa, T., Observational Studies on Large-Scale Traveling Ionospheric Disturbances Using GPS Receiver Networks, Ph. D Thesis, Kyoto University, Kyoto, Japan, 2004.
- 26) Nishioka, M., Otsuka, Y., Shiokawa, K., Tsugawa, T., Effendy, Supnithi, P., Nagatsuma, T. and Murata, K., T., On Post-Midnight Field-Aligned Irregularities Observed with a 30.8-MHz Radar at a Low Latitude: Comparison with F-Layer Altitude near the Geomagnetic Equator, Journal of Geophysical Research, Vol.117, 2012, A08337.
- 27) Tsugawa, T., Saito, A., Otsuka, A., Nishioka, M., Maruyama, T., Kato, H., Nagatsuma, T. and Murata, K., T., Ionospheric disturbances detected by GPS total electron content observation after the 2011 off the Pacific coast of Tohoku Earthquake, Earth Planets Space, Vol.63 (7), 2011, pp.875-879.
- 28) Otsuka, Y., Suzuki, K., Nakagawa, S., Nishioka, M., Shiokawa, K. and Tsugawa, T., GPS Observations of Medium-Scale Traveling Ionospheric Disturbances over Europe, Annales Geophysicae, 31, 163-172, doi:10.5194/angeo-31-163-2013, 2013.
- 29) Dagik Web サイト, <http://www.dagik.net>.
- 30) 松本 紘, 宇宙開拓とコンピュータ, 情報フロンティアシリーズ, 情報処理学会編, 共立出版, 1996.

# NICT サイエンスクラウドによる大規模シミュレーションデータ 分散可視化処理

村田健史<sup>\*1</sup>, 磯田総子<sup>\*2</sup>, 渡邊英伸<sup>\*1</sup>, 深沢圭一郎<sup>\*3</sup>,  
山本和憲<sup>\*1</sup>, 建部修見<sup>\*4</sup>, 田中昌宏<sup>\*4</sup>, 木村映善<sup>\*5</sup>

## High Performance Visualization Processing of Large-Scale Computer Simulation Data via NICT Science Cloud

Ken T. MURATA<sup>\*1</sup>, Fusako ISODA<sup>\*2</sup>, Hidenobu WATANABE<sup>\*1</sup>, Keiichiro FUKAZAWA<sup>\*3</sup>,  
Kazunori YAMAMOTO<sup>\*1</sup>, Osamu TATEBE<sup>\*4</sup>, Masahiro TANAKA<sup>\*4</sup> and Eizen KIMURA<sup>\*5</sup>

### Abstract

Science cloud is a cloud system designed for scientific researches, and expected as a new infrastructure for big data sciences. Not only parallelization of CPU as in super-computers, but I/O and network throughput parallelization are crucial for the big data sciences. One of the typical structures of science cloud is a scalable cluster in which multiple clusters in a cloud are connected with high-speed network. In the present study, we examine performance of parallelization of both CPU and I/O in a cloud system as the first step to high performance scalable clusters. In case with few processes executed on each computational node (server), parallelization efficiency is almost 100%. This high efficiency is expected to maintain in larger-scale cluster systems such as those with 100 servers. On the condition of multi-processes on each node, the present parallelization does not show good performance due to the congestions of I/O. Parallelization efficiency (speed-up) is as low as 20.6%. New techniques of decentralization of I/O within each node are required in the next step.

Keyword: Science Cloud, Gfarm, Pwroke, Big data, Magnetospheric MHD simulation, I/O parallelization

### 概 要

現在、欧米や日本において、科学研究向けクラウドとしてサイエンスクラウドが提案され、システムや利活用についての議論が進められている。サイエンスクラウドはビッグデータ時代のデータ指向型科学研究基盤として提案されている。クラウド上でのビッグデータ処理では、CPUの並列分散化だけではなくネットワークやI/Oの並列分散が重要である。本研究ではクラウド環境で高い計算効率を得るための基礎研究として、CPUおよびI/Oを同時に分散処理するシステムを提案する。8コアを有する6台のクラスタ計算機で基礎実験を行った結果、ノード内のプロセス数が少ない場合はほぼ100%の並列化効率が達成され、100台を超える大規模クラスタ環境でも高いスケーラビリティが予測された。一方、ノード内で複数プロセスが処理を行う場合にはディスクアクセスの輻輳が発生し、必ずしも高いスケーラビリティを得ることができない。8コアを有する6台の計算サーバで最も高い並列化効率を得たのは6プロセス並列の場合であり、高速化率は20.6であった。今後は、同ノード内のI/O処理を分散化することで、並列化効率をさらに向上させることが期待される。

### 1. はじめに

多くの科学研究分野でのデータのデジタル化・大規模化を背景として、データ指向型科学は、実験科学・観測科学、理論科学、シミュレーション科学に続く科学研究手法の第4のパラダイムと言われている<sup>1)</sup>。データ指向型科学の中でもビッグデータ科学は注目を浴びているが、近年、京コンピュータ<sup>2)</sup>に象徴されるように第3の研究手法である数値シミュ

\*1 情報通信研究機構 (National Institute of Information and Communications Technology)

\*2 株式会社サイエンス・サービス (Science Service Co., LTD)

\*3 九州大学情報基盤研究開発センター (Research Institute for Information Technology, Kyushu University)

\*4 筑波大学計算科学研究センター (Center for Computational Sciences, University of Tsukuba)

\*5 愛媛大学医学部 (Department of Medical Informatics Ehime University)

レーションの大規模化が進んでおり、スーパーコンピュータに代表される HPC (High Performance Computing) 技術の進歩は目覚ましい。一方、数値シミュレーションが出力する数値データは同様に大規模化しているが、出力データのポスト処理技術については十分に発展しているとは言い難い。

宇宙物理学分野においても、流体系や粒子系の数値シミュレーションが行われている。その場観測 (in-situ 観測) が容易ではない宇宙科学においては、数値シミュレーションの果たす役割は大きい。宇宙物理シミュレーションコード (プログラム) は、その多くが 3 次元化されており、大規模化が進んでいる。対象とする現象の複雑さに伴い、空間方向の大規模化と同時に時間方向の大規模化 (長期シミュレーション) が進んでいる。

HPC の様に CPU リソースを極限まで活用して高速計算を行う計算指向型の研究に対して、生成されたデータを最も高速に処理するためのデータ指向型の技術開発が必要とされている。このような背景のもと、HPC に対する概念として MTC (Many-Task Computing) や HTC (High Throughput Computing) が登場した<sup>3)</sup>。多種多様で大規模なデータを目的に応じて高速処理するための技術が MTC という概念である。MTC では、多様な計算機リソースを融合し、データ分散、並列データ処理やコンピュータとデータファイルをローカライズする工夫など、総合的なデータ処理環境をめざす。大規模数値シミュレーションデータ処理においては、HPC よりも MTC が有効である。また、同時に、HPC や MTC と言った計算指向 (Computing Oriented) ではなく、データ配置や並列データ処理、データ局所性などに着目したデータ指向計算である Data Intensive Computing (DIC) と言う概念も提案されている<sup>1)</sup>。

本研究では、10G ネットワークで接続された 6 台のクラスタ計算機環境において、MTC 型または DIC 型の並列分散データ処理システムを試験的に構築し、データ処理の並列化効率について調べる。これまでのクラスタ計算機では、計算指向型 (HPC 型) の並列処理研究が多いが、本研究ではデータ処理時間に対してデータ I/O 時間が無視できない場合について考える。この場合、ネットワークを介して分散ファイルサーバ (たとえば NAS ストレージ) のデータファイルにアクセスする場合、スループットのオーバーフローが生じるため、一定以上の並列化分散処理は効率が極端に悪くなる。この場合に有効となるのは、I/O の分散化と CPU の並列化を両立するデータ処理システムである。

そこで本研究では、分散ファイルシステム Gfarm<sup>4)</sup> とワークフローシステム Pwrake<sup>5)</sup> を用いてクラスタ計算機環境に I/O と CPU を並列分散処理するシステムを構築し、実際の科学研究データ処理における並列化効率やスケーラビリティについて基本性能を検討する。さらに検討結果をもとに実データファイルの処理を行い、その有効性を示すとともに、今後の NICT サイエンスクラウド<sup>7)</sup> をはじめとする科学研究向けクラウドシステムで主流となると考えられるスケーラブルクラスタ環境での並列処理について議論する。

## 2. 実験の目的と実験方法

### 2.1. Gfarm/Pwrake

1 節で述べたとおり、MTC および DIC では、処理全体の効率向上のためには、HPC で考える CPU の分散化だけではなく、I/O やネットワークの分散化が必要である。本実験では、分散ファイルミドルウェア Gfarm<sup>4)</sup> および Gfarm 用のワークフローシステム Pwrake<sup>5)</sup> を用いた CPU 分散および I/O 分散により実際に科学研究で用いている数値シミュレーションデータ処理の並列処理を行う。

図 1 に Gfarm および Pwrake の機能について説明する。Gfarm は HPCI<sup>9)</sup> などでも用いられている大規模分散ストレージに適した分散ファイルシステム (ミドルウェア) であり、メタデータサーバ (MDS) とファイルシステムノード (FSN) から構成される。複数のファイルシステムノードは L2 または L3 ネットワーク上で分散して配置され、すべてメタデータサーバにより管理される。クライアントノード (CN) は、分散ファイルシステムを利用するデータ処理ノードである。それぞれの CN は、複数の FSN から構成される分散ファイルシステムを共通ファイルシステムとして利用することができる。なお、Gfarm では特定のノード (計算機) がファイルシステムノードとクライアントノードを兼ねることができる。(図 1 は、FSN3 と FSN4 をクライアントノードとしても利用する場合である。)

Gfarm の特徴の一つは、データファイルの分散管理である。ユーザが分散ファイルシステムにファイル (たとえばファイル①) を保存する。Gfarm はあらかじめ設定した数の複製を作成し、異なる FSN に保存する。(図 1 の場合は、ファイル①は FSN1、FSN2 および FSN3 に保存される。) クライアントノードは、複製ファイルのうち、最も高速にアクセス

1) なお、HPC や MTC と並んで HTC (High-Throughput Computing) という用語が用いられることがあるが、HTC は HPC や MTC と比べると長期 (数週間、数か月など) にわたってデータを処理する環境・技術を意味する。

できる最適データファイルを参照する。(たとえば, 図 1 の場合にはファイル①は CN4 が処理し, ファイル②は CN5 が処理する。) データファイルごとの最適クライアントノードの選択は, ノード負荷やネットワークトラフィックによりメタデータサーバが行う。

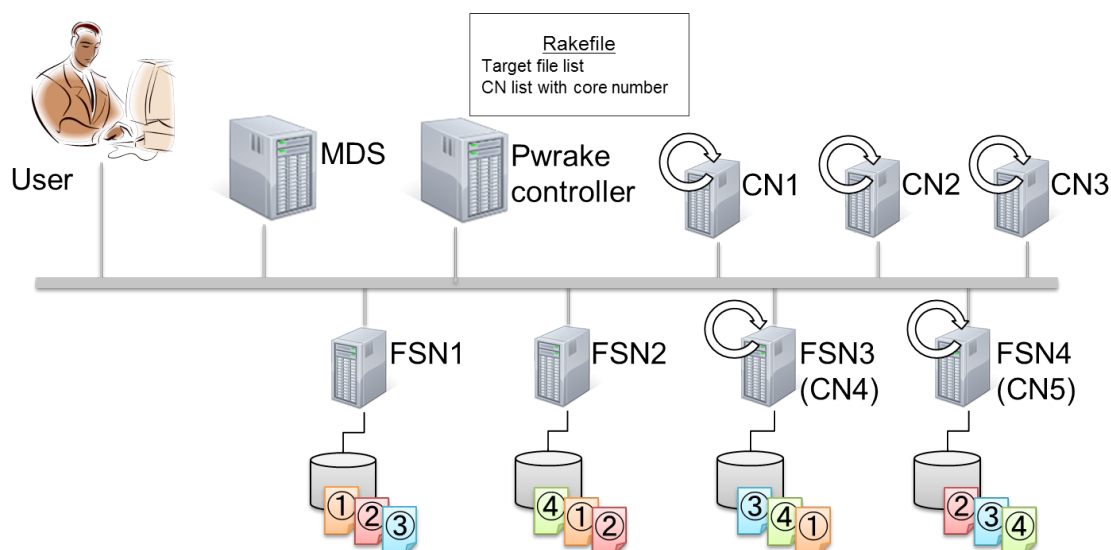


図 1 Gfarm と Pwrake の概念図 : MDS (メタデータサーバ) / FSN (ファイルシステムノード) / CN (クライアントノード) / Pwrake controller (Pwrake 制御サーバ)

Pwrake は, 複数のクライアントノードを用いて高速な並列分散処理を行うことを目的に開発されたワークフローシステムである。ユーザは, まず, Pwrake コントローラが管理する Rake ファイルに処理したいファイル一覧を指定し, 使用するクライアントノードとコア数のリストをノードファイルに記述する。Pwrake は処理の対象となるファイルリストを参照し, ノードファイルを読み込む。Pwrake は, 参照した各データファイルについて, 処理を行うクライアントノードとコアを指定する。その際, CN4 および CN5 のようにファイルシステムノードを兼ねたクライアントノードについては, クライアントノードがファイルシステムノードとして管理するファイルを優先的に処理に割り当てる。これにより, ネットワーク上にファイルアクセス (I/O) のトラフィックが流れることを回避できる。また, 多くの場合は, ネットワークファイル転送よりもディスクアクセス速度の方が高速であるため, データ処理 (ファイル読み込みおよび書き出し) の高速化が期待できる。

Pwrake のワークフローの原則の例を図 2 に示す。図 2 は, 2 台のクライアントノード (CN) がともにファイルシステムノード (FSN) を兼ねており, 2 台の CN がそれぞれ 3 プロセスと 2 プロセスで①～⑩の 20 のファイル処理を実行する場合である。CN が処理を行うプロセス数は, 図 1 の Rake ファイルが管理している。図 2 では簡単のため, 2 台の FSN は 20 ファイル全てを管理しているものとする。図の右側が, 各プロセスに割り当てられる処理対象ファイルである。図 1 の Pwrake コントローラはファイル処理が終わった各プロセスに, 次のファイルを割り当てる。割り当てるデータファイルは図 1 の Rake ファイルが管理している。処理の順序はほぼデータファイルの順序となるが, 処理時間が一定ではない場合には両者は必ずしも完全に一致はしない。

## 2.2. 実験の目的

1 節で述べたとおり, MTC および DIC では, 処理全体の効率向上のためには, HPC で考える CPU の分散化だけではなく, I/O やネットワークの分散化が必要である。本実験では, Gfarm を用いた CPU 分散および I/O 分散により実際に科学研究で用いている数値シミュレーションデータ処理の並列処理を行う。

本研究では Gfarm/Pwrake を用いることで高い並列化効率を目指すが, そのためには CPU の分散と I/O の分散のバランスと最適化が必要である。Gfarm/Pwrake による詳細な処理効率に関する報告は少なく<sup>5)</sup>, 本研究ではまず予備実験, 基礎実験により基本的な情報を収集する。それをもとに実際のデータ処理のためのモデルを作成し, 本実験 (実用化実験) を実施する。



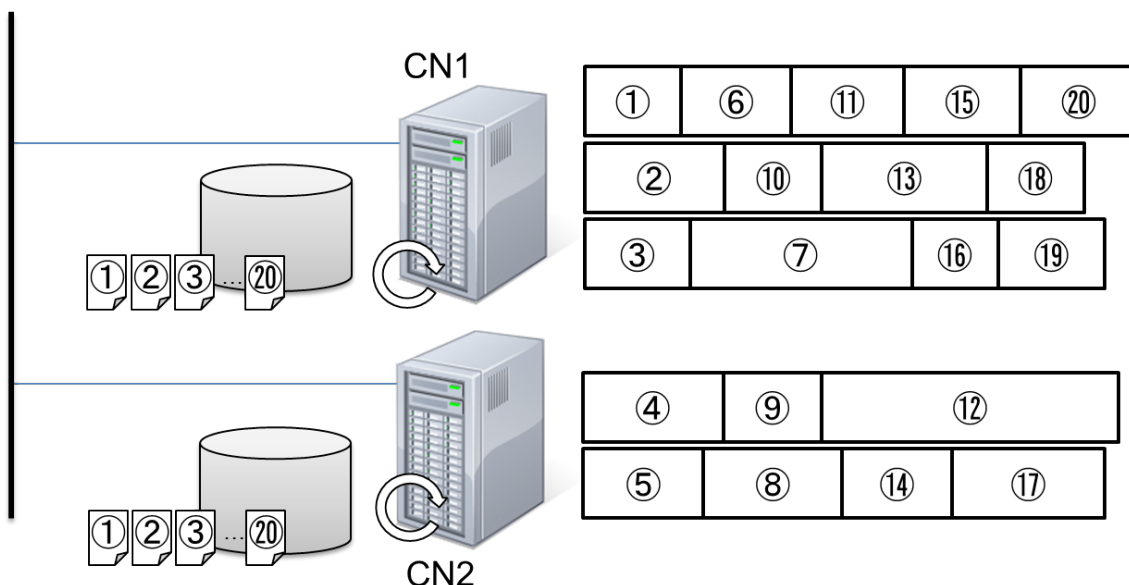


図 2 Pwrake によるワークフロー（クライアントノードが 2 台で、3 プロセスと 2 プロセスで処理する場合）：  
処理が終わったノード（コア）に次のデータファイルの処理を割り当てる。当該ファイルが FSN として管理するローカルファイルでない場合には、そのファイルは割り当てない。

### 2.3. データセットと実験環境

本研究が対象とするデータセットを図 3 に示す。データはスーパーコンピュータにより計算された時系列 3 次元地球磁気圏グローバル MHD シミュレーションデータ<sup>8)</sup>である。本シミュレーションコードは、 $450 \times 300 \times 300$  グリッドの 3 次元時系列数値計算であり、0.5 秒の時間分解能で、各時刻 2.2GB の数値データを出力する<sup>9)</sup>。出力データ形式は、数値シミュレーションで最もよく利用される形式の一つである HDF5 である。NICT が開発した可視化ツール（バーチャルオーロラツール）は HDF5 形式で記述されたこれらの数値データを読み込み、あらかじめ決めたパラメータにより可視化する。各時刻ステップのデータファイルの可視化には依存性がないため、本可視化処理を並列化する場合には可視化処理プロセス間でのメッセージ通信は発生しない。バーチャルオーロラツールは GUI を有するインタラクティブなアプリケーションであるが、あらかじめパラメータセットすることでコマンドラインからの処理が可能である。

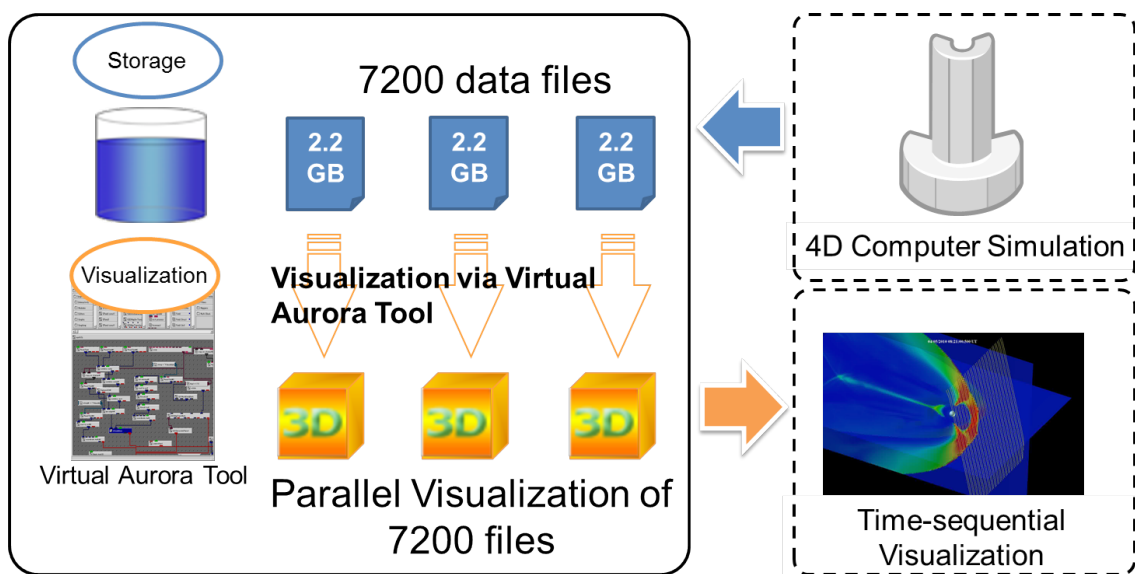


図 3 本研究が対象とする Global MHD シミュレーションデータ：時間分解能 0.5 秒で 1 時間分（7200 ステップ）の地球磁気圏現象のシミュレーションを行った結果である。実験では 7200 ステップデータのうちの一部を用いる。

図4に、本研究で行うMHDシミュレーションの可視化出力例を示す。図4では、磁力線(200本)およびxz面とxy面(座標軸はGSM座標系に準拠する)のプラズマ圧力のコンタープロットを出力している。本研究では、この数値データをすべてこの数値データ1時間分(7200ステップ分)の一部を並列処理することを目標とする。

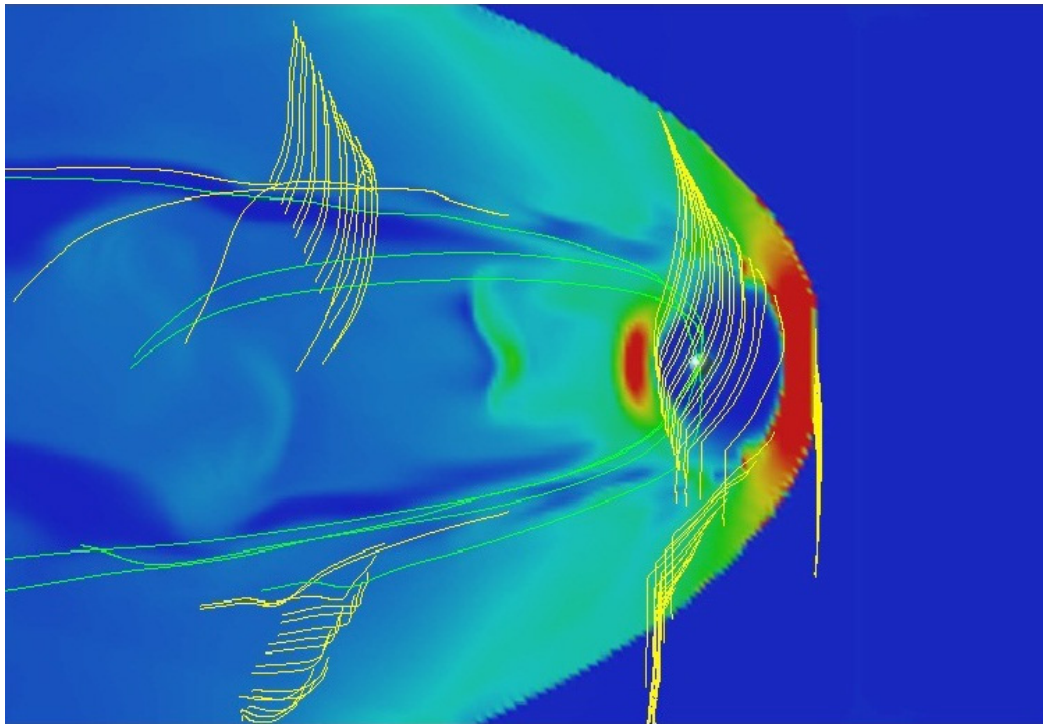


図4 本研究の可視化結果(3D可視化を特定の方向からキャプチャしたもの): 描画磁力線数は200本, コンターは2面(1面は垂直面であるため表示されていない)でありプラズマ圧力を示している。

図5に本研究の実験システムの概要図を示す。6台のデータ処理サーバが10Gスイッチによりネットワーク接続されている。各データ処理サーバは10Gネットワークカードを有する。本実験システムは外部ネットワークから独立しており、本実験システムに外部トラフィックが流れることはない。なお、6台のデータ処理サーバは図1のGfarmシステムのファイルシステムノード(FSN)とクライアントノード(CN)を兼ねている。システムにはメタデータサーバ(MDS)が1G接続されている。

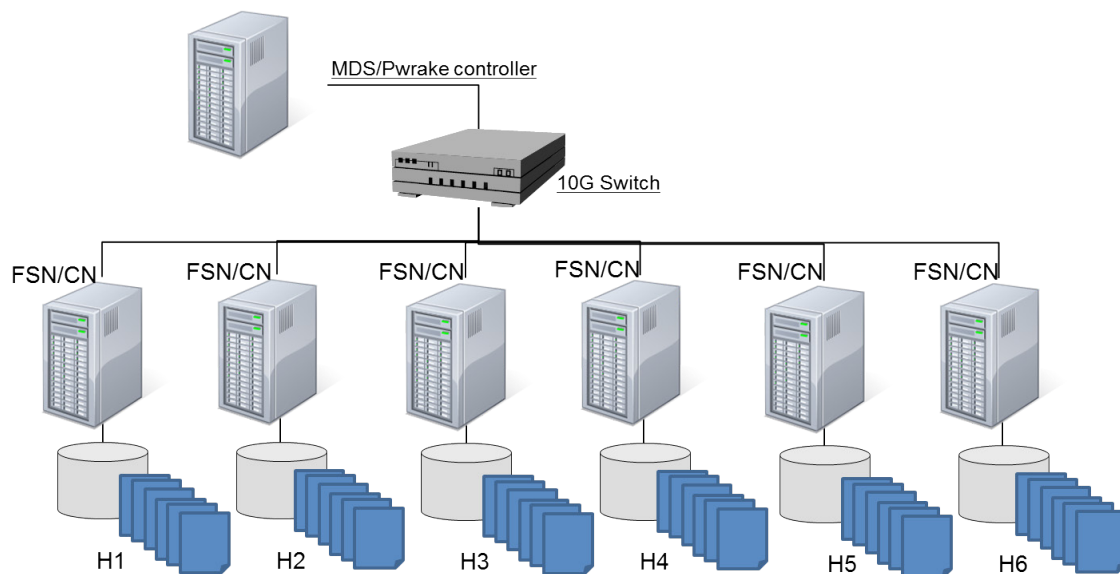


図5 本研究の実験環境: 並列処理サーバ・メタデータサーバ・10Gスイッチから構成されている。

2 図4に示す各時間ステップの3D可視化結果を一つの時系列可視化データファイルに統合する後処理は本研究対象には含まない。

### 3. 予備実験

本節では、2節で述べた実データ実験のための予備実験を行う。予備実験は、図5のシステムでGfarmのクライアントノード上でバーチャルオーロラツールを用いて対象データを読み込み、2.3節で述べた可視化処理を行う。図6に本研究のデータ処理時間の定義を示す。予備実験では、1データファイルに必要な読み込み時間（データI/O時間）とデータ処理時間（可視化処理時間）を比較する。なお、全処理時間は、これらにバーチャルオーロラツール（アプリケーション）の起動時間及び終了時間を加えた時間となる。

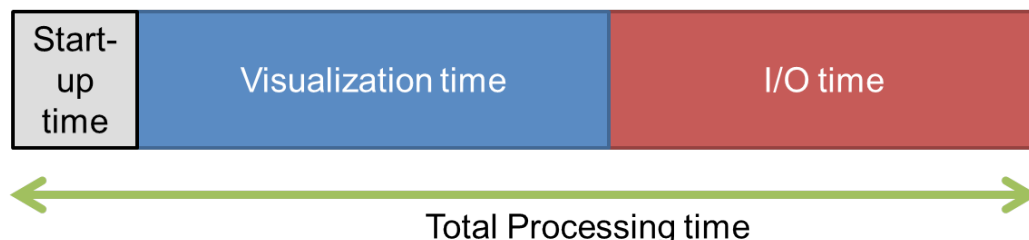


図6 (1) データI/O時間（I/O time）と(2) 可視化処理時間（Visualization time）：両者の合計にアプリケーション起動・終了時間（Start-up time）を加えた時間が全処理時間（Total Processing time）となる。

予備実験は、データ処理（可視化）に図5のうちの1台の計算機（クライアントノード）を用い、対象とするデータファイルと同じノードに保存する。これは、図1（FSN3およびFSN4）で示したFSNがCNと一致する場合に該当する。本節の予備実験では処理データファイル数を図3に示すシミュレーションの時間ステップ番号1から144までの144とし、1ファイル当たりの平均処理時間を求める。

図7(1)に、本予備実験の結果を示す。図7は144回の測定の平均値を示している。これまで、多くの大規模数値データ処理ではデータ処理の並列化が重視され、データI/O時間は無視されることが多かった。しかし、図7(1)では可視化処理時間（データ処理時間）とデータ読み込み時間（データI/O時間）は比較的同規模であり、I/O時間が無視できないことが分かる。

予備実験として、図7(1)と同様のデータに対してデータファイルをCNと異なるFSNに保存した場合についても調べる。すなわち、図1のCN1とFSN1のように、FSNがCNと異なる場合である。この場合の1データファイルに必要な読み込み時間（データI/O時間）とデータ処理時間（可視化処理時間）を図7(2)に示す。図7(1)と図7(2)を比較すると、両者の可視化処理時間はほとんど変わらないが、Gfarmのリモートストレージ（FSN）から参照する場合は、ローカルストレージを参照するよりもI/O時間が約1.6倍大きくなる。

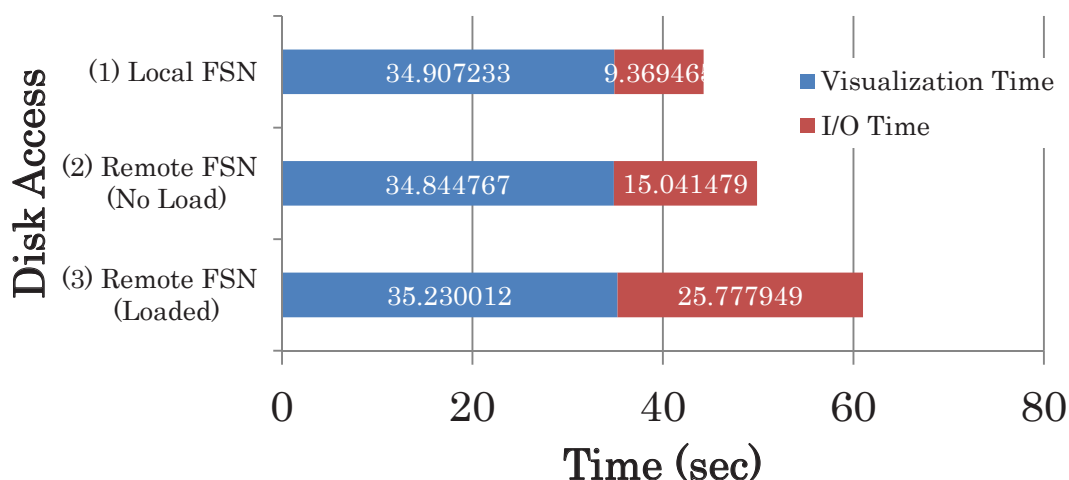


図7 予備実験の結果：本研究の対象となるデータファイル（1ファイル）を読み込み（I/O）後に可視化（Visualization）した場合のそれぞれの時間。(1) CNがFSNを兼ねる場合（ローカルFNS）（上図）、(2) CNが異なるFSNからデータファイルを読み込む場合（リモートFSN）（中図）、(3) リモートFSNに外部から他の負荷がかかっている場合（下図）。可視化処理時間はどの場合もほぼ一定であるが、データI/O時間はリモートFSNの場合は、外部負荷がない場合に1.61倍、外部負荷がある場合は2.75倍となる。

リモート FSN からアクセスする場合の I/O 時間増加の傾向は、計算の大規模並列化においてはさらに大きくなる。計算期間に流れる I/O データ量が増大し、ネットワークの輻輳が発生するためである。また、リモート FSN の負荷が上昇する場合にも、I/O 速度は低下する。FSN は任意の CN からのアクセスが常に起こり得るため、負荷を予測することができない。図 7(3) は、測定条件を図 7(2) と同様に設定し、さらにリモート FSN の全コアに外部から負荷をかけた場合の CN ノードの可視化処理時間および I/O 時間である。本実験では、CN からのアクセスとは別にリモート FSN に対して CPU コア数と同じだけのディスク I/O プロセス（データ読み出し）を外部から継続的に行うことにより、FSN を高負荷状態にした。その結果、図 7(2) と比して図 7(3) の I/O 時間はさらに長くなり、図 7(1) と比べると 2.75 倍の処理時間がかかっている。

本実験の可視化処理では図 6 に示すようにデータ読み込み（I/O）と可視化処理はシーケンシャルに実行されるため、同時に行われることがない。図 7(1) のように CN がローカル FSN のみを参照する場合には他の CN からのアクセスがないため、I/O 処理と可視化処理はプロセスごとに独立する。全処理時間の高速化のためには、図 7(3) のようにリモートアクセスがある場合は望ましくないことが分かる。本研究では、すべてのデータ処理を図 7(1) のような CN がローカル FSN のみを参照することとする。

2.3 節で述べたように、本実験の可視化処理はプロセス間通信が発生しないため、可視化並列処理のスケラビリティ向上は難しい。近年、コンピュータシミュレーションの出力データファイルサイズは増大化する傾向にあり、シミュレーションポスト処理（データ可視化）時間に比して I/O 時間が無視できない。全処理時間の短縮のためには I/O 時間の短縮が必須であり、図 7 で議論したようにローカルディスクを優先的に処理するアルゴリズムが有効である。

## 4. 基礎実験

### 4.1. 基礎実験の目的

本節では、3 節の予備実験結果をもとに、Gfarm/Pwroke による並列分散処理で高いスケラビリティを達成するための基礎実験を行う。3 節の予備実験では、Gfarm のクライアントノード（CN）を用いてデータ処理を行う場合には、ローカル FSN と比較してリモート FSN アクセスによる処理速度の低下が指摘された。Gfarm/Pwroke では、2.1 節で述べたように、複数 FSN においてファイル複製が保存される場合にはローカル FSN 上のファイルアクセスが優先する設定が可能である。言い換えると、図 5 の様な CN と FSN がすべて一致するシステムの場合、すべての FSN に対象ファイルの複製が保存されている場合には、CN は常にローカル FSN のファイルにアクセスする。そこで、本実験ではストレージコストは高いがすべての FSN にすべてのファイルを配置することとする。（本研究では、最高ストレージコストファイル配置と呼ぶ。）

本基礎実験では、図 5 の全クライアントノード（すなわち全ファイルシステムノード）を用いてデータ処理を実験ごとに 1 回行う。プロセス数によるスケラビリティを調べるため、図 5 のクライアントノード数  $N$  と各クライアントノードにおけるデータ処理プロセス数（用いるコア数） $M$  を変更して処理速度の計測を実施する。なお、処理データファイル数を図 3 に示すシミュレーションの時間ステップ番号 1 から 144 までの 144 とする。

### 4.2. 基礎実験①（ $M=1$ の場合）

基礎実験①では、複数のクライアントノード上でコア数  $M$  を 1 と固定して 1 ファイル当たりの平均処理時間を測定する。データファイル配置は、4.1 節で述べた最高ストレージコスト配置である。図 8 に測定結果を示す。2.3 節で示したとおり本可視化処理にはプロセス間通信がないため、ノードごとにコア数（プロセス数） $M$  を 1 と固定した場合、 $N=1 \sim 6$  までほぼ完全なスケラビリティを達成していることが分かる。

本基礎実験結果から、最高ストレージコスト配置を行う場合には、CN 数を 10 台、100 台、1000 台と増やしても高い高速化が期待できる。並列化する場合にオーバーヘッドになるのは、メタデータサーバへのアクセストラフィックや CPU 負荷、DB 検索速度である。NICT サイエンスクラウドでは、Gfarm による広域分散ファイルシステムを実験的に運用しているが、2 億ファイルを超える場合でもメタデータベースの応答には影響がないことが分かっている。

3 なお、本研究における I/O 時間測定は、バーチャルオーラツールのデータ読み込みモジュールの先頭と末尾で測定している。したがって、測定結果は厳密なディスク I/O 速度には一致しない。

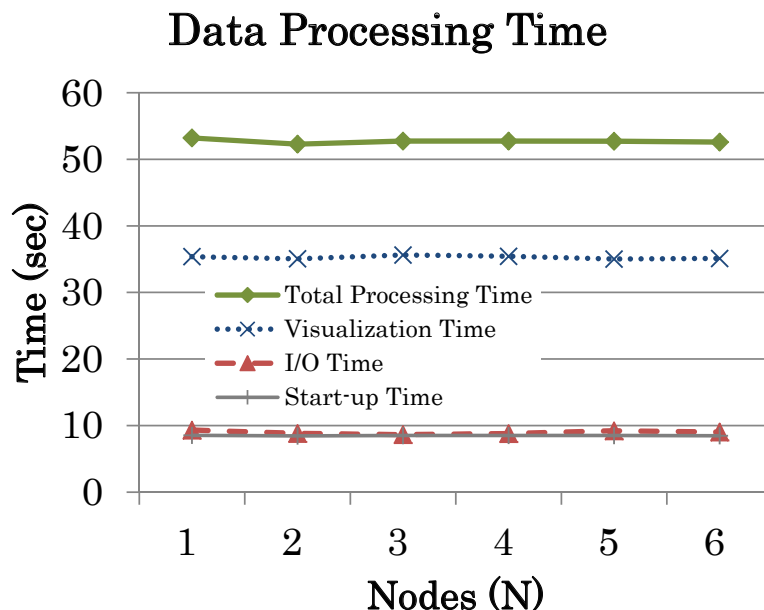


図 8 基礎実験①の結果：1ファイルのデータ処理時間（データI/O 時間，可視化処理時間，アプリケーション起動・終了時間および全処理時間）のノード数  $N$  依存性（各ノードのプロセス数（コア数） $M$  を 1 に固定した場合）。

#### 4.3. 基礎実験②

基礎実験②では，基礎実験①の結果を踏まえてノード数  $N$  を 1 および 6 に固定し，プロセス数  $M$  を 1 から 12 まで変更して処理を行う．データファイル配置は，基礎実験①と同様に最もストレージコストが高い配置である．図 9 および図 10 に，データ処理時間のプロセス数  $M$  への依存性を示す．

図 9 は，ノード H2 におけるプロセス数  $M$  に対する 1 ファイル当たりのデータ I/O 時間，可視化処理時間，アプリケーション起動・終了時間および全処理時間を示している．図 9 はノード H2 の結果であるが，他のノードでも同様の結果である．可視化処理時間については， $M=8$  まではほぼ一定の並列化効率を保っているが， $M=8$  を超えると低下する．これは，表 1 より対象となる CN のコア数が 8 であることから予想されるとおりである．しかし，可視化処理時間はアプリケーション起動・終了時間（図 6 の Start-up time）とともにプロセス数  $M$  への依存性は小さく，全処理時間の増加には寄与していない．

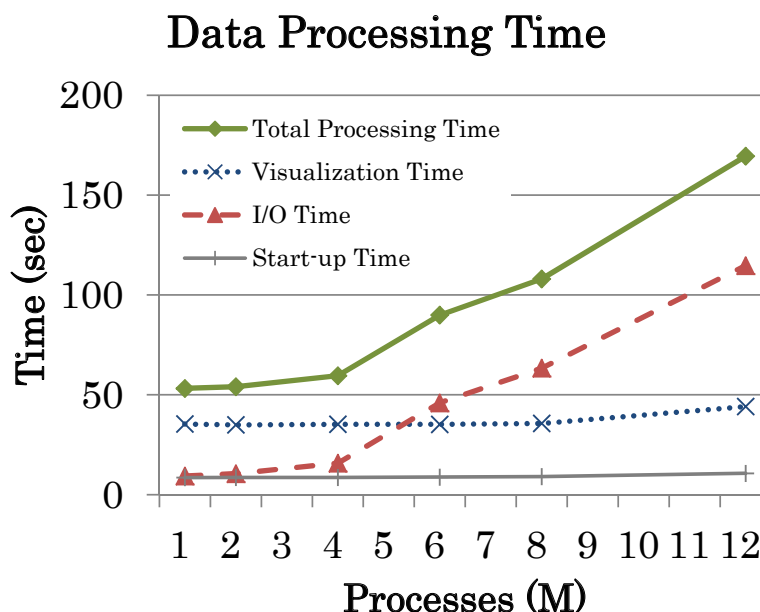


図 9 基礎実験②の結果：ノード数 1 のときの 1 ファイルあたりの平均データ処理時間（データ I/O 時間，可視化処理時間，アプリケーション起動・終了時間および全処理時間）のプロセス数  $M$  の依存性。



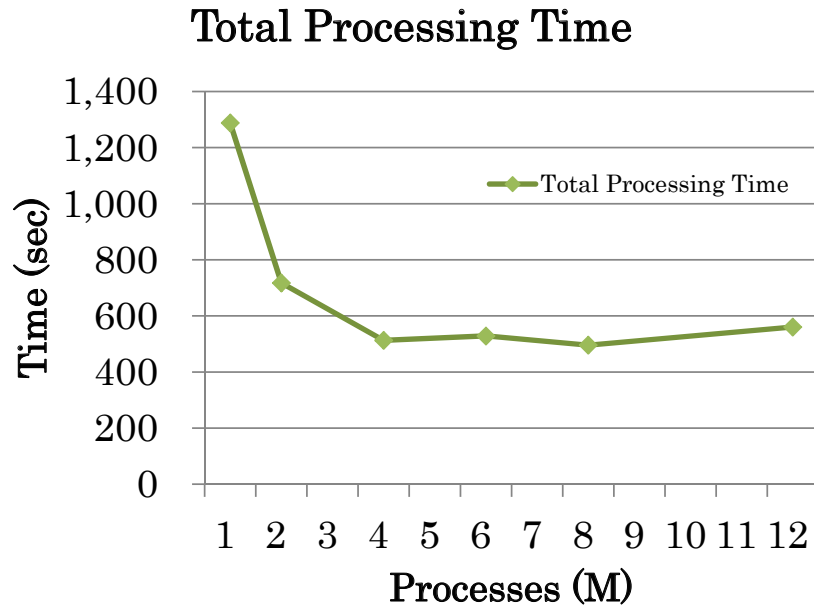


図 10 基礎実験②の結果（縦軸は全処理時間，横軸はプロセス数）：ノード数 6 の場合の並列処理による全処理時間。

表 1 本研究のデータ処理サーバ緒元

CPU コア数	8
CPU	Intel Xeon X5550 2.67 GHz
メインメモリ	144 GB
オペレーションシステム	OpenSUSE 11.1 (x86_64)
ハードディスク構成 (Read および Write はカタログスペック)	SATA 3 × 4 (RAID5) Read: 371 MB/sec Write: 137 MB/sec
NIC	10GbE

一方，データ I/O 時間はプロセス数  $M$  が増加するにしたがって増加する．これは，同じ CN で同時に複数のプロセスがローカルディスクを並列参照する場合にはディスクアクセス帯域の共有を行っていることによる．I/O 速度の低下は，特に  $M > 4$  で顕著である．一般に，SATA 系の HDD へのマルチアクセスにおいては，アクセス数が 2 ～ 8 以上に増加するほど I/O 性能は低下する傾向がある．

図 10 に，ノード数  $N$  を 6 としたときの全データ処理時間の  $M$  への依存性を調べた結果を示す．図 10 の全データ処理時間は図 8 や図 9 のファイル単位での処理時間ではなく，6 クライアントノードを用いた全 144 ファイル全体の処理時間（処理開始時刻と処理終了時刻の差）を表している．処理時間が最も短かったのはプロセス数  $M=8$  の場合であり，1 ノード 1 コアで必要であった全処理時間 7741.7 秒に対して 495.7 秒で処理が終了した．図よりプロセス数  $M=4$  までは処理時間が短縮されるが， $M > 4$  で処理時間は横ばいとなる．これは，図 9 に示すように  $M > 4$  では I/O の処理時間が増加することによる．

図 11 は，図 10 から求めた高速化率を示す．高速化率はノード数  $N = 1$  でプロセス数  $M=1$  の場合の全処理時間を基準値として，図 10 の全データ処理時間よりノード数  $N=6$  における各プロセス数  $M$  での処理時間から求める．その結果，最も高い高速化を達成したのはプロセス数  $M=8$  の場合であり，高速化率は 15.6 であった．

## 5. 実利用実験

### 5.1. 実利用実験の目的

前節の基礎実験では，6 台のクライアントノード（ファイルシステムノードを兼ねている）から構成される図 5 のシステムで Gfarm/Pwrake により 144 ステップ分の数値シミュレーションデータを並列処理する実験を行った．本節では，

このパラメータにより図5のシステムにおいて多ステップシミュレーションデータ処理を1回行い、その結果を評価する。処理データは、2.3節で述べたデータファイルのうち、ステップ1からステップ780まで780ファイルを対象とする。本Global MHDシミュレーションの780ステップ(時間分解能0.5秒)は、実時間で390秒(6分30秒)に該当する。6分30秒は地球磁気圏現象としては短時間であるが、0.5秒は磁力線追尾<sup>9)</sup>などに必要な時間分解能である。

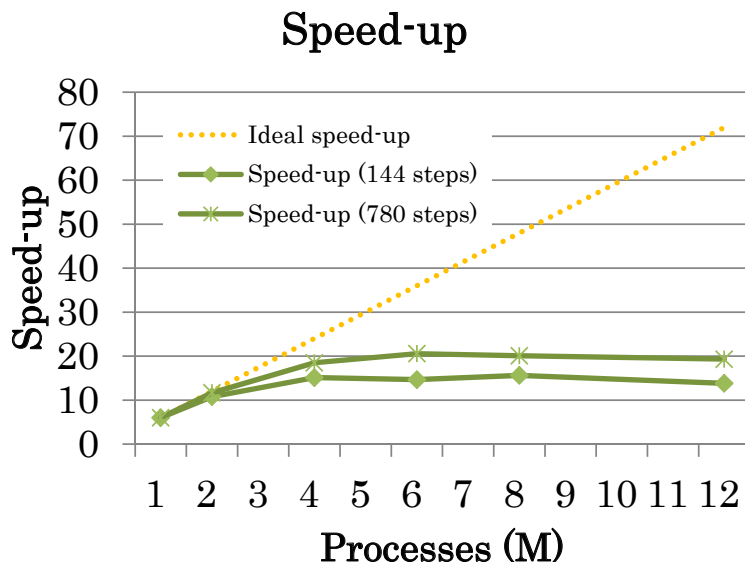


図 11 基礎実験②(144 ステップ)と実利用実験(780 ステップ)の結果: ノード数  $N=6$  の場合の並列処理による高速化率(Speed-up)。横軸はプロセス数。

## 5.2. データファイル処理時間と高速化率

図 12 に、実利用実験のファイルごとの処理時間を示す。図 12 は、処理を行った全 780 ファイルについて、それぞれの可視化処理時間とデータ I/O 時間(図 6)のデータファイルごとの時間変化を示している。横軸はファイル番号(計算機シミュレーションデータの時間ステップ番号)に対応する。基礎実験②の高速化率結果である図 11 には実利用実験の高速化率も示されており、最高の高速化率は  $M=6$  の場合で 20.6 である。実利用実験の高速化率は、図 12 の結果をもとに基礎実験②と同じ予備実験における 1 ノード 1 プロセスでの 1 ファイル処理時間から求めた。なお、図 12 の横軸は処理時刻ではなくデータファイルのステップ番号である。各処理時刻の可視化処理時間とデータ I/O 時間については、5.3 節において議論する。

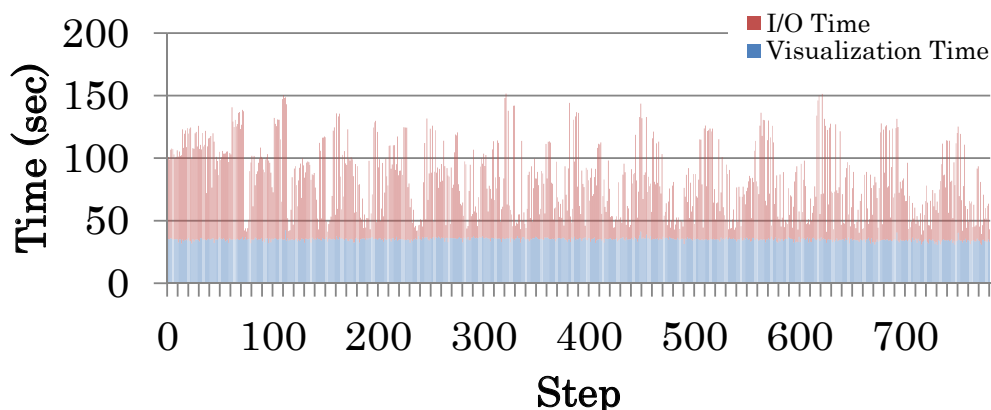


図 12 実利用実験のファイルごと処理時間: 処理を行った全 780 ファイルについて、それぞれの可視化処理時間(青色)とデータ I/O 時間(赤色)を示している。

<sup>4</sup> 2.3 節で述べたとおり、本研究のデータ処理ではプロセス間通信は発生しない。また、ファイル配置は 4.1 節で述べた最高ストレージコスト(すべての時間ステップのシミュレーションデータが全ての FSN に配置されている)を採用している。2.1 節で述べたとおり、図 12 の横軸はファイル処理の順序とは必ずしも一致していない。

図 11 より、実利用実験の高速化率はプロセス数  $M$  に関わらず基礎実験②の高速化率よりも高いことが分かる。実利用実験と基礎実験②との違いは処理したデータファイル数のみであり、その他の条件は同じである。基礎実験②の 144 ステップ処理よりも実利用実験の 780 ステップ処理の場合の方が高い高速化率を示す理由について、以下に検討・考察する。

### 5.3. 高速化率向上の検討と考察

図 12 によると各データファイルの全処理時間（図 6）において可視化処理時間はほぼ一定であり、780 ファイル処理の平均値は 35.07 秒である。これは、予備実験である図 7(1) の値とほぼ等しい。一方、データ I/O 時間は先頭から 60 ステップ目までと 60 ステップ以降で傾向が異なっている。60 ステップまでは 60 ～ 80 秒程度と大きい<sup>5</sup>が、ファイル処理が進むにしたがってばらつきが大きくなる傾向が読み取れる。

これを確認するため、6 台のクライアントノード（図 5）のコアごとのデータファイル全処理時間（図 6）のヒストグラムを図 13 に示す。図 13 の各ヒストグラムの色は処理したデータファイルの順序を示している。横軸は処理時刻であり、本図ではデータファイルごとにデータ I/O 時間と可視化処理時間をあわせた全処理時間を一つの色で示している。どのノードにおいても先頭または 2 番目のファイルについては全処理時間が各ノード内でほぼ一致しているのに対し、処理が進むにつれて全処理時間にずれが発生している<sup>6</sup>。また、全処理時間は処理が進むにつれて短くなる傾向にある。図 12 より可視化処理時間はほぼ一定であるので、全処理時間の短縮は I/O 時間の短縮を意味している。図 12 の後半の時間帯のように各コアでの処理時刻にずれが発生すると、ファイルアクセスを行う時刻がコアごとに異なってくるため、I/O の分散化が達成されたと考えられる。

この効果を確認するため、実利用実験において全クライアントノードの可視化処理時間とデータ I/O 時間の時間変化を示したのが図 14 である。図 14 では、図 12 に示した 6 ノードの全 6 プロセスについて、先頭から処理するファイル番号が等しい可視化処理時間およびデータ I/O 時間のステップごとの平均値を求めた結果である。たとえば、左から 2 つめの点は、各 36 プロセスが 2 番目に処理したデータファイルの可視化処理時間とデータ I/O 時間を示している。横軸の時刻は、36 プロセスの各処理時刻の平均で定義した<sup>7</sup>。

図 14 では可視化処理時間はほぼ一定であり、図 12 の結果と一致している。一方、データ I/O 時間は横軸の処理時刻により異なっており、全体に処理が進むごとに短くなる。これは、処理が進むごとに I/O 効率が上がっていることを示唆している。この理由を以下に考察する。

図 9 より、クライアントノード内の I/O の並列数が増えるほど I/O 時間は増加し、I/O 効率は低下することが分かっている。図 12 の最初のステップでは各ノードにおいて 6 プロセス全てで同時にファイル読み込みが始まっており、ディスクへの同時アクセスが行われるため I/O 効率は悪い。後半のステップでは、図 12 から分かるようにプロセス間での処理のばらつきが発生するためディスクへの同時アクセス数が減少し、結果的に I/O 効率の向上の理由となっている。処理の後半（時刻 1500 秒以降）では、I/O 時間がさらに減少している。これは、すべての処理が終了したプロセス（コア）が増え始めていることを示している。

本研究では、自発的なデータ処理時間のばらつきによる I/O の分散と効率化が実現した。一般に、Global MHD シミュレーションでは時間ステップごとのデータサイズは一定であり、ファイルごとに I/O 処理と可視化処理は独立に行われることが多い（図 6）。Gfarm/Pwrake によるこのようなシミュレーションデータの並列可視化処理（ポスト処理）では、意図的にプロセスごとに I/O 処理を分散化するアルゴリズムを導入することで、さらに高速化が可能となることが示唆された。

5 図 13 のフォーマットは図 2 右側のフォーマットに該当する。

6 シミュレーションデータの可視化処理では、データサイズは全てのデータファイルで等しいが、描画する磁力線の長さや複雑さが変動する。そのため、データ可視化処理時間（データ処理時間）はファイルごとに異なってくる。5.1 節で述べたとおり実利用実験の 780 ステップは 6 分 30 秒に該当するが、この期間は地球磁気圏構造に大きな変化がないため、可視化処理時間はほぼ一定である。

7 2.1 節で述べたとおり、Gfarm/Pwrake のワークフローでは、シミュレーションデータファイル順にデータ処理が行われるとは限らない。そのためにグラフの点が横軸方向に逆転することがある。

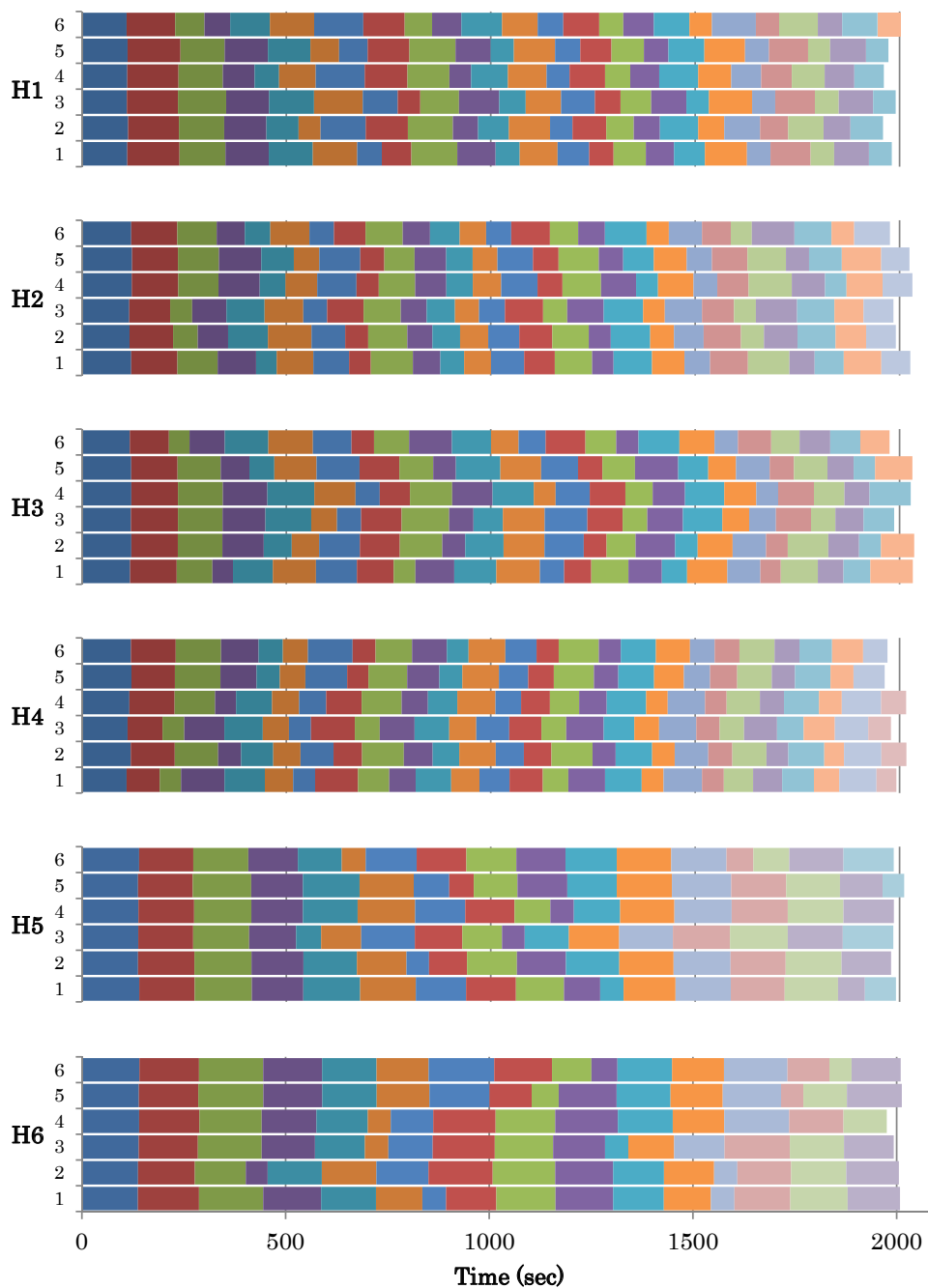


図 13 実利用実験の結果：各クライアントノード（H1 ～ H6）のコアごとのデータファイル処理時間のヒストグラム.

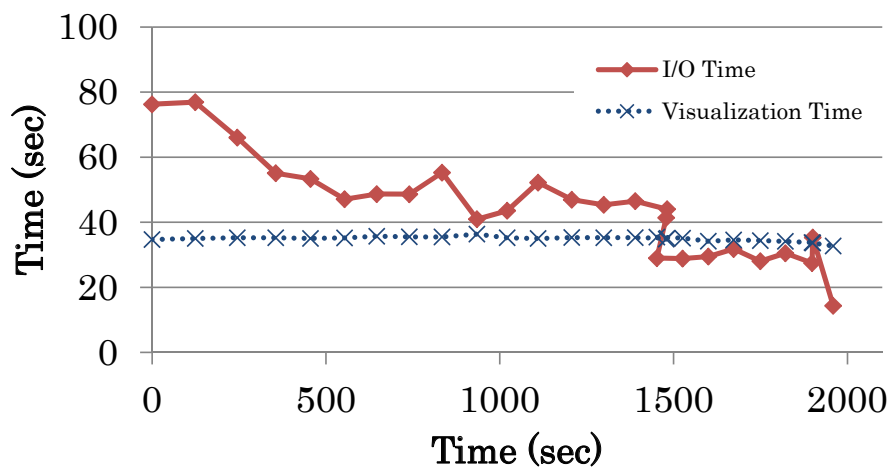


図 14 実利用実験の結果：全クライアントノードの可視化処理時間とデータ I/O 時間の平均値の時間変化

## 6. まとめ

計算指向型の HPC (High Performance Computing) は数値計算に有効な技術であるが、出力された大規模データのポスト処理に対して有効とは限らない。すなわち、大規模数値データを解析するための新しい科学研究手法が必要であり、Gray はデータ指向型科学研究手法を第 4 のパラダイムと呼んだ<sup>1)</sup>。

Many-Task Computing (MTC) はこのような背景において、HPC と対をなす概念として提案された。MTC では、多様な計算機リソースを融合し、データ分散、並列データ処理やコンピュータとデータファイルをローカライズする工夫など、総合的なデータ処理環境をめざす。すなわち、スーパーコンピュータが HPC 的であるとする、MTC 的であるのがサイエンスクラウドであると筆者らは考える。

MTC を実現するために我々が解決せねばならないサイエンスクラウド技術はまとめると次のようになる。① MTC を実現するために CPU、I/O およびネットワークスループットなどの分散化とその融合を行う技術、②クラウド内のヘテロ環境における分散データ処理のスケラビリティの達成、③目的達成のためのクラウドコストの最適化の解決手法の確立。

本研究の実験では①の CPU と I/O の分散を試みた。分散ファイルシステムと分散処理ワークフローツールにより、ネットワークスループットを最低にすることで、大規模ポスト処理に有効な並列分散処理を提案した。

一方で、③の視点からは本提案が最も有効であるとは限らない。本研究は閉じたネットワーク内において行ったが、これを複数ユーザのデータ処理が融合するサイエンスクラウドで行った場合には、コスト問題が発生する。たとえば、本研究ではストレージコストが最も高い方法（すべてのファイルシステムノードにすべてのデータファイルを配置する）を採用した。これは、例えば、100TB のデータを 100 サーバで処理する場合には合計で 10PB のデータ領域を必要とするため、現在のシステムでは現実的ではなくなる。また、分散ファイルシステムをクラウド内で複数ユーザが利用している場合には、データ処理ノード（クライアントノード）がファイルシステムノードとして機能することがボトルネックになる。すなわち、別ユーザがクライアントノードを兼ねたファイルシステムノードにアクセスすると、そのクライアントノードの負荷が上昇し、分散処理の最適化が崩れることがある。

②のヘテロ環境における分散データ処理を模式図にしたのが図 15 である。図 15 は複数のクラスタ（または HPC サーバ）がネットワーク接続されており、クラスタを随時追加することができる可用性の高いスケラブルクラスタシステムのイメージ図である。NICT サイエンスクラウドを含む多くのサイエンスクラウドでは、このようなスケラブルクラスタシステムであり、今後のサイエンスクラウドの主流となる。図 15 の様なヘテロ計算機環境においてスケラブルなデータ処理を実現するためには、サーバ内の分散データ処理の高速化（HPC 的高速化）、クラスタ内でのデータ処理の高速化（MTC 的高速化）、クラスタ間での分散データ処理の高速化（HTC 的高速化）の段階を経ることが有効な手順であろう。本研究は、クラスタ内でのデータ処理の高速化を行う MTC 的高速化研究である。Gfarm/Pwake を用いたクラスタ間での高速分散処理の研究もある<sup>9)</sup>が、本研究成果を活かし、クラスタ内で最高の処理効率を拡張した MTC 的高速分散データ処理技術の発展が今後は求められるであろう。NICT サイエンスクラウドは図 15 の環境をすでに実現しており、MTC 研究と HTC 研究のテストベッドとしての機能が期待される。

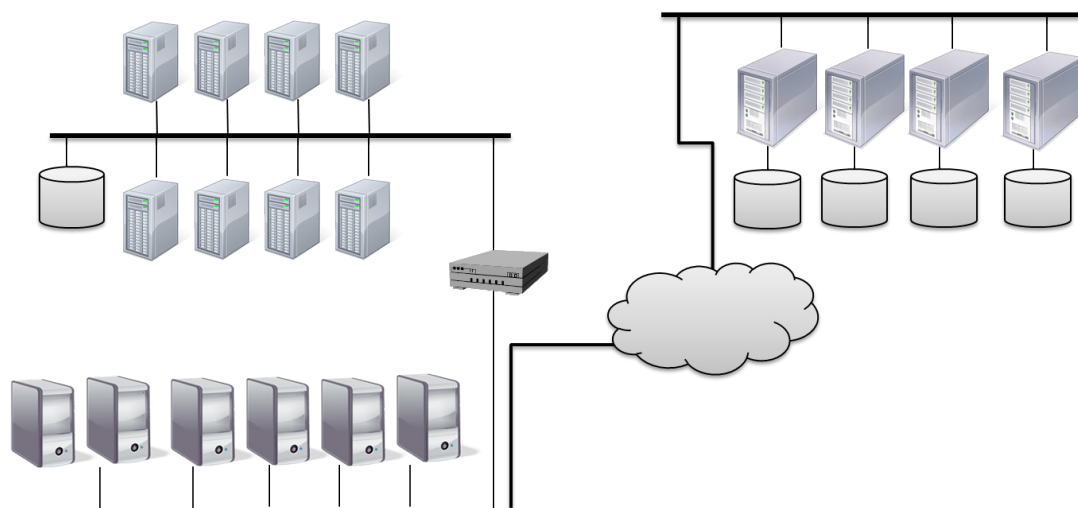


図 15 可用性の高いスケラブルクラスタシステムのイメージ図：複数のクラスタ（または HPC サーバ）がネットワーク接続されており、クラスタ単位で随時追加することができる。



### 参考文献

- 1) Hey, T., S. Tansley, and K. Tolle. The Fourth Paradigm: Data-Intensive Scientific Discovery, Microsoft Research, Washington, 2009.
- 2) 平尾公彦, 横川三津夫, 京コンピュータと計算科学研究機構 (<特集>次世代スーパーコンピュータ「京」:動き出した大型プロジェクトの全体像), NII 論文 ID(NAID):110008686987, 日本物理學會誌 66(7), 2011-07-05, 524-528.
- 3) 2012 年第 3 回 ScienceCloud ワークショップ <http://ceng.usc.edu/~simmhan/ScienceCloud2012/>, 2012.
- 4) Tatebe,O., K. Hiraga and N. Soda. Gfarm Grid File System. New Generation Computing,vol.28, no.3, 2010, p.257-275.
- 5) 田中昌宏, 建部修見, 並列分散ワークフローシステム Pwrake による大規模データ処理, 宇宙科学情報解析論文誌, 1, 2012, p.67-75.
- 6) HPCI コンソーシアム Web サイト, <http://hpci-c.jp/>.
- 7) Murata, K. T., S. Watari, T. Nagatsuma, M. Kunitake, H. Watanabe, K. Yamamoto, Y. Kubota, H. Kato, T. Tsugawa, K. Ukawa, K. Muranaga, E. Kimura, O. Tatebe, K. Fukazawa and Y. Murayama, A Science Cloud for Data Intensive Sciences, Data Science Journal, Vol. 12, 2013, pp.WDS139-WDS146.
- 8) Fukazawa, K., Y. Aoyama, T. Ogino and K. Yumoto, Response of Cross Polar Cap Potential to IMF and Velocity of Solar Wind, J. Atmos. Solar-terr. Phys. 72, 2010, doi:10.1016/j.jastp.2010.06.002.
- 9) Kubota, Y., K. Yamamoto, K. Fukazawa and K., T. Murata, Visualization of the Flux Rope Generation Process Using Large Quantities of MHD Simulation Data, Data Science Journal, Vol. 12, 2013, pp. WDS134-WDS138.

# 地球惑星試料デポジトリ DREAM の構想と現状

北川宙, 谷内勇介, 国広卓也, 中村栄三

## Depository for references of earth and analytical materials : its concept and status on 2013

Hiroshi Kitagawa, Yusuke Yachi, Tak Kunihiro, and Eizo Nakamura

平成 25 年 11 月 27 日

### Abstract

Institute for Study of the Earth's Interior, Okayama University is constructing DREAM (Depository for references of earth and analytical materials). As a core of the depository, software dedicated for analytical laboratories to guarantee traceability and accessibility of rocks with their geochemical properties has been developed. The software documents the sample origin, current sample location, the location of any sample sub-sets (e.g. thin section, solution, etc.) as well as archives all associated geochemical datasets. In this article, concept and status of DREAM is described.

### 概要

岡山大学地球物質科学研究センターでは、岩石試料を分析データとともに将来引き継ぐべく、地球惑星試料デポジトリ DREAM を構築中である。分析実験室で変化を続ける物質の保守の結果、試料の親子関係と格納場所を追跡することが物質管理における重要二要素であるとの認識に至る。この二要素を追跡し試料とデータを紐付ける専用データベースシステムの開発を行った。DREAM はこのシステムを用いて恒常的に運用されており、現在 12,000 試料の形態、採取場所、現保管場所のみならず、薄片や分解溶液といった子試料の所在、そして試料の化学組成データ追跡している。本稿では DREAM の構想と現状について述べる。

### 1. 構想

サンプルリターン試料を地上で分析する際、探査機搭載装置のような制約を受けない。そして物質が存在する限り、最先端の分析技術を適用できる。これまでに人類はアポロ、ルナ、ジェネシス (1)、スターダスト (2)、そしてはやぶサ (3) といったサンプルリターンミッションを遂行した。ジェネシスは太陽系の平均酸素同位体組成決定を第一目標にかかげ、実験室における分析装置 MegaSIMS 開発をミッション中核に位置付けた。太陽風が埋めこまれた試料が予想外の手順で地球に帰還した時、分析装置はまだ完成していなかった。しかし先に述べたよう、サンプルリターン試料は、汚染対策のされた環境で保管されている限り、分析の時期を問わない。三年後に分析が行われ、太陽系の平均同位体組成が“隕石に観察される異常と認識していたものであること”が判明する。惑星こそが太陽から進化した異端的存在であり、人類はガリレイ・ガリレオから再び学ぶこととなった (4)。岡山大学地球物質科学研究センターは、物質分析を基軸として地球・惑星の起源、進化およびダイナミクスに関する先進的かつ実証的な研究を行っている。総合的地球化学分析と年代測定技術により、46 億年におよぶ地球を含む惑星の歴史を紐解き理解する上で、最も基本となる物理化学的情報を提供する、というのがスタンスである。本センターは、先進的地球惑星物質科学を展開すると同時に、そこで用いた重要な物質科学的試料 (岩石、薄片、パウダー等の形態を持つ地球試料や隕石試料) と公表されたデータを、統一のプロトコルによって管理し、世界の研究者に供することで学問の体系化と継承を実現することを目指すこととした。着実な教育研究成果の積み重ねによって、それを国際的に還元し社会的な「知」のニーズに答えることは学術研究の責務であって、その継続発展を可能とする体制が確立することの意義は極めて大きい。地球惑星試料デポジトリ (DREAM : Depository for

the Reference Earth and Analytical Materials) は、単なる博物学的な収集にとどまらず、学術的に価値の高いデータが付随した物質試料をデータベース化して保管し、将来の研究課題に役立てるとともに共同研究に供することを目的とする。地球惑星物質のみならず、病態生理関連物質等にもその運用を拡張し、これが機能することによって全学的な取り組みへ、さらには大学組織を超えた共通の物質科学デポジトリへと発展することも期待できよう。さてこのようなデポジトリに格納された物質を将来の研究課題において再分析するにあたり、相応量と出版された分析データが残されているだけでは不十分である。それまでに獲得された分析データに加え、試料の生い立ちや加工状況、関連試料の所在が伴っていることが必要である。近年、地球惑星物質の分析データのアーカイブが進められている。例えば GEOROC(5), PetDB(6), GeoReM(7), NAVDAT(8) といったオンラインシステムでは、出版された岩石の化学・同位体組成に参照できる。これらは、多種多様な分析データに系統のかつ迅速にアクセスする機会を与え大変有用であるが、試料の再分析にはさほど役立たない。我々が構築を目指すのは、出版されたデータを格納するデジタルアーカイブではない。実験室で継続して試料を管理(収集、記載、管理、配付、公開、すなわちキュレーション)し、試料の再分析そして科学の継続発展を支援するデポジトリである。このデポジトリは物質試料そのもの、それを格納するハードウェア、そして情報を管理するデータベースから構成されることとなる。データベースは組成や物性情報のみならず、物質の処理法や機器分析の条件、分析箇所を記録する画像といった公開を前提としない情報をアーカイブする。このようなデポジトリは現存せず、研究者の退官と同時に、試料整理と称し大半の試料が破棄されている。

## 2. メデューサ

我々がこのようなデポジトリを構想している中、ハヤブサミッションにおける初期分析を担当した(9)。試料は分析に応じさまざまに加工され、短期間に多くの実験室を回遊した。このような動的状況を記録し、試料状態を共有し、試料を継承するには、試料管理と化学分析データのカタログ化を統合的に行う専用データベースシステムの開発が必須であるとの認識に至った。

DREAM(<http://dream.misasa.okayama-u.ac.jp/>) は専用データベースシステム・メデューサ(10)を中核に据えた、物質と分析データを継承するためデポジトリである。本章ではデータベースシステム・メデューサを簡単に紹介する。

メデューサはウェブアプリケーションフレームワークである Ruby on Rails を用いて開発された。ユーザーはウェブブラウザもしくはウェブ API を用いたクライアントプログラムによってアクセスする。

メデューサは、出版されたデータを扱うのみならず、物質が変化し移動する様を追跡する。これは試料加工に伴い発生する試料-子試料-孫試料という親子関係と、建物-部屋-棚-容器などの格納状況を表現すると換言できる。試料の親子関係は stone クラスの内部リンク、格納状況は box クラスの内部リンクによって記述する。この二つを含む代表的なクラスを表に、それらの相互関係を図1に示す。以下に採取・加工・格納・分析において発生するレコード関係を記す。

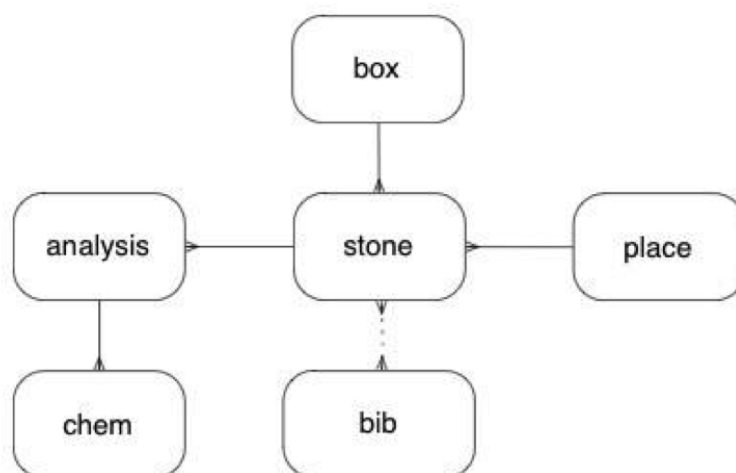


図1 試料を記述するクラスの関係。線端の三本足は複数のレコードとリンクされる関係性を示す。例えば analysis は複数の chem とリンクすることが多い。破線は研究成果が公表された時に発生する文献へのリンクである。

表 1 メデューサの扱うレコードの種類

クラス	説明
stone	試料そのもの
box	試料を保管する箱, 棚, 部屋, 建物, またはマウント
place	試料の採取地点
analysis	試料を分析した手法
chem	分析によって得られた化学特性 ( 元素・同位体存在度 )
bib	分析結果を公表した論文

採取時に試料 (stone) と場所 (place) の関係が発生する ( 図 2). 実験室に運ばれた試料は薄片・粉末・溶液へと加工される. これら子試料に対応する stone レコードが作成され, リンクにより親子関係が表現される ( 図 2). 同様の試料の所在は box レコードのリンクにより表現される. 粉末試料は密閉式ボトルに封入され, 他試料と共にクリーンルームの棚に納められる ( 図 3). 続いて分析が行われる. 図 4a は全岩分析 (analysis) によって決定した複元素濃度 (chem) が試料 (stone) に関連付けられている様子である. スポット分析では試料表面の多点から複数の元素濃度を得て, 図 4b に示す分析点への多重リンクを加えて表現される.

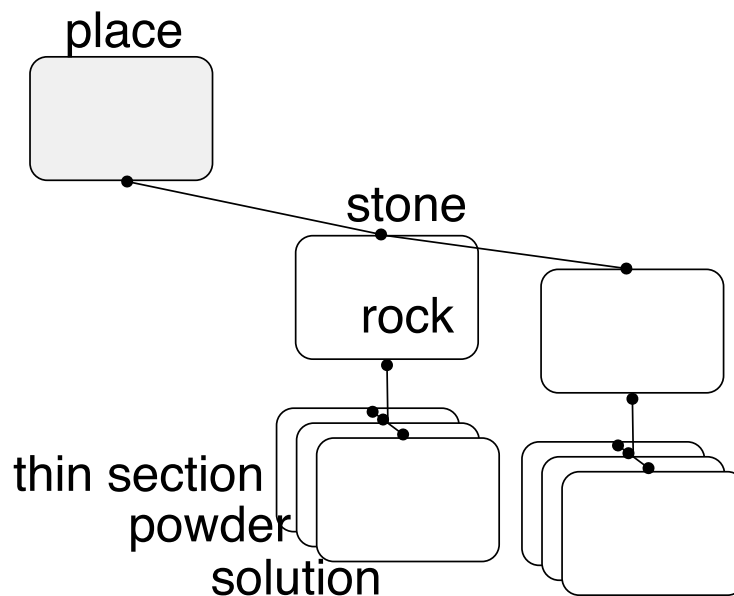


図 2 試料の親子関係の表現するレコード: 採取点から 2 つの岩石が採取され, それぞれから薄片, 粉末, 溶液が作成された状況を表現している.

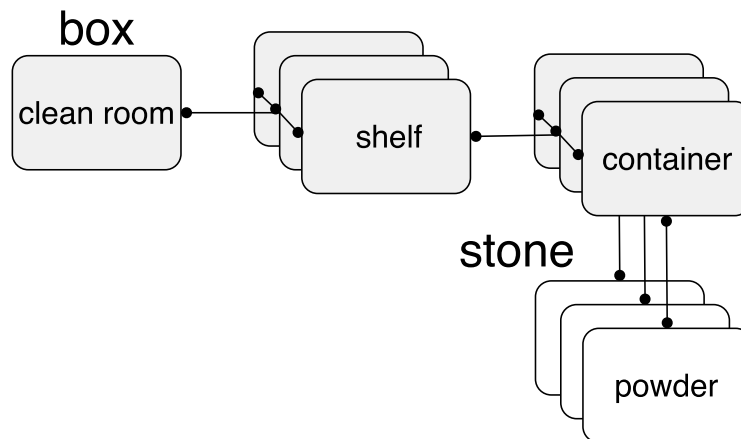


図 3 試料の格納状況の表現するレコード: 複数の粉末試料が個別に容器に保存され, それらがクリーンルーム内の棚に格納されている状況を表現している.

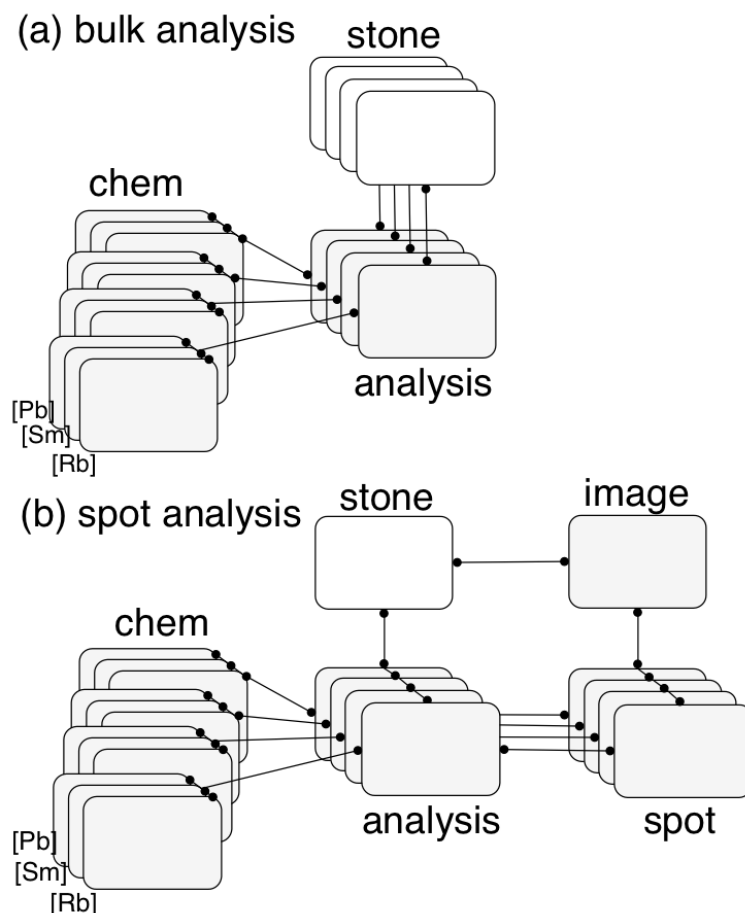


図4 試料の分析状況の表現：(a) 全岩試料の分析によって Pb,Sm,Rb 濃度を得た状況表現する；(b) スポット分析によって試料表面の多点から Pb,Sm,Rb 濃度を得た状況表現する。

上記の入りくんだレコード関係はメデューサによってダイナミックに辿られ、親子関係、位置関係、分析結果がウェブブラウザにて表示される。イトカワ回収試料を観閲したものを図5に示す。イトカワ試料は8つの標準試料と共にマウント“mount-K002-b”に搭載され、棚“Hayabusa1”に保管される。イトカワ試料の家系と、8つの標準試料が一つのマウントに格納される状況が画面左手上下の樹形図にて表現される。鉾物片に対して行われた分析の結果は、右手にテーブルとして示される。

### 3. 現状

DREAM には年 3,000 の試料が追加され現在 12,000 点が登録されている。試料のみならず、備品、消耗品、そしてソフトウェアのライセンスも管理され、DREAM は既に実用段階にあるといえる。しかしいくつかのパイロットデータを除き公開に至っていない。DREAM の現在の課題を以下に述べる。

#### 3.1 作業インフラの充実

試料は ID により識別される。実空間の物質試料とデータベース空間のレコードはこの ID で連結するため、試料は ID を記したメディア（ラベルと呼ぶ）と共に行動する必要がある。子試料が発生する度にレコードのラベルを作成する必要がある、試料を移動する度にデータベースを更新する必要がある。

従ってレコードとラベルの作成と、現在位置の更新を迅速に行う環境の整備は DREAM 運用の必要条件のひとつである。環境不備は、本来の作業に影響を与え、即試料の管理放棄につながる。ウェブブラウザのみで試料管理を行うことは事実上難しい。



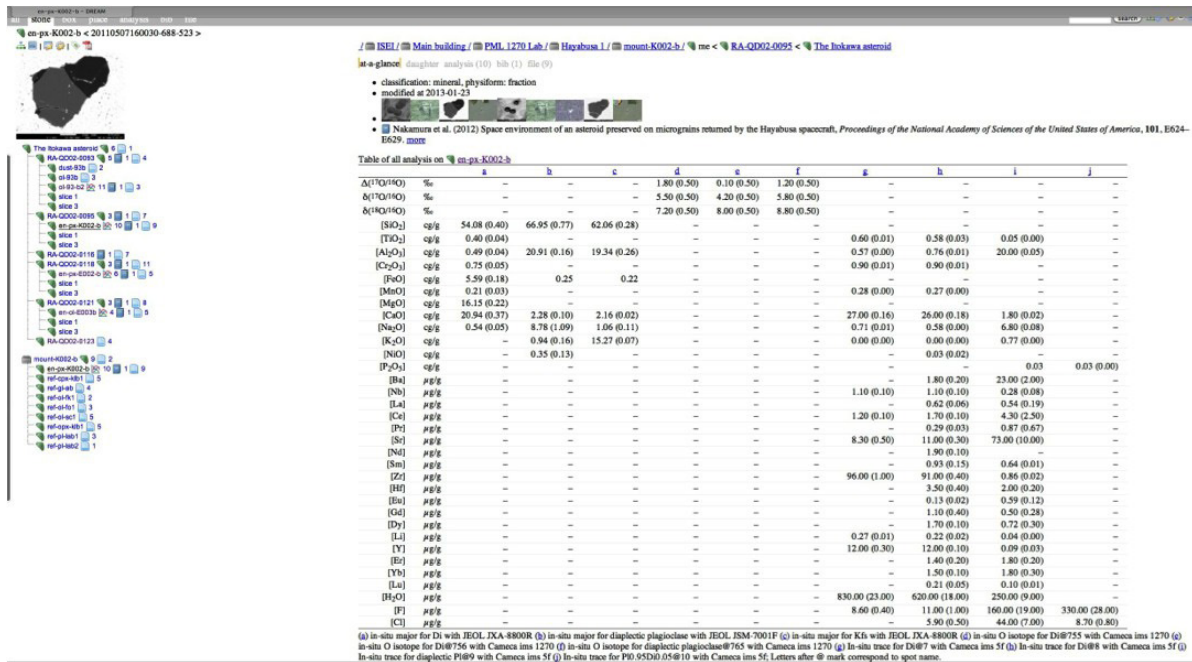


図 5 ウェブブラウザを用いて小惑星イトカワから回収された鉱物片 (low-Capryroxene) を観閲している様子: イトカワ試料は 8 つの標準試料と共にマウント “mount-K002-b” に搭載され, 棚 “Hayabusa1” に保管される. イトカワ試料の家系と, 8 つの標準試料が一つのマウントに格納される状況が画面左手上下の樹形図にて表現される. 鉱物片に対して行われた分析の結果は, 右手にテーブルとして示される.

現在は 10 セットのラップトップコンピュータ, テーププリンタ, バーコードリーダー, カメラ, ハンドヘルドコンピュータを各現場に配備している (図 6). これらの作業インフラを拡大し, 保守性, 機能性, 使い勝手と安定性を高めることは緊急の課題である.



図 6 作業現場に配置される “DREAM 屋台” と呼ばれるポータブル作業環境: ラップトップコンピュータ, テーププリンタ, バーコードリーダー, カメラ, ハンドヘルドコンピュータから構成される. 移動させて利用することが前提となるため, UPS で駆動する.

### 3.2 メデューサの最適化

現在のメデューサはパフォーマンス、ユーザーインターフェース、そしてセキュリティに不安を持つ。

メデューサはリンクをダイナミックに辿る。重要な試料や標準試料などリンクを多く持つ試料を閲覧する際にはレスポンスが問題となる。また、ルーチンをスピードを持ってこなす機能に足りない。例えば試料を6つ載せたマウントを登録するにはよくある作業なのだが表2に示すよう48行程が必要である。他にも例えば現場での操作ミスをリカバーする方法に足りない。誤って消去したレコードを復旧することは可能であるが、あらゆるデータを同時に一日前に戻すことになる。

これらの問題を解消するため、クライアントの開発を含め、ソフトウェアとしてのメデューサの最適化をはかる必要がある。

### 3.3 アロケーション体制/データ登録ガイドラインの整備

現在出版されたものについて“試料の存在と分析データ”の公開を目指している。特定試料を公開することに技術的困難はない。しかし公開後は、サンプルリクエストに対応する必要がある。アロケーション委員会の整備が公開の必要条件だが、現在はメデューサの開発に注力する段階にあり、組織整備に至っていない。メデューサは、出版以前の分析データの共有を可能とする。研究の渦中にある関連試料の存在やその格納場所、また分析データを系統的に示す機能は非常に有用である。ただし(出版されたデータと異なり)出版前のデータは処理の再検討等による変更がありえるため、取り扱いに注意が必要である。まず、出版前のデータ登録について、ガイドラインを整備する予定である。その後、生データから動的にデータ処理を行う機能の実装を計画している。

表2 試料を6つ載せたマウントをメデューサに登録に必要な作業

作業	手数
新規マウントを登録	1
マウントのグループとパーミッションの設定	1
マウントの格納場所(例:クリーンベンチ)を入力	1
マウントのラベルを印刷	1
マウントの写真を撮影・アップロード	1
マウント写真のグループとパーミッションの設定	1
それぞれの試料を新規登録	6
それぞれの試料のグループとパーミッションの設定	6
それぞれの試料の格納場所(マウント)を入力	6
それぞれの試料の親を入力	6
それぞれの試料の写真をアップロード	6
それぞれの試料のマウント内位置を入力	6
それぞれの試料の写真のグループとパーミッションの設定	6
計	48

## 4. まとめ

岡山大学地球物質科学研究センターでは、岩石試料を分析データとともに将来引き継ぐべく、地球惑星試料デポジトリ DREAM を構築中である。本稿ではその構想、現状について述べた。DREAM は試料管理のための開発されたデータベースシステム・メデューサを核とする。メデューサは分析データのみならず加工履歴、保管場所、分析条件など情報を統合し、試料を追跡する。

現在メデューサのドキュメント整備を終え、安定版の開発に着手したところである。安定版では、パフォーマンスが向上すると同時にセキュリティの強化がなされ、また簡便な導入と運用が可能となる。

## 参考文献

- [1] Burnett D.S., Barraclough B.L., Bennett R., Neugebauer M., Oldham L.P., Sasaki C.N., Sevilla D., Smith N., Stansbery E., Sweetnam D. and Wiens R.C. (2003) The Genesis Discovery Mission: Return of solar matter to Earth. *Space Science Reviews*, 105, 509-534.
- [2] Brownlee D.E., Burnett D., Clark B., Hanner M.S., Horz F., Kissel J., Newburn R., Sandford S., Sekanina Z., Tsou P. and Zolensky M. (1996) STARDUST: Comet and interstellar dust sample return mission. In: Gustafson B.A.S. and Hanner M.S. (eds), *Physics, Chemistry, and Dynamics of Interplanetary*, IAU Colloquium 150, Astronomical Society of the Pacific Conference Series, 104, 223-226.
- [3] Fujiwara A., Kawaguchi J., Yeomans D.K., Abe M., Mukai T., Okada T., Saito J., Yano H., Yoshikawa M., Scheeres D.J., Barnouin-Jha O., Cheng A.F., Demura H., Gaskell R.W., Hirata N., Ikeda H., Kominato T., Miyamoto H., Nakamura A.M., Nakamura R., Sasaki S. and Uesugi K. (2006) The rubble-pile asteroid Itokawa as observed by Hayabusa. *Science*, 312, 1330-1334.
- [4] McKeegan K.D., Kallio A.P.A., Heber V.S., Jarzebinski G., Mao P.H., Coath C.D., Kunihiro T., Wiens R.C., Nordholt J.E., Moses Jr. R.W., Reisenfeld D.B., Jurewicz A.J.G. and Burnett D.S. (2011) The oxygen isotopic composition of the Sun inferred from captured solar wind. *Science*, 332, 1528-1532.
- [5] Sarbas B. (2002) Geochemistry of oceanic island-arc and active continental margin volcanic suites: Some statistical evaluations and implications using the database GEOROC. *Eos Transactions AGU*, 83, Fall Meeting Supplement, Abstract V62B-1401.
- [6] Lehnert K., Su Y., Langmuir C.H., Sarbas B. and Nohl U. (2000) A global geochemical database structure for rocks. *Geochemistry Geophysics Geosystems*, 1, 1012, doi:10.1029/1999GC000026.
- [7] Jochum K.P., Nohl U., Herwig K., Lammel E., Stoll B. and Hofmann A.W. (2005) GeoReM: A new geochemical database for reference materials and isotopic standards. *Geostandards and Geoanalytical Research*, 29, 333-338.
- [8] Carlson R.W., Walker J.D., Black R.A., Glazner A.F., Farmer L. and Grossman J. (2001) NAVDAT: a western North American volcanic and intrusive rock geochemical database. *Geological Society of America, Abstracts with Programs*, 33, 175.
- [9] Nakamura, E., Makishima, A., Moriguti, T., Kobayashi, K., Tanaka, R., Kunihiro, T., Tsujimori, T., Sakaguchi, C., Kitagawa, H., Ota, T., Yachi, Y., Yada, T., Abe, M., Fujimura, A., Ueno, M., Mukai, T., Yoshikawa, M. and Kawaguchi, J. (2012). Space environment of an asteroid preserved on micrograins returned by the Hayabusa spacecraft. *PNAS Plus*, 109 (11), 4031-4032.
- [10] Yachi Y., Kitagawa H., Kunihiro T. and Nakamura E. (2013) Software dedicated for curation with geochemical datasets in an analytical laboratory. *Geostandards and Geoanalytical Research*, in-press, DOI: 10.1111/j.1751-908X.2013.00205.x

# ALMA キューブデータ簡易解析 Web アプリケーションの開発

江口智士\*

## Development of a Web Application for Quick Analysis of ALMA Cube Data

Satoshi Eguchi \*

### Abstract

Virtual Observatory (VO) is a collection of various astronomical databases of various wavelengths which are connected by "VO interface". Volume of astronomical observational data has been increasing exponentially, and the Atacama Large Millimeter/submillimeter Array (ALMA) is estimated to generate a 2 TB standard calibrated data cube even for one target. It is hard to handle such a huge data cube in a personal computer. We develop a VO system which dynamically converts a huge data cube into a small one with low resolution so that it meets users' scientific interests. The core component of the system is ALMAWebQL, which is a typical client-server application built on Google Web Toolkit (GWT). We developed a new endian conversion technique named "Just-in-Time endian conversion method" to reduce the reading time since the FITS format adopts big endian, and the endian conversion time is comparable to the (pure) reading time and the data extraction time on the server side. By applying this new method in native implementation to ALMAWebQL, it becomes about 3.5 (2.5 by switching GWT (= pure Java) to native implementation and 1.4 by application of multithreaded Just-in-Time endian conversion method) faster than the original GWT implementation.

Keywords: Astronomical databases: miscellaneous -- Virtual observatory tools -- Methods: data analysis

### 概要

仮想天文台 (VO) は、様々な波長の様々な天文データベースの集合体であり、それらは「VO インターフェース」という規格で結ばれている。天文の観測データは指数関数的に増加し続けており、Atacama Large Millimeter / submillimeter Array (ALMA) に至っては 1 天体につき 2TB もの標準較正済みデータが生成されると見積もられている。このようなサイズのデータは、もはや通常のパーソナルコンピュータでは処理することができない。そこで我々は、利用者の科学的関心を満足する小さなデータキューブを巨大なデータキューブから動的に生成する VO システムを構築した。この VO システムの核となるのが ALMAWebQL であり、これは Google Web Toolkit (GWT) を用いて書かれた典型的なクライアント-サーバ型アプリケーションである。また、我々はファイルの読み込みにかかる時間を短縮するために「Just-in-Time エンディアン変換法」と呼ばれる新しいエンディアン変換法を開発した。これは、FITS ファイルがビッグエンディアンを採用しており、サーバ側のデータ処理において純粋なファイル読み込みにかかる時間とエンディアンの変換にかかる時間、データを加工する時間が拮抗するためである。この方法をネイティブコードで実装して ALMAWebQL に適用したところ、GWT(=Java) 実装と比較して約 3.5 倍 (GWT のみの実装からネイティブコードへの修正で 2.5 倍、マルチスレッド化された Just-in-Time エンディアン変換の適用で 1.4 倍) 高速化された。

---

\* 国立天文台天文データセンター (Astronomy Data Center, National Astronomical Observatory of Japan)

## 1. はじめに

天文の観測データは現在、指数関数的に増加している。図1は、様々なサーベイデータアーカイブの較正済みデータの総容量の時系列変化と今後の見通しをプロットしたものである<sup>1</sup>。図中の赤線で示したように、データサイズは「1年半で2倍」という猛烈な勢いで増加の一途を辿っており、2020年代に予定されている Large Synoptic Survey Telescope (LSST) に至っては、一晩の観測で30TBもの生データが生成されると見積もられている<sup>2</sup>。いっぽうCPUの性能向上はここ10年ほど鈍化しており<sup>3</sup>、これまでの「観測装置から得られたデータをパーソナルコンピュータで解析する」というスタイルがまもなく破綻することは明白である。したがって、新しい解析手法の確立が急務である。

Atacama Large Millimeter / submillimeter Array (ALMA) は、日米欧の国際協力のもと南米チリに建設中の世界最大の電波望遠鏡である<sup>4</sup>。ALMAが生成するデータは非常に巨大であり、完成時には年間200TBもの生データが生成される。ここから研究者が解析に使用できる標準較正済みデータに処理したものでも、1天体につき2TB以上のサイズにもなる<sup>3</sup>。これを現在の世界のインターネットの平均速度2.6Mbpsで各自のローカルにダウンロードしようとする、約75日かかる計算になる。では、適当な媒体にコピーして空輸すれば万事巧いくかという、先述の通り計算機の性能向上は鈍化しており、2TBものファイルを個人のローカルの環境で処理することはほぼ不可能である。そこで国立天文台では、ALMAを来たるべき天文ビッグデータ時代のプロトタイプと位置づけ、我々 Japanese Virtual Observatory (JVO) グループとALMAグループが密接に連携して、VOの技術をベースにした新しいデータ解析手法の開発を精力的に行っている。

本論文では、Webブラウザ上でALMAのデータを簡易解析するためのアプリケーション「ALMAWebQL」の設計と、JVOポータルサイト<sup>5</sup>で運用する過程で判明した問題点とその改善方法について紹介する。

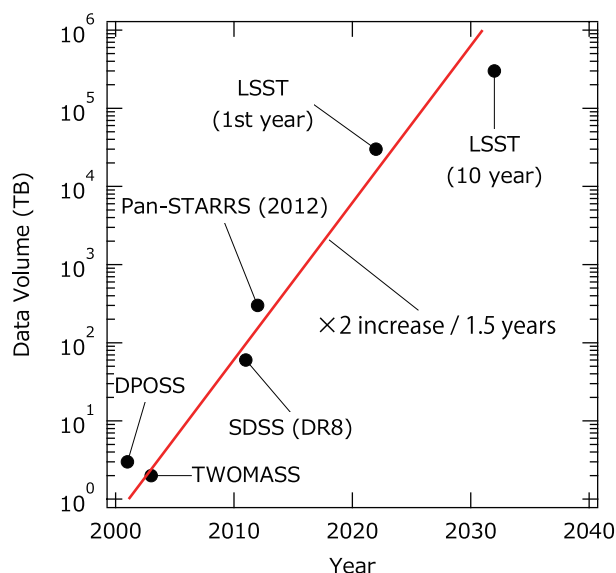


図1 様々なサーベイの較正済みデータの総容量の時系列変化および今後の見通し

## 2. JVO ポータルサイトの設計

### 2.1 データの最適化

天文データの場合、解析する研究者の専門次第で必要とする中身が異なる。例えば、ブラックホールの観測的研究を行っている研究者は、スペクトル情報は不要だが非常に限られた領域について超高空間分解能のデータを必要とするかもし

1) 1)2MASS: <http://www.ipac.caltech.edu/2mass/overview/about2mass.html>, 2)DPOSS: [http://www.ifa.hawaii.edu/~rgal/science/dposs/dposs\\_frames\\_scan.html](http://www.ifa.hawaii.edu/~rgal/science/dposs/dposs_frames_scan.html), 3)SDSS(DR8): [http://www.sdss3.org/dr8/data\\_access.php](http://www.sdss3.org/dr8/data_access.php), Pan-STARRS: Alex Szalay 氏の講演資料 <http://labs.yahoo.com/files/11SzalaySciAppofLargeDB.pdf> より推計, 4)LSST: [http://www.lsst.org/lsst/science/concept\\_data](http://www.lsst.org/lsst/science/concept_data) より推計

2) <http://www.lsst.org/lsst/science/development>

3) <http://preshing.com/20120208/a-look-back-at-single-threaded-cpu-performance>

4) <http://alma.mtk.nao.ac.jp/j/>

5) <http://jvo.nao.ac.jp/portal/>



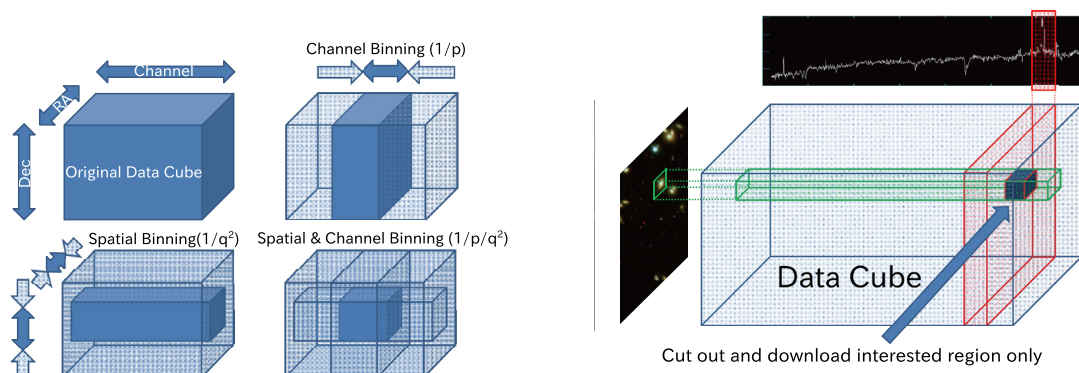


図2 ビニングによるデータ圧縮の原理（左）と切り出し（cut-out）によるデータサイズの削減（右）

れない。いっぽう、星間物質の研究を行っている研究者の場合、ある輝線について、空間分解能はほどほどある広い領域の情報を必要とするかもしれない。このように考えると、ALMA のデータでも全部をそっくりそのまま研究に使うケースは非常に希であることがわかる。つまり、

1. ALMA の 3 次元<sup>6</sup> データを、様々なビンングパラメータで予めビンングしておく（図 2 左）
2. Web ブラウザでビンングパラメータを選択できるようにする。
3. 利用者の興味に応じて、選択したビンングパラメータのデータから必要な部分のみ切り出せるようにする（cut-out, 図 2 右）

というシステムを構築すれば、科学的価値を損なうことなく ALMA のデータをパーソナルコンピュータで処理できるサイズまで小さくできる。

## 2.2 ポータルサイトの全体像および利用方法

以上を踏まえ、我々は JVO ポータルサイト <http://jvo.nao.ac.jp/portal/> を図 3 で示される構成にすることにした<sup>1)</sup>。ここでは ALMA のデータだけでなく、すばる望遠鏡の較正済みデータを検索・取得することができる。利用者はポータルサイトの「ALMA」のリンクから、必要なデータを観測 ID、天体名、座標等で検索する。検索結果はテーブル形式で一覧表示され、その中のうちどれかひとつのリンクをクリックすると、その観測データの詳細情報が記載されたページが表示される。このページには「WebQL」というボタンがあり、このボタンをクリックすると § 2.1 の処理をグラ

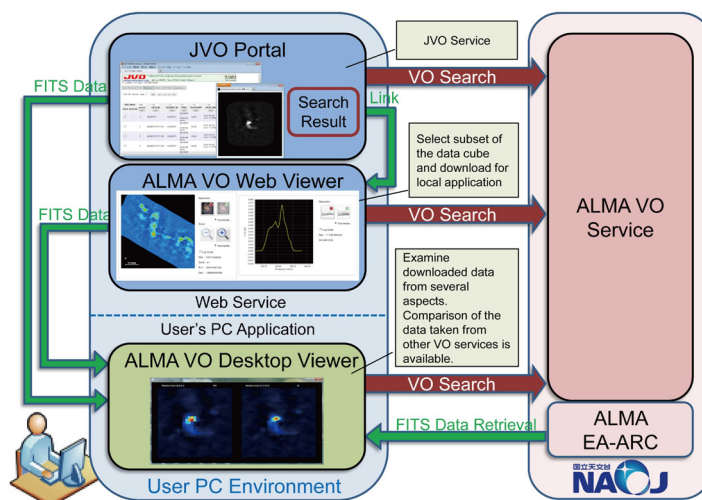


図3 JVO ポータルサイトの全体像

6 将来的には偏光の情報を含んだ 4 次元のデータになるが、現時点で公開されているデータは 3 次元データである。

フィカルに行うための Web アプリケーション「ALMAWebQL」が起動する。利用者は ALMAWebQL を利用して、空間分解能や周波数分解能、すなわちビニングパラメータを変更したり、空間イメージを拡大・縮小したり、空間イメージの中心位置や周波数範囲を変更したりすることができる (図 4)<sup>7</sup>。ALMAWebQL 上で満足 of the いく結果が得られたら、「Download」ボタンをクリックすることで現在表示中のデータを FITS 形式で利用者のローカル環境にダウンロードすることができる。

ALMAWebQL からダウンロードしたデータは、利用者のローカル環境で動作する「Vissage」<sup>8</sup>を使用することにより、より詳細な解析を行うことができる。Vissage は ALMA のキューブデータを詳細に閲覧するためのソフトウェアであり、モーメント図、チャンネルマップおよび P-V 図を作成する機能を搭載している。Vissage は Java アプリケーションであるため、OS を問わず使用できる。Vissage には上記機能に加えて JVO ポータルサイトと連動する機能が実装されており、ALMA のイメージ上に JVO ポータルサイトで公開しているすばる望遠鏡のイメージをオーバーレイ表示することや、Vissage で詳細解析中にデータの不足を感じた場合に、Web ブラウザを起動してそのデータを切り出したパラメータで ALMAWebQL を直接呼び出したり、あるいは違うビニングパラメータのデータを直接ダウンロードしたりすることができる。したがって、ALMAWebQL と Vissage を行き来することにより、最終的に利用者は研究に必要な十分なデータを短時間で得ることができる。

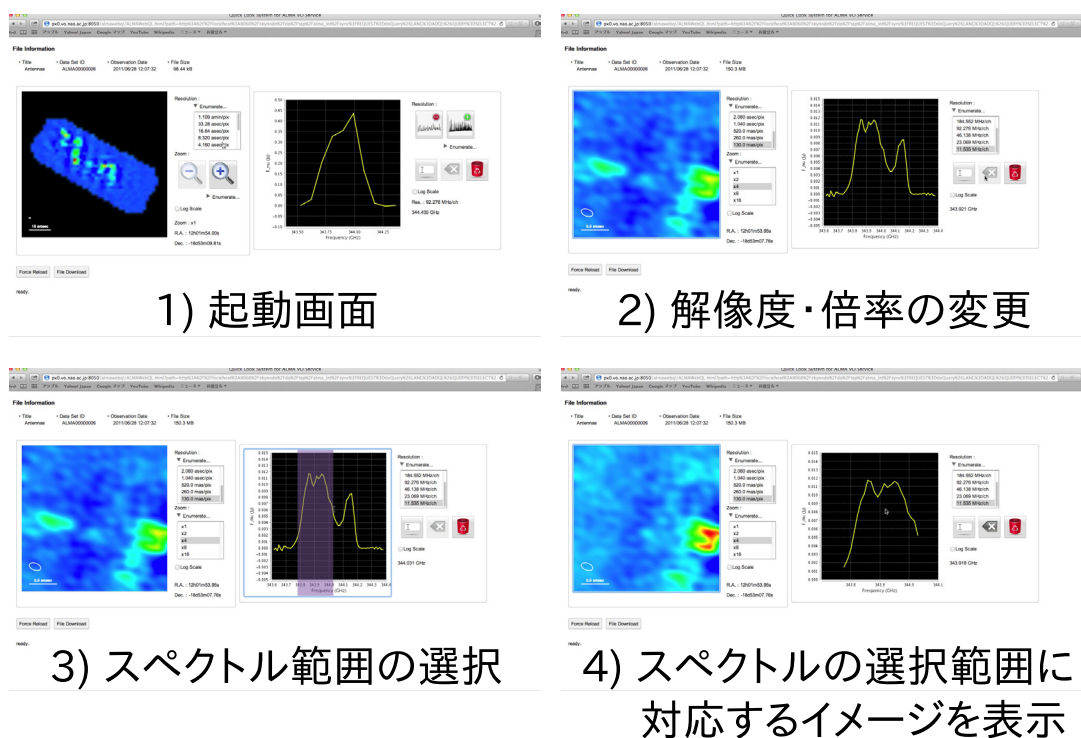


図 4 ALMAWebQL のスクリーンショット

### 3. ALMAWebQL の設計

以上のように、ALMA のデータを JVO ポータルサイトで配信する上で ALMAWebQL は極めて重要な役割を担っている。ALMAWebQL の開発目的が「ALMA の巨大データを利用者の研究目的に必要なサイズまで小さくすること」であるので、すべての処理を Web ブラウザだけで行うのは本末転倒であり、ALMAWebQL の構造が JVO ポータルサイトを運営するサーバ上で動くコンポーネントと Web ブラウザ上で動くコンポーネントに分離されるのは自明である。このようなクライアント-サーバ型アプリケーションを実装する場合、両者で異なる開発言語を使用するのは保

<sup>7</sup> 現時点では、ALMAWebQL が表示できるデータは JVO ポータルサイトで公開されている ALMA のデータのみである。将来的には JVO スペース (JVO ユーザ用のオンラインディスクスペース) にアップロードされたデータも ALMAWebQL で閲覧可能になる予定である。

<sup>8</sup> <http://jvo.nao.ac.jp/download/Vissage/>

守性という観点から好ましくない．というのは，クライアント側とサーバ側で全く同じ機能を実装するケースを考えたとき，片方のバグ修正をもう片方にも手作業で反映しなければならないからである．

クライアント・サーバ型アプリケーションを同じ言語で取り扱えるフレームワークの代表的なものに，Microsoft が提供している「ASP.NET」と Google が提供している「Google Web Toolkit (GWT)」の2つがある．前者は同社が提供する .NET Framework をベースとしており，開発言語の選択肢が複数ある．しかし，すべての処理をサーバ側で行うという欠点がある．対して後者は，クライアント側・サーバ側ともに Java 言語で記述し，コンパイル段階でコンパイラがクライアント用コードを JavaScript に変換するアプローチを取っている．ALMAWebQL では，Web ブラウザ側で表示しているイメージやスペクトルに対して物理的な値を対応させる，すなわち，マウスカーソルのある位置のピクセルが天球座標のどこに対応するのか，あるいはスペクトルのどの周波数に対応するのかをリアルタイムに計算する必要がある．仮に ASP.NET を採用した場合，ユーザがマウスを動かす度に画面上の座標と物理的値とを対応させる計算処理のために，サーバとの通信が発生する．しかし GWT を採用した場合には，画面上の座標と物理的値との対応を Java コードから生成された JavaScript コードで処理させられるため，サーバとの通信を必要最低限に抑えることができる．そのため，ALMAWebQL では GWT を採用することにした．

図 5 は，ALMAWebQL の設計を概念的に図示したものである．

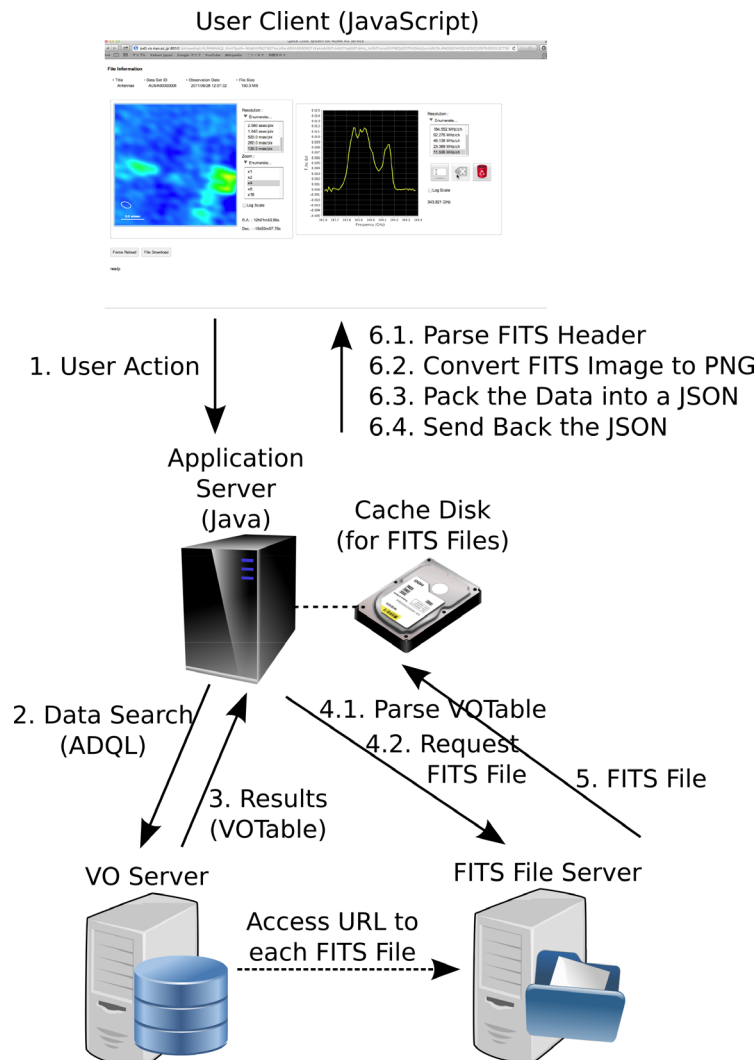


図 5 ALMAWebQL の設計概念図

1. ユーザの操作により FITS ファイルへのアクセスが必要になると，Web クライアントは Remote ProcedureCall(RPC) という仕組みを使ってサーバ側の FITS 処理メソッドを呼び出す．
2. このメソッドが呼び出されると，アプリケーションサーバは ALMA 用 VO サーバに対して ADQL と呼ばれる VO

用検索言語を用いて該当の FITS ファイルの検索を行う。

3. 検索結果は VOTable という XML 形式のテーブルで返されるので、
4. アプリケーションサーバは
  - 4.1. VOTable をパースして、FITS ファイルが置いてある場所（サーバ）の URL を取り出す。
  - 4.2. 続いてこの URL に対して HTTPGET クエリを送信して、FITS ファイルが送信されるのを待ち受ける。
5. FITS ファイルがファイルサーバから送信開始されると、アプリケーションサーバは送られてくるデータをローカルのキャッシュ用ディスクに書き込む。
6. FITS ファイルの転送が終了すると、アプリケーションサーバは
  - 6.1. FITS ファイルのヘッダ情報を解析すると同時に、
  - 6.2. イメージ HDU の中からユーザが指定した部分のみ取り出し、周波数方向あるいは空間方向にデータを積分して 2 次元イメージあるいはスペクトルに変換する<sup>9</sup>。できた 2 次元イメージおよびスペクトルを PNG 形式<sup>10</sup>で圧縮した上で Base64 でエンコードして<sup>11</sup>、
  - 6.3. 両者を JavaScriptObjectNotation(JSON) 形式に変換する。
  - 6.4. そして、JSON 化した FITS ヘッダ、イメージおよびスペクトルを Web クライアントに送信する。
7. Web クライアントはアプリケーションサーバから送られてきたデータを元に画面を更新して、次のユーザ操作を待つ。

表 1 各種記憶装置の典型的な読み書き速度

Hardware	Read / Write Performance
Hard Disk Drive(HDD)	~ 50 MB/s
Solid State Disk(SSD)	~ 250 MB/s
Random Access Memory(RAM)	~ 10 GB/s

#### 4. ファイル I/O の最適化

ALMA のデータキューブを周波数方向あるいは空間方向に積分して 2 次元イメージあるいはスペクトルを生成する処理に要する時間は、データサイズに比例する。この処理を最近の 3GHz 前後の CPU で行うと、シングルスレッドで ~ 8GB/s の処理速度となる。表 1 は、ハードディスク、SolidStateDisk(SSD)、メインメモリの典型的な読み書き速度を表したものである。したがって、SSD を用いた RAID アレイ<sup>12</sup>や RAM ディスクを使用するケース、加えてオペレーティングシステム (OS) によりファイルがメモリ上にキャッシュされているケースでは、データキューブの処理速度とファイルの読み込み速度が拮抗する。特に ALMAWebQL では、FITS ファイルサーバからファイルをダウンロードしてアプリケーションサーバのローカルディスクに書き込む過程で、その内容が OS によりメモリ上にキャッシュされるため、ファイル I/O 処理を工夫することにより、ALMAWebQL のレスポンスを向上させられると考えられる。そこで我々は、エンディアン変換に注目した。

エンディアンとは、コンピュータが数値データをメモリに格納するときの順番を表す言葉で、メモリの内容を 16 進ダンプしたときに大きい位から小さい位という、我々人間が通常使っている順番で数値が表示される方式をビッグエンディアン、その反対をリトルエンディアンと言う。現在稼働しているサーバやパーソナルコンピュータではほとんどがリトルエンディアンを採用している一方、天文の世界で広く使われている FITS 形式はビッグエンディアンを採用している。したがって、FITS データを処理する過程で必ずエンディアンの変換を行う必要が出てくる。我々はこの変換過程を最適化できないかと考えて試行錯誤した結果、「Just-in-Time エンディアン変換法」という新しいアルゴリズムを編み出した<sup>2)</sup>。

9 ALMA の FITS ファイルの構造については図 2 を参照。

10 JPEG ではなく PNG フォーマットを採用したのは、JPEG 形式は離散コサイン変換を施した後にバンドパスフィルタを通すため、いわゆる「モアレ」が発生する可能性があり、データ圧縮による偽の空間構造と本物の空間構造との区別が困難になるからである。PNG フォーマットは JPEG よりも圧縮率は悪いものの可逆圧縮アルゴリズムを採用しており、このような問題は発生しない。

11 こうすることで、アプリケーションサーバ側に外部公開用ディレクトリを作成して、そこに一時的な PNG ファイルを置くことを避けられるため、セキュリティ的に安全性が確保される。

12 市販（民生品）の SSD を 16 個で RAID0 のシステムを構成して限界までチューニングを施すと、シーケンシャルアクセスで ~ 4GB/s の速度が出る（<http://www.fumi.org/pc/ultra40/index.html>）

ここで便宜上、FITS ファイルを取り扱うライブラリ、例えば CFITSIO<sup>4)</sup> や SFITSIO<sup>13</sup> 等で使われている方法を「On Ahead エンディアン変換法」と呼ぶことにする。

double 型のデータのエンディアン変換を行う関数を

```
double convert_endian(double x);
```

とすると、FITS ファイルの全要素を足し合わせるコードは次のように書ける：

•On Ahead エンディアン変換法

```
{
double *v;      // ファイルの中身を保持する配列
size_t len;     // 配列の要素数
size_t i;
double sum = 0.0;

// ファイルの読み込み（省略）

// エンディアン変換（注：FITS ライブラリ内部）
for (i=0; i<len; ++i){
    v[i] = convert_endian(v[i]);
}

// 総和を求める
for (i=0; i<len; ++i){
    sum += v[i];
}
}
```

•Just-in-Time エンディアン変換法

```
{
double *v;      // ファイルの中身を保持する配列
size_t len;     // 配列の要素数
size_t i;
double sum = 0.0;

// ファイルの読み込み（省略）

// 総和を求める
for (i=0; i<len; ++i){
    sum += convert_endian(v[i]);    // エンディアン変換
}
}
```

3.4GB の 2 次元 FITS イメージを使って、上記の「総和を求める」というコメントのあるループの処理時間と、ファイルの読み込み部分も含めて全部の処理を完了するのに要した時間を測定した（図 6）。総和を求めるループのみを比較すると、Just-in-Time エンディアン変換法ではエンディアン変換を行う処理が加わっているため、On Ahead エンディアン変換法よりも有意に遅くなりそうなものであるが、実際は予想に反して両者に違いは認められない。対して全体の処理時

13 <http://www.ir.isas.jaxa.jp/~cyamauch/sli/index.ni.html>



間を比較すると、On Ahead エンディアン変換法のみエンディアン変換のループが存在するため、このループ分だけ Just-in-Time エンディアン変換法よりも遅くなる。結果的に、Just-in-Time エンディアン変換法を用いることで、シングルスレッドで約 20%、マルチスレッド化した場合には約 40% もパフォーマンスが向上することが判明した。なお、この比較はネイティブコードで行った。

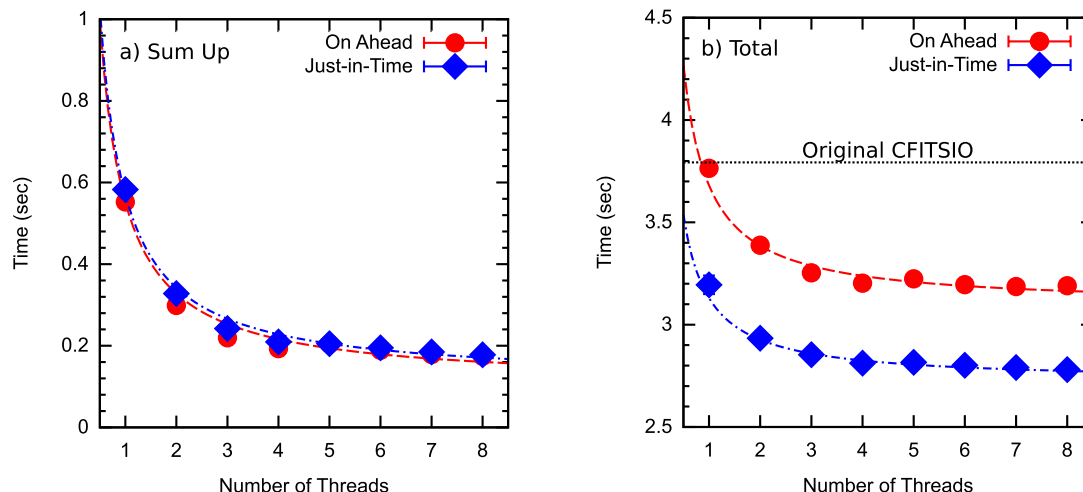


図 6 3.4GB の 2 次元イメージ FITS ファイルの全要素を足し上げるのに要した時間のうち、総和を求めるループ部分のみを終了するのに要した時間（左）と、ファイルの読み込み時間も含めた処理時間の合計（右）

いっぽう、Java はビッグエンディアンを採用しており、ソースコード上はエンディアン変換を行う必要はないため、仕組みを正しく理解していないとネイティブコードよりも高速に FITS ファイルを処理できそうに思えてしまう。そこで、Java と `nom.tam.fits`<sup>14</sup> ライブラリを用いて上と同じ FITS ファイルの全要素を足し上げるのに要する時間を測定したところ、ネイティブコードより約 2.5 倍遅いことがわかった。これは `double` 型のデータを CPU のレジスタ (Java の場合は「スタック」) に読み込むのに、ネイティブコードでは 1 オペレーションコードの実行で済むところを、Java の仮想マシンでは 4 つのオペレーションコードを実行する必要があるからである<sup>15</sup>。

続いて、Just-in-Time エンディアン変換法を ALMAWebQL に適用して、(シングルスレッドで) ファイルサイズに対してどのように振る舞うかを調べた (図 7)<sup>16</sup>。イメージの生成およびスペクトルの生成ともに、100MB 付近で On Ahead エンディアン変換法よりも Just-in-Time エンディアン変換法が優勢になり、200MB 付近で予測通り約 20% 高速であることがわかる。したがって、Java から Just-in-Time 変換法を用いた並列化されたネイティブコードに切り替えることで、200MB 以上のファイルに対して

- Java からネイティブコード (On Ahead エンディアン変換法) に切り替えることで約 2.5 倍
- ネイティブコード (On Ahead エンディアン変換法) に対して並列化された Just-in-Time エンディアン変換を適用することで約 1.4 倍

の合計約 3.5 倍の高速化が達成できた計算になる。

<sup>14</sup> NASA/GoddardSpaceFlightCenter で開発された Java 用 FITS ライブラリ

<sup>15</sup> <http://docs.oracle.com/javase/specs/>

<sup>16</sup> GWT では、アプリケーションサーバで動くコードは純粋な Java サーブレットであるので、JavaNativeInterface(JNI) を介することで、ネイティブコードで書かれた Just-in-Time エンディアン変換法のコードを使用することができる。

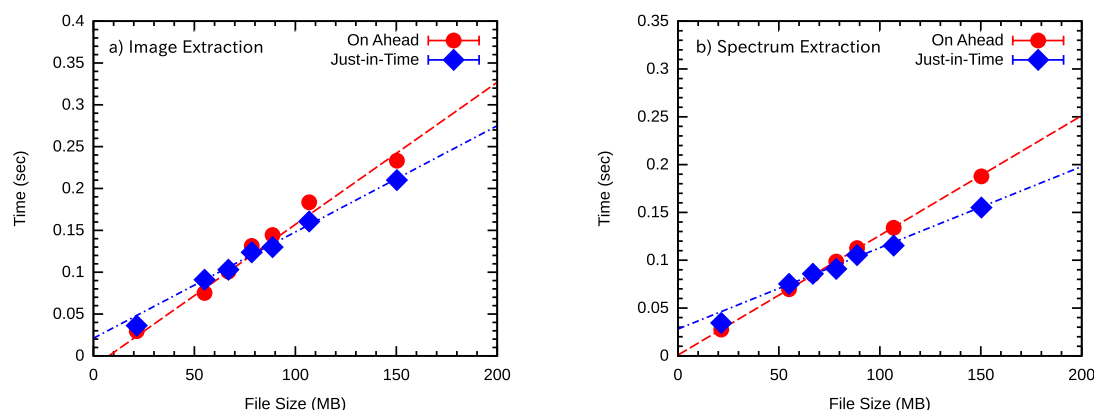


図 7 Just-in-Time エンディアン変換法を適用した ALMAWebQL の、2 次元イメージの生成時間（左）とスペクトルの生成時間（右）のファイルサイズ依存性

## 5. まとめ

現在、天文データのサイズは指数関数的に増加を続けており、現在の「すべての解析作業を個人の計算機環境で完結させる」スタイルは、非常に高い確率で近い将来破綻する。これは昨年観測が始まった ALMA 望遠鏡のデータで現実のものとなりつつあり、1 天体につき 2TB を超える ALMA のデータを解析する工夫が必要である。そこで我々は、巨大データをビンニングで事前に小さくしておき、それを Web ブラウザ上で動くアプリケーション「ALMAWebQL」で簡易解析して利用者の目的に合致するデータを動的に生成する VO システムを構築した。ALMAWebQL は GWT を使って書かれた Java プログラムである。

ALMAWebQL では、データキューブを処理する時間とファイルを読み込む時間が拮抗するため、ファイル読み込みを高速化することでアプリケーションのレスポンスを向上させるアプローチを採用することにした。FITS ファイルを読み込む際にはエンディアン変換を行う必要があり、ここを新開発の「Just-in-Time エンディアン変換法」を用いたネイティブコードに置き換え、マルチスレッド化することにより、GWT だけ (=Java) で記述したときよりも約 3.5 倍ものパフォーマンス向上が得られた。

今後も ALMAWebQL の開発は精力的に行う予定であり、新しい機能ができ次第順次公開する予定である。是非一度 JVO のポータルサイト <http://jvo.nao.ac.jp/portal/> を訪れて実際に ALMAWebQL を使って頂き、機能改善・要望等を送って頂けると幸いである。

## 参考文献

- 1) Eguchi, S., Kawasaki, W., Shirasaki, Y., Komiya, Y., Kosugi, G., Ohishi, M., Mizumoto, Y., Pro-totype Implementation of Web and Desktop Applications for ALMA Science Verification Data and the Lessons Learned, arXiv:1211.3790 (2012 November)
- 2) Eguchi, S., "Superluminal" FITS File Processing on Multiprocessors: Zero Time Endian Conversion Technique, Publications of Astronomical Society of the Pacific, Vol. 125, No. 927 (2013 May), pp.565-579
- 3) Lucas, R., Richer, J., Shepherd, D., Testi, L., Wright, M., & Wilson, C., Estimation of ALMA Data Rate, ALMA Memo #501 (2004 June)
- 4) Pence, W. D., CFITSIO: A FITS File Subroutine Library, Astrophysics Source Code Library, record ascl:1010.001 (2010 October)

# 超高層物理学分野のメタデータ・データベースへの連想検索の適用

小山 幸伸<sup>\*1</sup> 阿部 修司<sup>\*2</sup> 八木 学<sup>\*3</sup> 梅村 宜生<sup>\*4</sup> 堀 智昭<sup>\*4</sup>  
新堀 淳樹<sup>\*5</sup> 佐藤 由佳<sup>\*6</sup> 家森 俊彦<sup>\*1</sup> 田中 良昌<sup>\*6</sup> 橋口 典子<sup>\*5</sup>  
上野 悟<sup>\*7</sup> 谷田貝 亜紀代<sup>\*5</sup>

## Application of associative search to the metadata database of the upper atmosphere

Yukinobu KOYAMA<sup>\*1</sup>, Shuji ABE<sup>\*2</sup>, Manabu YAGI<sup>\*3</sup>, Norio UMEMURA<sup>\*4</sup>, Tomoaki HORI<sup>\*4</sup>  
Atsuki SHINBORI<sup>\*5</sup>, Yuka SATO<sup>\*6</sup>, Toshihiko IYEMORI<sup>\*1</sup>, Yoshimasa TANAKA<sup>\*6</sup>, Noriko HASHIGUCHI<sup>\*5</sup>,  
Satoru UeNo<sup>\*7</sup>, and Akiyo YATAGAI<sup>\*5</sup>

### Abstract

In order to understand the mechanism of global-scale phenomena in the upper atmosphere, multidisciplinary researches using many kinds of data are important. An infrastructure to access to many kinds of data on the Internet is one of the keys to the multidisciplinary researches. The Inter-university Upper atmosphere Global Observation NETwork (IUGONET) project solved this problem by developing a metadata database to provide information such as URL of database or each data file, which are managed dispersively by several institutes. Because the metadata database covers a wide scientific field, it is likely that non-specialized users can not easily select a right search phrase. We connected an associative search engine which is called GETAssoc with the metadata database, in order to get the related words and to generate re-query phrases by using them automatically. We collected metadata from SAO/NASA Astrophysics Data System to create the dictionary for associative search. We show that the prepared dictionary gives the related terms as expected for generating re-query phrases.

keywords Database, Metadata, Associative Search, Information Retrieval, Interdisciplinary Study, Upper Atmosphere

### 概要

超高層大気に於ける地球規模の物理現象の機構を解明する為には、様々なデータを用いた分野横断的な研究推進が重要である。インターネット上の様々なデータを直ちに取得する為の基盤が、分野横断的な研究の1つの鍵となる。IUGONET プロジェクトは、複数の研究機関によって分散管理されている、様々なデータの URL 等の所在情報等を提供可能なメタデータ・データベースを構築することによって、この課題に対処した。しかしながら、このメタデータ・データベースは広範な研究領域を対象としている為、専門分野外のユーザーにとって単語検索時における適切な検索語句の選択が容易でないことが指摘された。これを補助するべく、検索語句の関連語の取得とそれらを用いた再検索クエリ文字列の自動生成を行う為に、そのメタデータ・データベースと連想検索エンジン GETAssoc を連携させた。そして、この連想検索用の辞書を作成する為に、SAO/NASA Astrophysics Data System からメタデータを抽出した。再クエリ文字列を生成する為に用意した辞書を用いることで、期待通りの連想検索結果を得ることが出来た。

キーワード データベース, メタデータ, 連想検索, 情報検索, 学際的研究, 超高層大気

---

\*1 京都大学大学院理学研究科附属地磁気世界資料解析センター

(Data Analysis Center for Geomagnetism and Space Magnetism, Graduate School of Science, Kyoto University)

\*2 九州大学国際宇宙天気科学・教育センター (International Center for Space Weather Science and Education, Kyushu University)

\*3 東北大学大学院理学研究科惑星プラズマ・大気研究センター

(The Planetary Plasma and Atmospheric Research Center, Graduate School of Science, Tohoku University)

\*4 名古屋大学太陽地球環境研究所 (Solar-Terrestrial Environment Laboratory, Nagoya University)

\*5 京都大学生存圏研究所 (Research Institute for Sustainable Humanosphere, Kyoto University)

\*6 国立極地研究所 (National Institute of Polar Research)

\*7 京都大学大学院理学研究科附属天文台 (Kwasan and Hida Observatories, Graduate School of Science, Kyoto University)

## 1. はじめに

高度数十 km より上空の大気は「超高層大気」と呼ばれており、中性大気と電離大気が混じり合う「熱圏・電離圏」領域から、太陽風と地球磁場の相互作用によって生じ、無衝突プラズマが物理を支配する「磁気圏」領域を含む。図 1 に示す様に、超高層大気は、下層大気からの波動や対流による運動量・エネルギーの流入に加え、太陽放射や太陽風の影響、そして化学反応・光化学反応といった物理過程が複雑に絡み合う領域であることが知られている。

従って、超高層物理学の研究においては、全球規模の観測データを用いた多角的なデータ解析が必要であり、磁力計、レーダー、そして太陽望遠鏡など、様々な観測器を用いた地上観測が継続的に行われてきた。こうした観測データは、観測を行った各研究機関ごとにデータベース化ならびに公開されてきたが、分散管理されたこれらの観測データを横断的に検索するシステムは過去に存在しなかった為、多くの種類のデータを必要とする地球規模の学際的研究推進に多大な労力を要した。この様な背景から、大学間連携プロジェクト「超高層大気長期変動の全球地上ネットワーク観測・研究 (IUGONET: Inter-university Upper atmosphere Global Observation NETwork)」においては、観測データの流通を促進する為に、IUGONET 参加各機関において分散管理されている観測データに関するメタデータ・データベースを構築した<sup>1)2)3)4)5)6)7)</sup>。

### ▶ 複雑に絡み合う超高層大気の特徴

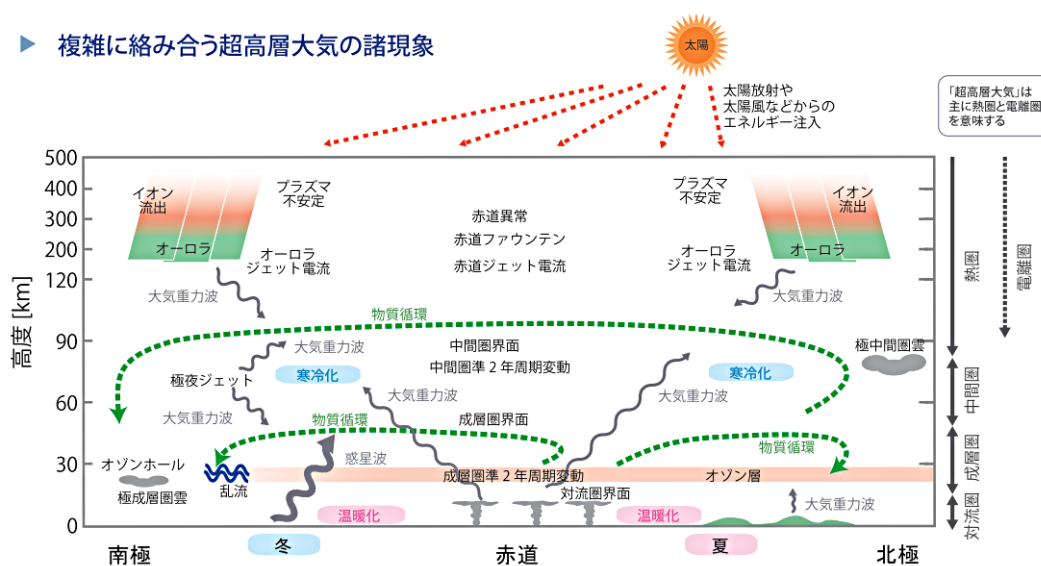


図 1 複雑に絡み合う超高層大気の特徴

図 2 に示した Web ベースの IUGONET メタデータ・データベース<sup>8)</sup>は、時刻検索、領域検索、そして単語検索が可能である。IUGONET メタデータ・データベースによって、各機関が管理する観測データに関するメタデータを横断的に検索出来る仕組みを構築したものの、超高層物理学を中心としつつ、隣接する天文学や気象学に渡る、広範な研究領域を登録メタデータの対象にしているが故に、専門分野外のユーザーにとって単語検索時における適切な検索語句の選択が容易でないことが指摘された。そこで、連想検索エンジン GETAssoc<sup>9)</sup>を導入することによって、ユーザーが入力した検索語句の関連語を複数個提示し、そのリンクをクリックすることにより、再検索を行う機能を実装し、メタデータ・データベースに組み込むことを検討した。

## 2. IUGONET メタデータ・データベース

超高層物理学分野の上位概念である太陽地球系物理学分野で使用されており、主に衛星データ向けに開発された Space Physics Archive Search and Extract (SPASE) メタデータ・フォーマット<sup>10)11)</sup>をベースとして、地上観測データに関する要素を拡張したメタデータ・フォーマットが、IUGONET 共通メタデータ・フォーマット<sup>12)</sup>である。

このメタデータ・フォーマットに則って作成された XML 形式のメタデータは、機関リポジトリ<sup>13)</sup>等で利用される DSpace<sup>14)</sup>を改変した IUGONET メタデータ・データベースにインポートされる。標準で Dublin Core メタデータ・フォーマット<sup>15)</sup>を取り扱う DSpace からの主要なカスタマイズ箇所は、1. 観測日時や緯度経度の範囲を取り扱う点、2. 機関リ

ポジトリにおいては PDF ファイル等のデジタルコンテンツが保持される Bitstream<sup>16)</sup> に、XML 形式のメタデータを保持する点、である。観測データの所在情報がメタデータ内に含まれているため、分散管理されている観測データに到達可能である。2013 年 4 月時点において、データセットのメタデータを 900 件程度を含む、合計 800 万件を超えるメタデータが IUGONET メタデータ・データベースに登録されている。

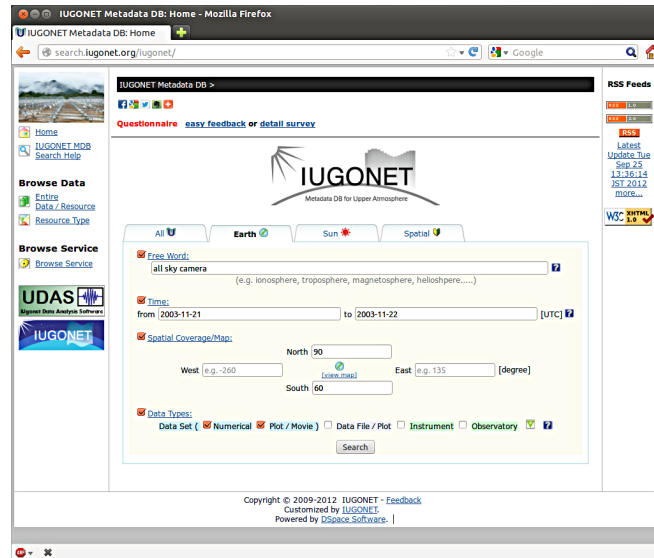


図 2 IUGONET メタデータ・データベース。機関リポジトリ等で利用される DSpace を改変した。時刻検索，領域検索，そして単語検索が可能である。

### 3. IUGONET メタデータ・データベースと連想検索エンジン GETAssoc の連携

図 3 に IUGONET メタデータ・データベースと連想検索エンジン GETAssoc の連携システム図を示す。

GETAssoc は、国立情報学研究所連想情報学研究開発センターで開発された連想検索エンジンである。IUGONET メタデータ・データベースと GETAssoc は、HTTP プロトコルを用いて接続される。GETAssoc のシステムの構築には、GETAssoc 以外に外部ツールがいくつか必要であるが、それらの外部ツールは、相互にバージョン依存している為、インストール手順が複雑であった。そこで、外部ツールを含めた GETAssoc インストール手順を自動化する為に、Ant のビルドファイルである build.xml を作成した。このビルドファイルは、Scientific Linux 6.4 (32bit) 上で動作確認済みであり、GitHub 上の IUGONETAssociativeSearch リポジトリ<sup>17)</sup>において公開している。なお、GETAssoc は標準で 32bit OS 用に最適化された設定であるので、32bit OS 上で動作させることとした。2013 年 5 月 31 日現在、IUGONET メタデータ・デー

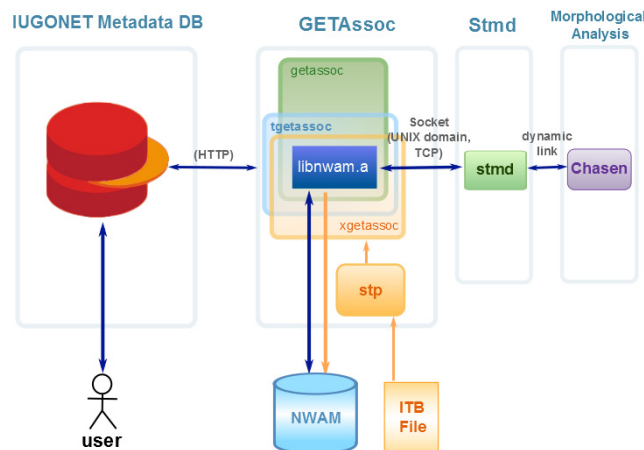


図 3 IUGONET メタデータ・データベースと連想検索エンジン GETAssoc の連携システム図。形態素解析システムには CHASEN を用いた。



データベースは名古屋大学において<sup>8)</sup>、また、GETAssoc システムは東北大学において運用している<sup>18)</sup>。IUGONET プロジェクトは大学間連携事業であるが、特定の研究機関にサーバー管理の負荷が集中しない様に、両システムを別々の機関で管理している。IUGONET メタデータ・データベースから、GETAssoc システムの利用時にはタイムアウトを設けている為、GETAssoc サーバーの不調やネットワーク負荷が高い場合には、連想検索を行わずに直ちに通常の検索を行う。

図4に IUGONET メタデータ・データベースと GETAssoc の連携システムにおける検索シーケンス図を示す。検索シーケンスは、1. ユーザーが入力した検索語句を読み込み、2. 通常のメタデータ検索を行い、3. ユーザーが入力した検索語句を元に、メタデータの該当有無に関わらず GETAssoc を用いて連想検索を行い、4. メタデータならびにユーザーが入力した検索語の関連語を表示する、5. ユーザーは、必要に応じて関連語を用いて再検索を行う、という流れを想定している。ブラウザ上で表示された関連語は、HTML で記述されたクエリ文字列へのリンクになっている為、関連語をブラウザ上でクリックすると、直ちに再検索される。

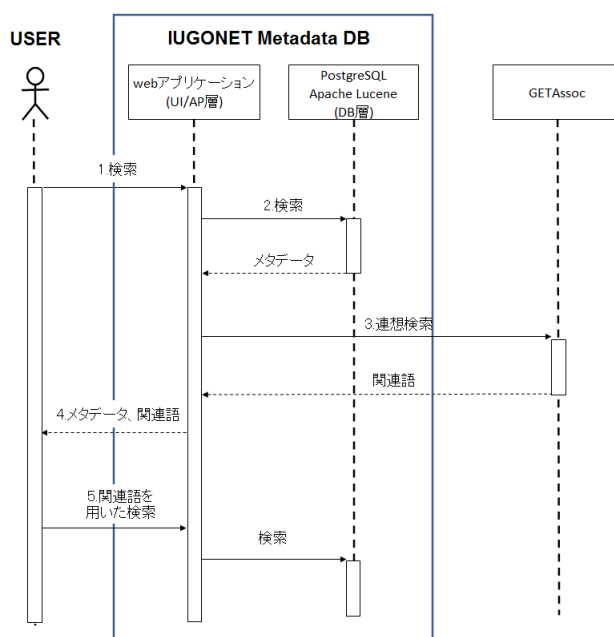


図4 IUGONET メタデータ・データベースと GETAssoc の連携システムにおける検索シーケンス図。

#### 4. ADS を用いた辞書の作成

GETAssoc では、図5に例示した ITB ファイル形式によって関連語を結びつける為の辞書を記述し、図3に示した通り、インデックス作成コマンド `stp` を用いて、連想検索に必要な NWAM インデックス・ファイルに変換する必要がある。つまり、超高層物理学分野のメタデータ・データベースにとって有益な連想検索を行う為には、如何にして効果的な ITB ファイルを作成するかという問題に帰着したことになる。我々は、1. Wikipedia のノートページと利用者ページを除くデータ<sup>19)</sup>を用いた辞書作成、2. Wikipedia の要約を用いた辞書作成、3. Wikipedia を基にした人力による辞書作成、の3つの辞書作成の試みを過去に行った<sup>20)</sup>。1. はその元データが 40GB にも及び取り扱いが困難だったことから、2. は連想検索対象のドメインを超高層物理学分野に絞込む為の要素が無かったから、それぞれ断念し3. に当時は帰着したが、依然として辞書の自動生成の要求がある。1. と関連した他研究者による取り組みとして、NASA の Semantic Web for Earth and Environmental Terminology (SWEET) オントロジーを用いて、超高層物理学や太陽地球系物理学の範疇に留まらない、多様な事象を包含する Wikipedia<sup>19)</sup> から、地球環境に関連した用語を抽出して利用する取り組み<sup>21)22)</sup>もある。情報検索に加えて新たにオントロジーを導入することは、我々にとって高コストであったので、この手法を踏襲せずに、情報検索の技術内で収めることとした。以上の議論を踏まえ、過去に行った多様な事象を包含する Wikipedia の利用から方針転換し、対象範囲が天文から物理の範囲に限定された文献メタデータ・データベースである SAO/NASA Astrophysics Data System (ADS)<sup>23)</sup>を用いて辞書を作成することにした。

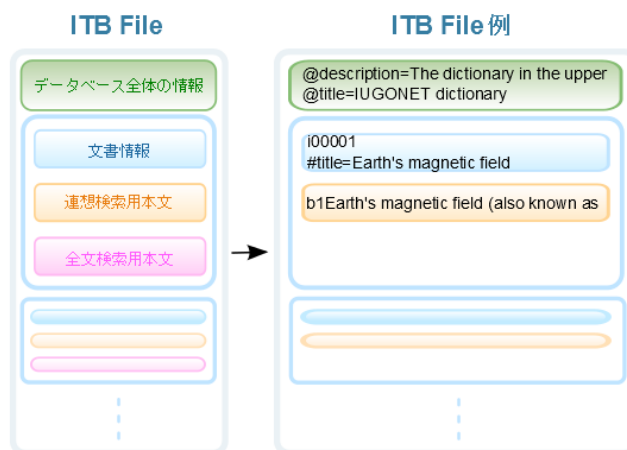


図5 ITB ファイルの概要と例。連想検索のみに着目している為、本件では全文検索用本文は作成しない。

#### 4.1 ADS 文献メタデータの収集

ADS 文献メタデータ・アーカイブは一括ダウンロード出来ず、外部からメタデータ要素を抽出する為の Web API も整備されていない。この為、ユーザーが通常のブラウザを用いてアクセスするのと同様の手順でメタデータを収集しなければならない。全ての ADS 文献メタデータを取得するのは膨大な工数がかかる上、その中に不要なメタデータもあるので非効率である。そこで、IUGONET プロジェクトへ参加している、約 50 名の研究者らが書いた論文や学会予稿の ADS 文献メタデータのみを採取し、それらを元に連想検索用の辞書を作成することを試みた。

#### 4.2 ADS 文献メタデータの収集事例

まず初めに、本論文の共著者である家森 俊彦 (Iyemori, T.) に関連した ADS 文献メタデータ検索を例示する。名字の “Iyemori” で検索した結果、119 件が該当し、“Iyemori, T.” で検索した結果、115 件が該当した。これら全てのメタデータが、本人に関連した著作のメタデータであることを確認した。次に、本論文の共著者である佐藤 由佳 (Sato, Y.) に関連した ADS 文献メタデータ検索を例示する。名字の “Sato” で検索した結果、26861 件が該当し、“Sato, Y.” で検索した結果、2094 件が該当した。http://adsabs.harvard.edu/abs/2008EP%26S...60..207S を含む計 6 件が本人の論文のメタデータであることが確認された。ADS は、ユニークな著者 ID によって管理されていないため、名前による文字列検索を行うことになり、名寄せ問題に直面する<sup>24)</sup>。次に、本論文の著者である小山 幸伸 (KOYAMA, Y.) に関連した ADS 文献メタデータ検索を例示する。名字の “Koyama” で検索した結果、4507 件が該当し、“Koyama, Y.” で検索した結果、514 件が該当した。前出と同様に名寄せ問題に直面した。本著者による ADS 文献メタデータの例として、http://adsabs.harvard.edu/abs/2005SSCom.133..785K が挙げられるが、これは、本著者による光物性理論に関する論文のメタデータであり、超高層物理学とは無関係である。ADS では、検索対象のデータベースを、“Astronomy”、“Physics”、“arXiv e-prints” の 3 つのデータベースから絞り込むことが可能であるので、これ以降は “Astronomy” に限定した。

名前による文字列検索を用いた文献メタデータの絞り込みでは、上記の 3 例からうまくいかないことが分かった。そこで、どの文献メタデータを収集するかは手動で行うこととした。具体的には、下記の手順で文献メタデータを収集した。

1. IUGONET プロジェクトへ参加している、約 50 名の研究者の人名 (例, “Iyemori, T.”) を検索キーとして ADS 上で検索を行い、
2. 該当した文献メタデータの URL を、内容を吟味した上で人力で抽出し、
3. 2 を Google スプレッドシート<sup>25)</sup> 上でデータベース化し、
4. 3 に公開属性を付与した上で、google-spreadsheet-ruby ライブラリを用いた Ruby プログラムを用いて、インターネット越しにその Google スプレッドシート内のシートにアクセスして、文献メタデータの URL リストを抽出し、
5. 当該 URL にアクセスして、メタデータを抽出した。

#### 4.3 ADS 文献メタデータから ITB ファイルの生成

得られた ADS 文献メタデータの一例を図 6 に示す。XSLT を用いて HTML から要素を取り出して ITB へ変換しよう

と試みたが、この HTML は妥当でなかった為、xsltproc で処理することが出来なかった。その為、Ruby を用いて、この HTML を通常のテキストデータとして扱い、パターンマッチングによって “citation\_title” と “Abstract” を切り出して、ITB ファイルを生成した。その後、インデックス作成コマンド stp によって、連想検索可能な状態にした。

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.0 Transitional//EN"
"http://www.w3.org/TR/REC-html40/loose.dtd">
<HTML>
<!--RAW FILE-->
<head>
<base href="http://ads.nao.ac.jp/abs/1996AnGeo..14..608I" />
<link rel="canonical" content="http://ads.nao.ac.jp/abs/1996AnGeo..14..608I" />
<meta name="robots" content="noarchive" />
<meta http-equiv="Content-Type" content="text/html; charset=UTF-8" />
<TITLE>Decay of the Dst field of geomagnetic disturbance after substorm onset and its i</TITLE>
<meta name="citation_language" content="en" />
<meta name="citation_doi" content="10.1007/s00585-996-0608-3" />
<meta name="citation_abstract_html_url" content="http://ads.nao.ac.jp/abs/1996AnGeo..14..608I" />
<meta name="citation_title" content="Decay of the Dst field of geomagnetic disturbance after substorm onset and its
implication to storm-substorm relation" />
<meta name="citation_authors" content="Iyemori, T.; Rao, D. R. K." />
<meta name="citation_issn" content="0992-7689" />
<meta name="citation_date" content="06/1996" />
<meta name="citation_journal_title" content="Annales Geophysicae" />
<meta name="citation_volume" content="14" />
<meta name="citation_firstpage" content="608" />
<meta name="citation_lastpage" content="618" />
<link title="schema(PRISM)" rel="schema:prism" href="http://prismstandard.org/namespaces/1.2/basic/" />
<meta name="prism:publicationName" content="Annales Geophysicae" />
<meta name="prism:issn" content="0992-7689" />
<meta name="prism:volume" content="14" />
<meta name="prism:startingPage" content="608" />
<meta name="prism:endingPage" content="618" />
<link title="schema(DC)" rel="schema:dc" href="http://purl.org/dc/elements/1.1/" />
<meta name="dc:language" content="en" />
<meta name="dc:identifier" content="doi:10.1007/s00585-996-0608-3" />
<meta name="dc:date" content="1996-06" />
<meta name="dc:source" content="Annales Geophysicae, vol. 14, Issue 6, pp.608-618" />
<meta name="dc:title" content="Decay of the Dst field of geomagnetic disturbance after substorm onset and its
implication to storm-substorm relation" />
<meta name="dc:creator" content="Iyemori, T." />
```

図 6 文献メタデータの例.

## 5. ADS 文献メタデータを元にした辞書を用いた連想検索事例

連想検索の動作確認を Web 上で行う為のプログラムである gss3 protocol analyzer<sup>9)</sup> が、GETAssoc 同様に国立情報学研究所で開発・配布されている。図 3 に示した連携システムの全体構築に先立ち、gss3 protocol analyzer の Web サイトを立ち上げ公開した<sup>26)</sup>。この gss3 protocol analyzer 上で、ADS 文献メタデータを元にした辞書を用いた連想検索例を図 7 に示す。“Fretext” 欄に入力した “magnetism” に関連した単語、例えば地磁気指数のひとつである “K-index” 等が、その関連度と共に出力された。文献メタデータ収録対象を天文から物理の範囲に限定された文献メタデータ・データベースである ADS を利用した上に、超高層物理学分野の研究者による ADS 文献メタデータにまで絞り込んだ上で辞書を作成している為、的確な関連語が上位に選択されている。以前行った多様な事象を包含する Wikipedia を元にした辞書の利用で得られた連想検索結果は、専門用語と一般用語が入り乱れていた為、以前より格段の進歩があった。

次に、IUGONET メタデータ・データベースと GETAssoc サーバーの連携事例を図 8 に示す。なお、IUGONET メタデータ・データベースの本公開サーバー上では未連携であり、九州大学における非公開のテスト環境上で本テストを行った。

GSS3 PROTOCOL ANALYZER This demonstration is powered by GETAssoc.

Assoc Full text Getprop Catalogue Request Response 0.042s

Freetext:

Keyword vector:

Article name:

Filter by phrase: Include  Exclude

Filter by keyword: Include  Exclude

Target:

---

Stage1

niwords:

cutoff-df:

stage1-sim:

Stage2

narticles:

from 1 > next 10 24 articles 78 keywords

Score	Name	Title	Score	Name
0.611 11	K-index		2.722	e
0.610 14	International Association of Geomagnetism and Aeronomy		2.476	t
0.608 18	Intermagnet		2.436	a
0.598 1	Earth's magnetic field		2.220	n
0.577 21	International Geomagnetic Reference Field		2.185	i
0.571 15	Committee on Data for Science and Technology		2.133	o
0.554 24	COSPAR international reference atmosphere		2.098	r

図 7 ADS 文献メタデータを元にした辞書を用いた連想検索例.

IUGONET Metadata DB. >

[Questionnaire](#) [easy feedback](#) or [detail survey](#)

---

**Search Results**

Time from:  to  [UTC]

Data Set ( ☒ Numerical ☒ Plot / Movie ☐ Data File / Plot ☐ Instrument ☐ Observatory )

Search produced no results.

Results/Page  | Sort items by  In order

**Did you mean:**

[CME](#)  
[AE index](#)  
[Ionosphere](#)  
[Intergovernmental Panel on Climate Change](#)  
[Exosphere](#)

This associative search is powered by [GETAssoc](#)

1

Copyright © 2009-2012 IUGONET - [Feedback](#)  
 Customized by [IUGONET](#)  
 Powered by [DSpace Software](#).

図 8 IUGONET メタデータ・データベースと GETAssoc サーバーの連携事例.

IUGONET メタデータ・データベースにおいて、“substorm”という検索語句で検索したところ、該当メタデータは0件であった。超高層物理学分野において“substorm”は重要な専門用語である為、この単語で何もヒットしないことを不思議に思われるかも知れない。現在の IUGONET のメタデータは、観測条件等の1次データに紐づく客観的事実を淡々と記載している段階であり、いつ substorm が起こったのかといった、1次データから得られる知見情報は殆ど記述されていないことが、0件の検索結果の理由である。“substorm”を含むメタデータは0件であったものの、“substorm”と関連した専門用語が5つ推薦されており、さらにはこのリンクを選択することで再検索が行われた為、連想検索はメタデータ・データベースの検索の補助として有効なことを確認した。連想検索を検索の補助として利用する一方で、今後、検索ログを元にしたメタデータの見直し、修正も行う。

## 6. まとめ

超高層物理学を対象とした IUGONET メタデータ・データベースでは、専門分野外のユーザーにとって単語検索時における適切な検索語句の選択が容易でないことが指摘されていた。そこで、再検索文字列を自動生成する為の関連語を得る為に、IUGONET メタデータ・データベースと連想検索エンジン GETAssoc を連携させた。対象範囲が天文から物理の範囲に限定された論文メタデータ・データベースである SAO/NASA Astrophysics Data System から、IUGONET 参加研究者らによる ADS 文献メタデータを抽出した辞書を作成・使用した。GETAssoc サーバーを立ち上げ、適切な連想検索が出来ていることを確認した。テスト環境下において、GETAssoc サーバーと IUGONET メタデータ・データベースを連携させ、GETAssoc サーバーが返す連想検索結果が、メタデータ・データベースの検索の補助になることを確認した。今

後は、IUGONET メタデータ・データベースの本運用機での連想検索の導入が課題である。

## 謝辞

大学間連携プロジェクト「超高層大気長期変動の全球地上ネットワーク観測・研究」は、文部科学省特別教育研究経費（研究推進）[平成21年度]および特別経費（プロジェクト分）[平成22年度～]の交付を受けて、平成21年度より6ケ年計画で実施している事業である。IUGONET プロジェクトで使用される観測データのデータベース構築ならびにメタデータ作成に携わった全ての方々に感謝する。

## 参考文献

- 1) 林 寛生, 小山 幸伸, 堀 智昭, 田中 良昌, 新堀 淳樹, 鍵谷 将人, 阿部 修司, 河野 貴久, 吉田 大紀, 上野 悟, 金田 直樹, 米田 瑞生, 田所 裕康, 元場 哲郎, 大学間連携プロジェクト「超高層大気長期変動の全球地上ネットワーク観測・研究」, 宇宙科学情報解析論文誌, 第1号, JAXA-RR-11-007 (ISSN 1349-1113), pp. 113-120, 2012.
- 2) 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 Sci. J., 12, p.WDS179-WDS184, 2013.
- 3) 小山 幸伸, 河野 貴久, 堀 智昭, 阿部 修司, 吉田 大紀, 林 寛生, 田中 良昌, 新堀 淳樹, 上野 悟, 金田 直樹, 米田 瑞生, 元場 哲郎, 鍵谷 将人, 田所 裕康, 超高層物理学分野のメタデータ・データベースの開発, 宇宙科学情報解析論文誌, 第1号, JAXA-RR-11-007 (ISSN 1349-1113), pp. 91-98, 2012.
- 4) 堀 智昭, 鍵谷 将人, 田中 良昌, 林 寛生, 上野 悟, 吉田 大紀, 阿部 修司, 小山 幸伸, 河野 貴久, 金田 直樹, 新堀 淳樹, 田所 裕康, 米田 瑞生, IUGONET 共通メタデータフォーマットの策定とメタデータ登録管理システムの開発, 宇宙科学情報解析論文誌, 第1号, JAXA-RR-11-007 (ISSN 1349-1113), pp. 105-111, 2012.
- 5) 小山 幸伸, 河野 貴久, 林 寛生, 堀 智昭, 田中 良昌, 鍵谷 将人, 吉田 大紀, 上野 悟, 阿部 修司, 三好 由純, 金田 直樹, 能勢 正仁, 岡田 雅樹, 超高層物理学分野におけるメタデータ・データベースの構築, Proc. of the 2nd Forum on Data Engineering and Information Management, 2010, F4-3, <http://db-event.jp.org/deim2010/proceedings/files/F4-3.pdf>.
- 6) 河野 貴久, 小山 幸伸, 堀 智昭, 阿部 修司, 吉田 大紀, 林 寛生, 新堀 淳樹, 田中 良昌, 鍵谷 将人, 上野 悟, 金田 直樹, 田所 裕康, DSpace を用いた超高層物理学のためのメタデータ・データベースの構築, Proc. of the 3rd Forum on Data Engineering and Information Management, 2011, C8-5, <http://db-event.jp.org/deim2011/proceedings/pdf/c8-5.pdf>.
- 7) 梅村 宜生, 小山 幸伸, 堀 智昭, 阿部 修司, 林 寛生, 新堀 淳樹, 田中 良昌, 上野 悟, 米田 瑞生, 金田 直樹, 元場 哲郎, 超高層物理学のための分野横断型メタデータ・データベースの構築, Proc. of the 4th Forum on Data Engineering and Information Management, 2012, A7-1, <http://db-event.jp.org/deim2012/proceedings/final-pdf/a7-1.pdf>.
- 8) <http://search.iugonet.org/iugonet/>
- 9) <http://getassoc.cs.nii.ac.jp>
- 10) Todd King, James Thieman, and D. Aaron Roberts, SPASE 2.0: a standard data model for space physics, Earth Sci Inform, 3:67-73, 2010.
- 11) Thieman, J. R., D. A. Roberts, T. A. King, C. C. Harvey, C. H. Perry, and P. J. Richards, SPASE AND THE HELIOPHYSICS VIRTUAL OBSERVATORIES, Data Sci. J., 9, p.IGY85-IGY93, 25 February, 2010.
- 12) <http://www.iugonet.org/data/schema/iugonet.xsd>
- 13) <http://www.nii.ac.jp/irp/list/>
- 14) <http://www.dspace.org/>
- 15) <http://dublincore.org/>
- 16) <https://wiki.duraspace.org/display/DSDOC17/Architecture>
- 17) <https://github.com/iugonet/IugonetAssociativeSearch>
- 18) <http://getassoc.iugonet.org/>
- 19) <http://dumps.wikimedia.org/enwiki/>



- 20) 八木 学, 小山 幸伸, 阿部 修司, 梅村 宜生, 堀 智昭, 田中 良昌, 新堀 淳樹, 上野 悟, 佐藤 由佳, 谷田貝 亜紀代, Bernd RITSCHHEL, 連想検索エンジン GETAssoc の超高層物理学におけるメタデータ・データベースへの適用, Proc. of the 5nd Forum on Data Engineering and Information Management, 2013, A10-1, <http://db-event.jp/2013/proceedings/pdf/a10-1.pdf>.
- 21) 梅原 佳宏, 清水 敏之, 吉川 正俊, 地球科学データと学術論文の関連付け, Proc. of the 4th Forum on Data Engineering and Information Management, 2012, A7-3, <http://db-event.jp/2012/proceedings/final-pdf/a7-3.pdf>.
- 22) 梅原 佳宏, 清水 敏之, 吉川 正俊, 地球科学データと学術論文の対応付け獲得に向けた事例調査, Proc. of the 5th Forum on Data Engineering and Information Management, 2013, D4-4, <http://db-event.jp/2013/proceedings/pdf/d4-4.pdf>.
- 23) <http://adsabs.harvard.edu/index.html>
- 24) 蔵川 圭, 武田 英明, 研究者識別子 ORCID の取り組み, 情報管理, Vol. 54 (2011) No. 10 P622-631.
- 25) <https://docs.google.com/spreadsheet/pubkey=0AgbaOOcOsZncdE5sOFJySXFOdnNRMXBoMnQ5MFJLRmc&output=html>
- 26) [http://getassoc.iugonet.org/geta\\_client/gss3.html](http://getassoc.iugonet.org/geta_client/gss3.html)

# CALET の運用及びデータ解析・管理システムの概要

赤池 陽水<sup>\*1</sup>, 浅岡 陽一<sup>\*2</sup>, 上野 史郎<sup>\*3</sup>, 田村 忠久<sup>\*4</sup>, 寺澤 敏夫<sup>\*1</sup>,  
富田 洋<sup>\*3</sup>, 鳥居 祥二<sup>\*2</sup>, 中川 友進<sup>\*3</sup>, 仁井田 多絵<sup>\*2</sup>

## Data Analysis and Handling for CALET

Yosui AKAIKE<sup>\*1</sup>, Yoichi ASAOKA<sup>\*2</sup>, Shiro UENO<sup>\*3</sup>, Tadahisa TAMURA<sup>\*4</sup>, Toshio TERASAWA<sup>\*1</sup>,  
Hiroshi TOMIDA<sup>\*3</sup>, Shoji TORII<sup>\*2</sup>, Yujin E. NAKAGAWA<sup>\*3</sup> and Tae NIITA<sup>\*2</sup>

### Abstract

CALorimetric Electron Telescope (CALET) is a new observatory to be placed on the International Space Station in 2014 to carry out the accurate measurements of electrons in 1 GeV – 20 TeV, gamma-rays in 10 GeV – 10 TeV and protons and nuclei in several 10 GeV – 1000 TeV. The scientific objectives of the mission are mainly to search for nearby cosmic ray sources and dark matter signatures. CALET, with its imaging and deep calorimeter, provides excellent energy resolution and high background rejection, which are confirmed by Monte Carlo simulations and accelerator beam experiments. In this paper, we present the data analysis methods based on simulations and experiments with the ground system for CALET operation and data handling.

Key words : cosmic ray, calorimeter, International Space Station, ground analysis system

### 概 要

CALorimetric Electron Telescope (CALET) は国際宇宙ステーションの「きぼう」に搭載予定の宇宙線観測装置で、2014年から5年間の観測を実施予定である。CALETの主検出器は、電子（1 GeV – 20 TeV）、ガンマ線（10 GeV – 10 TeV）の観測に最適化したカロリメータで、優れたエネルギー分解能と強力な粒子識別能を有し、宇宙線近傍加速源と暗黒物質候補の探索が主な目的である。観測性能については、これまでに気球実験や加速器実験で検証が行われており、シミュレーション計算や加速器実験結果から高い精度が確認されている。本論文では、これらの性能評価を基にCALETが取得するデータの解析手法について解説する。また運用中におけるCALETの地上システムや、データの取り扱いについてその概要を紹介する。

### 1. はじめに

CALorimetric Electron Telescope (CALET) 実験は、国際宇宙ステーション（Internal Space Station: ISS）の日本実験棟「きぼう」の船外プラットフォームに搭載予定の宇宙線観測装置で、2014年度から運用を開始し、5年間の観測を行う予定で開発が進められている。CALETはJAXA宇宙環境利用センターと早稲田大学を中心とした、イタリア（ASI）、アメリカ（NASA）との共同ミッションとして開発が実施されている。高エネルギー電子成分（1 GeV – 20 TeV）、ガンマ線（10 GeV – 10 TeV）、陽子・原子核成分（数 10 GeV – 1000 TeV）の観測により、近傍加速源や暗黒物質候補の探索という高エネルギー宇宙物理学の主要な課題に挑戦するとともに、陽子・原子核成分の高精度観測や太陽変動、ガンマ線バーストのモニタを実施し、体系的な高エネルギー宇宙の解明を目指している [1]。

CALETの主観測対象である高エネルギー電子成分の観測を実現するためには、優れたエネルギー分解能と強力な粒子識別能を有する検出器による、長期間の観測が必要になる。我々はこれまでに、気球実験 [2] や加速器実験 [3, 4] を通じてその観測性能を実証しており、シミュレーション計算を駆使して解析手法を含む装置性能の最適化を図ってきた [5]。

---

\*1 東京大学 宇宙線研究所

\*2 早稲田大学 理工学研究所

\*3 宇宙航空研究開発機構 宇宙科学研究所

\*4 神奈川大学 工学部

ここでは、CALET のエネルギー測定や粒子種選別などのデータ解析手法について、シミュレーション計算や加速器実験で得られた実績を基に詳述する。

CALET が ISS の軌道上で取得するデータは、これまでの船外プラットフォームにおける実験と同様に、ISS が持つ通信経路を利用して筑波宇宙センターで受信し保存される。CALET のデータは運用やモニタのために早稲田大学に設置される CALET オペレーションセンター (WCOC: 仮称) でほぼリアルタイムに受信し、軌道上データの地上モニタを実施するとともに、早稲田大学を中心とする国際共同チームによりサイエンスデータの解析を行う。このような地上系運用システムの概要についても紹介する。

## 2. CALET の概要

### 2.1. 観測目的

高エネルギーの電子は、宇宙空間を伝播中に、シンクロトロン放射と逆コンプトン散乱の影響を受けて、エネルギーの 2 乗に比例してエネルギーを失う。このため、エネルギーの高い電子ほど寿命が短いという特徴を有する。従って、高エネルギー領域では地球から比較的近傍の加速源の影響を強く受けたエネルギースペクトルが理論的に予想される。最近の観測によると、PAMELA[6] や AMS-02[7] が、図 1 に示すようなエネルギーと共に増大する陽電子比を報告した。これはそれまで標準的に考えられてきたような、陽子が星間物質との相互作用で陽電子を生成する伝播モデル [8] だけでは説明が不可能で、その起源が大きな注目を集めている。ATIC[9] や PPB-BETS[10] の気球実験、及び Fermi[11, 12] による衛星実験では、電子 + 陽電子のエネルギースペクトルを観測し、600 – 800 GeV のエネルギー領域において、程度の違いはあるものの、やはり過剰な成分を報告している。これらは、近傍パルサーなどの天体起源の可能性とともに、暗黒物質の対消滅・崩壊で生成される（陽）電子の兆候とも考えられており、今後より詳細な観測が望まれている状況にある。

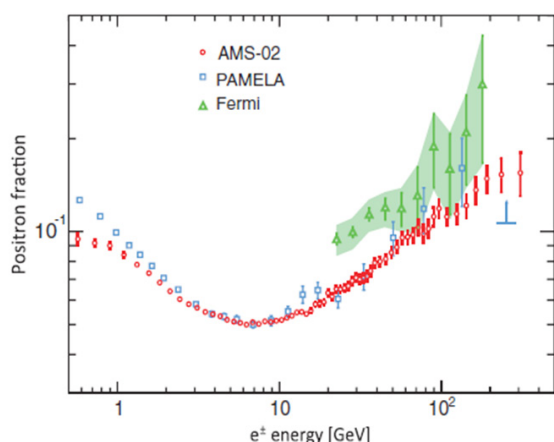


図 1 陽電子 / (電子 + 陽電子) のエネルギースペクトル [7]

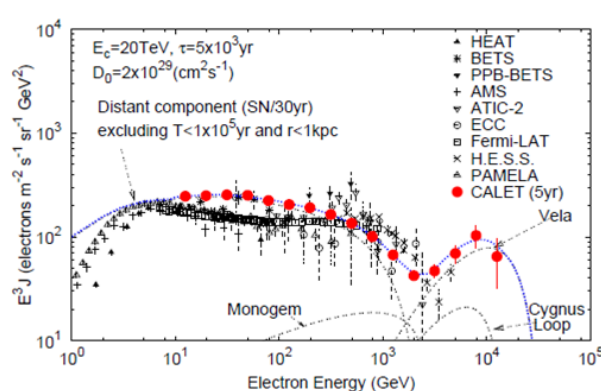


図 2 電子 + 陽電子 のエネルギースペクトルと CALET の観測予測

また 1 TeV 以上のエネルギー領域は、直接観測ではエマルジョン・チェンバーによる気球実験 [13] の観測例があるだけで、まだほとんど観測が行われていない。このエネルギー領域に飛来する電子の候補となる加速源は、地球近傍 1 kpc 以内、1 万年以下の加速源に限られるため、図 2 に示すように超新星残骸 Vela などの直接的な影響が明確にエネルギースペクトルに現れることが理論的に示唆されている [14]。CALET は、電子 + 陽電子のエネルギースペクトルを 1 GeV – 20 TeV の広範囲かつ約 2 % の高精度で観測することで、新たな発見に挑戦するとともに宇宙線の起源や加速・伝播機構の体系的な解明を目指している。

### 2.2. 観測機器

CALET の観測装置を図 3 に示す。CALET の観測器は主検出器であるカロリメータ (CAL) と、ガンマ線バーストモニタ (CGBM) から構成され、補助的な装置としてスターセンサ (ASC) や GPS 受信機 (GPSR)、ミッションデータ処理回路 (MDC) などから構成される。図 4 に、主検出器であるカロリメータの概要を示す。カロリメータは、入射粒子

の電荷を測定する Charge Detector (CHD)、粒子シャワーの初期発達の様子を検出し、入射粒子の到来方向や電荷の有無を計測する Imaging Calorimeter (IMC)、エネルギー測定や粒子バックグラウンド除去を担う Total Absorption Calorimeter (TASC) から構成される。

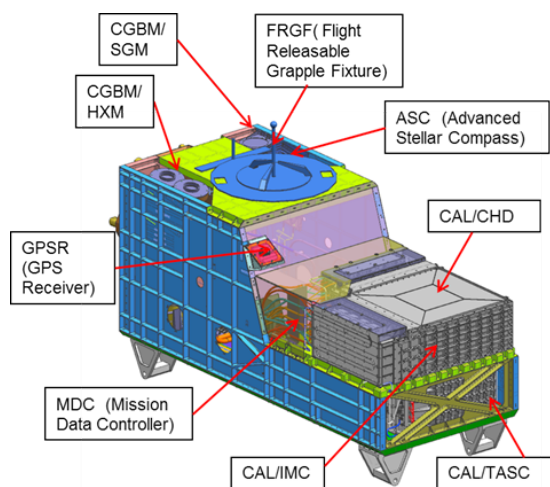


図 3 CALET 観測装置の概要

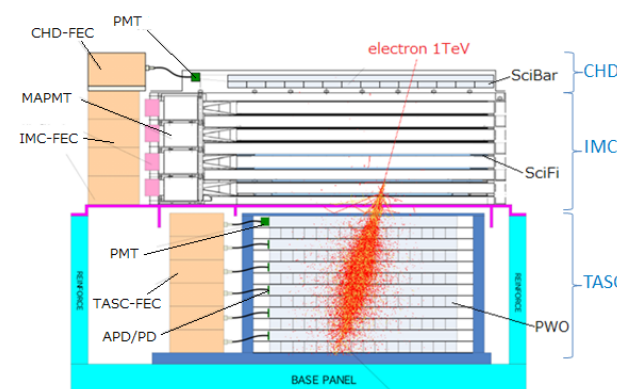


図 4 カロリメータの概要と電子 1 TeV の観測例

### 2.2.1. Charge Detector

CHD は検出器最上部に位置する検出器で、電離損失量の値から入射粒子の電荷を測定する。検出器は、幅 32 mm、長さ 450 mm、厚さ 10 mm のプラスチックシンチレータを 14 枚並べたものを X、Y 方向に 1 層ずつ重ねた構造をもつ。IMC から得られる入射位置の情報と組み合わせることで電荷を決定することができる。また、入射位置に CHD の信号の有無を検知することで、ガンマ線の識別に使用することもできる。

### 2.2.2. Imaging Calorimeter

IMC は、8 層のシンチレーティングファイバーベルトと、7 層のタングステン板を交互に積層したサンプリング型のカロリメータである。入射粒子のシャワーの初期発達形状を詳細に計測し、到来方向の推定や電荷の判別を行う。タングステン板は、0.2 放射長 ( $X_0$ ) の板が 5 枚、1  $X_0$  の板が 2 枚の計 3  $X_0$  であり、各シンチレーティングファイバーベルトは、断面 1 mm<sup>2</sup>、長さ 448 mm のシンチレーティングファイバー (SciFi) を 448 本平行に並べたものを 2 層直交して並べた構造となる。これらを組み合わせることで、3 次元的なシャワー発達の様子を仔細に捉えることが可能である。

### 2.2.3. Total Absorption Calorimeter

TASC は検出器下部に位置する検出器で、阻止能の高い無機シンチレータ (PW0) を用いた不感領域が非常に小さい全吸収型のカロリメータである。各 PW0 の大きさは幅 19 mm、長さ 326 mm、厚さ 20 mm であり、1 層 16 本並べたものを X、Y 交互に計 12 層重ねた構造を持つ。TASC の厚みは 27  $X_0$  あり、TeV 領域の電子シャワーエネルギーを捉える必要十分な厚みを有する。この厚い検出器を採用することで、高精度なエネルギー測定性能を実現し、細分化された構造によりシャワー形状を捉えることで、強力な粒子識別性能を発揮する。

## 2.3. カロリメータの観測性能

検出器の観測性能については、シミュレーション計算 [5] や加速器実験 [3, 4] によって性能検証を行っている。検出器シミュレーションには、EPICS[15] を用いた。表 1 に、CALET の基礎的な観測性能を示す。検出器の幾何学的因子は 1200 cm<sup>2</sup>sr あり、これまでの観測結果 [11] から外挿すると、1 TeV 以上の電子を 5 年間の観測で約 1000 例収集することが可能である。また、TeV 領域の電子観測を実現するために必要な 10<sup>5</sup> の陽子除去性能や、精確なエネルギースペクトル測定に不可欠な 2~3 % のエネルギー分解能を有する点が、CALET の大きな特徴である。

表 1 カロリメータの観測性能

	幾何学的因子 $S \Omega$ [cm <sup>2</sup> sr]	エネルギー分解能 $\Delta E$ [%]	角度分解能 $\Delta \theta$ [degree]	陽子除去性能 e/p	電荷分解能 $\Delta Z$ [e]
電子	1200	2 (>100 GeV)	1.1 ~ 0.11 (1 GeV – 20 TeV)	10 <sup>5</sup>	–
ガンマ線	1000	2 (> 100 GeV)	0.54 ~ 0.22 (10 GeV – 10 TeV)	–	–
陽子・原子核	600	30 ~ 20 (p ~ Fe)	< 1	–	0.15 – 0.3 (p ~ Fe)

### 3. CALET のデータ解析

CALET は、検出器に入射する宇宙線が検出器中で引き起こすシャワー粒子を可視化することによって、粒子の到来方向やエネルギーを測定し、そのシャワー形状から粒子種を特定する。ここでは、加速器実験やシミュレーション計算を例に、粒子識別やエネルギー測定に関する CALET の主要な解析手法について述べる。

#### 3.1.1. エネルギー較正

各検出器からのシグナルは、最小電離損失粒子（Minimum Ionizing Particle: MIP）によるシグナルを基準として絶対値を較正する。最小電離損失とは、相対論的速度を持つ荷電粒子の電離損失によるエネルギー損失量である（厳密にはその最小値）。図 5 は、TASC を構成する PWO のミュオン粒子通過時における ADC 分布で、CERN-SPS 加速器実験の測定例である。この分布の最頻値を 1 MIP と定義する。エネルギー損失量はシンチレータごとに既知であるので、ADC 値を粒子数（=ADC 値 /MIP）に換算することで、シミュレーション計算との比較からエネルギーの推定が可能になる。図 6 は、CERN-SPS の加速器実験で得られた電子 10 GeV のシャワー粒子を観測した際の TASC 中における全粒子数分布である。EPICS[15] を用いた検出器シミュレーションが、実験データをよく再現していることが確認できる。

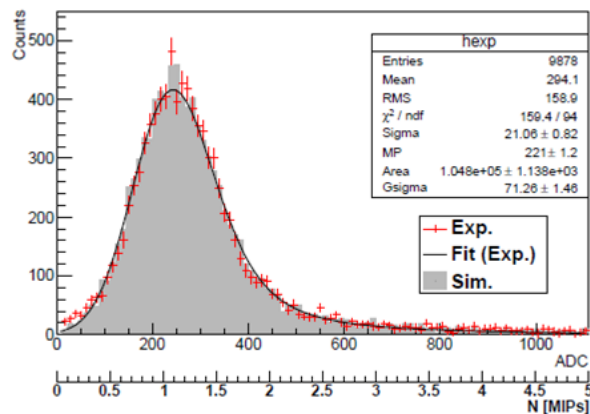


図 5 ミュオン粒子による最小電離粒子の TASC-PWO の ADC 分布（CERN-SPS 加速器実験 [3]）

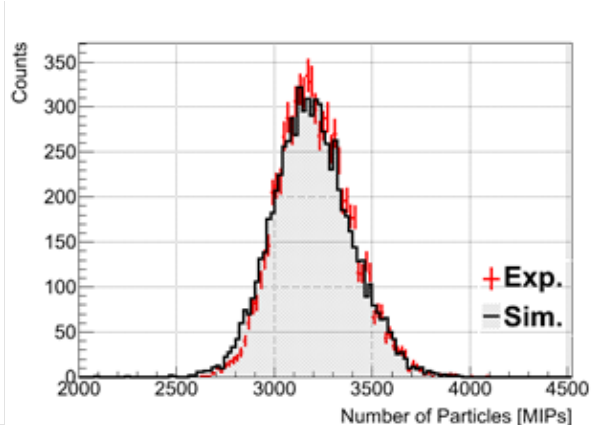


図 6 TASC における粒子数分布（電子 10 GeV: CERN-SPS 加速器実験 [3]）

#### 3.1.2. 到来方向測定

粒子の到来方向は、IMC 中のシャワー粒子によるシグナルをフィッティングすることにより推定する。シグナルは、後方散乱の影響でシャワー軸近辺以外からも検出されるため、シャワーが十分に発達した大きなシグナルとして出力される最下層の SciFi において、まずシャワー軸の位置をエネルギー重心から決める。そして、ここを基準の一つずつ上層に位置を推定して、各層のシャワーのエネルギー重心を求める。それらを直線でフィッティングしてシャワー軸を決定し、粒子の到来方向を求める。



### 3.1.3. 電荷測定

入射粒子の電荷測定には、検出器最上段に位置する CHD、及び IMC を用いる。荷電粒子の電離損失量は入射電荷の 2 乗に比例するため、シャワー軸で推定した入射位置のシンチレータのシグナルの大きさを調べることで粒子の電荷を決定することができる。厳密には、シンチレータの発光量は消光（クエンチング）効果のために、電荷の 2 乗の比例関係からは外れる特性を持つが、この消光効果の度合いを含め、加速器実験（HIMAC、GSI 加速器、CERN-SPS）で電荷測定性能を検証し、入射電荷  $Z = 40$  までの荷電粒子の測定が可能であることを確認している。電荷分解能は、 $0.15 \sim 0.3e$  である [4]。

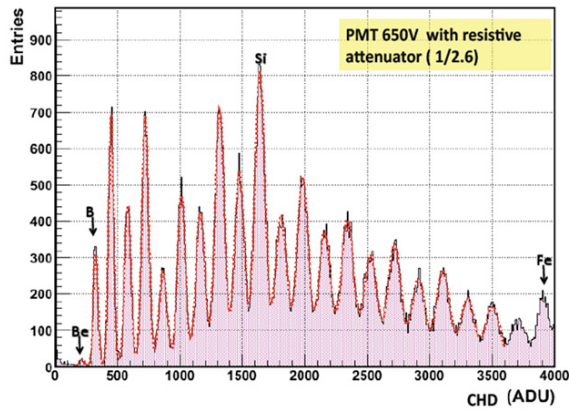


図 7 CHD で測定した電荷分布（GSI 加速器実験 [4]）

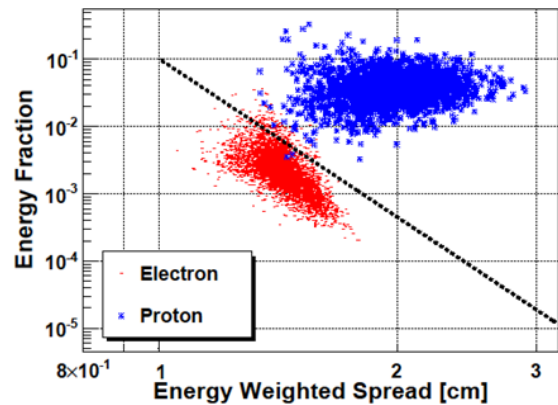


図 8 TASC におけるシャワー形状による電子陽子識別 [5]

### 3.1.4. 電子 / 陽子識別

電子の観測では、陽子が大きなバックグラウンドになる。その存在比はエネルギーが高くなるほど大きくなり、1 TeV 領域における陽子の存在比は、電子に対しておよそ 1000 倍になる。従って、TeV 領域において電子観測における陽子の影響を数 % 程度以下に抑えるためには、検出器には  $\sim 10^5$  の陽子除去性能が要求される。

CALET では、この電子と陽子の識別を、検出器中のシャワー発達の差異から判別する。電子は電磁相互作用により検出器中で電磁シャワーを引き起こすのに対して、陽子は核相互作用によるハドロン多重発生からシャワーが始まるため、その発達の形状が電子とは大きく異なる。検出器の垂直方向の吸収層の厚みは電子に対して  $30 X_0$  あるのに対して、陽子に対しては  $1.3 \text{ m.f.p.}$  (平均自由行程) である。従って、電子は TeV 領域でも検出器中でほぼすべてのエネルギーを失い、検出器下層部でシャワー粒子数が衰退するのに対して、陽子は 2 次ハドロンの逐次相互作用等により、検出器下部においてもシャワー粒子数は衰退しない。この深さ方向のシャワー発達特性の違いと、検出器中のシャワー粒子の横拡がりの大きさの差を用いて電子と陽子の識別を高精度に行うことが可能である。図 8 に、横軸にシャワー粒子の横拡がり（平均値）、縦軸は TASC 中の全エネルギー損失と最下層のエネルギー損失の比をとって、シミュレーション計算から期待される電子と陽子の分離の様子を示す。1 TeV の電子と、そのバックグラウンドとなる 1 TeV 以上の陽子を、観測結果に基づくエネルギースペクトルを仮定して生成したイベントをプロットしている。この散布図を利用して、1 TeV 領域で  $10^5$  の陽子除去性能を達成可能であることを確認している [5]。これは、CALET が TeV 領域の電子シャワーに対して十分な厚さを持つ検出器であることと、シャワー形状を高精度に可視化できる多チャンネル構造を持つことから実現される。

## 4. 観測データ処理システム

### 4.1. データ内容

CALET のデータは、カロリメータのサイエンスデータに加えて、CGBM の定時データ、House Keeping (HK) データ、GPSR や ASC の定時データなどから構成される。ここで定時データとは、一秒周期で取得されるデータの総称である。HK データも定時データとして周期的に取得されている。多チャンネルのサイエンスデータは、ペDESTAL 相当の信号を削除する「ゼロサプレッス」処理を行い可変長のデータとすることで、全データ量を抑える。表 2 に、CALET の主なデータ内容を示す。

表 2 CALET のデータ内容

<b>■ CAL サイエンスデータ</b> CHD (全 28ch) : 最大 1094 byte ~ 1 kB IMC (全 7168ch) : 最大 28972 byte ~ 29 kB TASC (全 192ch) : 最大 2124 byte ~ 2 kB Total ~ 32 kB <b>■ CGBM の定時 (サイエンス) データ</b> パルスハイト : 6608 byte ~ 7 kB タイムヒストリ : 4976 byte ~ 5 kB	<b>■ House Keeping データ</b> 温度、電流、電圧、カウントレートなど : 1094 byte ~ 1 kB <b>■ GPSR の定時データ</b> <b>■ ASC の定時データ</b> <b>■ その他</b> ISS の軌道情報など
--	--

#### 4.2. 軌道上におけるイベントトリガーシステム

ISS 軌道上における運用中、CALET では観測対象に応じて以下のような 3 種類のトリガーモードを使用する。これらは、各トリガー用の検出器からのシグナルのロジックの組み合わせで実現する。

- **High Energy Shower Trigger** : 10 GeV 以上のシャワー粒子が観測対象である。CALET が通常使用するシャワートリガーとして運用する。このトリガーは CHD と IMC 上層の信号を要求しないので、荷電粒子に加えてガンマ線の取得も可能である。シミュレーションによる予測では、バックグラウンドを含め平均 12 イベント/sec でイベントが取得できる見込みである。
- **Low Energy Shower Trigger** : 1 GeV 以上の電子を観測対象とする。これは、太陽モジュレーションの研究のため、より低エネルギー領域の一次宇宙線を測定できるように、地磁気緯度の影響が小さい高緯度地域 (cutoff rigidity < 2 GV) で使用する。
- **Single Trigger** : 検出器校正用のトリガーである。宇宙線の最小電離粒子が取得できるように、各トリガー検出器の閾値を下げ、定期的に使用する。

この他にも、MDC は CGBM でのガンマ線バースト検出時に予め指定しておいたコマンド列をマクロコマンドとして自動実行する機能を有しており、ガンマ線バーストの候補天体が視野に入っている可能性のある 10 分程度の期間、低エネルギーのガンマ線取得に特化したトリガーモードを実現することができる。また、検出器保護のため、CHD において指定した閾値を超える高カウントレートが発生した場合にはマクロコマンドの実行による対応が可能になっている。高カウントレートが予測される South Atlantic Anomaly や高緯度地方では、保護処理用マップを適用することで、マクロコマンドの実行による処理が可能である。これらの機能は、軌道上の観測環境に応じて柔軟な運用が行えるように自由度の高い設計となっている。

#### 4.3. データのダウンリンク

CALET が取得するデータは、これまでの他の ISS 船外実験の運用形態に倣い、ISS から NASA と JAXA が有する 2 種類の経路を利用し、JAXA の筑波宇宙センター (TKSC) に設けられた User Operations Area (UOA) で受信する。データ通信には、それぞれ中速系 (~300 kbps) と低速系 (~35 kbps) を使用する。UOA では、データの保存やモニタ、コマンドの作成・送信を行う。CALET に関するすべてのデータはその後、早稲田大学に設置される Waseda CALET Operations Center (WCOC: 仮称) へ配信される。UOA から WCOC へ配信されるデータには二つの系統が存在する。一つは、ISS の可視時間に装置の状態やデータ取得状況をリアルタイムにモニタリングするための送信速度を最優先した生データ配信で、もう一つはデータ解析のための時間情報、イベント情報などを整理した Level0 データ配信である。WCOC では生データを常時監視するとともに、Level0 データを物理量に変換し、Level1 データとして CALET の国内外の各研究機関へデータ解析用に配信する。図 9 に、データダウンリンクの概念図を示す。CGBM のデータについては速報性が要求されるため、バースト検出の時間情報などの即時配信を行う予定である。10 秒未満の遅延で WCOC へ配信されるリアルタイムデータを用いた The Gamma-ray Coordination Network (GCN)[16] や The Astronomer's Telegram (ATel)[17] への即時配信、及び The Interplanetary Network (IPN)[18] への参加が CGBM チームによって具体的に検討されている。

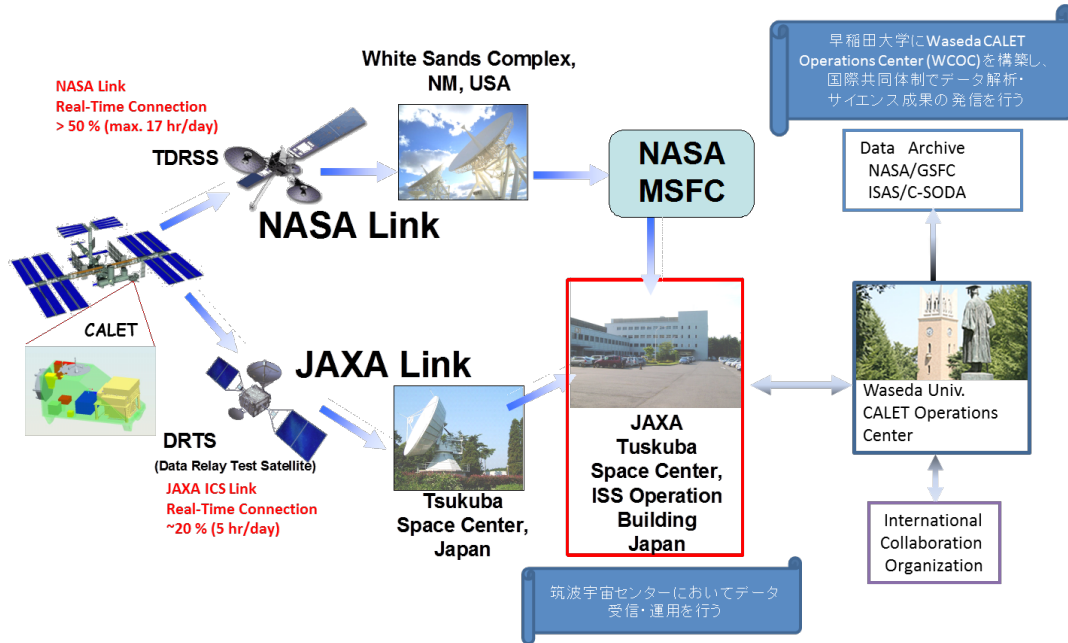


図 9 データリンクの概念図

#### 4.4. 軌道上運用計画

CALET の運用は、TKSC/UAO 及び WCOC にて行う。JAXA は装置健全性の確保に、早稲田大学はサイエンスデータの質の確保に責任を持ち、明確な役割分担の下、密接に協力しながら高効率運用を支える。そのため、WCOC では、UAO と同様に CALET メンバーの研究者が常時 HK データモニタやイベントモニタ等のクイックルック (QL) によりリアルタイムデータを監視し、必要に応じて UOA からコマンドの配信を行う。緊急な対処が必要な事態が発生した場合に、UAO と WCOC 間での情報伝達を迅速に行うため、両者では同一の QL が監視可能となるようなシステム設計がなされている。図 10 は、現在開発中の各チャンネルの温度等の HK データのモニタ表示の一例である。規定の範囲を逸脱した項目を検知した際は、アラートを出すなどのシステムを構築して常時監視できる設計になっている。図 11 は、サイエンス（観測されたシャワー粒子）データに対する QL システムのイベント・ディスプレイの例であり、30 GeV の陽子検出時の例である。各コンポーネントの ADC 値に応じて色分けしており、トリガーモードと各トリガー信号のヒットパターンを表示している。このモニタ画像などを用いてサイエンスデータの監視を行う。

リアルタイムデータ配信に対して、科学解析用のデータとなる Level0 データでは、時系列補正とデータ欠損の補完が行われている。Level0 データは中速系リンクを介して送出されたデータ packets を全て含み、データ形式は可変長のバイナリとなっている。ファイルサイズやファイル数を考慮して 1 時間に 1 ファイルが生成される。Level0 データは CALET データ解析における実質的な一次データとなるため、完備性が確実に確認できるように設計されている。

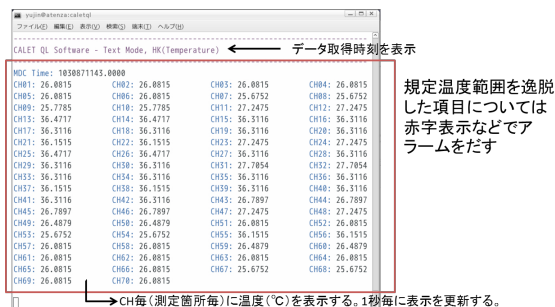


図 10 HK データモニタ例

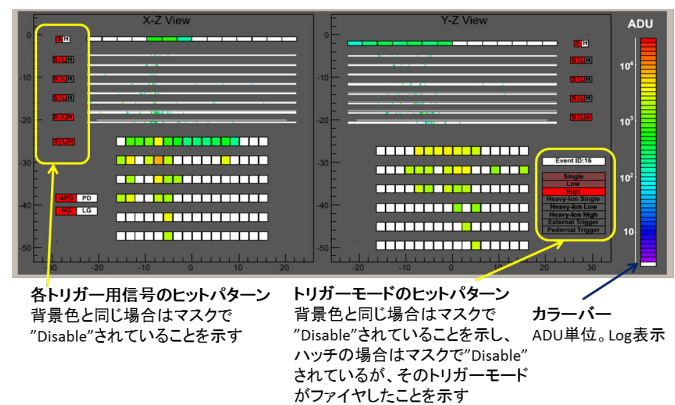


図 11 観測データのモニタ画像表示例

#### 4.5. データ解析・アーカイブ

CALET の科学解析用データの取り扱い、国際研究チーム間で合意されている DPAP (Data Processing and Analysis Plan) と呼ばれる文書に基づいて行われる。WCOC では、Level0 データに各 ADC 値を物理量に変換するなどの処理を加えた Level1 データを生成し、国内外の各研究機関へ配信する。その後、各機関は DPAP の規定に従って高次のレベルのデータを生成し、研究チーム内で合意を前提としてエネルギースペクトルや到来方向分布などの観測結果の導出を行う。CALET が取得するデータは、一部速報する CGBM のデータを除き、観測後 2 年間 CALET の共同研究者が優先権をもってデータ解析を行う。また科学解析結果を含む高次レベルのデータに関しては、日米伊の各データアーカイブサイトに恒久的な保存を行う計画である。アーカイブサイトの候補の絞り込みや、アーカイブ用データの形式やアーカイブ過程の詳細について現在計画中であり、日本では ISAS C-SODA の利用を予定している。

### 5. まとめ

CALET は、国際宇宙ステーションにおける本格的宇宙線観測プロジェクトであり、2014 年の打ち上げ後 5 年間の観測により、これまで実現できなかった高精度かつ高統計な宇宙線諸成分とガンマ線バーストの観測を目指している。CALET が取得するデータは、筑波宇宙センター (TKSC) を経由して、早稲田大学に設置する CALET オペレーションセンター (WCOC) へほぼリアルタイムで送信する。WCOC で各データを物理量に変換した後、サイエンスチームの各研究機関へ配信し、データ解析を実施する。現在、これまでの気球実験や加速器実験の経験を基に、観測運用やデータ解析に向けたハードウェアとソフトウェア両面について開発を行い、2014 年度予定の打ち上げを目指してデータ解析システムの構築を実施している。

### 参考文献

- [1] S. Torii et al., “Overview of the CALET Mission to the ISS”, Proc. of International Cosmic Ray Conference, Vol.6, 347-350, 2011
- [2] 仁井田多絵 他, “気球搭載型 CALET プロトタイプ (bCALET-2) による電子・ガンマ線観測”, 宇宙航空研究開発機構研究開発報告: 大気球研究報告, JAXA-RR-11-008, 17-46, 2012
- [3] Y. Akaïke et al., “CALET observational performance expected by CERN beam test”, Proc. of International Cosmic Ray Conference, CR-IN 0726, 2013
- [4] M. S. Marrocchesi et al., “Beam test performance of a scintillator-based detector for the charge identification of relativistic ions”, Nucl. Instrum. Meth. Phys. Res. A 659, 477-483, 2011
- [5] Y. Akaïke, “Expected CALET Telescope Performance from Monte Carlo Simulation”, Proc. of International Cosmic Ray Conference, Vol.6, 367-370, 2011
- [6] O. Adriani, et al., “An anomalous positron abundance in cosmic rays with energies 1.5-100 GeV”, Nature, 458, 607-609, 2009
- [7] M. Aguilar et al., “First Result from the Alpha Magnetic Spectrometer on the International Space Station: Precision Measurement of the Positron Fraction in Primary Cosmic Rays of 0.5 – 350 GeV”, Phys. Rev. Lett. 110, 141102, 2013
- [8] I. Moskalenko and A. Strong, “Production and Propagation of Cosmic-Ray Positron and Electrons”, Astrophys. J., 493, 694-707, 1998
- [9] J. Chang et al., “An excess of cosmic ray electrons at energies of 300-800 GeV”, Nature, 456, 362-365, 2008
- [10] K. Yoshida, et al., “Cosmic-ray electron spectrum above 100 GeV from PPB-BETS experiment in Antarctica”, Adv. Space Res., 42, 1670-1675, 2008
- [11] A. A. Adbo et al., “Measurement of the Cosmic Ray  $e^+ + e^-$  Spectrum from 20 GeV to 1 TeV with the Fermi Large Area Telescope”, Phys. Rev. Lett. 102, 181101, 2009
- [12] M. Ackermann et al., “Fermi LAT observations of cosmic-ray electrons from 7 GeV to 1 TeV”, Phys. Rev. D 82, 092004, 2010
- [13] T. Kobayashi et al., “Observations of High Energy Cosmic-Ray Electrons from 30 GeV to 3 TeV with Emulsion Chambers”, Astrophys. J., 760, 146-158, 2012

- [14] T. Kobayashi, et al., “The Most Likely Sources of High-Energy Cosmic-Ray Electrons in Supernova Remnants” , *Astrophys. J.*, 601, 340-351, 2004
- [15] EPICS HP: <http://cosmos.n.kanagawa-u.ac.jp/EPICSHome/index.html>
- [16] GCN HP: <http://gcn.gsfc.nasa.gov/>
- [17] ATel HP: <http://www.astronomerstelegam.org/>
- [18] IPN HP: <http://www.ssl.berkeley.edu/ipn3/index.html>



# SMILES Level 2 データ処理における逆問題の解析手法

眞子直弘<sup>\*1</sup>, 鈴木睦<sup>\*2</sup>, 佐野琢己<sup>\*2</sup>, 今井弘二<sup>\*2</sup>, 光田千紘<sup>\*3</sup>,  
山田道夫<sup>\*4</sup>, 竹広真一<sup>\*4</sup>, 塩谷雅人<sup>\*5</sup>

## Inverse analysis methods for SMILES Level 2 data processing

Naohiro Manago<sup>\*1</sup>, Makoto Suzuki<sup>\*2</sup>, Takuki Sano<sup>\*2</sup>, Koji Imai<sup>\*2</sup>, Chihiro Mitsuda<sup>\*3</sup>,  
Michio Yamada<sup>\*4</sup>, Shin-ichi Takehiro<sup>\*4</sup>, Masato Shiotani<sup>\*5</sup>

### Abstract

In the SMILES Level 2 data processing, an inverse problem occurs when deriving mixing ratios of trace gas species from observed brightness temperature spectra. If the information content from the observation is not enough, it is necessary to give additional information from outside to stabilize the solution, and there exist a lot of mathematical techniques for such purposes. In this research, we compared three regularization methods, namely Optimal Estimation Method (OEM), Tikhonov Regularization Method (TRM), and Maximum Entropy Method (MEM), and investigated the most favorable regularization method for SMILES. It turned out that in most cases a hybrid method of OEM + TRM gives the best solution.

**Keywords**— Inverse Analysis, Regularization

### 概要

SMILES 地上データ処理では校正済み輝度温度スペクトル (Level 1B データ) から微量気体分子の混合比鉛直分布等 (Level 2 データ) を導出する際に逆問題を取り扱う。測定から得られる情報が十分でない場合、外部から付加的な情報を与えて解を安定化させる必要があるが、そのために様々な数学的手法が考案されている。本研究では Optimal Estimation Method (OEM), Tikhonov Regularization Method (TRM), Maximum Entropy Method (MEM) の3つの正則化法について比較を行い、SMILES のデータ解析に適した方法を調べた。その結果、OEM + TRM のハイブリッド法を用いると多くの場合において最も良い結果が得られることが分かった。

### 1. はじめに

Superconducting Submillimeter-Wave Limb-Emission Sounder (SMILES) は国際宇宙ステーションの日本実験棟暴露部に取り付けられ、2009 年 10 月から 2010 年 4 月までの約半年間に渡り太陽非同期軌道上において地球大気から放射される 600 GHz 帯サブミリ波のリム観測を行った<sup>1)</sup>。SMILES は冷凍機を使って 4K まで冷却した超伝導ミキサを受信機に用いることにより従来の衛星搭載センサに比べて 1 桁程度小さいノイズレベル (輝度温度 0.4K/single-scan) を実現した。これによって成層圏・中間圏に存在する O<sub>3</sub>, HCl, ClO, HNO<sub>3</sub>, HOCl, HO<sub>2</sub>, CH<sub>3</sub>CN, BrO といった微量気体分子の情報を豊富に含む輝度温度スペクトルが高精度に測定され、既にいくつかの科学的成果が得られている<sup>2,3)</sup>。表 1 に SMILES 観測データの主なパラメータをまとめる。

表 1 SMILES 観測データの主なパラメータ

観測緯度帯	38° S – 65° N (nominal)	観測高度範囲	10 – 80 km (nominal)
周波数帯	Band A (624.32 – 625.52 GHz)	高度分解能	3.5 – 4.1 km (nominal)
	Band B (625.12 – 626.32 GHz)	積算時間	0.5 s / 観測高度
周波数サンプリング	0.8 MHz / チャンネル	データ取得時間	53 s / スキャン
周波数分解能	1.8 MHz (FWHM)	システム雑音温度	0.4 K / スキャン

\*1 千葉大学環境リモートセンシング研究センター (Center for Environmental Remote Sensing, Chiba University)

\*2 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)

\*3 富士通エフ・アイ・ピー株式会社 (Fujitsu FIP Corporation)

\*4 京都大学数理解析研究所 (Research Institute for Mathematical Sciences, Kyoto University)

\*5 京大大学生存圏研究所 (Research Institute for Sustainable Humanosphere, Kyoto University)

SMILES のデータセットには RAW（未処理データ）、Level 0（地上で編集された未処理データ）、Level 1B（校正された輝度温度スペクトルおよび関連データ、以降単に L1B と表記）、Level 2（観測地点において導出された微量気体分子混合比の鉛直分布等の変数、以降単に L2 と表記）、Level 3（一様な時空間グリッド上に補間された変数）のようにレベルが定義されている。L2 データ処理では L1B データに含まれる校正済み輝度温度スペクトルをシミュレーション結果と比較することにより逆問題を解いて様々な変数を導出する。この過程では従来より Optimal Estimation Method<sup>4)</sup>と呼ばれる逆解析アルゴリズムが用いられてきた<sup>5,6)</sup>。しかしながら、得られる解には不自然な振動が含まれることもあり、数学的に必ずしも最善のアルゴリズムではなかった。一方、工学や応用数学の広い分野では一般に、非線型方程式の逆問題に対して Thikhnov Regularization Method<sup>7)</sup>等の様々な手法が用いられている。本研究では SMILES の高精度データに含まれる情報を最大限に引き出すため、逆解析アルゴリズムの高度化について検討した。

## 2. 逆解析アルゴリズム

### 2.1 線型代数モデル

SMILES で測定される輝度温度スペクトルは、大気中の微量気体分子からの放射や背景放射の伝搬を記述した大気モデル、およびアンテナや分光器等の特性を記述した装置モデルを使った放射伝達シミュレーションで再現できる。ここで、SMILES で測定される輝度温度スペクトルを  $\mathbf{y}$ 、導出したい微量気体分子の混合比を  $\mathbf{x}$  とすると、 $\mathbf{y}$  は要素数  $m$ （観測接線高度数×周波数グリッド数 = 約  $40 \times$  約  $1600$ ）の成分を持つ測定ベクトル、 $\mathbf{x}$  は要素数  $n$ （分子種の数×導出高度グリッド数 = 約  $10 \times$  約  $30$ ）の成分を持つ状態ベクトルとみなせる。これらのベクトルは Forward モデル  $\mathbf{F}$  によって関係付けられるが、実際の測定には測定誤差が加わり、Forward モデルには近似等による解析誤差が加わるので、これらをまとめて  $\epsilon$  とすると次のように記述できる。

$$\mathbf{y} = \mathbf{F}(\mathbf{x}) + \epsilon \quad (1)$$

我々の目標は式(1)を逆算して  $\mathbf{x}$  を求めることである。今、真値に十分近い先験値  $\mathbf{x}_a$  の周辺を考えると、 $\mathbf{x}$  と  $\mathbf{y}$  の微小な変化は比例関係にあると考えられ、この比例係数を  $m \times n$  行列  $\mathbf{K} = \partial \mathbf{F} / \partial \mathbf{x} = \nabla \mathbf{F}$ （Weighting Function と呼ぶ）で表すと式(2)のような線型近似が成り立つ。

$$\mathbf{y} = \mathbf{F}(\mathbf{x}_a) + \mathbf{K}(\mathbf{x} - \mathbf{x}_a) + \epsilon \quad (2)$$

式(2)は一見、単純な連立 1 次方程式であるがいくつか注意点がある。まず、一般的にリム観測では  $m \gg n$  であり、式(2)は条件過剰になっている。また、式(2)は正確な値が分からない誤差を含んでおり、これを厳密に満たす解は存在しない。このような問題は測定値と計算値の差を最小化する最小二乗法を使って解くことができる。ただし、例えば隣り合う周波数グリッドの測定値が完全に独立ではないように、式(2)から得られる独立な条件式の数が  $m$  より多くなることはなく、逆に例えば状態ベクトルの高度グリッドを細かく取り過ぎた場合等は独立な条件式の数が  $m$  より少なくなることがある。この場合、状態ベクトルを一意に決めることができない不良設定問題になる。さらに、逆解析で得られる解がノイズの影響を受けやすい悪条件問題にもなっている。リム観測で得られる放射輝度はある広がった高度範囲に分布する微量気体分子からの寄与を積算したものであり、微量気体分子の高度プロファイルがランダムに変化したとしても測定値は平均化されてあまり変化しない。逆に、放射輝度の測定値から微量気体分子の高度プロファイルを導出する場合、放射輝度の測定値に小さいノイズが加わったとしても大きな影響を受ける。このような逆問題を解くために、解を安定化させる正則化を導入する。

通常の最小二乗法では式(3)のように標準偏差で規格化したスペクトル残差の二乗和 ( $\chi^2$ ) を最小化する。

$$\chi^2(\mathbf{x}) = (\mathbf{y} - \mathbf{F}(\mathbf{x}))^T \mathbf{S}_y^{-1} (\mathbf{y} - \mathbf{F}(\mathbf{x})) \quad (3)$$

ここで、 $\mathbf{S}_y$  は測定ベクトルに対応する分散共分散行列であり、SMILES ではシステム雑音温度から計算される輝度温度スペクトルの分散を成分に持つ対角行列を使用する。不良設定問題ではスペクトル残差が小さい解が無数に存在し、解が一意に求まらない。そこで、スペクトル残差と同時に正則化項  $c$  を最小化することを考える。これは以下の式(4)で表される Cost Function を最小化することにより実現できる。正則化項  $c$  の違いにより、次節で説明する Optimal Estimation Method、Tikhonov Regularization Method、Maximum Entropy Method のようなバリエーションが考えられる。

$$M(\mathbf{x}) = \chi^2(\mathbf{x}) + c(\mathbf{x}) \quad (4)$$

解が先験値から遠い場合は式(2)で省いた高次項が無視できずに非線型問題となるが、Levenberg-Marquardt 法を用いて解の変動  $\delta \mathbf{x}$  が十分小さくなるまで式(5)を反復計算することで高速かつ安定に解を求められる。

$$\delta \mathbf{x}_i = \mathbf{x}_{i+1} - \mathbf{x}_i = -[\mathbf{H}_i + \lambda \mathbf{D}]^{-1} \nabla M(\mathbf{x}_i) \quad (5)$$

式(5)において、勾配 $\nabla M$ は次の式(6)を計算して得られる。

$$\nabla M(\mathbf{x}_i) = -2\mathbf{K}_i^T \mathbf{S}_y^{-1} (\mathbf{y} - \mathbf{F}(\mathbf{x}_i)) + \nabla c(\mathbf{x}_i) \quad (6)$$

式(5)のパラメータ $\lambda$ を0にすると Gauss-Newton 法の式に一致する。Gauss-Newton 法は以下の式(7)のように Hessian  $\mathbf{H}_i$  を近似して計算が困難な $\mathbf{F}$ の2次微分を避けているが、Newton 法と同様に2次収束の高速性を持つ。

$$\nabla(\nabla M(\mathbf{x}_i)) \approx 2\mathbf{K}_i^T \mathbf{S}_y^{-1} \mathbf{K}_i + \nabla(\nabla c(\mathbf{x}_i)) \equiv \mathbf{H}_i \quad (7)$$

ただし、初期値が悪いと発散することがあるので、 $M(\mathbf{x})$ が減少しない場合（または $\delta\mathbf{x}$ の大きさが減少しない場合）は式(5)のパラメータ $\lambda$ の値を大きくして適当な正定値行列 $\mathbf{D}$ （例えば恒等行列）を加え、 $\delta\mathbf{x}$ を最急降下法と同様の方向（ $\nabla M$ ）に向ける。

## 2.2 Optimal Estimation Method

Optimal Estimation Method (OEM) は次式のように標準偏差で規格化した状態ベクトルと先験値の差分の二乗和を小さくするような正則化項を用いる。

$$c(\mathbf{x}) = (\mathbf{x} - \mathbf{x}_a)^T \mathbf{S}_a^{-1} (\mathbf{x} - \mathbf{x}_a) \quad (8)$$

ここで、 $\mathbf{S}_a$ は先験値の分散共分散行列である。この正則化項は、状態ベクトルの確率密度が $\mathbf{x}_a$ を中心に $\mathbf{S}_a$ で与えられる分散を持って正規分布していると仮定し、Bayes 理論に基づいて測定ベクトル $\mathbf{y}$ を得る条件で事後確率を最大化することによって導ける<sup>4)</sup>。例えば1次元の場合、状態量 $x$ の確率密度、および状態量が $x$ である条件での測定量 $y$ の確率密度はそれぞれ式(9)および式(10)のようになる。

$$P(x) = \frac{1}{\sigma_a \sqrt{2\pi}} \exp \left[ -\frac{(x - x_a)^2}{2\sigma_a^2} \right] \quad (9)$$

$$P_x(y) = \frac{1}{\sigma_y \sqrt{2\pi}} \exp \left[ -\frac{(y - Kx)^2}{2\sigma_y^2} \right] \quad (10)$$

Bayes の定理より、測定量が $y$ である条件での状態量 $x$ の確率密度は式(11)で与えられる。

$$P_y(x) = \frac{P_x(y)P(x)}{P(y)} \propto \exp \left[ -\frac{1}{2} \left( \frac{(x - x_a)^2}{\sigma_a^2} + \frac{(y - Kx)^2}{\sigma_y^2} \right) \right] = \exp \left[ -\frac{1}{2} M(x) \right] \quad (11)$$

ここで、測定量 $y$ の確率密度 $P(y)$ は $x$ に依らない定数であり、事後確率を最大化することは式(8)を正則化項とする Cost Function を最小化することに等しい。このような正則化によって最小二乗法の解が修正される様子を図1に示す。図1(b)のように測定誤差が小さい（ $\sigma_y \ll K\sigma_a$ ）場合には解は正則化にあまり影響されずに求まるが、図1(c)のように測定誤差が大きい（ $\sigma_y \gg K\sigma_a$ ）場合には解は先験値に縛られる。このように、正則化によって測定から十分な情報が得られない範囲の解を選択的に修正することができる。

一般的に先験値の分散を正確に知ることは難しく、誤差が正規分布しているとも限らないが、ここではおおまかな見積もり値を使用する。SMILES では  $\text{O}_3$ ,  $\text{HCl}$  等の微量気体分子の混合比や温度の高度プロファイル、アンテナの仰角オフセットといった様々な物理量の同時リトリバルを行っているが、 $\mathbf{S}_a$ を使って規格化することでこれらを無次元量にすることができる。Levenberg-Marquardt 法の正定値行列 $\mathbf{D}$ には $\mathbf{S}_a^{-1}$ の非対角項を0にした行列 $\mathbf{D}_a^{-1}$ が利用できる。また、 $\mathbf{S}_a$ の非対角項に例えばスケール高度 $z_s$ で指数関数的に減衰する値（ $[\mathbf{S}_a]_{ij} = [\mathbf{S}_a]_{ii} \exp[-|z_i - z_j|/z_s]$ ）を入れて高度相関を持たせることができる。OEM ではこのような高度相関を入れたり高度グリッド $\mathbf{z}$ の調整を行わないと多くの場合に解が振動する。

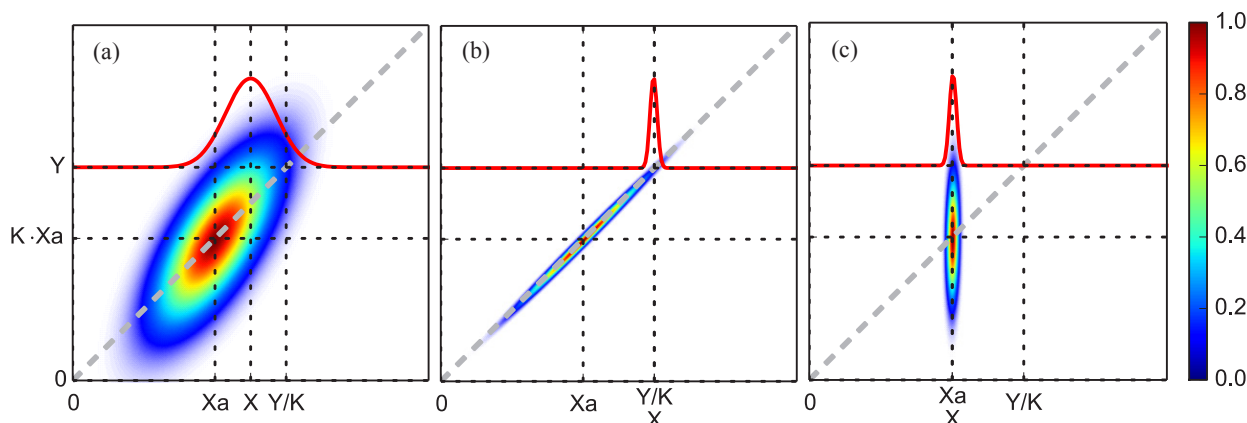


図1 OEM の正則化によって最小二乗法の解が修正される様子. カラースケールは Cost Function  $M(x)$  に対応する指数関数  $\exp[-M(x)/2]$  の大きさを表す. 破線は  $y = Kx$  の直線である. 実線は  $y$  の値が  $Y$  であるときの指数関数の大きさであり, この値が最大となるときの  $x$  の値が  $X$  (解) である. (a) 一般的な場合 (b) 測定誤差が小さい場合 (c) 測定誤差が大きい場合

### 2.3 Tikhonov Regularization Method

Tikhonov Regularization Method (TRM) は OEM と同じく状態ベクトルと先験値の差分が小さいことを条件にするが, OEM と異なり式(12)のような正則化項を用いる.

$$c(x) = (x - x_a)^T \alpha L_k^T L_k (x - x_a) \quad (12)$$

ここで,  $L_k$  は  $k$  次の差分を表す正則化行列であり,  $\alpha$  は正則化の強度を調整する正則化パラメータである. (複数分子種の同時リトリバルを行う場合はスカラー  $\alpha$  の代わりに分子種ごとの正則化パラメータを要素に持つ対角行列  $M_a$  を用いる.) 例えば次数が 0 であれば  $L_k = L_0$  は恒等行列であり, 正則化は状態ベクトルを先験値そのものに近づけるように働く. OEM と同様に 1 次元の例を考えると,  $\alpha = 1/\sigma_a^2$  の置き換えにより TRM は OEM と一致することが分かる. TRM の  $\alpha$  を大きくすることは OEM の  $\sigma_a$  を小さくすることに相当し, 図 1(c) のように測定誤差が大きい ( $\sigma_y \gg K/\sqrt{\alpha}$ ) 場合には解は先験値に縛られる.

次数が 1 であれば  $[L_1 x]_i = x_i - x_{i-1}$  となり, 状態ベクトルに任意のオフセットを加えても Cost Function の値は変わらないことから, 状態ベクトルのオフセットは先験値に依らず, 正則化とは無関係にスペクトル残差最小の条件から決まる (ただし, 以下で述べる境界条件の影響を受ける). 同様に, 次数が 2 であれば状態ベクトルのオフセットと傾きは先験値に依存しない. 次数が大きくなるに従って先験値への依存度は少なくなるが, より多くの成分が混ざって高度分解能が悪化するため, 適当な次数を選ぶ必要がある. 次数  $k$  の正則化行列には  $(n-k) \times n$  行列と  $n \times n$  行列の 2 形式が考えられる<sup>7)</sup>. 前者は正則化が効かない状態ベクトル (例えば  $L_1$  に対する状態ベクトルのオフセット, すなわち定数ベクトル) を持つが, 後者は  $k$  行が境界条件として働くために解を一意に決めることができる. 一般に, 正則化行列には次数の異なる差分行列の線型結合を用いても構わない. OEM に前述した高度相関を入れることは  $L_0$  と  $L_1$  の線型結合を使うことに相当する. (ただし, OEM の高度相関では正則化の強度を調整することができない.)

正則化パラメータは Cost Function に含まれるスペクトル残差と正則化項のバランス調整の役割を果たし, 小さ過ぎると雑音が多い解, 大き過ぎると先験値に縛られた解になるため, 適した値を選ぶ必要がある. 正則化パラメータの決定方法には  $\alpha$  を徐々に大きくしていったスペクトル残差が増加を始める点に設定する方法 (L-curve 法), 平滑化誤差とノイズ誤差を合わせたトータル誤差が最小となるように設定する方法等がある<sup>8)</sup>. L-curve 法は正則化パラメータを徐々に大きくしながら逆問題を解いた時, スペクトル残差と正則化項の散布図を描くと L 字状の軌跡が得られることを利用するが, このような図を描かなくても得られた解を見るとスペクトル残差が増加を始める点でそれまで安定していた解が先験値に向かって動き出すことから判断できる. このような状態になる前の解を見て, 振動が適度に抑制される正則化パラメータを選ぶのが良いと考えられる. 正則化の効果は正則化パラメータが多少 (1 桁程度) 変化してもあまり変わらないため, 厳密に最適化する必要はない. ただし, 分子種が異なれば最適値は異なり, 同じ分子種でも日変動するものは昼と夜で最適値が異なる. この問題に対する簡便策として,  $L_k^T L_k$  の代わりに  $L_k^T S_a^{-1} L_k$  を使用して状態ベクトルの残差を規格化する方法が挙げられる. この規格化により残差はだいたい 1 のオーダーになると考えられ, どの分子種にも同じような正則化パラメータを使うことができる. (次数が 0 で正則化パラメータが 1 であれば OEM と等価である.) 状態ベクトルの大きさに応じた先験値の分散を用いると, 正則化パラメータを定数にする場合と比べて状態ベクトルが小さい範囲でより強く正則化が作用する.

TRM のメリットとして, OEM のような高度グリッドの最適化が不要で等間隔グリッドを使用すること, 解の滑らかさや高度分解能を正則化行列や正則化パラメータの選び方によって調整できることが挙げられる.

## 2.4 Maximum Entropy Method

Maximum Entropy Method (MEM) は状態ベクトル  $\mathbf{x}$  のベクトル関数  $\mathbf{q}(\mathbf{x})$  について次のように計算される Shannon の情報量を最大化する.

$$S(\mathbf{x}) = - \sum_{i=1}^n q_i \ln(q_i) \quad (13)$$

MEM の正則化項には以下の式(14)を用いる.

$$c(\mathbf{x}) = -\alpha S(\mathbf{x}) \quad (14)$$

この式から分かるように, MEM は先験値を必要としないことが特長である. TRM と同様に正則化パラメータ  $\alpha$  を使って正則化の強度を調整することにより, 観測結果を束縛条件にして情報量 (不確かさ) が最大になる解が得られる. ここでは文献<sup>9)</sup>にない,  $\mathbf{x}$  のベクトル関数として以下のように2次差分を規格化したものを用いる.

$$q_i = p_i/p \quad (15)$$

$$p_i = x_{i-1} - 2x_i + x_{i+1} + 2x_{\max} - 2x_{\min} + \zeta \quad (16)$$

$$p = \sum_{i=1}^n p_i = -(x_1 + x_n) + n(2x_{\max} - 2x_{\min} + \zeta) \quad (17)$$

式(16)で, 定数  $2x_{\max} - 2x_{\min} + \zeta$  ( $x_{\max}$ ,  $x_{\min}$  はそれぞれ  $\mathbf{x}$  の成分の最大値, 最小値,  $\zeta = 10^{-15}$ ) は  $\mathbf{p}$  の成分が必ず正となるように加えてある. 最小二乗法の計算に必要な Shannon の情報量の1次および2次微分はそれぞれ式(18), 式(19)のようになる.

$$\begin{aligned} \frac{\partial S}{\partial x_k} &= - \sum_{i=1}^n \left( \frac{1}{p} \left( 1 + \ln \left( \frac{p_i}{p} \right) \right) \left( A_{ik} - \frac{p_i}{p} p_k^x \right) \right) \\ \frac{\partial}{\partial x_m} \left( \frac{\partial S}{\partial x_k} \right) &= - \sum_{i=1}^n \left[ \frac{1}{p p_i} \left( A_{im} - \frac{p_i}{p} p_m^x \right) \left( A_{ik} - \frac{p_i}{p} p_k^x \right) - \frac{1}{p^2} \left( 1 + \ln \left( \frac{p_i}{p} \right) \right) \left( A_{ik} p_m^x + A_{im} p_k^x - \frac{2p_i}{p} p_m^x p_k^x \right) \right] \end{aligned}$$

ここで,  $A_{ik} = \partial p_i / \partial x_k$  は  $i = k$  の時  $-2$ ,  $i = k \pm 1$  の時  $1$ , それ以外は  $0$  であり,  $p_k^x = \partial p / \partial x_k$  は  $k = 1$  または  $n$  の時  $-1$ , それ以外は  $0$  である.

## 2.5 ハイブリッド法

これまでに説明した OEM, TRM, MEM はスペクトル残差にそれぞれの正則化項を加えたものを最小化する方法であるが, これらの正則化項は組み合わせることもできる. 実際, NASA の Aura 衛星に搭載された Microwave Limb Sounder (MLS) などのリム観測では OEM と TRM を組み合わせ使用している<sup>10)</sup>.



### 3. 正則化アルゴリズムの比較

SMILES に適した正則化アルゴリズムを探すため、シミュレーションによって作成した輝度温度スペクトルの疑似データから微量気体分子の混合比を逆算し、真値との比較を行った。問題に適したアルゴリズムは正則化が必要な高度における高度プロファイル（絶対量や微分量）によってケース・バイ・ケースであると考えられるので、様々な分子種（Band A の  $O_3$ 、Band B の  $HCl$ 、Band C の  $ClO$ 、Band C の  $HNO_3$ 、Band B の  $HO_2$ 、Band A の  $HOCl$  の 6 種類）について比較した。真値のプロファイルにはモデル計算（WACCM<sup>11)</sup>）で得られた赤道域の帯状平均（南緯  $10^\circ$  ~ 北緯  $10^\circ$ 、2009 年 11 月 17 日、ローカルタイム 11 ~ 13 時）を使用し、先験値および初期値には真値を 1.5 倍にしたものを用いた。高度グリッドの間隔は 2.5 km とした。比較した逆解析アルゴリズム（16 種類）を表 2 に挙げる。TRM、MEM の正則化パラメータは常用対数スケールで 0.1 刻みずつ値を変化させ、有効高度範囲における平均 2 乗誤差の平方根（RMSE）が最小になる値を選択した。

いくつかの逆解析アルゴリズム（OEM、TRM\_k1、TRM\_k2、TRM\_k2\_oem）について、各分子混合比の誤差と真値の高度プロファイルを図 2 に示す。この図から、OEM 単独では真値が大きい高度範囲でランダム誤差が大きくなることが分かる。これは高度グリッド間隔（2.5 km）が解析の高度分解能に比べて細かいためだと考えられる。TRM\_k1 と TRM\_k2 は似たような誤差プロファイルを示し（例外：高高度の  $HCl$ ）、それぞれ単独では主に真値が小さい高高度で誤差が大きくなっている。TRM\_k2\_oem は OEM と TRM\_k2 の良い部分を合わせた形になっており、全高度範囲で誤差が比較的小さくなっている。

全 16 種類の逆解析アルゴリズムについて、図 2 に示した各分子種の有効高度範囲で計算した RMSE（OEM\_10km の値で規格化している。TRM、MEM については正則化パラメータごとの値）を図 3 に示す。状態ベクトルの残差を規格化した TRM、規格化しない TRM、MEM の 3 グループ内では、それぞれ同じような正則化パラメータで RMSE が最小になっており、各アルゴリズムで得られた RMSE の最小値は図 4 のようになった。図 3、図 4 から以下のことが分かる。

- i) TRM\_k1 と TRM\_k1\_mxn の差はほとんどない（例外： $O_3$ 、 $HCl$ ）
- ii) TRM\_k2 と TRM\_k2\_mxn の差はほとんどない（例外： $O_3$ ）
- iii) TRM\_k1\_nrm と TRM\_k1\_hyb の差は小さい
- iv) TRM\_k2\_nrm と TRM\_k2\_hyb の差は小さい（例外： $HCl$ ）
- v) TRM\_k1\_hyb より TRM\_k1\_oem の方が RMSE の最小値が小さい（例外： $HO_2$ ）
- vi) TRM\_k2\_hyb より TRM\_k2\_oem の方が RMSE の最小値が小さい

項目 i, ii より、TRM の境界条件は多くの場合、解に影響を与えないことが分かる。ただし、 $O_3$ 、 $HCl$  については境界条件が無いと高高度で相対誤差が大きくなることが認められた。項目 iii, iv より、状態ベクトルの残差を規格化した場合、TRM に OEM を組み合わせても効果が薄いことが分かる。これは残差の規格化により高高度の正則化が強く働いて誤差が抑えられるためと考えられる。ただし、TRM\_k2\_nrm では  $HCl$  の高高度で大きな誤差が見られ、TRM\_k2\_hyb では誤差が抑えられていた。項目 v, vi からは、TRM に OEM を組み合わせれば状態ベクトルの残差を規格化しない方が良いことが分かる。この特徴が顕著な  $HNO_3$  について調べてみると、TRM に OEM を組み合わせれば状態ベクトルを規格化した場合、低高度および高高度の真値が小さい範囲において、正則化が強すぎて解が先験値に縛られるために誤差が大きくなっていることが分かった。

各分子種ごとに RMSE の小さい方から順位を付け、各アルゴリズムごとに全分子種の順位を足し合わせると、TRM\_k2\_oem (14)、TRM\_k1\_oem (17) で小さい値が得られた。以下、順位の和が 39 ~ 48 のグループ（TRM\_k2\_hyb、TRM\_k0\_hyb、TRM\_k2\_nrm、TRM\_k1\_nrm、TRM\_k1\_hyb、MEM\_k2\_hyb、TRM\_k1）、順位の和が 56 ~ 65 のグループ（TRM\_k0、TRM\_k2、OEM\_10km、TRM\_k1\_mxn、TRM\_k2\_mxn）と続き、OEM (85)、MEM\_k2 (88) で大きい値が得られた。このことから、誤差をなるべく小さくするという観点では TRM\_k2\_oem または TRM\_k1\_oem が良いアルゴリズムであると言える。正則化パラメータの設定が楽であることを加味すると、TRM\_k2\_hyb も良いアルゴリズムである。理想的にはアルゴリズムを一つに限定せず、分子種ごとに最適なアルゴリズムを選択するのが良いと考えられる。MEM\_k2 は先験値の情報を利用しないために誤差が比較的大きくなっているが、信頼できる先験値が得られない場合は選択肢に入れても良いと思われる。

表 2 比較を行った正則化アルゴリズム

OEM					TRM				
アルゴリズム	高度相関				アルゴリズム	次数	境界条件 <sup>a</sup>	規格化 <sup>b</sup>	OEM <sup>c</sup>
OEM	×				TRM_k1_hyb	1	○	○	○
OEM_10km	○ (10 km)				TRM_k2	2	○	×	×
TRM					TRM_k2_mxn	2	×	×	×
アルゴリズム	次数	境界条件 <sup>a</sup>	規格化 <sup>b</sup>	OEM <sup>c</sup>	TRM_k2_nrm	2	○	○	×
TRM_k0	0	○	×	×	TRM_k2_oem	2	○	×	○
TRM_k0_hyb	0	○	×	○	TRM_k2_hyb	2	○	○	○
TRM_k1	1	○	×	×	MEM				
TRM_k1_mxn	1	×	×	×	アルゴリズム	次数	OEM <sup>c</sup>		
TRM_k1_nrm	1	○	○	×	MEM_k2	2	×		
TRM_k1_oem	1	○	×	○	MEM_k2_hyb	2	○		

a: 次数 $k$ の正則化行列に $n \times n$ 行列（正方行列）を用いて $k$ 行のベクトルを境界条件とする

b:  $\mathbf{L}_k^T \mathbf{L}_k$ の代わりに $\mathbf{L}_k^T \mathbf{S}_a^{-1} \mathbf{L}_k$ を用いて状態ベクトルの残差を規格化する

c: OEMを組み合わせたハイブリッド法を用いる

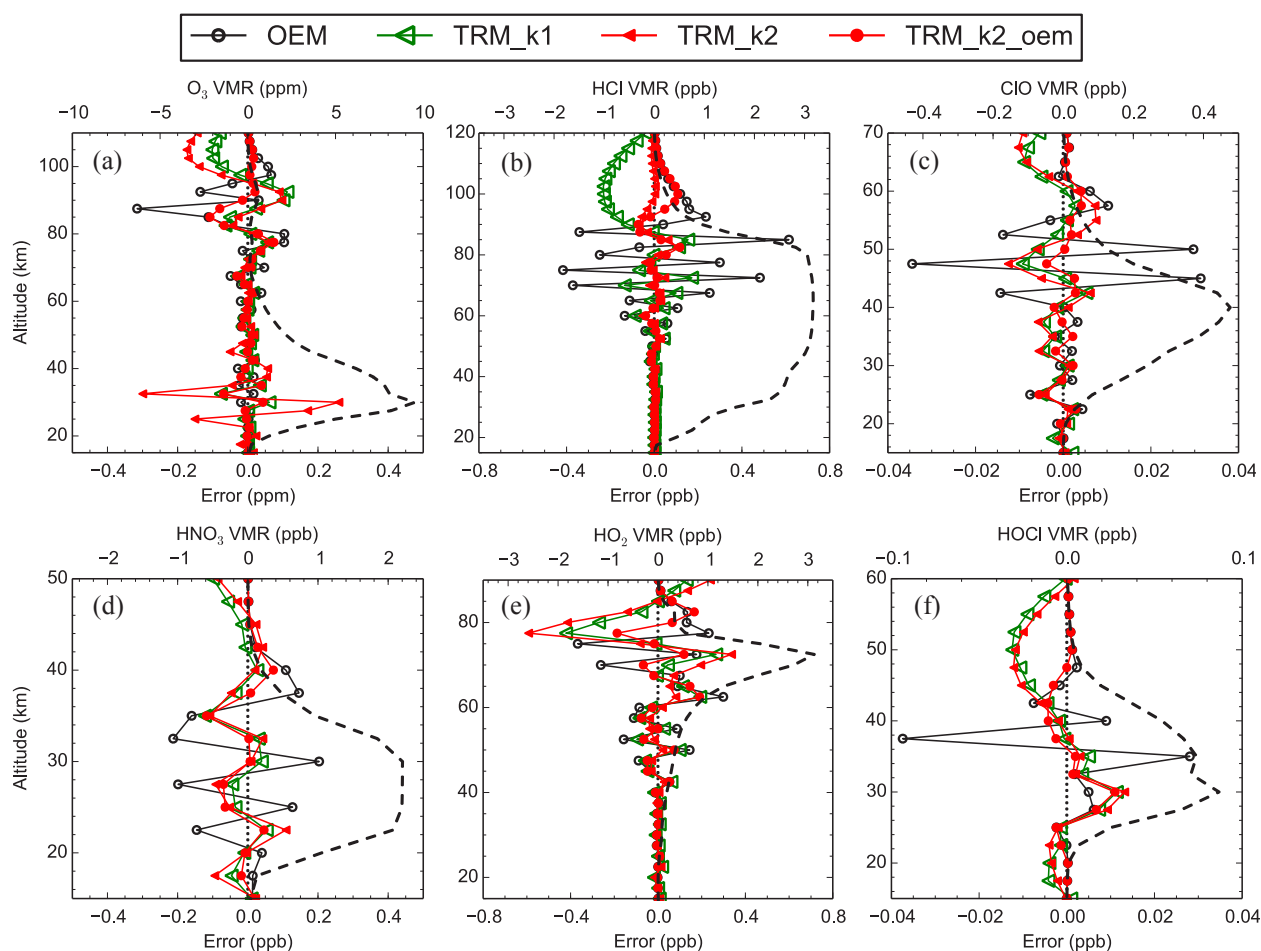


図 2 逆解析結果の誤差(実線)と真値(破線)の高度プロファイル. (a)  $\text{O}_3$  (b)  $\text{HCl}$  (c)  $\text{ClO}$  (d)  $\text{HNO}_3$  (e)  $\text{HO}_2$  (f)  $\text{HOCl}$ .

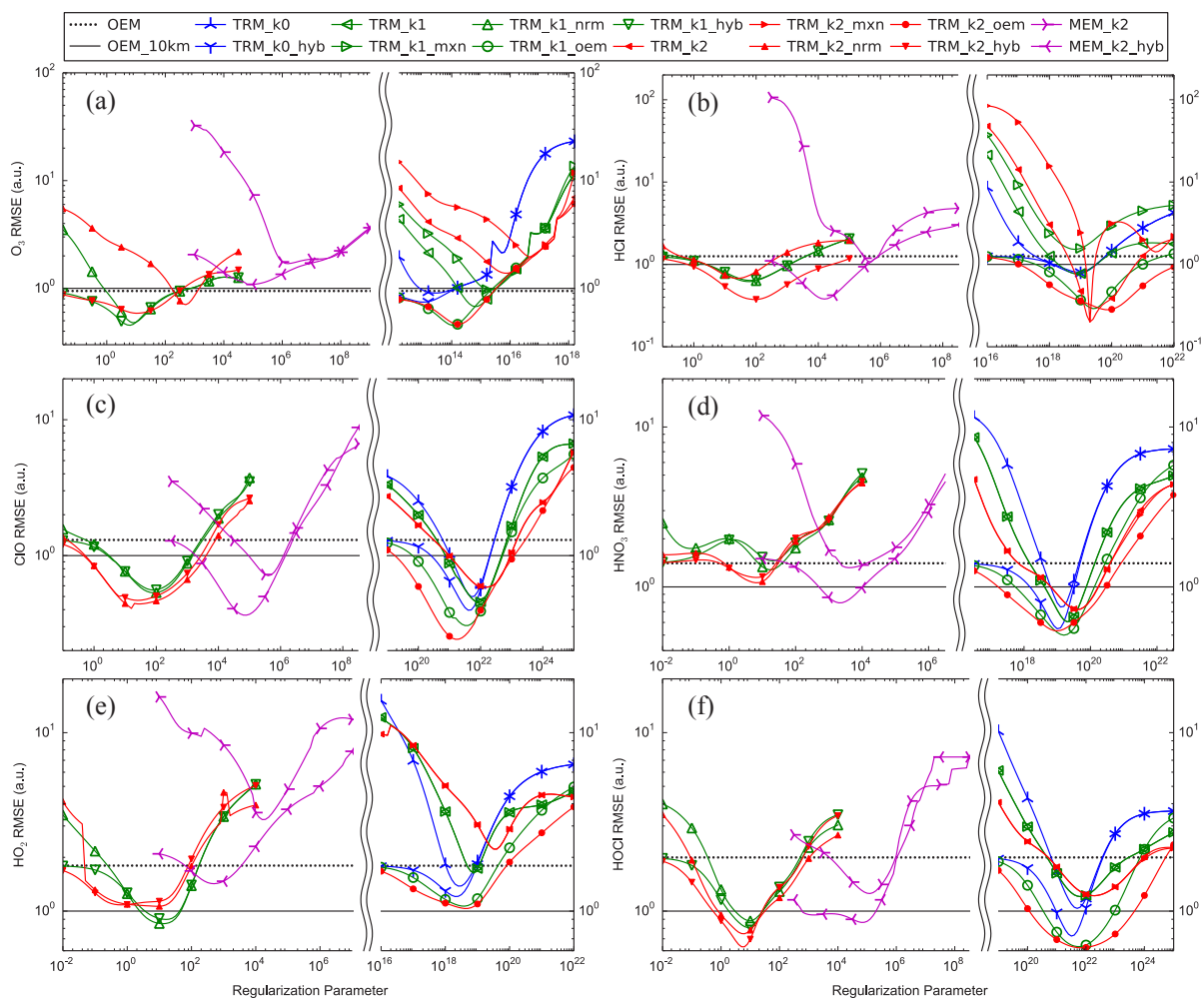


図3 逆解析誤差の正則化パラメータ依存性(OEM\_10km の値で規格化, 各パネルの意味は図2と同じ)

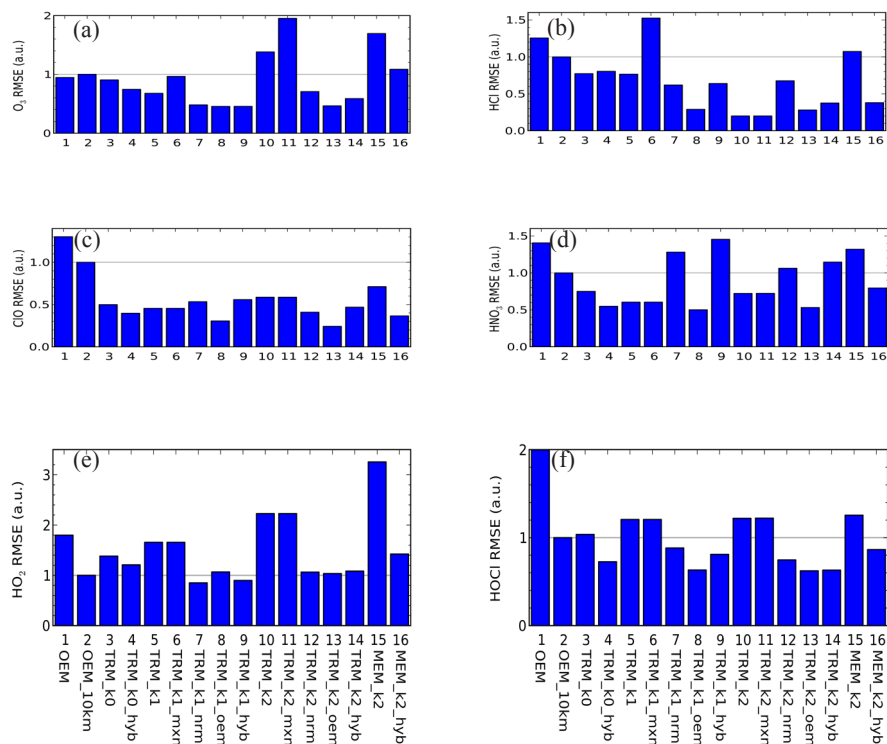


図4 逆解析誤差の最小値(OEM\_10km の値で規格化, 各パネルの意味は図2と同じ)

#### 4. SMILES の逆解析手法

SMILES L2 データ処理では、比較的誤差が小さく、正則化パラメータの設定が楽なハイブリッド法 (TRM\_k2\_hyb) を逆解析に使用する。このハイブリッド法では以下のような正則化項を使用する。

$$c_{\text{hyb}}(\mathbf{x}) = (\mathbf{x} - \mathbf{x}_a)^T \mathbf{S}_a^{-1} (\mathbf{x} - \mathbf{x}_a) + (\mathbf{x} - \mathbf{x}_a)^T \mathbf{L}_2^T \mathbf{S}_r^{-1} \mathbf{L}_2 (\mathbf{x} - \mathbf{x}_a) \quad (20)$$

$$\mathbf{S}_r^{-1} = \mathbf{M}_a \mathbf{S}_a^{-1} \quad (21)$$

対角行列  $\mathbf{M}_a$  の対角要素は分子種ごとの正則化パラメータである。正則化パラメータの値は基本的に 10 とするが、解の振動を強く抑える必要がある HOCl および  $\text{HNO}_3$  に対してはそれぞれ 50 および 1000 とする。

逆問題の解は式(20)を式(5)に代入して得られる以下の式を反復計算することで求められる。

$$\delta \mathbf{x}_i = [\mathbf{K}_i^T \mathbf{S}_y^{-1} \mathbf{K}_i + \mathbf{S}_c^{-1} + \lambda \mathbf{D}_a^{-1}]^{-1} [\mathbf{K}_i^T \mathbf{S}_y^{-1} (\mathbf{y} - \mathbf{F}(\mathbf{x}_i)) - \mathbf{S}_c^{-1} (\mathbf{x}_i - \mathbf{x}_a)] \quad (22)$$

$$\mathbf{S}_c^{-1} = \mathbf{S}_a^{-1} + \mathbf{L}_2^T \mathbf{S}_r^{-1} \mathbf{L}_2 \quad (23)$$

解が収束したことの目安は解の変動が期待される誤差と同程度になることであり、式(24)の値が状態ベクトルの要素数  $n$  と同程度かどうかで判断できる。

$$\delta \mathbf{x}_i^T \mathbf{S}_x^{-1} \delta \mathbf{x}_i = \delta \mathbf{x}_i^T [\mathbf{K}_i^T \mathbf{S}_y^{-1} (\mathbf{y} - \mathbf{F}(\mathbf{x}_i)) - \mathbf{S}_c^{-1} (\mathbf{x}_i - \mathbf{x}_a)] \quad (24)$$

$$\mathbf{S}_x^{-1} = \mathbf{K}_i^T \mathbf{S}_y^{-1} \mathbf{K}_i + \mathbf{S}_c^{-1} \quad (25)$$

導出解  $\hat{\mathbf{x}}$  の特性を評価するために、真値  $\mathbf{x}$  に関する偏微分  $\mathbf{A} = \partial \hat{\mathbf{x}} / \partial \mathbf{x}$  (Averaging Kernel Matrix, 以降単に AKM と表記) が有用である。AKM の行ベクトル (Averaging Kernel)  $\partial \hat{x}_i / \partial x_j$  は高度グリッド  $i$  の導出解  $\hat{x}_i$  が各高度グリッドの真値にどれほど感度があるかを表し、AKM の列ベクトル  $\partial \hat{\mathbf{x}} / \partial x_j$  は高度グリッド  $j$  の真値  $x_j$  の変化が各高度グリッドの導出解にどれほど影響を与えるかを表す。理想的な逆解析では AKM は単位行列となるが、現実的には Averaging Kernel の半値全幅で表される有限の高度分解能を持ち、感度が過小 (過大) である場合には Averaging Kernel の和は 1 より小さく (大きく) なる。SMILES では以下の式を使って収束後の結果から AKM の見積もりを行っている。

$$\mathbf{A} = \mathbf{S}_x \mathbf{K}^T \mathbf{S}_y^{-1} \mathbf{K} \quad (26)$$

エラー比  $r_e = \sqrt{[\mathbf{S}_x]_{ii} / [\mathbf{S}_c]_{ii}}$  は正則化によって与えられた観測以外の情報が解の正確さに寄与する割合を表し、これまで SMILES では  $r_e$  が 0.5 より小さい高度を有効と判定してきた。しかしながら、OEM と TRM のハイブリッド法を導入した場合、OEM に高度相関を入れないと有効高度範囲が狭く見積もられる傾向があった。そこで、有効高度範囲の判定基準を Averaging Kernel の和が 0.6 以上となるように変更した。図 5(b)に見られるように、TRM の正則化が強く働くと導出解の高度から遠く離れた高度に感度を持つことがあるため、Averaging Kernel の和を取る際は各高度グリッド  $\pm 2$  グリッドの範囲 (合計 5 グリッド) のみを考慮することにした。これにより OEM の高度相関を省き、TRM のみで解の振動を制御することができるようになった。

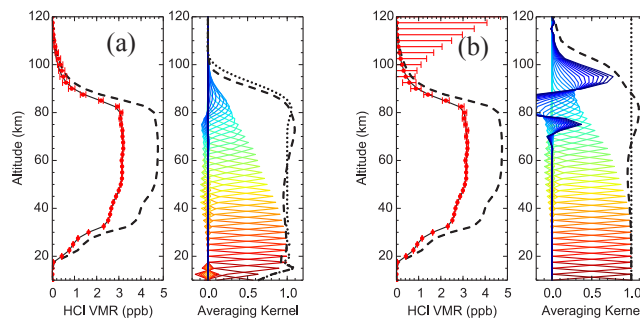


図5 逆解析で得られた HCl の導出解と Averaging Kernel の例。(a) TRM\_k2\_oem (b) TRM\_k2\_mxn . 各アルゴリズムについて、左パネルの実線は真値、破線は先験値および初期値、点と誤差棒は導出解およびその不確実性を示す。右パネルの実線は Averaging Kernel, 点線は Averaging Kernel の和, 破線は各高度グリッド  $\pm 2$  グリッドの範囲における Averaging Kernel の和を示す。

OEM から OEM+TRM のハイブリッド法に変更したことによる計算量の増加は Forward 計算等に必要な計算量に比べて僅かであり、SMILES L2 データ処理の所要時間への直接的影響は無視できる。ただし、高度グリッドの間隔を 2.5 km に統一したことで状態ベクトルのサイズが増加したため、処理時間が少し増加した。

## 5. まとめ

SMILES L2 データ処理における逆解析を高度化するため、これまで用いられてきた Optimal Estimation Method (OEM) に加え、Tikhonov Regularization Method (TRM)、Maximum Entropy Method (MEM) についての調査を行った。6つの分子種 (SMILES 観測 Band A の O<sub>3</sub>, Band B の HCl, Band C の ClO, Band C の HNO<sub>3</sub>, Band B の HO<sub>2</sub>, Band A の HOCl) について、シミュレーションを用いて様々な正則化法を比較した結果、多くの場合に高度相関を入れた OEM よりも OEM と TRM を組み合わせた方が良い結果が得られることが分かった。これは、OEM では先験値と観測値が与えられた条件で最適な解が得られるものの、OEM の高度相関では正則化の強度が調整できないために解の振動を十分に抑えられないのに対し、TRM では正則化パラメータの調整によって振動が効果的に抑えられるためと考えられる。MEM に関しては、適当な先験値がある場合は OEM や TRM に比べて誤差が大きくなることが確かめられた。

OEMとTRMのハイブリッド法をO<sub>3</sub>, HCl, HNO<sub>3</sub> の 3 分子種に適用したSMILES L2 v2.3, v2.4 はすでにデータが公開され、様々な研究で使用されている<sup>12,13)</sup>。現在はOEMとTRMのハイブリッド法を全ての分子種に適用したSMILES L2 v3.0 の処理中であり、このデータは 2013 年の秋にリリースされる予定である\*。

## 参考文献

- 1) K. Kikuchi et al., "Overview and early results of the Superconducting Submillimeter-Wave Limb-Emission Sounder (SMILES)," *J. Geophys. Res.*, vol. 115, no. D23306, 2010.
- 2) K. Imai et al., "Validation of ozone data from the Superconducting Submillimeter-Wave Limb-Emission Sounder (SMILES)," *J. Geophys. Res.*, vol. 118, no. 11, pp. 5750 - 5769, 2013.
- 3) T. Sakazaki et al., "Diurnal ozone variations in the stratosphere revealed in observations from the Superconducting Submillimeter-Wave Limb-Emission Sounder (SMILES) on board the International Space Station (ISS)," *J. Geophys. Res.*, vol. 118, no. 7, pp. 2991 - 3006, 2013.
- 4) C. D. Rodgers, *Inverse Methods for Atmospheric Sounding: Theory and Practice*, World Scientific, 2000.
- 5) C. Takahashi, S. Ochiai and M. Suzuki, "Operational retrieval algorithms for JEM/SMILES level 2 data processing system," *J. Quant. Spectrosc. Radiat. Transfer*, vol. 111, no. 1, 2010.
- 6) C. Mitsuda et al., "Current status of level 2 product of Superconducting Submillimeter-Wave Limb-Emission Sounder (SMILES)," *Proc. SPIE 8176*, 2011.
- 7) A. Doicu, T. Trautmann, F. Schreier, *Numerical Regularization for Atmospheric Inverse Problems*, Springer, 2010.
- 8) M. Ridolfi and L. Sgheri, "A self-adapting and altitude-dependent regularization method for atmospheric profile retrievals," *Atmos. Chem. Phys.*, vol. 9, pp. 1883-1897, 2009.
- 9) J. Steinwagner, G. Schwarz and S. Hilgers, "Use of a Maximum Entropy Method as a Regularization Technique during the Retrieval of Trace Gas Profiles from Limb Sounding Measurements," *J. Atmos. Oceanic Technol.*, vol. 23, no. 12, pp. 1657-1667, 2006.
- 10) N. J. Livesey and W. Van Snyder, "EOS MLS Retrieval Processes Algorithm Theoretical Basis," 2004.
- 11) A. Kunz, L. L. Pan, P. Konopka, D. Kinnison and S. Tilmes, "Chemical and dynamical discontinuity at the extratropical tropopause based on START08 and WACCM analyses," *J. Geophys. Res.*, vol. 116, no. D24, 2011.
- 12) R. A. Stachnik et al., "Stratospheric BrO abundance measured by a balloon-borne submillimeterwave radiometer," *Atmos. Chem. Phys.*, vol. 13, pp. 3307 - 3319, 2013.
- 13) A. K. Smith et al., "Satellite observations of ozone in the upper mesosphere," *J. Geophys. Res.*, vol. 118, no. 11, pp. 5803 - 5821, 2013.

\* <http://darts.isas.jaxa.jp/iss/smiles/>



# 小型望遠鏡を用いた突発天体の自動広視野サーベイ

前原裕之<sup>\*1</sup>

## Automated Wide-field Survey for Transient Objects with a Small Telescope

Hiroyuki Maehara<sup>\*1</sup>

### Abstract

We report an outline of the automated wide-field optical survey system for bright transients and variable stars using a small CCD camera and a telephoto lens. Our survey system named “Kiso/Kyoto Wide-field Survey (KWS)” has a field-of-view of 5 degrees x 7.5 degrees and could cover a survey area of about 100 degrees x 120 degrees per night (8 hours). We achieve a photometric precision of <5% for stars with a V magnitude of between 5 and 11. We also developed a web-based tool for data retrieval. We discovered more than 100 new variable stars and obtained pre-discovery images of 2 new transients with the KWS.

### 概要

明るい突発天体および変光星の探査を目的として、可視光で広い視野を自動的にサーベイするシステムを小型の CCD カメラと写真レンズを用いて開発したので、その概要を報告する。本システムは  $5^{\circ} \times 7.5^{\circ}$  の視野を持ち、1 晩で  $100^{\circ} \times 120^{\circ}$  の範囲を観測することができる。V 等級で 5-11 等の天体について、5% 以下の測光精度を達成している。また、サーベイデータを取得できる Web ベースのプログラムの開発も行った。本システムを用いて、100 以上の未知変光星と、2 つの突発天体の発見前の様子を観測することができた。

Keyword: sky survey, variable star

### 1. はじめに

新星や超新星、ガンマ線バーストといった突発天体はいつどこに出現するか予測できず、また一般に増光後短時間で暗くなってしまう。このため、それらの天体の性質や増光の物理機構を探するためには、(1) 広い範囲をモニターし、(2) 早期に検出してアラートを出し、(3) より大きな望遠鏡を用いた詳細な観測（分光など）に繋げる必要がある。また、突発天体をターゲットとした、広視野・高頻度のサーベイ観測では膨大な数の天体の測光データが得られるため、本来の目的以外の様々な天体の中長期的な変動の研究にも非常に有用なデータとして利用することができる。近年では、このような突発天体や変光天体をターゲットとして、広い範囲を可視光でサーベイするプロジェクトとして、All Sky Automated Survey (ASAS)<sup>1)</sup>、Pi of the sky<sup>2)</sup>、Catalina Sky Survey<sup>3)</sup> などが多くの成果を上げている。

突発天体や変光天体を検出する方法には大きく分けて、(1) 基準画像を作成し、撮影画像と基準画像の差分から検出する方法、(2) 画像中の星をすべて検出して測光してデータベース化し、個々の天体の測光データから検出する方法とがある。前者は超新星サーベイのように、銀河の非常に近い場所に出現した天体であっても、超新星が写っている撮影画像から写っていない基準画像を引き算することで、超新星を検出することができるという利点がある。また、後者はすべての天体を測定するため、データ量が非常に多くなるので、突発天体だけを効率的に検出する目的からするとやや不利な面もあるが、写っているすべての天体の測光データが得られるという利点がある。

突発天体の増光メカニズムの解明には、前述のとおり検出後の高分散分光や高時間分解能の連続分光観測など、詳細な観測が不可欠である。しかし、大きな望遠鏡を用いた深いサーベイでないと検出できない暗い突発天体では、そのような観測を行うことは 8-10m クラスの望遠鏡であっても難しい。一方で、中小口径の望遠鏡で詳細な観測が行える明るい突

---

<sup>\*1</sup> 東京大学大学院理学系研究科附属天文学教育研究センター木曾観測所  
(Kiso Observatory, Institute of Astronomy, School of Science, The University of Tokyo)

発天体は発生頻度が極めて小さいと予想されるため、空の広い範囲を監視する必要がある。本研究では、明るい突発天体のうち特に突発的な短時間の増光を起こす晩期型星であるフレア星の検出および、大きな黒点を持つフレア星の自転による光度変化を観測することを目的に、市販の CCD カメラと写真用のレンズを用いた広視野サーベイシステムを開発した。突発的な増光だけでなく、増光を起こす天体の普段の明るさの変化も同時にとらえる必要があるため、本研究では前述の 2 つの方法のうち、(2) の全天体の測光カタログを作成する方法を用いた。2010 年から焦点距離 50mm のレンズと CCD カメラを使った WindowsPC ベースの自動観測と Pyraf, SExtractor, MySQL 等を用いた解析システムを立ち上げ、このシステムで 2010 年 12 月から 2012 年 3 月まで (Phase I), その後 2012 年 3 月から 2013 年 4 月まではカメラのみ焦点距離 105mm のレンズとより大面積のチップの CCD カメラへ更新したシステム (Phase II), さらに 2013 年 5 月以降は自動撮像システムを LinuxPC ベースとして複数台のカメラの制御に対応したシステム (Phase III) を用いて、定常的な観測を行なった (現在も観測継続中)。本論文では、第 2 節で自動撮像、解析システムの概要および処理のフローを説明し、第 3 節で測光精度や解析パイプライン等の処理時間などの性能評価および、検出された変動天体を紹介する。第 4 節ではまとめと今後の課題について述べる。

## 2. 構築システム

### 2.1. 自動撮像

本研究では、観測を完全に自動化するため、あらかじめ指定した座標を順次撮影するシステムを制作した。使用した赤道儀や撮像・望遠鏡制御ソフトウェア等を表 1 にまとめた。また、その晩に観測できる座標リストを自動的に生成し、使用したソフトウェアで利用できるように、CSV 形式で座標リストを出力するプログラムを Python 言語で作成した。このプログラムでは、月と撮影する座標の角距離を計算し、月の 20 度以内の座標は観測しないようになっている。なお、いずれのシステムも Bessell V バンドのフィルターを組み込み撮像している。

表 1 ハードウェア、ソフトウェアの組み合わせ

	カメラ+レンズ	赤道儀	カメラ / 望遠鏡制御ソフトウェア	FOV	積分時間
Phase I (2010/12-2012/03)	f=50mm F/2.0 + SBIG ST-7E	VixenSP+Skysensor 2000PC	StellaGear <sup>1</sup> (WindowsPC)	5.1° × 7.7°	30 秒
Phase II (2012/03-2013/04)	f=105mmF/2.0 + SBIG ST-8XME	同上	同上	5.0° × 7.5°	15 秒
Phase III (2013/05-)	同上	Vixen SXW	Nightview <sup>2</sup> / 自作 Python script (LinuxPC)	同上	同上

### 2.2. 解析パイプライン

解析パイプラインは Python および shell script で作成した。一次処理を行う image\_reduction.py, WCS の書き込みおよび測光を行う autophoto の 2 つからなる。

- image\_reduction.py

FITS ヘッダーの DATA\_TYP を見て、OBJECT になっている画像の場合、ヘッダーの EXPTIME に対応するダークフレームおよび予め登録されているフラットフレームを用いて一次処理を行う。処理には PyRAF<sup>3</sup> を用いた。

- autophoto

天体検出に SExtractor<sup>4</sup>, WCS 書き込みに WCSTools<sup>5</sup> および SCAMP<sup>6</sup> を用いる。処理は以下の手順で実行し、最終的に天体の座標 (天球座標および CCD チップ上の座標) と機械等級、等級エラーをファイルとして出力する。

① SExtractor で星を検出し、CCD チップ上の座標および機械等級、等級のエラー (X\_IMAGE, Y\_IMAGE, MAG\_

1 <http://www.astroarts.co.jp/products/stlgear/>

2 <http://www.physics.muni.cz/mb/nightview/>

3 [http://www.stsci.edu/institute/software\\_hardware/pyraf](http://www.stsci.edu/institute/software_hardware/pyraf)

4 <http://www.astromatic.net/software/sextractor>

5 <http://tdc-www.harvard.edu/wcstools/>

6 <http://www.astromatic.net/software/scamp>

APER(1), MAGERR\_APER(1)) をテキスト形式で出力する。

- ② 前の手順で作成したファイルを WCSTools の imwcs コマンドで, Tycho2 カタログ<sup>4)</sup>を用いて WCS のパラメータを決め, FITS ヘッダーに追加する。
- ③ 再度, SExtractor で星を検出し, XWIN\_IMAGE, YWIN\_IMAGE, ERRWIN\_IMAGE, ERRAWIN\_IMAGE, ERRBWIN\_IMAGE, ERRTHRAWIN\_IMAGE, FLUX\_AUTO, FLUXERR\_AUTO, FLUX\_APER, FLUXERR\_APER の各値を FITS-LDAC 形式<sup>7)</sup>で出力する。
- ④ SCAMP を用いて前段で作成した FITS-LDAC 形式のファイルから再度 WCS のパラメータを求める。基準星の位置のカタログには, 固有運動分を補正し, 観測年の位置 (ICRS 座標) に変換した Tycho2 カタログを用いる。これは, Hipparcos/Tycho カタログ元となった観測から既に 20 年以上が経過し, 本研究で対象としている 5-6 等星のように, 比較的近距离にあり見かけの固有運動が大きな天体 (100-1000 mas 程度) では, 数秒から数十秒程度カタログ位置からずれていることが見込まれ, 50-100mm 程度の焦点距離のレンズであっても, その影響が無視できないためである。SCAMP を用いて求めた WCS のパラメータを再度 FITS ヘッダーに書き込む。
- ⑤ WCS が書き込まれた画像に写っている天体の明るさを SExtractor を用いて測定し, 天体の座標 (天球座標および CCD チップ上の座標) と機械等級, 等級エラーを出力させる。等級は aperture size を固定した aperture 測光の等級 (MAG\_APER(1)) を用いる。

## 2.3. 測光・解析用データベース

測光データの保存や, 等級ゼロ点の決定, 既知天体との照合のためのデータベースを MySQL を用いて作成した。データベースには以下のテーブルを含む。

- Frame\_Info

frame\_id を主キーとするテーブルで, 画像のファイル名, 撮影日時や露出時間, フィルター, pixel scale などの FITS ヘッダーに含まれる情報の他, 等級ゼロ点や等級ゼロ点の誤差, 等級ゼロ点の決定に使用した標準星の数, 視野中心の airmass を保管するテーブル。

- Object\_Info

oid を主キーとするテーブルで, 検出した画像の frame\_id, SExtractor で検出した全天体の天球座標, CCD チップ上の xy 座標, 機械等級, 機械等級の誤差, airmassなどを保管するテーブル。frame\_id をもちいて, Frame\_Info テーブルに保管されている情報との対応をとることができる。天体の位置については後述する通り, Yamauchi (2011)<sup>5)</sup>の方法でカタログマッチングを行うため, 以下の方法で求めた  $x, y, z$  の値も保持している。

$$x = \cos \alpha \cos \delta \quad (1)$$

$$y = \sin \alpha \cos \delta \quad (2)$$

$$z = \sin \delta \quad (3)$$

- HIP

Hipparcos カタログ<sup>6)</sup>のデータを保持している。ただし, 天球座標はオリジナルの座標ではなく, 固有運動を補正した 2011 年の ICRS 座標とし, 式 (1) ~ (3) から計算した  $x, y, z$  の値も保持する。

- TYC

Tycho2 カタログのデータを保持している。HIP テーブルと同様に座標は固有運動を補正した 2011 年時点の ICRS 座標で,  $x, y, z$  の値も保持する。

- GCVS

変光星総合カタログ (GCVS+NSV)<sup>7)</sup>と ASAS の変光星カタログ<sup>8)</sup>のデータを保持する。座標はオリジナルのカタログの値をそのまま使用し,  $x, y, z$  の値も保持する。また, 本研究で新たに変光していることが分かった天体についても, 逐次追加している。

## 2.4. データベースへの登録および等級ゼロ点の決定

一次処理および測光の終わったファイルおよび検出天体の情報は以下の手順でデータベースに登録され, Hipparcos カタログの V 等級を基準に用いて各画像ごとに等級ゼロ点を決定する。

<sup>7</sup> [http://marvinweb.astro.uni-bonn.de/data\\_products/THELIWWW/index.html](http://marvinweb.astro.uni-bonn.de/data_products/THELIWWW/index.html)

## ① 画像ファイルの登録 : insert\_file.py

WSC を書き込み済みの FITS ファイルのファイル名および FITS ヘッダの内容、撮影日時と視野中心の座標から計算した airmass を Frame\_Info テーブルに登録する。

## ② 検出天体の登録 : scat2id.py

SExtractor で検出・測光した天体の座標、機械等級などが保存されているファイルの内容および、各天体の airmasse を計算し、それらを Object\_Info テーブルに格納する。

## ③ カタログマッチングと等級ゼロ点の決定 : identify\_sql.py

Hipparcos カタログ (HIP テーブル) から等級ゼロ点決定用の標準星として、測光精度が十分確保できる明るさでかつ CCD のリニアリティが保証される範囲内にあり、変光しておらず、極端に赤くない天体を選択する。また、近くに星があり重なって分離できないと正確な測定ができないため、近くに別な星がある天体は標準星として用いないようにする必要がある。本システムでは以下の A~E 全ての条件を満たす天体を選択し、標準星として用いた。

A)  $5.5 < V_{\text{mag}} < 9.5$

B)  $V_{\text{scatter}} < 0.02$

C)  $B - V < 1.2$

D) 3 pixel 以内に Tycho2 カタログの天体がない

E) 変光星カタログに登録されている天体ではない

実際には、上記の条件を満たす HIP テーブルの天体を含む view を HIP\_obs という名称で作成する。その後、Object\_Info に含まれる検出天体と、HIP\_obs の天体を下記のような SQL クエリーでマッチングした結果を HIP\_Reference という view として作成する。

```
CREATE OR REPLACE VIEW HIP_Reference AS SELECT h.hid as hid, o.obj_id as obj_id, (h.vmag - o.phot_i_mag)
as diffmag, o.phot_i_mag_err as phot_i_mag_err, h.vmag as vmag, h.bv as bv, h.flg_tmp as flg_tmp from HIP_obs h,
Object_Info o WHERE h.flg_tmp=1 AND o.frame_id=%s AND o.flg_ok=1 and (1.0 - o.x*h.x + o.y*h.y + o.z*h.z)<%s and
o.phot_i_mag<0.0 and o.flg_ok=1
```

※ %s にはマッチングする画像の frame\_id と、同一天体であるとみなす閾値 ( $\sigma$ ) が入る。

カタログマッチングは、検出天体の座標方向の単位ベクトルと基準星方向の単位ベクトルの内積と同一天体であるとみなす閾値 ( $\sigma$ ) を比較することで行っている。検出天体の座標方向の単位ベクトル  $\vec{o}$ , 基準星となる Hipparcos カタログの天体の座標方向の単位ベクトル  $\vec{h}$ , 2つの単位ベクトルのなす角度を  $\theta$  とすると、この2つの単位ベクトルの内積  $\vec{o} \cdot \vec{h} = \cos \theta$  は  $\theta$  が十分に小さければ  $\vec{o} \cdot \vec{h} \approx 1 - \theta^2/2$  となる。したがって、検出天体とカタログ天体が閾値よりも接近している条件は、

$$1 - \vec{o} \cdot \vec{h} \approx \frac{\theta^2}{2} < \sigma \quad (4)$$

となる。本システムでは、経験的に検出天体の測定位置とカタログ位置の角度差  $\theta$  が 0.7 pixel に相当する角度よりも小さい場合に同一天体とみなすように設定した。1 pixel あたりの角度を  $s [\text{arcsec}] = 4.8 \times 10^{-6} [\text{rad}]$  とすると、閾値は  $\sigma = 5.9 \times 10^{-12} \text{ rad}^2$  となる。

次に、SExtractor で測定した機械等級と対応する Hipparcos 天体のカタログにある V 等級の差の平均値および標準偏差を求める。いったん求めた平均値および標準偏差を使い、(機械等級 - カタログ等級) が  $3\sigma$  の範囲に収まっていない天体を除外し、再度 (機械等級 - カタログ等級) の平均値および標準偏差を計算し、この値を等級ゼロ点およびその誤差として、Frame\_Info テーブルに保存する。天体の等級は Frame\_Info テーブルの等級ゼロ点の値と、Object\_Info テーブルの機械等級を足すことで求められる。

## 2.5. Web インターフェース

観測データのチェックや公開のため、Web ブラウザから特定の天体を検索してデータおよび光度曲線を表示する CGI を作成した。入力フォームに GCVS テーブルにある天体名または、天体の座標をもとにした ID (例えば天体の赤経を

01h23m45s, 赤緯を  $-10^{\circ} 20' 30''$  とすると, ID は 012345-1020.5 となる) を入力することで, その天体の位置に対応する検出天体のデータおよび光度曲線を表示する (図 1). また, 天体の oid のリンクをクリックするとその天体を中心とする  $300 \text{ pixel} \times 300 \text{ pixel}$  の範囲を JPEG 画像として表示することができる. この Web フォームは <http://kws.cetus-net.org/~maehara/VSdata.py> で公開している.

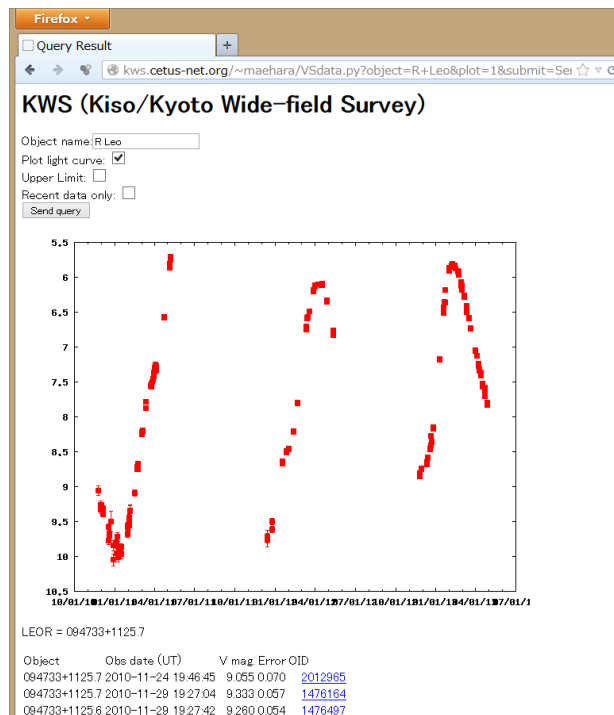


図 1 データ表示用の Web フォーム

### 3. 運用状況と性能評価

2010 年 12 月の運用開始から 2013 年 4 月末までの 266 夜の観測で 103,053 枚の画像を撮像し, 延べ 1 億 9694 万天体を検出してデータベースに登録した. 図 2 に 2010 年 12 月から 2013 年 4 月までの観測夜数の分布を示す. 現在のサーベイは子午線付近の赤緯  $-33^{\circ} \sim +66^{\circ}$  の範囲を南から撮像し, 続いて赤経を 28 分ずらして再び赤緯  $-33^{\circ} \sim +66^{\circ}$  を撮像するようになっており, 一晩 (8 時間) で  $100^{\circ} \times 120^{\circ}$  程度の範囲を撮像することができる.

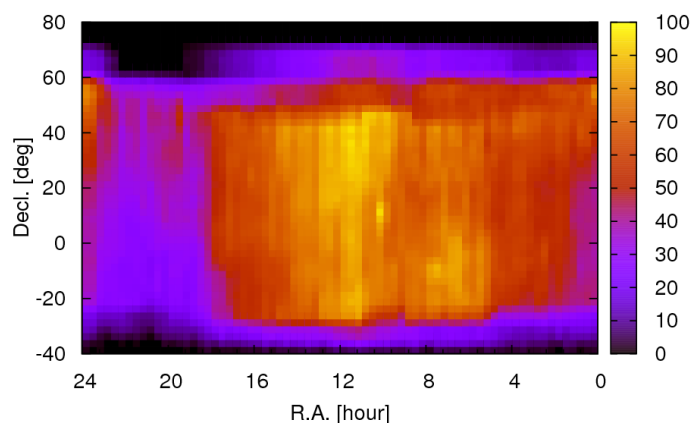


図 2 天球面上での観測夜数の分布 (2010 年 12 月 - 2013 年 4 月)



### 3.1. 測光精度

本研究で作成した測光データベースを利用すると、検出された天体ごとに位置の誤差範囲内のある天体を同定して、その測光値の時系列データを得られる。変光していない天体であれば、測光値は理想的には一定であり、そのばらつきは測光精度を表すと考えることができる。図3, 4に各天体の明るさと、測光値の標準偏差の散布図を示す。図3は2012年3月から2012年11月までの観測で得られたもので、表1のPhase IIの機器を京都大学花山天文台に設置して観測を行ったデータから作成したもので、図4は同じ機材を東京大学木曽観測所に移設して、2012年12月から2月までの観測で得られたデータから作成したものである。花山天文台(京都市内)と比べると木曽観測所のほうが空が暗いため(背景の明るさが1/30程度)、花山天文台に設置していた時には5%以下の標準偏差となる明るさは5-10等であったのに対し、木曽観測所設置時には5-11等で5%以下の標準偏差となった。5等よりも明るいところではCCDの飽和のために正しい測定値が得られていない。

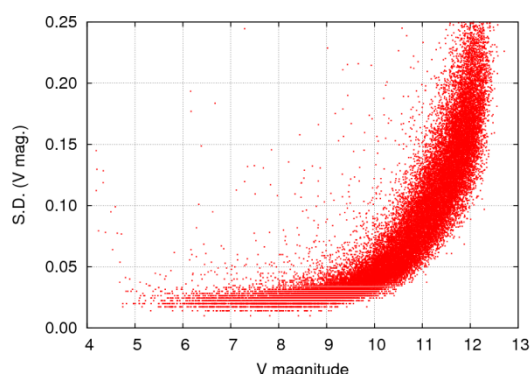


図3 天体の明るさと測光値の標準偏差  
(京都; 2012/03-2012/11)

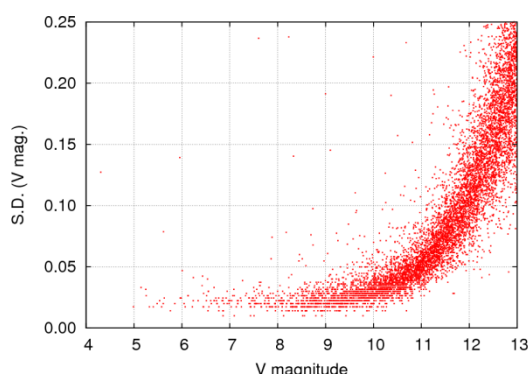


図4 天体の明るさと測光値の標準偏差  
(木曽; 2012/12-2013/02)

### 3.2. 処理速度

突発天体の観測には検出した画像を即時に検出する必要がある。表2に本システムで撮像したデータを処理し、データベースへの登録および過去のデータとの比較を行って、新天体および新変光星候補をリストアップするまでにかかる時間をまとめた(すべてシングルスレッドで処理)。解析およびデータベースには以下の環境を用いた。

- ハードウェア : Intel Xeon E5-2620 (2.0GHz), 32GB memory, Intel 335 240GB SSD
- OS: Scientific Linux 6.3 (kernel 2.6.32) x86\_64
- データベース : MySQL 5.1.69
- 解析パイプライン : IRAF 2.16, PyRAF 2.1, WCSTools 3.8.7, SExtractor 2.8.6, SCAMP 1.7.0

表2 各領域のデータ処理にかかる所要時間

視野中心座標	14h48m, +16°	16h12m, +12°	17h30m, -19°
検出天体数	2588	3520	4871
一次処理 + WCS	6.4 [sec]	7.6 [sec]	8.7 [sec]
DB insert (file/object)	1.1/2.2 [sec]	1.1/2.7 [sec]	1.1/3.2 [sec]
等級ゼロ点決定	1.5 [sec]	1.8 [sec]	2.2 [sec]
新天体・変光星検出	397 [sec]	541 [sec]	776 [sec]

観測領域によって検出天体数が異なるため、1視野の天体数が増加するにしたがって、処理の所要時間も増加する。積分時間は15秒なので、撮像後次のフレームの撮影が終わるまでの間に、一次処理とデータベースへの登録、および等級ゼロ点の決定までは終了する。また、おおむね10-15分以内には新天体と新変光星のチェックまで終了し、結果をHTMLファイルとして出力できる。

### 3.3. 変動天体の検出例

2013 年 4 月までに観測したデータから、本研究のデータから変光が検出された天体のうち、既知の変光星ではない天体は 109 個であった。これらのうち大部分が IRAS ソースであり、光度曲線から RGB/AGB 星の脈動変光星であると考えられるものが 98 天体、残りの 11 天体は食連星であると考えられる。図 5, 6 に検出された新変光星の光度曲線の例を示す。

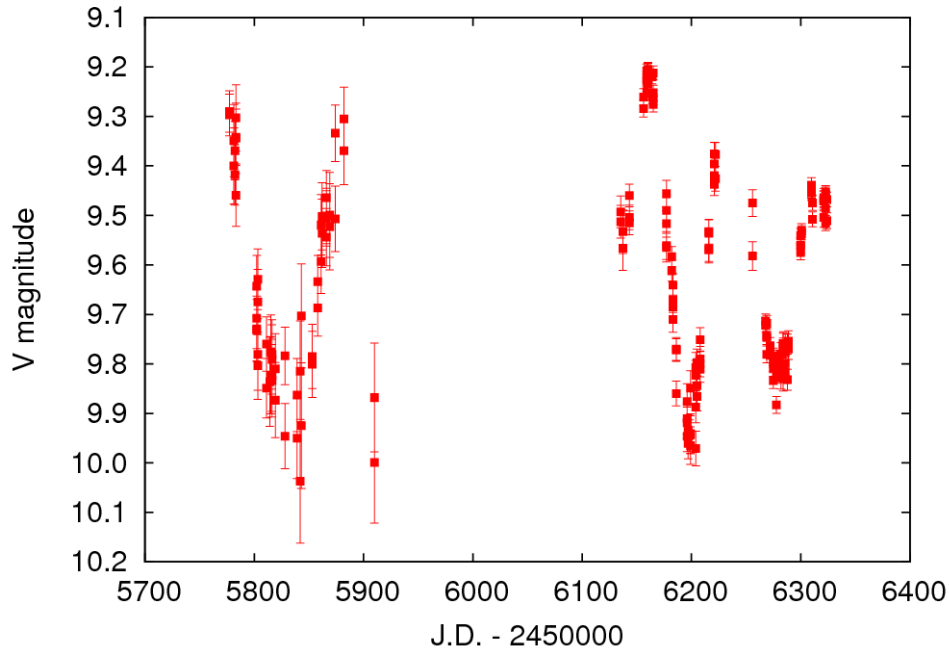


図 5 J235339+5335.3 (脈動星候補) の光度曲線

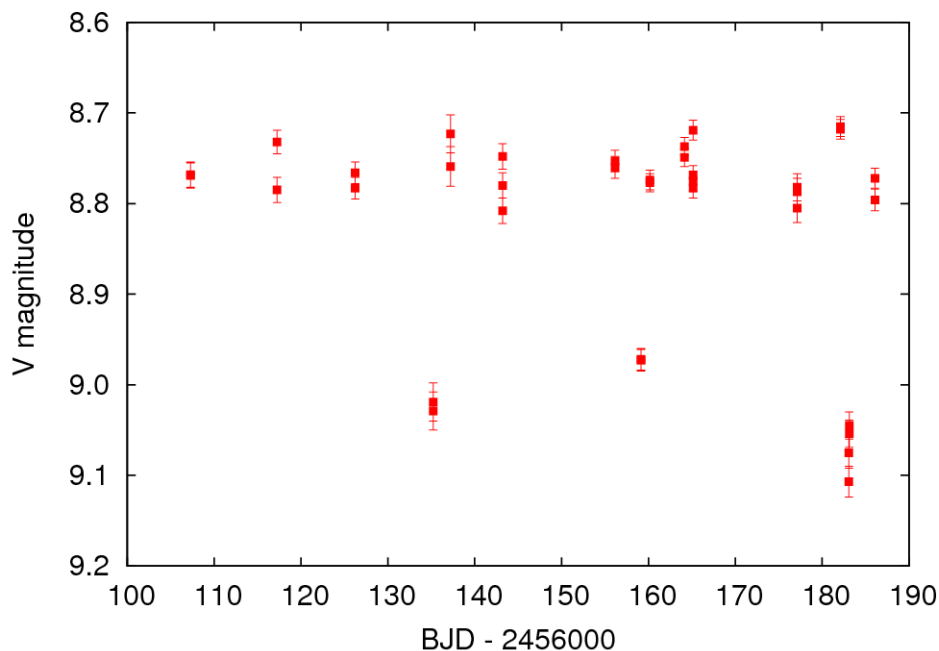


図 6 J 231202+5320.5(食連星候補)の光度曲線

新天体については、撮影地（木曽）と解析用計算機の設置場所（京都）までのデータの転送に時間がかかったために、いずれも他のサーベイでの発見後の確認となったが、矮新星 1 天体<sup>9)</sup>と新星 1 天体<sup>10)</sup>の発見前の状態が捉えられていた。2013 年 3 月以降は解析用計算機も木曽に設置し、撮影後ただちに解析が行われるようにした。

#### 4. まとめと今後の展望・課題

市販の CCD カメラおよびカメラレンズ、PC からの制御に対応した赤道儀を組み合わせることで、完全自動のサーベイシステムを作成することができた。また、得られたデータを自動解析し、データベース化することで、変動天体や突発天体の観測的研究に有用な測光データベースを構築することができた。本システムで使用した CCD カメラとレンズは約 40 万円程度、PC からの制御に対応した赤道儀も 20 万円程度であり、制御用 PC や解析・データベース用の計算機も含めても 100 万円未満のコストに抑えることができた。

現在のシステムでは、撮像後のデータ処理およびデータベースへの登録等は撮影時間内に終了しているが、突発天体・変動天体の検出に 1 視野の撮影にかかる時間の 10 倍程度の時間がかかっている。この時間の大部分は、約 2 億件のデータからなる検出天体テーブルから、該当する天体の過去のデータを取り出すために費やされており、今後もデータが増えるのに伴って処理時間が増大すると考えられる。突発天体をほぼ撮影時間と同程度の時間で検出するために、検出天体テーブルから測光値の平均値や標準偏差、最大・最小値などをカタログ化し、生の観測データではなく、作成したカタログを検索することや、適切な領域分割を行うことで、テーブルサイズを縮小する等により、検索速度を上昇させる必要がある。今後は撮像後ただちに検出が終了するようにすることで、新星・矮新星等の突発天体の他、恒星フレアのような短時間のイベントについても、暗くなる前に増光していることが判定できるように、システムの改良を進める予定である。

本システムで検出できる突発天体は 11 等よりも明るい天体であると考えられる。このような明るい突発天体の分光等による詳細な観測は口径 1-2m クラスの望遠鏡でも可能であり、例えば光・赤外線天文学大学間連携<sup>8</sup>に参加している機関の望遠鏡で、追跡観測が行えると考えられる。また、本システムは市販の機材およびオープンソースのソフトウェアを利用しているので、開発した解析パイプラインを天文アマチュアも含め広く利用できるように配布や利用のためのドキュメントの整備等を進める予定である。

#### Acknowledgment

本研究の一部は JSPS 科研費 23740150 の助成を受けて行ったものである。また、本研究で作成したシステムのテストには日本変光星研究会の清田誠一郎氏にも協力いただいた。

#### References

- 1) Pojmanski, G. “The All Sky Automated Survey. Catalog of Variable Stars. I. 0h - 6h Quarter of the Southern Hemisphere” *Acta Astronomica*, 52, (2002), pp. 397-427
- 2) Burd, A., et al., “Pi of the Sky - all-sky, real-time search for fast optical transients” , *New Astronomy*, Volume 10, (2005), pp. 409-416
- 3) Drake, A.J., et al. “First Results from the Catalina Real-Time Transient Survey” *The Astrophysical Journal*, 696, (2009), pp. 870-884
- 4) Høg, E., et al., “The Tycho-2 catalogue of the 2.5 million brightest stars” , *Astronomy and Astrophysics*, v.355, (2000), pp. L27-L30<sup>9</sup>
- 5) Yamauchi, C., “Development of 2MASS Catalog Server Kit” , *Publications of the Astronomical Society of the Pacific*, Volume 123, (2011), pp.1324-1333
- 6) Perryman, M.A.C., & ESA, *ESA Special Publication*, 1200, (1997)<sup>10</sup>
- 7) Samus, N.N., Durlevich, O.V., Kazarovets, E.V., Kireeva, N.N., Pastukhova, E.N., Zharova, A.V., et al., “General Catalog of Variable Stars (GCVS database, Version 2012 Jan)” , *CDS B/gcvs*, (2012)<sup>11</sup>
- 8) Kazarovets, E.V., Samus, N.N., Durlevich, O.V., Kireeva N.N., Pastukhova, E.N., “The 80th Name-List of Variable Stars. Part III -- RA 16h to 24h” , *IBVS No.6052*, (2013)

8 <http://oister.oao.nao.ac.jp/>

9 <http://cdsarc.u-strasbg.fr/viz-bin/Cat?I/259>

10 <http://cdsarc.u-strasbg.fr/viz-bin/Cat?I/239>

11 <http://cdsarc.u-strasbg.fr/viz-bin/Cat?B/gcvs>

- 9) vsnet-alert 15240, <http://ooruri.kusastro.kyoto-u.ac.jp/mailarchive/vsnet-alert/15240>
- 10) Nishiyama, K. et al., “NOVA CEPHEI 2013 = PNV J23080471+6046521”, Central Bureau Electronic Telegrams, 3397, 1, (2013)

# 地球磁気圏 3 次元グローバル MHD シミュレーションの磁力線 追尾可視化技術

久保田康文<sup>\*1</sup>, 村田健史<sup>\*1</sup>, 山本和憲<sup>\*1</sup>, 深沢圭一郎<sup>\*2</sup>, 坪内健<sup>\*3</sup>

## Visualization technique using a system of Magnetic Field Tracing in Global MHD simulations

Yasubumi Kubota<sup>\*1</sup>, Ken T. Murata<sup>\*1</sup>, Kazunori Yamamoto<sup>\*1</sup>, Keiichiro Fukazawa<sup>\*2</sup>, Ken Tsubouchi<sup>\*3</sup>

### Abstract

We developed a system of Magnetic Field Tracing in Global MHD simulations in order to understand magnetosphere convection. To trace magnetic flux tube with high precision, we need to process big data such as all time-series simulation data. In this paper, we show a parallel distribution visualization technique for magnetic field tracing by using NICT science cloud.

### 概要

地球磁気圏と太陽風の相互作用を理解するため、3 次元グローバル MHD シミュレーションを用いて研究が行われてきた。MHD シミュレーションデータを可視化する際に、磁力線の可視化は物理を理解する上で重要である。しかし精度よく磁力線を追尾し可視化することは、大規模な時系列データを処理する必要があるため困難であった。我々は大規模データを処理することができる NICT サイエンスクラウドを利用することで磁力線追尾可視化技法の開発を行ってきた。本論文ではこれまでの開発の現状について紹介する。

### 1. はじめに

磁気圏対流を理解することは太陽風、磁気圏、電離圏のエネルギー輸送を理解する上で重要である。Dungey は地上で観測された電離圏の対流から、磁気圏の境界で太陽風磁場と地球磁場が繋ぎ変わり、太陽風が磁気圏の対流を駆動していると考えた<sup>1)</sup>。現在では地上の電離圏観測と人工衛星による磁気圏の観測から、Dungey の考えた磁気圏対流が大まかには合っていると考えられている。しかしながら、磁気圏の観測が人工衛星観測のようにローカルな領域しか観測できないことから、今でも複雑な磁気圏対流の詳細は明らかになっていない。その中で磁気圏全体を解くことができる 3 次元グローバル MHD シミュレーションコードが開発され多くの成果を上げてきた<sup>5,6)</sup>。同時に複雑な 3 次元シミュレーションデータを理解するために、3 次元の可視化技術も必要とされてきた。松岡らは 3 次元グローバル MHD シミュレーションで再現された磁気圏尾部の複雑な磁気フラックスロープの構造について 3 次的に磁力線を可視化することで構造を明らかにした<sup>7)</sup>。

3 次元 MHD シミュレーションで磁力線を可視化するためには、ある始点を与えその始点から磁場に沿って積分を行い、磁力線を描くのが一般的である。これまで、磁力線の時系列 3 次元可視化は磁力線を描くための始点を空間に固定して連続するスナップショットを繋げることで可視化を行ってきた。このような磁力線の可視化は固定された始点から描かれる磁力線の変化を見ることはできるが、磁気圏対流のような磁力線の輸送は分からない。そこで我々はある流体要素を速度に沿って時間方向に積分をして追尾し、その流体要素を始点として磁力線を描くことで磁気圏対流の可視化を行った。

磁力線を追尾し可視化をすれば直接、磁気圏対流を見ることができるようになる。しかしながら、このような可視化は以下の技術的な問題で困難であった。1 つは精度よく流体要素を追尾し可視化するためにはシミュレーションの全時系列データを処理し、その流体要素を始点として磁力線を可視化する必要がある。このようなスーパーコンピュータで計算された大量データの処理と可視化をするためには、CPU の並列分散化だけでなくネットワークや I/O の並列分散化が

\*1 情報通信研究機構

\*2 九州大学

\*3 東京大学



必要である。また、磁力線を追尾するためにはプラズマの磁場凍結が成り立つ必要があり、磁場凍結の検証をしながら追尾をする必要がある。磁場凍結が成り立たない領域では、磁場が拡散し磁力線が繋ぎ変わり磁力線のトポロジーが変化する。以下ではこれまで開発を行ってきた磁力線追尾可視化技術の概略の紹介をする。セクション2では磁力線追尾と可視化に必要な大規模データ処理について述べ、セクション3ではプラズマの磁場凍結の検証について述べる。

## 2. 磁力線追尾のための NICT サイエンスクラウドによる大規模データ処理

磁力線追尾で用いた3次元グローバルMHDシミュレーションデータとNICTサイエンスクラウド<sup>3)</sup>による大規模データ処理について述べる。

### 2.1 3次元グローバルMHDシミュレーションデータ

使用したシミュレーションデータは地球磁気圏と太陽風の相互作用についてグローバルにMHD方程式系を解いている。解像度は0.2Reで、計算格子は $450 \times 300 \times 300$ のデカルト座標系を用いた。1ステップのデータ量は1.3GBである。時間刻みは0.5秒で2時間分のデータを用いた。全データ量は約18TBとなる。太陽風パラメータはGalaxy15イベント(2010/4/5)のACEで観測された値を用いた。

### 2.2 NICTサイエンスクラウドによる大規模データ処理

磁力線を追尾するためには磁力線を描く際の始点となる流体要素を追尾する必要がある。時間方向に精度よく追尾するためには、最小の時間分解能で積分をするのがよい。つまり時系列の順に14400個ファイルを読み込み処理する必要がある。また流体要素を追尾した結果を始点として磁力線を可視化する必要がある。このような大規模データ処理を行う場合、ボトルネックになるのは可視化処理と共にディスクI/Oである。NICTサイエンスクラウドではCPU分散をすると同時に、並列分散大規模ストレージにGfarmファイルシステム<sup>4)</sup>を用いている。Gfarmを用いることで、アクセス集中によるディスクI/O性能の低下を回避することができる。またGfarmによるスケーラブルな並列I/O性能を活用し、Rakeによるワークフロー記述性を用いるため開発されたPwrake<sup>8)</sup>を用いて磁力線追尾と可視化のためのワークフローを記述した。

可視化にはNICTサイエンスクラウドのアプリケーションであるAVSベースで開発されたバーチャルオーロラを用いた。バーチャルオーロラはHDF形式の地球磁気圏シミュレーションデータを読み込み、磁気圏断面、等数値面、流線、ベクトル、点の球表示、2次元プロット等、解析に必要な可視化をインタラクティブに操作できる。図1はバーチャルオーロラの実行画面である。インタラクティブに操作し、磁力線追尾の初期状態を可視化、初期の追尾データを作成、その後、並列分散可視化を行う。

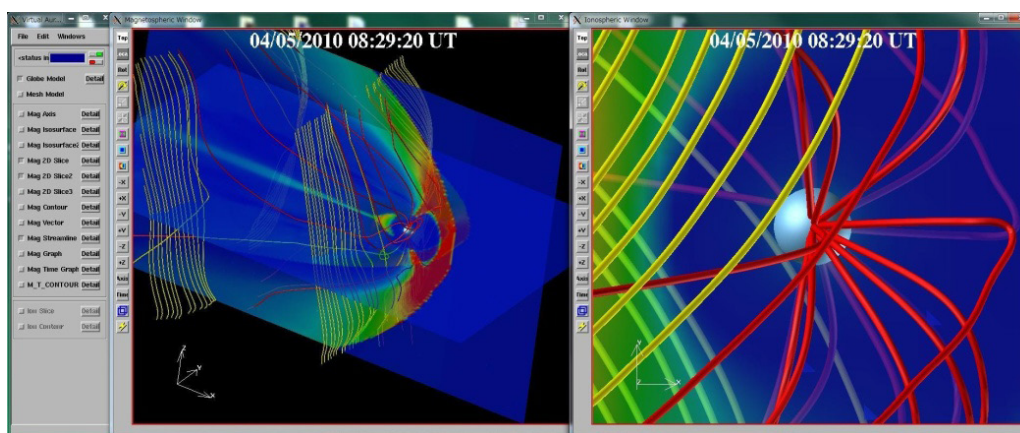


図1 バーチャルオーロラ実行画面。磁気圏の断面と磁力線を描いている。

図2は磁力線追尾のワークフローである。青枠で囲まれている処理過程は並列分散処理を行った処理過程である。最初に、全シミュレーションデータをバーチャルオーロラで読み込めるHDF形式に変換する。また、シミュレーションデータを時系列の順で読み込み流体要素の追尾計算を行う。その後、HDF形式のデータと流体要素の追尾データをバーチャルオーロラに読ませて磁力線を描き、1時刻ステップ毎にHDF形式の3次元オブジェクトの出力を行う。最後に複数のHDF形式の3次元オブジェクトをAVSの幾何形状のアニメーションデータフォーマットであるGFAファイルに結合をする。36コアで並列分散処理を行うことで、数時間で磁力線を追尾し可視化することができるようになり、インタラクティブに研究を進めることが可能となった。

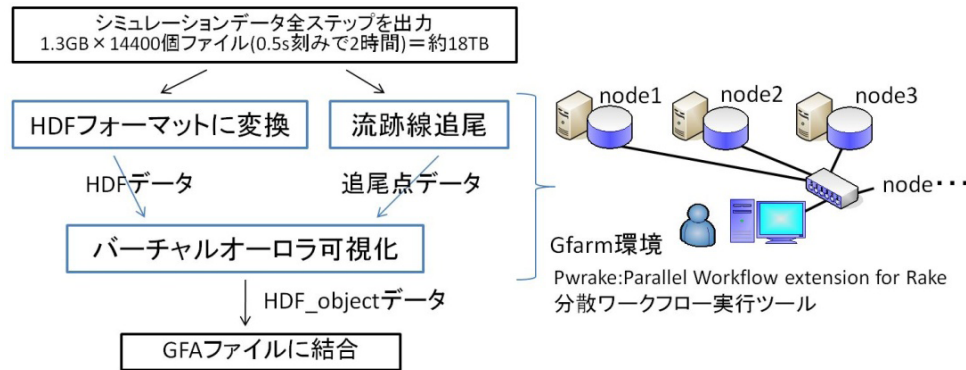


図2 磁力線追尾のワークフロー。青枠の処理過程について gfarm と pwrake を用いて並列分散処理を行っている。

表1に1stepあたりの磁力線追尾処理と可視化処理のそれぞれの演算時間とI/O時間を示した。磁力線追尾処理では、I/O時間が主な処理時間である。可視化処理では、演算時間がI/O時間に対して約6倍となっている。処理の種類により演算時間とI/O時間に偏りがある。最適にリソースを分配するため、ノードに対してコア数を変えて計算をした。本計算の処理に用いたコア数は磁力線追尾処理にはI/O処理を分散させるため6ノード×1コアを用いた。可視化処理にはI/O処理と演算処理を分散させるため6ノード×6コアを用いた。表2に磁力線追尾処理と可視化処理について1000stepの処理時間と1コアの処理時間との比を示した。まず可視化処理では1コアの処理時間との比が19.64倍となっている。6ノード×6コアで全コア数が36に対して効率が出ていないのはI/O処理で輻輳が生じているためである<sup>7)</sup>。磁力線追尾処理では1コアの処理時間との比が2.79倍となっている。6ノード×1コアで全コア数が6に対して効率が出ていないのは、追尾処理は追尾点データの受け渡しのためファイル間の処理に依存関係があり、待ち時間が生じるためである。磁力線追尾処理の並列化効率を高くするのは今後の課題である。

表1 1ファイルあたりの演算時間とI/O時間。750本の磁力線について処理を行った。

	演算時間 [sec]	I/O 時間 [sec]
磁力線追尾処理 (1.3GB)	0.33	4.88
可視化処理 (2.3GB)	47.50	7.57

表2 1000stepの処理にかかった時間と1コアの処理時間との比。

	1000stepの処理時間 [sec]	1コアの処理時間との比 [倍]
磁力線追尾処理 (6ノード×1コア)	1869.29	2.79
可視化処理 (6ノード×6コア)	2804.77	19.64

大規模なGFAファイルを見るためにはビューアが必要である。NICTでは64bit版GFAプレーヤーを開発した。ビューアのメモリーモードでのGFAデータのレンダリングにはパソコンのメモリの制限があるため14400stepを表示するのは難しいが、これまで2000stepのGFAデータを可視化した実績がある。

### 3. 磁力線追尾技法の検証

地球磁気圏 3 次元グローバル MHD シミュレーションでは、磁場の誘導方程式を解いており、物理的または計算を安定させるために人工的に磁場の拡散項を与えている。磁場の拡散項が移流項と比較して小さい領域でプラズマの磁場凍結を仮定できる。拡散項が大きな領域ではプラズマの磁場凍結が破れ、磁力線を追尾することは意味をなさない。以下では、シミュレーション中でプラズマの磁場凍結の定義である微小面積内の磁束の変化をモニターすることで、磁力線追尾ができているかを検証する。

#### 3.1 プラズマの磁場凍結と磁束保存

いまプラズマ中の微小面積  $dS$  を通る磁束  $\Phi$  は、磁束密度を  $B$  とすると

$$\Phi = \oint B \cdot dS \quad (1)$$

と書ける。微小面積  $dS$  が時間と共に移動する場合を考える。磁場の誘導方程式は速度を  $v$  とすると

$$\frac{\partial B}{\partial t} = \nabla \times (v \times B) + \eta \nabla^2 B \quad (2)$$

と書ける。 $\eta$  は抵抗係数で 0.002 である。式 (1) を時間微分し式 (2) を代入すると、磁束の時間変化は

$$\frac{d\Phi}{dt} = \eta \oint \nabla^2 B \cdot dS \quad (3)$$

となる。式 (3) の右辺が 0 の場合、時間に対して微小面積を通る磁束は変化しない。このため、微小面積と共に、磁場が凍結したように移動する。

#### 3.2 磁束の計算と評価方法

磁力線追尾は 2.2 節で述べた処理を行い、流体要素を追尾し、その後流体要素を始点として磁力線を描く。時刻  $t$  の流体要素のある位置  $x(t)$ ,  $x(t)$  の位置の速度を  $v$  とすると、 $dt$  後の流体要素の位置  $x(t+dt)$  は

$$x(t+dt) = x(t) + vdt \quad (4)$$

である。 $v$  は一番近い格子点の値から線形に補間し求めた。

ある微小面積の磁束を計算するため、図 3 の様に微小な円の円周上に 29 点と中心に 1 点の流体要素を配置し、それぞれの流体要素を追尾する。円周の隣り合った 2 点と中心点からできる三角形の面積を計算し、式 (1) から磁束を計算する。29 個の三角形の磁束を足し合わせ微小面積内の磁束を計算する。

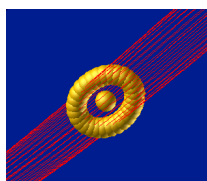


図 3 円周上に 29 点と中心に 1 点の流体要素から磁力線を書いた磁気フラックスチューブ

以下では、初期に配置した微小面積内の磁束からの磁束の変化率でプラズマの磁場凍結を評価する。磁束の変化率の定義は

$$(\Phi(t) - \Phi(0)) / \Phi(0) \times 100 \quad (5)$$

である。磁束の変化率が 0% の場合は、完全に磁場凍結が成り立っている。磁束の変化率が -100% の場合は、磁場が拡散し微小面積の法線方向の磁束密度がなくなる場合を表しており、この領域で磁気フラックスチューブは他の磁気フラックスチューブと繋ぎ変わることができることを意味している。

### 3.3 磁気圏内の磁束保存の検証

磁気圏内の磁束保存について、特徴的な領域を通過する微小面積を追尾することで磁束保存の特徴を調べた。半径  $0.1R_E$  の微小面積をシース、ローブ、プラズマシートの磁気フラックスチューブに配置して、480 秒間追尾を行い磁束の保存を調べた。図 4 は地球磁気圏 MHD シミュレーション結果の子午面の圧力断面上に、調べた微小面積と磁気フラックスチューブを書いた図となっている。左図が初期状態 ( $t=0$  秒)、右図が  $t=370$  秒の図である。図中の橙の円内に微小面積の位置を示している。図中の英字は個々の微小面積を示しており、左図と右図で対応している。磁気フラックスチューブの色はトポロジーを表しており、紫が detached、赤が open の磁気フラックスチューブである。初期状態で (a), (b), (b') は一本の磁気フラックスチューブ中に配置してある。 $t=370$  秒では、1 本であった磁気フラックスチューブが 3 本に分かれることが分かる。つまり、(b), (b') を始点とした磁気フラックスチューブは地球磁場と繋ぎ変わり、南北対象に夜側のローブ領域に輸送される。(a) は後に述べるが磁場の拡散領域を通過したため、追尾することに意味がなくなった磁力線である。(c) は磁場の繋ぎ換えが起き、open から detached の磁気フラックスチューブとなる。(d) はローブ領域を輸送される。

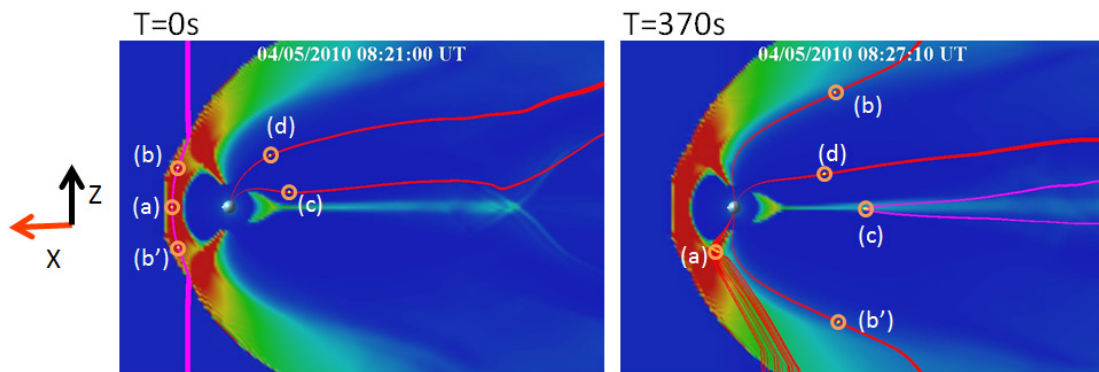


図 4 圧力の子午面断面図上に磁気フラックスチューブを書いた図。橙の円は微小面積の位置を示している。英字は個々の微小面積を示している。磁気フラックスチューブの色はトポロジーを表しており、紫が detached (地球と繋がっていない)、赤が open (一方の端が地球と繋がっている) の磁気フラックスチューブである。

図 5 は初期状態からの磁束、法線方向の磁束密度、微小面積の変化率を表しており、黒線が磁束、青線が法線方向の磁束密度、緑線が微小面積を示している。法線方向の磁束密度と微小面積の変化率の定義は式 (5) の磁束を、それぞれ法線方向の磁束密度と微小面積に置き換えたものである。(a), (b), (c), (d) は図 4 と対応している。最初に (d) を見ると、プラズマシートに向けて輸送される間、磁束密度が減少するが、面積が増えて、磁束はほぼ一定となり、プラズマの磁場凍結が成り立っていることを示している。(a) を見ると、磁束の変化率は  $t=80$  秒で  $-100\%$  を超える。 $t=80$  秒で磁場が拡散し、法線方向の磁束密度がほぼなくなることを示している。ここで磁力線の繋ぎ換えが起きる。磁場の拡散領域を通過した後は、プラズマの磁場凍結が破れているため、図 4 の右図の (a) のように流体要素を追尾することは意味をもたない。 $t=80$  の前に法線方向の磁束密度が上昇し面積が減少するが、磁気圏境界の前に停留点があり太陽風磁場が強くなるためである。(b) を見ると、 $t=0$  から 180 秒に、法線方向の磁束密度が減少し、次に増加する。これは流体要素がカusp領域を通過するためである。面積は逆相関となっており、磁束は保存する。磁束密度と面積の変化率は  $t=180$  秒でそれぞれ  $-50\%$ ,  $100\%$  を超えるが磁束は  $20\%$  程度の変動となっている。磁束が時間に対しておおよそ保存をしている。(b), (b') を始点とした磁気フラックスチューブは (a) の磁束の変化率が  $-100\%$  を超えた領域で地球磁場と繋ぎ変わるが、それぞれ磁気フラックスチューブを形成したまま、ローブ領域に輸送される。(c) はプラズマシート通過時に磁束の変化率が  $-50\%$  を超える。しかし、磁束の変化率が  $-100\%$  を超えないため、(b), (b') と同様に磁気フラックスチューブ上の他の領域で繋ぎ換えがあり、繋ぎ変わった後も磁気フラックスチューブを維持する。

磁力線の繋ぎ換えが起こる領域は (a) のように磁束の変化が  $-100\%$  を超える領域である。3 次元グローバル MHD シミュレーションでは、磁場の拡散項が大きいいため、(b), (b'), (c) のような磁力線の繋ぎ換えが起きない領域でも磁場が拡散していることが確認された。そのため、磁気フラックスチューブを追尾する場合、どの程度磁場が拡散をしているかモニターをしながら追尾する必要があることがわかった。



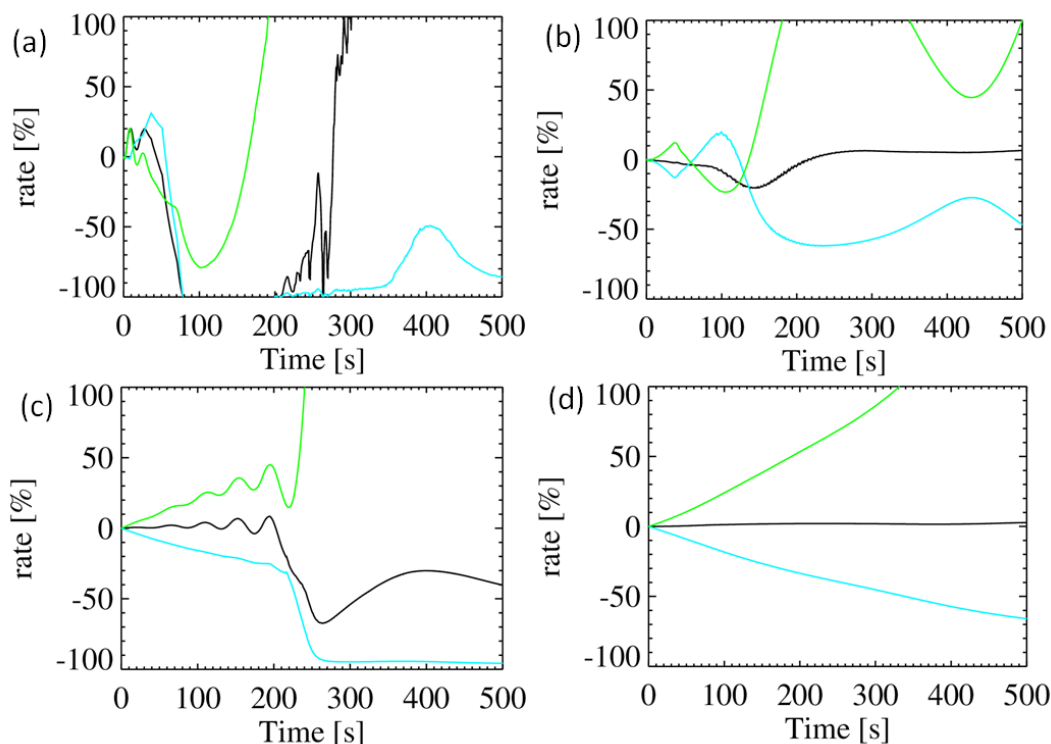


図 5 縦軸が初期状態からの変化率，横軸が時間．黒線は磁束，青線が法線方向の磁束密度，緑線が微小面積の変化率である．(a), (b), (c), (d) は図4と対応している．

#### 4. まとめ

NICT サイエンスクラウドを用いて地球磁気圏 3 次元グローバル MHD シミュレーションの大規模データを並列分散可視化することで、磁束の保存をモニターしながら磁気フラックスチューブを追尾し可視化することができた。可視化処理では CPU と I/O の並列分散化を行い、約 20 倍の高速化が実現できた。磁力線追尾処理では I/O の並列分散化を行い約 3 倍の高速化を実現した。磁力線を追尾するためにはプラズマの磁場凍結が成り立つ必要がある。磁場凍結が成り立たない領域では、磁場が拡散し磁力線が繋ぎ変わる。そのため 30 点の追尾点から成る微小面積を用いて磁束を計算し、磁束の保存をモニターしながら磁気フラックスチューブを追尾し可視化した。その結果、初期状態からの磁束の変化率が -100% になる場合に磁力線が繋ぎ変わることがわかった。磁力線の繋ぎ換えを含めて磁気フラックスチューブを追尾し可視化することができた。今後は本論文の技術を用いて磁力線の繋ぎ換えが起きる磁場の拡散領域を特定できることから、太陽風パラメータを変化させることによって磁場の拡散領域がどのように変化するかを調べる予定である。

#### 参考文献

- 1) Daisuke Matsuoka, Ken T. Murata, Shigeru Fujita, Takashi Tanaka, Kazunori Yamamoto, and Eizen Kimura, Analyses of 3D Structure of Magnetic Flux Ropes via Global MHD Simulations, Journal of Visualization, 28(6), 2008, 38-46.
- 2) Dungey, J. W., Interplanetary magnetic field and the auroral zones, Phys. Rev. Lett., 6, 1961, 47-48
- 3) Ken T. Murata, Shinichi Watari, Tsutomu Nagatsuma, Manabu Kunitake, Hidenobu Watanabe, Kazunori Yamamoto, Yasubumi Kubota, Hisao Kato, Takuya Tsugawa, Kentaro Ukawa, Kazuya Muranaga, Eizen Kimura, Osamu Tatebe, Keiichiro Fukazawa and Yasuhiro Murayama, A Science Cloud for Data Intensive Sciences, Proceedings of the 1st ICSU World Data System Conference, 2011.
- 4) Osamu Tatebe, Kohei Hiraga, and Noriyuki Soda. Gfarm Grid File System. New Generation Computing, Vol. 28, No. 3, pp. 257-275, 2004.



- 5) T. Ogino, R. J. Walker, M. Ashour-Abdalla, J. M. Dawson, An MHD Simulation of the Effects of the Interplanetary Magnetic Field  $B_y$  Component on the Interaction of the Solar Wind with the Earth's Magnetosphere during Southward Interplanetary Magnetic Field, *J. Geophys. Res.*, 91, 1986, 10,029-10,045.
- 6) T. Tanaka, A. Nakamizo, A. Yoshikawa, S. Fujita, H. Shinagawa, H. Shimazu, T. Kikuchi, and K. K. Hashimoto, Substorm convection and current system deduced from the global simulation, *J. Geophys. Res.*, 115, 2010, A05220
- 7) 磯田総子, 村田健史, 渡邊英伸, 深沢圭一郎, 山本和憲, 建部修見, 田中昌宏, 木村映善, NICT サイエンスクラウド ～ビッグデータ分散処理性能検証～, *信学技報*, vol. 113, no. 86, SC2013-4, 2013, pp. 19-24.
- 8) 田中昌宏, 建部修見, 並列分散ワークフローシステム Pwrake による大規模データ処理, *宇宙科学情報解析論文誌*, 1, 2012, 67-75

# 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

後藤 康路<sup>\*1</sup>, 山田 竜平<sup>\*2</sup>, 山本 幸生<sup>\*3</sup>, 横山 昌平<sup>\*1</sup>, 石川 博<sup>\*4</sup>

## A system for visualizing large-scale moonquake data considering waveform similarity using SOM

Yasumichi Goto<sup>\*1</sup>, Ryuhei Yamada<sup>\*2</sup>, Yukio Yamamoto<sup>\*3</sup>,  
Shohei Yokoyama<sup>\*1</sup>, Hiroshi Ishikawa<sup>\*4</sup>

### Abstract

Large-scale data of moonquake were obtained from seismometers located on the Moon by the NASA Apollo missions from 1969 to 1977. According to previous analysis of the lunar seismic data, previous studies found that deep moonquakes occur periodically from identical sources at a depth of about 700 to 1200km. The deep moonquakes occurred from identical sources have high similarities among each waveform. This similarity is important to classify the sources and investigate the generation mechanism of moonquakes. From the reason, classification of moonquakes has been processed. We, therefore, develop the web system for visualizing moonquakes considering waveform similarity to progress study of moonquake classification. Our system maps moonquakes data to two-dimensional output space using Self-Organizing Map (SOM). We embed Hadoop in the back-end system to apply SOM to enormous moonquakes data. In this paper, to select a feature for SOM, we evaluate several features based on classified data. Using selected feature, we perform SOM to moonquake data and discuss its result.

Keywords: Moonquake, Visualization, Self-Organizing Map

### 概要

1969 年から 1977 年の 7 年半の間、NASA のアポロミッションにおいて、月面に設置された地震計を用いた連続観測により、膨大な量の月地震データが取得された。これまでの月地震データの解析より、深発月震は、月深部の約 700-1200km で発生しており、同一の震源から周期的に発生する事が明らかにされている。同一震源から発生する深発月震間において、それらの波形は高い類似性が見られる。この波形の類似性は、震源の分類や、月地震の発生原因の究明において重要な情報となる。そのため、現在に至るまで、人手による月地震の分類が行われている。しかし、膨大な量の月地震データを人手によって分類することは困難である。そこで、本研究では、月地震データの分類研究を促進するため、波形の類似性を考慮して月地震データを可視化するためのシステムを実装する。本システムでは、月地震データを 2 次元空間上へマッピングするため、Self-Organizing Map (SOM) を用いる。膨大な量の月地震データを効率的に処理するため、本システムの処理のバックエンドに Hadoop を用いる。SOM に用いる特徴量を事前に選択するため、月地震データの人手による分類結果に基づいて、Support Vector Machine(SVM) により、月地震を適切に分類可能な特徴量を複数検証する。SVM の分類結果から選択された特徴量を用いて SOM を適用し、その結果について考察する。

### 1. 背景

NASA のアポロミッションにおいて、Apollo Lunar Surface Experiments Package(ALSEP) と呼ばれる観測装置群が設置された<sup>1)</sup>。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地震の連続観測を行う実験である。この実験により、約 7 年半に渡る月地震記録が取得された。これらのデータは、長周期地震計の 3 軸成分および短周期地震計の Z 軸成分で構成されており、公開されている<sup>2)</sup>。

\*1 静岡大学大学院情報学研究科 (Graduate School of Informatics, Shizuoka University)

\*2 国立天文台 RISE 月惑星探査検討室 (National Astronomical Observatory of Japan, RISE project)

\*3 宇宙航空研究開発機構宇宙科学研究本部 (The Institute of Space and Astronautical Science)

\*4 首都大学東京システムデザイン学部 (Faculty of System Design, Tokyo Metropolitan University)

PSE によって取得された月地震の観測データは、現在も解析が行われており、月地震の発生原因、活動度、および月の内部構造の研究において多くの知見が得られている<sup>3,4)</sup>。これらの解析の一環として、月地震のクラス分類が行われており、波形から予測された発生深度や発生要因をもとに、各月地震に深発月震、浅発月震、隕石衝突などのラベルが付与されている。特に、深発月震は同一震源から周期的に発生することが分かっているため<sup>5,6)</sup>、月-地球間に働く潮汐力と関連して、その発生原因について研究が進められている<sup>7,8)</sup>。これらの研究を通して、深発月震の震源にはラベルが付与されている。これらの付与は、人手によって行われ、月地震イベントカタログとして公開されている。この月地震イベントカタログは、現在でも多数の研究に基づいて更新が行われている<sup>9)</sup>。しかし、PSE によって取得された月地震のデータは、膨大な量である。そのため、未だにラベルが付与されていない月地震や人手ではラベルの付与が困難な月地震が存在する。また、人手によるラベルの付与の正確性の検証が困難であるという課題がある。そこで、本研究では、コンピュータを用いた月地震解析のファーストステップとして、月地震データの解析・可視化システムを構築する。本システムは、月地震の波形の類似関係を可視化し、人手により付与されたラベルと比較を行うことで、ラベルが付与されていない月地震の解析の補助、付与されたラベルの検証、および未知クラスの発見補助を目的とする。

本研究では、月地震解析基盤と、解析結果の可視化インターフェイスを提案する。月地震解析基盤では、Self-Organizing Map(SOM)<sup>10)</sup>を用い、高次元データである月地震波形を、波形の類似度に基づいて低次元空間へのマッピングを行う。また、SOM を適用するため、SOM に用いる特徴量の選択を事前に行う。そのため、教師あり機械学習手法のひとつである Support Vector Machine(SVM)を用いて、特徴量ごとに分類結果の評価を行い、最も人手によるラベルの付与に近いと評価された特徴量を用いる。可視化インターフェイスでは、月地震解析基盤で得られた SOM の結果と、人手による分類結果と合わせて提示する。解析結果と人手による分類結果を比較することにより、付与されたラベルの正確性の検証や、人手による分類結果に反するような性質を示す波形の発見を促進する。

## 2. 提案システム

本研究で提案するシステムの構成図を図1に示す。Web インターフェイスは Web サーバ、月地震解析基盤はクラスタによって構築されている。月地震データセットは、約 1TB の大規模な時系列データであり、13,058 件の月地震を含んでいる。そのため、最終的に全データの解析を目的とする本研究では、SOM や SVM の計算量が膨大になると考えられる。このような課題に対して、分散処理の導入が考えられる。しかし、分散処理プログラムの開発コストは非常に高価である。そこで本研究では、分散処理の導入を容易にするため、大規模データ処理基盤である Hadoop<sup>11)</sup>と、Hive<sup>12)</sup>を用いる。ユーザの要求を受け、Hadoop クラスタ内の多数のコンピュータを用い、月地震データの管理、SOM、および SVM 処理を行う。ユーザへの可視化結果の提示には、HTTPD を介した Web インターフェイスを用いる。可視化インターフェイスでは、ユーザの要求に応じて、SOM の結果に基づいた Web ページを生成する必要があるため、リアルタイムなデータ応答が必要になる。しかし、Hadoop はリアルタイムなデータ応答に適していない。そのため、予め Hadoop によってデータ処理を行い、その結果を RDBMS に蓄積するという形式をとる。Hadoop と RDBMS 間のデータ転送には Sqoop<sup>13)</sup>を用いる。

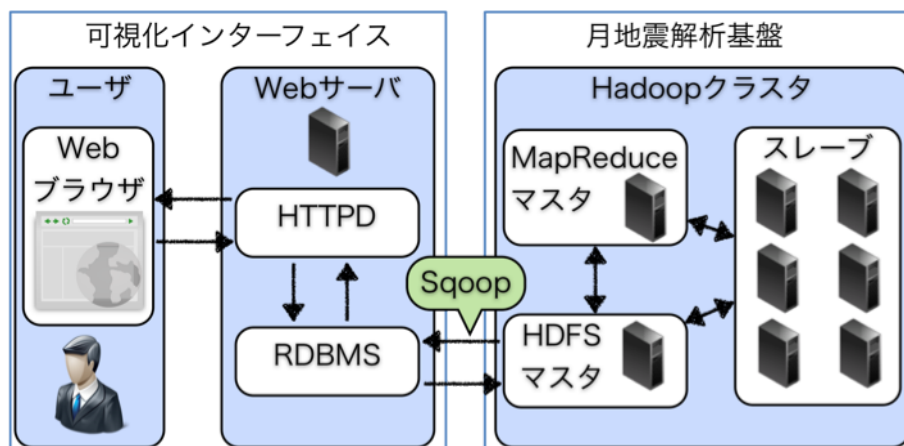


図1 提案システムの構成図

## 2.1. データ管理

月地震解析基盤で用いる Hadoop は、分散ファイルシステム HDFS と、分散処理フレームワーク MapReduce により構成されている。大規模月地震データを HDFS で管理し、MapReduce によってデータ処理を行うことによって、分散並列処理によるデータ解析の効率化を図る。また、Hive を用いることにより、これらのファイルを SQL ライクなドメイン固有言語を介して MapReduce 操作を行うことができる。Hive はユーザ定義関数の定義ができるため、MapReduce を用いたプログラムを直接作成する場合と比較して、高度なデータ処理を容易に記述および実行できるようになる。また、HDFS に格納された圧縮ファイルに対するデータ操作もサポートしており、MapReduce 処理におけるコンピュータ間の通信のオーバーヘッド削減や、専有するストレージ容量の削減の効果がある。

月地震データの管理では、RCFile<sup>14)</sup>形式を用いる。RCFile は、Hadoop 上でのデータ格納形式のひとつである。列指向データファイル形式という特徴を持つため、列単位でのデータの管理およびデータ操作を効率的に行うことができる。月地震データは、時系列データとして保存されるため、列指向データファイル形式が適していると考えられる。また、RCFile と合わせ、高速データ圧縮手法である Snappy 圧縮<sup>15)</sup>を用いる。これらの手法により、TSV 形式で約 1TB のデータセットに対して、HDFS 上において約 1/5 のサイズに圧縮し、データ管理を行った。

## 2.2. SOM

教師なし機械学習手法のひとつである SOM は、入力データの類似性に基づいて低次元の出力空間へ入力データをマッピングする手法である。各データは、出力空間上に配置されたセルにマッピングされる。各セルは、入力データと同じ次元の重みベクトルを持つ。SOM は、入力データの傾向に基づいてセルの重みベクトルに学習し、入力データを最も類似したセルに配置することで、低次元空間へのマッピングを行う。入力データ  $x_n$  に最も類似したセルを返す関数  $c(x_n)$  を式 (1) に示す。

$$c(x_n) = \arg \min_p d_{pn} \quad (1)$$

ここで、 $d_{pn}$  は、 $x_n$  とセル  $p$  の距離を表す。

学習時には、一度出力空間上に入力データをマッピングし、セルの周囲に配置された入力データを用いて、重みベクトルの更新を行う。学習は、学習係数  $\sigma$  を 1 から 0 になるまで単純減少させながら行う。セル  $r_i$ ,  $r_j$  間の影響度を表す近傍関数  $h(r_i, r_j)$  を式 (2)、学習の正則化項  $\gamma_p$  を式 (3)、重みベクトルの更新式を式 (4) に示す。

$$h(r_i, r_j) = \exp\left(-\frac{\|r_i - r_j\|^2}{2\sigma^2}\right) \quad (2)$$

$$\gamma_p = \frac{1}{\sum_i^N h(p, x_i)} \quad (3)$$

$$m_p(t+1) = \gamma_p \sum_i^N h(p, c(x_i)) x_i \quad (4)$$

$m_p(t+1)$  は、セル  $p$  の更新後の値、 $N$  は入力データ数を表している。これらの処理により、類似したデータを同じセルに、セル同士の距離が近いほど、セル間の類似度が高くなるようにデータを配置する。高次元特徴量を持つデータを各データ間の類似性を保ちながら低次元空間にマッピングすることで、ユーザはそれらのデータを容易に閲覧可能になる。

本システムでは、月地震データが線形分離不可能である場合を考慮し、データが線形分離不可能な場合にも適用可能なカーネル SOM<sup>16)</sup>を用いる。一般的な SOM は、距離関数を用いてデータ間の類似度を計算するのに対して、カーネル SOM は、カーネル法を用いてデータ間の類似度を計算する。カーネル法は、非線形データに対するデータ解析で広く用いられている手法であり、データを高次元へ非線形写像することで、解析手法を非線形データに対して拡張する。実際は、高次元での解析は膨大な計算量を伴うため、写像後の空間における内積を定義するカーネル関数を用いて類似度の計算を行う。カーネル関数を用いる場合、セルの重みベクトルを求めることができないため、直接セルと各データの距離を求める。式 (5) にセル  $p$  とデータ  $x_i$  の距離の式を示す。ここで、 $K(x_i, x_j)$  はカーネル関数を表す。

$$\begin{aligned}
d_{pn} = & K(\mathbf{x}_n, \mathbf{x}_n) \\
& - 2\gamma_p \sum_i^N h(\mathbf{p}, c(\mathbf{x}_i)) K(\mathbf{x}_n, \mathbf{x}_i) \\
& + \gamma_p^2 \sum_i^N \sum_j^N h(\mathbf{p}, c(\mathbf{x}_i)) h(\mathbf{p}, c(\mathbf{x}_j)) K(\mathbf{x}_i, \mathbf{x}_j)
\end{aligned} \tag{5}$$

### 3. SVM を用いた特徴量の評価と選択

月地震データセットを、カーネル SOM を用いて可視化する際に、各データ間の類似度を計算するために用いる特徴量を選択する。本研究では、月地震データの人手による分類結果に基づいた特徴量の選択を行うために教師あり機械学習手法のひとつである SVM を用いる。SVM は、教師あり機械学習手法のひとつであり、クラス分類によって予測ラベルが計算される。そのため、SVM によるクラス分類の結果と人手による月地震分類の結果を比較し、定量的に月地震の特徴量を評価することができる。本実験では、人手による月地震の分類結果を SVM のトレーニングデータとし、複数の特徴量について SVM を適用し比較・考察を行う。

今回の実験では、システムの適用例として、月地震データセットから波形を抽出し、実験用のデータセットとする。用いる波形は、人手により深発月震と分類された月地震のうち、震源が A1, A6, A7, A8, A9, および A10 と分類されている月地震波形 551 件である。これらの波形を選択した理由として、震源位置が良く決められており波形と震源との関連が明瞭である点と、各震源に対して地震波形のサンプル数が充分であるため、特徴量の評価が容易であるという点が挙げられる。特に、実験で使用した 551 件は S/N 比が高く、比較的 S 波が明瞭に識別できるものを選択している。各震源に含まれるデータ数の内訳を、以下の表 1 に示す。各波形データは、月地震の S 波を中心に、前後約 10 分程度になるように調整し、トレンド引きやノイズ除去といった前処理を適用したデータを利用する。

表 1 実験に用いる月地震データ数の内訳

震源クラス	A1	A6	A7	A8	A9	A10	合計
要素数	204	54	53	55	106	79	551

本実験において用いる特徴量算出手法を以下に示す。

- Power Spectral Density (PSD)
- フレーム毎に区切った波形の PSD
  - (フレーム長: 波形の 1/16, シフト長: 波形の 1/32, 窓関数: ハニング窓)
- PSD のエンベロップ
- フレーム毎に区切った波形の PSD エンベロップ

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。PSD から特徴量を算出する手法のうち、広く用いられる手法として、フレーム分割やローパスフィルタによるエンベロップを用いる手法が挙げられる。本実験では、これらの特徴量算出手法を用いて特徴量を複数の波形形式から抽出し、評価を行い、適切な特徴量を選択する。波形形式は、直接波形データの値を用いる場合と、波形データのエンベロップを用いる場合の 2 種類である。本実験では、4 種類の特徴量算出手法と 2 種類の波形形式の組み合わせである 8 特徴量について評価を行う。図 2 に、8 特徴量の導出手順を示す。本実験において、エンベロップは移動平均に基づいて算出する。

今回はカーネル関数として線形カーネルを用いる。評価指標としては、クラス毎に適合率、再現率、および F 値を用い、10 分割交差検定によって検証を行う。適合率は、予測結果に対して、正しく予測できた割合を示す値である。再現率は、正解に対して、予測結果が網羅できた割合を示す値である。F 値は、適合率と再現率の調和平均である。



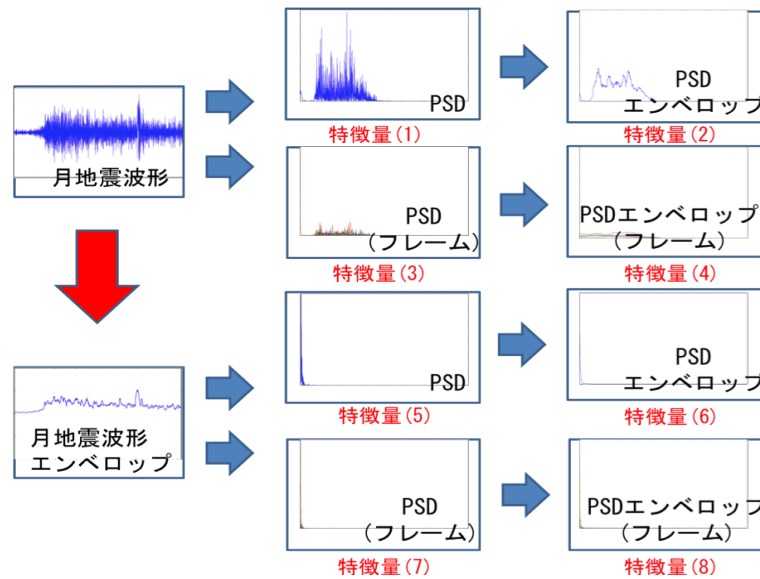


図2 特徴量抽出手順

図3、図4および図5に月地震波形を用いた場合の実験結果、図6、図7および図8に月地震波形のエンベロップを用いた場合の実験結果を示す。実験結果より、月地震波形を用いた場合と月地震波形のエンベロップを用いた場合の両方において、PSDを特徴量に用いた場合のF値の平均値が最も高く評価されたため、PSDがより適切な特徴量であることが確認された。また、図5の月地震波形を用いた場合と図8の月地震波形のエンベロップを用いた場合の実験結果の比較より、最もF値の平均が高い特徴量は、月地震波形を用いた場合のPSDである。そのため、次章では、月地震波形を用いた場合のPSDに基づいてカーネルSOMを適用する。

ここで、最もF値の平均値が高い、月地震波形を用いた場合のPSDの結果に着目する。各震源の分類精度を比較すると、A1のクラス分類精度は、他の震源のデータと比較して高い。この結果について、A1を震源とする波形は、他の震源とは異なる性質を含んでいるためF値が高い値になったと考えられる。また、A8について、他の震源の適合率と比較して、A8の適合率は低い。このことから、A8を震源とする月地震データが含む性質を、A8以外の震源の月地震データの一部も含んでいる可能性があると考えられる。しかし、このようなクラス分類の結果のみに基づいて、月地震に対する検証や考察を行うのは困難である。そこで、次章では、カーネルSOMを用いた可視化手法を適用することにより、月地震データの検証を行う。

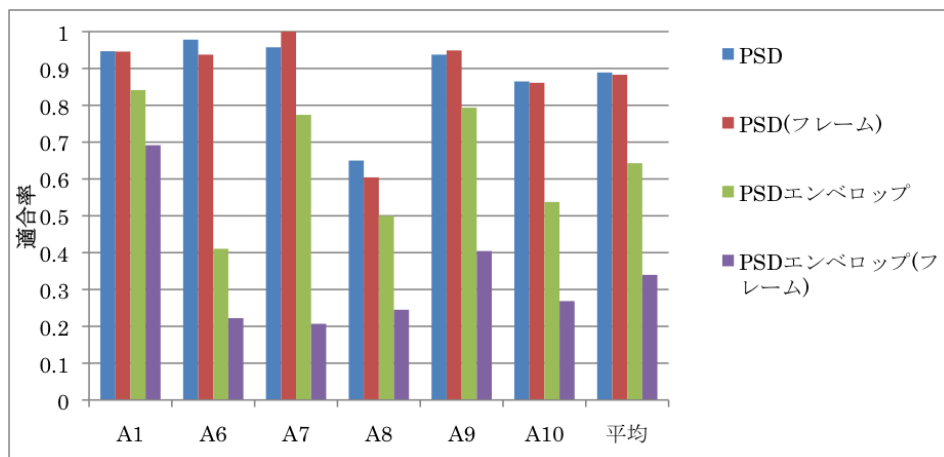


図3 特徴量選択のための評価実験結果（月地震波形を用いた場合）：適合率

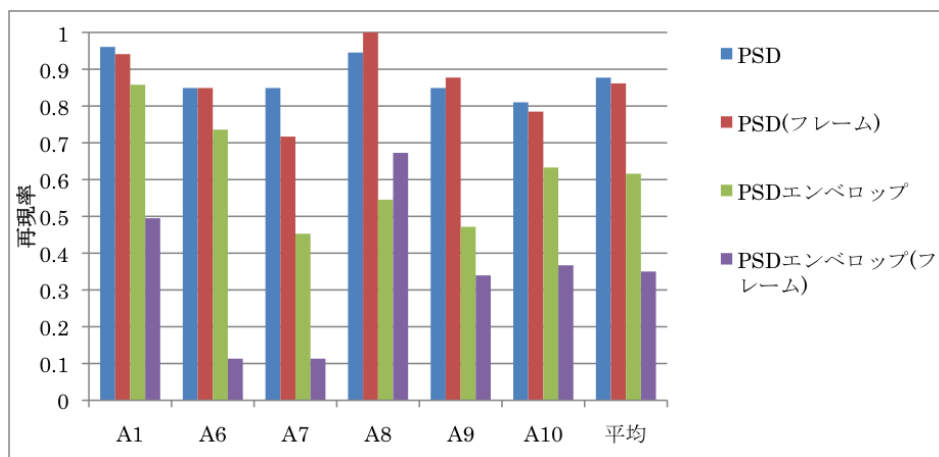


図4 特徴量選択のための評価実験結果(月地震波形を用いた場合): 再現率

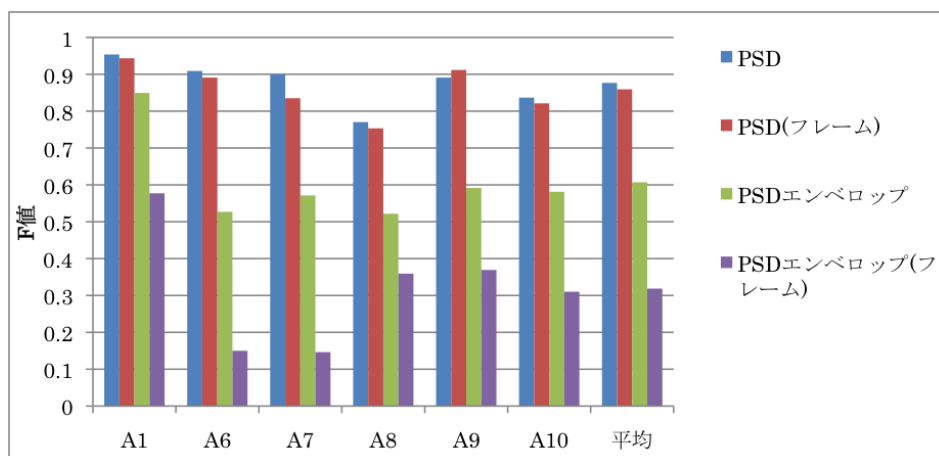


図5 特徴量選択のための評価実験結果(月地震波形を用いた場合): F 値

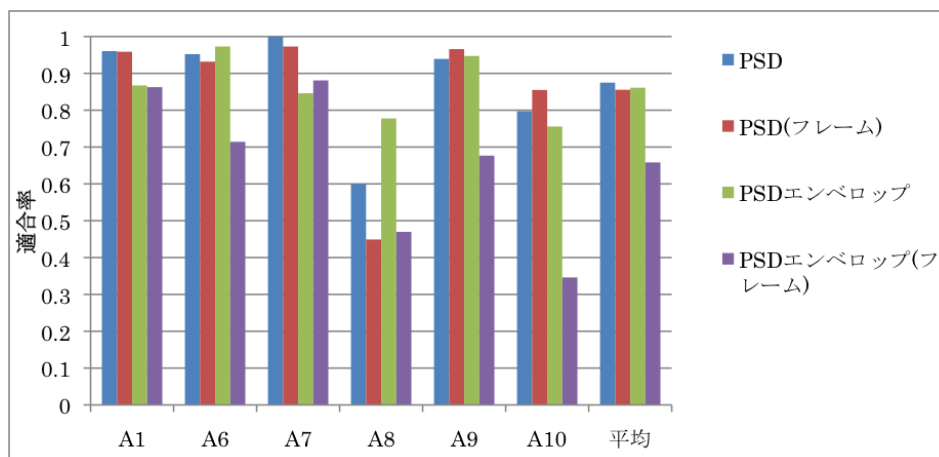


図6 特徴量選択のための評価実験結果(月地震波形のエンベロップを利用した場合): 適合率

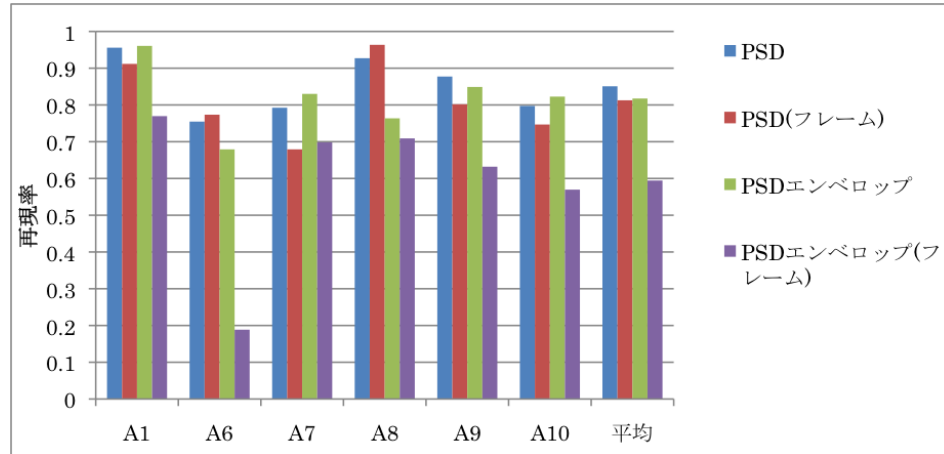


図 7 特徴量選択のための評価実験結果（月地震波形のエンベロップを利用した場合）：再現率

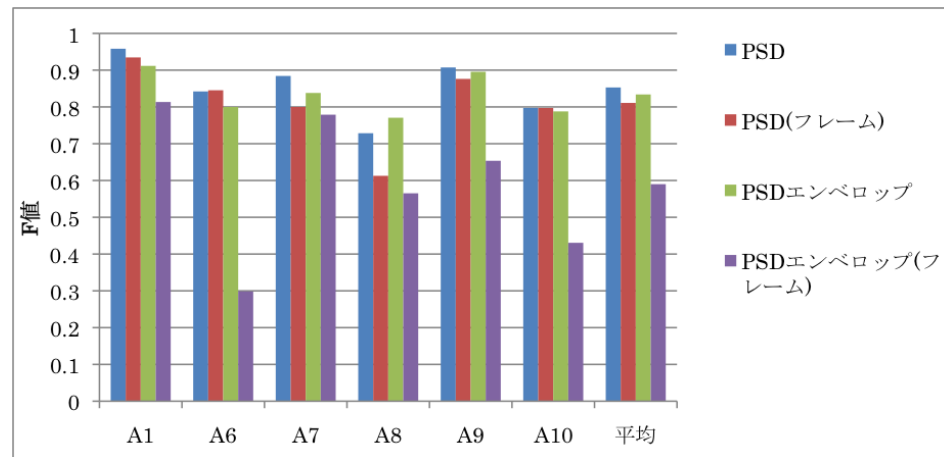


図 8 特徴量選択のための評価実験結果（月地震波形のエンベロップを利用した場合）：F 値

#### 4. カーネル SOM による月地震の可視化例

月地震解析基盤において、月地震データに対してカーネル SOM を適用する。また、SOM の結果に基づいて、可視化インターフェイスを用い、人手によって付与されたラベルが他のデータとどのように類似しているかを可視化する。カーネル SOM に用いる特徴量は、前章の実験において最も F 値の平均値が高い、月地震波形を用いた場合の PSD を用いる。カーネル SOM で用いるカーネル関数は、線形カーネルを用いる。カーネル SOM を適用するデータは、前章の実験に用いた月地震データと同様である。本章では、カーネル SOM のパラメータについて、複数のパラメータで実験を行い、適切と考えられる値を設定した。具体的なカーネル SOM のパラメータは以下の通りである。

- 出力空間：トラス空間
- セルの数：25(5 × 5)
- 学習のイテレーション回数：10 回

図 9 は、カーネル SOM の適用例である。各セルの円グラフは、セルに配置された月地震について、人手によって付与されたラベル毎に色分けし、内訳を表示している。

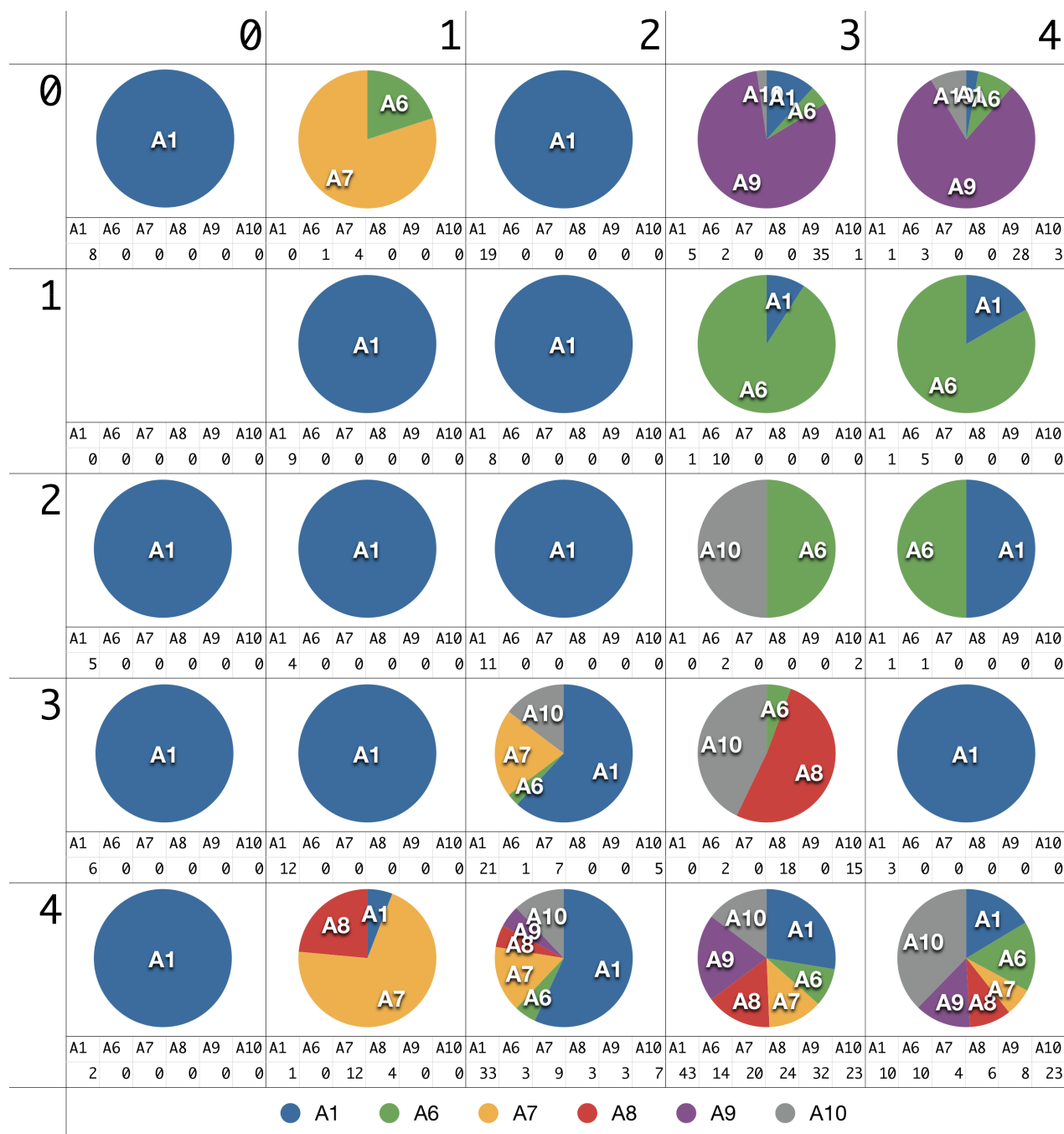


図9 深発月震に対するカーネル SOM 適用例

前章の結果において、F 値が高かった震源 A1 のデータ (図9の円グラフにおいて青色) に着目すると、可視化結果では、全 25 個のセルのうち A1 が震源である月地震データのみを含むセルが 11 個存在し、非常に広範囲のセルを A1 のみで専有している。この結果より、震源 A1 から発生したとされる月地震データは、他の震源との類似度が低く特徴的な性質の波形を含むことが分かる。震源 A6 のデータ (図9の円グラフにおいて緑色) に着目すると、セル 1 行 3 列目周辺に全体の約半数のデータが集まり、ほぼセル専有している。震源 A7 のデータ (図9の円グラフにおいて黄色) は、セル 4 行 1 列目周辺に集まり、セル 4 行 1 列目を、ほぼ専有している。震源 A8 のデータ (図9の円グラフにおいて赤色) は、セルを専有してはいないものの、セル 3 行 3 列目およびセル 4 行 3 列目に集中している。震源 A9 のデータ (図9の円グラフにおいて紫色) は、セル 0 行 3 列目周辺に集まっており、特にセル 0 行 3 列目と 0 行 4 列目を、ほぼ専有している。震源 A10 のデータ (図9の円グラフにおいて灰色) も、A8 と同様にセルを専有してはいないものの、セル 3 行

目 3 列目およびセル 4 行目 4 列目に集中している。セルの専有の有無は、3 章の実験において、A8 と A10 の F 値が比較的良かったことと対応していると考えられる。

また、右下のセル 4 行 3 列目とセル 4 行 4 列目において、多数の震源のデータが集まっていることが分かる。前章の月地震波形を用いた場合の PSD でのクラス分類結果において、A8 と誤分類された波形は全 28 件であり、その内訳は正解震源 A1 が 4 件、A6 が 3 件、A7 が 4 件、A9 が 7 件、A10 が 10 件である。これらの月地震に着目すると、セル 4 行 3 列目に 20 件、セル 4 行 4 列目に 8 件の月地震が集まっていることが分かった。これらの月地震の波形を図 10、図 11 に示す。また、図 12 にセル 4 行 1 列目に配置された A8 を震源とする波形を示す。分類結果の正誤に関わらず、セル 4 行 3 列目に配置された波形は、波形の振幅が小さいという共通点がある。そのため、波形の性質がノイズに埋もれ、分類に失敗した結果、A8 と誤分類されたものと考えられる。A8 の適合率は、A8 の月地震データが正しく分類された数 TP、A8 以外の月地震データが A8 と誤分類された数 FP より、 $TP/(TP+FP)$  として求められるため、誤分類の増加によって A8 の適合率が低下したと考えられる。

同様に、A10 と分類された月地震波形に着目する。図 13 と図 14 は、セル 4 行 3 列目に配置された波形である。また、図 15 に、セル 2 行 3 列目に配置された A8 を震源とする波形を示す。これらの波形についても、分類結果の正誤に関わらず、セル 4 行 3 列目に配置された波形は、波形の振幅が小さいという共通点がある。そのため、セル 4 行 3 列目に配置されたデータには、波形の性質がノイズに埋もれ、分類が困難である波形が集まったと考えられる。

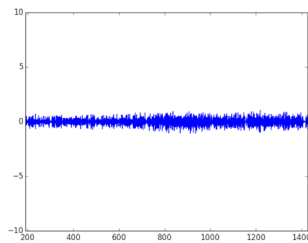


図 10 分類結果 A8  
正解震源 A6 の波形  
(セル 4 行 3 列目に配置)

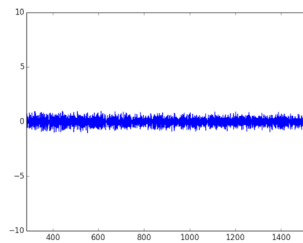


図 11 分類結果 A8  
正解震源 A8 の波形  
(セル 4 行 3 列目に配置)

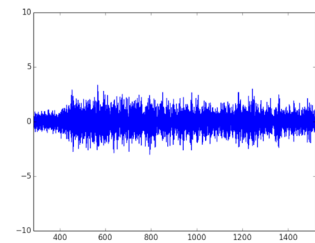


図 12 分類結果 A8  
正解震源 A8 の波形  
(セル 4 行 1 列目に配置)

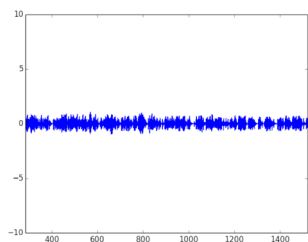


図 13 分類結果 A10  
正解震源 A8 の波形  
(セル 4 行 3 列目に配置)

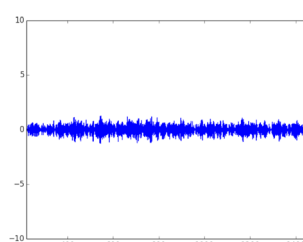


図 14 分類結果 A10  
正解震源 A10 の波形  
(セル 4 行 3 列目に配置)

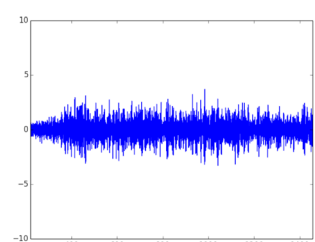


図 15 分類結果 A10  
正解震源 A10 の波形  
(セル 2 行 3 列目に配置)

## 5. まとめ

本研究では、波形の類似性を考慮した月地震データの解析・可視化システムの実装を行った。大規模データに適用するために、月地震解析基盤に Hadoop を用いた。可視化手法としてカーネル SOM を用い、その結果を提示する可視化インターフェイスを実装した。また、システムの適用例として、一部の深発月震の震源分類について、教師あり機械学習手法である SVM の結果と比較し、考察を行った。SOM に基づいた可視化結果より、震源 A1 について、他の震源と異なる性質を持つ波形が多いことや、震源 A8 について、ノイズにより波形の持つ性質が埋没し、そのままでは他の震源との分類が困難であることが分かった。しかし、このように人手による分類結果と機械的な分類結果との比較を円滑



に行うことができるという点において、本システムは、今後、人手による月地震分類結果の検証や、未分類とされている月地震の分類を行っていく上で、有効であると考えられる。

今後は、他のカーネル関数を用いた精度向上や、全月地震データセットについて、他の観測点で得られた月地震データ、震源位置および未知クラスの月地震波形を考慮した可視化を進めていく予定である。

### 参考文献

- 1) N. Toksöz, A. Dainty, C. Solomon and K. Anderson, “Structure of the Moon,” *Reviews of Geophysics and Space Physics*, 12, 1974, pp. 539-565.
- 2) 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, 第一号, 2012, pp. 121-131.
- 3) Y. Nakamura, G. Latham and J. Dorman, “Apollo lunar seismic experiment –final summary” , *Journal of Geophysical Research*, 87, 1982, pp. A117-A123.
- 4) P. Lognonné, J. Gagnepain-Beyneix and H. Chenet, “A new seismic model of the moon: Implication for structure, thermal evolution and formation of the moon” , *Earth and Planetary Science Letters*, 211, 2003, pp. 27-44.
- 5) D. R. Lammlein, “Lunar seismicity and tectonics” , *Phys. Earth Planet. Inter.*, 14, 1977, pp. 224-273.
- 6) R. C. Bulow, C. L. Johnson, B. G. Bills and P. M. Shearer, “Temporal and spatial properties of some deep moonquake clusters” , *Journal of Geophysical Research: Planets*, 112.E9, 2007.
- 7) H. Araki, “Focal processes of deep moonquakes” , *J. Geod. Soc Japan*, 47, 2001, pp. 508-513.
- 8) C. C. Hon and N. Toksöz, “Tidal stresses in the Moon” , *Journal of Geophysical Research*, 83.B2, 1978, pp. 845-853.
- 9) Y. Nakamura, G. Latham, J. Dorman and J. Harris, “Passive seismic experiment long-period event catalog” , *Galveston Geophysics Laboratory Contribution*, 491, 2008.
- 10) T. Kohonen, “The self-organizing map” , *Proceedings of the IEEE*, 78.9, 1990, pp. 1464-1480.
- 11) “<http://hadoop.apache.org/>” .
- 12) “<http://hive.apache.org/>” .
- 13) “<http://sqoop.apache.org/>” .
- 14) Y. He, R. Lee, Y. Huai, Z. Shao, N. Jain, X. Zhang and Z. Xu, “RCfile: A fast and space-efficient data placement structure in mapreduce-based warehouse systems” , *The IEEE International Conference on Data Engineering (ICDE)*, 4, 2011, pp. 1199-1208.
- 15) “<http://code.google.com/p/snappy/>” .
- 16) P. Andras, “Kernel-kohonen networks” , *International Journal of Neural Systems*, 12, 2002, pp. 117-135.

# 数値モデルを用いた惑星大気研究と地球流体電脳倶楽部 DCMODEL プロジェクト

竹広 真一<sup>\*1</sup>, 高橋 芳幸<sup>\*2</sup>, 西澤 誠也<sup>\*3</sup>, 林 祥介<sup>\*2</sup> 地球流体電脳倶楽部<sup>\*4</sup>

## Researches on planetary atmospheres using numerical models and GFD-Dennou Club DCMODEL project

Shin-ichi Takehiro<sup>\*1</sup>, Yoshiyuki O. Takahashi<sup>\*2</sup>, Seiya Nishizawa<sup>\*3</sup>, Yoshi-Yuki Hayashi<sup>\*2</sup>, and GFD-Dennou Club<sup>\*4</sup>

### Abstract

Space missions for planetary atmospheres and numerical simulations are fundamental components for researches on planetary atmospheres. Recently, numerical simulation models become essential to planning of missions, analyzing and providing observational data. On the other hand, the simulation models used in these researches have become so expanded and complicated that it is difficult for one researcher to understand, develop, maintain and provide their source codes.

“Model gap problem” is one of the difficulties with numerical simulation models. In order to check the validity of the simulation models and to develop them further, it is necessary to perform numerical calculations with numerical models with various complexities and to compare the numerical results mutually. Expansion of the simulation models makes difficult to perform such inter-comparison for checking of the validity. In an attempt to overcome the “model gap problem”, we have been proceeding “dcmmodel project”, where a series of hierarchical numerical models with various complexity is developed and maintained as a voluntary activity. The numerical models are developed with the following intentions: 1) Hierarchical models with various complexities. 2) Common “style” of programming in order to ensure readability. 3) Open source codes to the public. 4) Scalability of the models in order to ensure execution on various platforms. 5) Stressing the importance of documentation and presenting a method for producing reference manuals. The series of hierarchical models enable us to conduct multiple numerical experiments with multiple models easily, which will lead to a deeper understanding of various phenomena in planetary atmospheres.

### 概要

惑星大気探査と数値シミュレーションとは惑星大気研究に欠かせない手段である。さらに近年の惑星大気探査においては、数値シミュレーションモデルはその計画立案や観測データの整理提供において、不可欠な道具となりつつある。一方で、大気の数値シミュレーションモデルは複雑化巨大化しており、一個人の研究者が掌握し、開発・維持・提供していくことは困難なソフトウェアとなっている。その困難の一つに、いわゆる「モデルギャップ問題」がある。シミュレーションモデルの正当性を担保し維持向上させていくためには、様々な複雑度をもつ数値モデルによる数値計算を実行し相互比較することにより、計算の正当性を吟味することを行わなければならない。モデルの複雑化巨大化はこのような相互比較による検証作業を極めて困難にしている。「モデルギャップ問題」を克服することの試みとして、われわれは地球惑星流体研究のためのさまざまな複雑度を持つモデルを階層的に構築し、整備するボランティアベースの活動「dcmmodel プロジェクト」を行ってきた。ソフトウェア開発においては以下の点に留意している：1) 様々な複雑度のモデル群。2) 共通の「型」を持ったモデル群。3) 誰でも使えるモデル群。4) 様々な規模の計算機で実行できるモデル群。5) 解説文書の整備されたモデル群。このような階層的モデル群は、複数のモデルによる複数の数値実験実現を容易にし、惑星大気現象の理解と惑星大気シミュレーションモデルの開発につながる事が期待される。

---

\*1 京都大学数理解析研究所 (Research Institute for Mathematical Sciences, Kyoto University)

\*2 神戸大学地球惑星科学専攻 / 惑星科学研究センター

(Department of Earth and Planetary Sciences/Center for Planetary Sciences, Kobe University)

\*3 理化学研究所計算科学研究機構 (RIKEN Advanced Institute for Computational Science)

\*4 <http://www.gfd-dennou.org/>

## 1. 数値モデルを用いた惑星大気研究

近年、多くの探査機による惑星の観測が実施され、惑星に関する新たな情報が得られるようになってきている。探査機観測に基づいた惑星固体・大気および周辺環境の研究は、通常はそれらの観測結果だけでなく、さまざまなレベルの数値モデルを相補的に用いることにより進められる。得られた観測データの量が少ない時代には、その観測結果をモデルにより解釈することがまず行われる。新たなデータが加わり、データ量が増加し蓄積されてくると、それらを取り込んで数値モデルの改良・精緻化が行われるようになる。その結果としてモデルが発達し数値シミュレーションがもっともらしい結果を与えるようになると、数値モデルの出力がデータベース化され、惑星の平均的描像を提供できるようになる。さらには、観測データをモデルに取り込んでデータの再解析、あるいは次に述べるデータ同化を行うことにより、観測の穴を埋めるデータが生成できて、より詳細な惑星の状態を知ることができるようになる。すなわち、数値シミュレーションモデルと観測データの整理解析は不可分な関係となってくる。このように発達し洗練されたモデルの出力は惑星の状況を記述するだけでなく、次期探査計画へとフィードバックされることにつながる。例えば疑似観測データを用いた同化実験を行うことにより、探査計画が効率的であるのか否かを評価することが可能となる。

このように、数値シミュレーションモデルはいわば観測と種々の基礎過程に関する知見の集大成といえるだろう。

### 1.1 データ同化

上に述べたデータ同化とは、観測データと数値モデルから時間発展する系の状態を推定する方法のことである。良くできたシミュレーションモデルは、観測のない時空間領域における物理量や観測できない物理量の情報を提供できる。データ同化結果を用いてモデルの検証を行ったり、現象の理解を深化させることにつながる。そもそもデータ同化は数値天気予報の初期値生成のために開発・発達してきており、地球の気象海洋分野では同化されたデータが基本的なデータセットとして活用されてきている。過去約 50 年の長期間にわたる観測データの同化および再解析データは、それら自体を観測結果としてみなし、気象および気候研究に利用されている。その結果として研究者人口が拡大し、研究内容の進化につながっている。

惑星大気研究においても、次に紹介するように、火星大気の同化データが公開されるようになって来ており、地球大気と同様、同化データセットを利用した研究が広がってくるだろうと思われる。

### 1.2 海外での研究例

数値シミュレーションモデルと観測データの積極的な協同による惑星大気研究は、海外では普通に行われるようになりつつある。その一例として、Laboratoire de Météorologie Dynamique (LMD) (France) と Oxford Univ., Open Univ. (UK) のグループが協力して行ってきた活動が上げられる。火星の気候モデルデータベース、Mars Climate Database (MCD, <http://www-mars.lmd.jussief.fr/>), は大気大循環モデルの計算結果をコンパイルしたものであり、様々な季節、地方時、緯度、経度、高度における物理量(風速、温度、密度、物質の量など)のデータを集約している。このデータセットはインターネットで一般に公開されている<sup>1)</sup>。

火星大気データ同化、Mars Analysis Correction Data Assimilation (MACDA), は欧米の火星(惑星)大気探査の観測データを用いた同化計算である。ダストストーム時のダスト量と平均子午面場の時間変化の様子を見ることができたりする<sup>2)</sup>。

このような海外の情勢に対して、日本国内の惑星大気研究は、残念ながら惑星探査観測と数値モデルとの協同が密ではない。数値モデルは観測データの解釈を行うためのツールとしてしか主に使われておらず、データ同化や探査計画へのフィードバックなどは行われていない。研究体制も組織化されているとはいえず、個人や小規模のグループが別々に小さな予算規模で研究している。

## 2. 地球流体電脳倶楽部の活動

以下では、国内の惑星大気研究活動の一例として、地球流体電脳倶楽部がここ 10 数年にわたって行っている数値モデル開発活動である DCMODEL プロジェクトを紹介する。

地球流体電脳倶楽部は、地球惑星科学研究に必要な計算機(電脳)資源の開発や収集を行ってきている。それを通じて研究基盤の維持、知識・技術の伝承を試みている。インターネット上に設置したサーバーにおいて描画(DCL)やデータ解析ツールなど(Dennou-ruby)、データ収集(衛星画像、数値予報・再解析データなど)、数値モデル(地球流体・惑星大気

モデルなど), 関連セミナー資料(発表資料, ビデオなど), その他メンバーが開発した電腦資源など, を提供している。このような活動は, 大学等の組織を越えた有志によりボランティア的に支えられている。ソフトウェア開発やサーバーの維持管理等の活動は恒常的かつ大規模な予算に裏付けされているわけではなく, 有志の獲得した小規模な科学研究費等の一部を運用することにより維持してきている。

## 2.1 DCMODEL プロジェクト

先に述べたように, 近年の惑星大気の研究においては数値実験あるいはシミュレーションによるアプローチが盛んに行われるようになってきている。そこで用いられている数値モデルは単純な流体計算に加えて, 放射・乱流過程, あるいは雲・降水に代表される構成物質の相変化などの多くの物理素過程を取り入れている。そのため, 今やシミュレーション数値モデルは一人の研究者が容易には掌握できない大きさと複雑さを持ったソフトウェアとなっている。その結果として, 複雑なモデルによる計算結果の正当性の検討と現象をより深く理解するための, 簡略な概念モデルや素過程による比較計算を行うことが難しくなっておりモデルギャップ問題として認識されつつある<sup>3)</sup>。複雑なシミュレーションモデルと簡単な概念モデルあるいは素過程モデルのギャップを埋めるためには, 素過程モデルや概念モデルを種々の複雑度で階層的に用意する必要がある。これらによる数値実験を並列的に行って, それらの結果を比較検討していく研究を行う必要がある。このような背景を鑑み, われわれは地球惑星流体研究のためのさまざまな複雑度を持つモデルを階層的に構築・整備する活動「`dcmodel` プロジェクト」<sup>4)</sup>を行って来ている。

## 2.2 `dcmodel` のモデルの特徴

`dcmodel` プロジェクトでは, 次のような特徴を持つモデル群を構築することを目指している

### 1. 様々な複雑度のモデル

現象の理解と複雑なモデルの正当性のチェックのために, 簡単なモデル(素過程モデル・概念モデル)から複雑なモデルまでさまざまな複雑度をもつモデル群を階層的に用意する。

### 2. 共通の「型」を持ったモデル

複雑度の異なる多数のプログラムを用いた数値実験を同時に効率的に行うためには, それらのプログラムコードの可読性が重要な要素となる。`dcmodel` プロジェクトのモデル群では共通の書式・変数名規則・入出力ルーチン・データ書式を用いている。共通のスタイルを持ったプログラムコードにしておくことで, 一つのプログラムを学習すれば他のプログラムの理解が速く進むことが期待される。また, データ形式を共通にしておくことで, 複数のモデルの数値計算結果の後処理や可視化を同じような操作で行なうことができる。

プログラムの維持管理だけでなくモデルのユーザーにとっても可読性は重要な要件である。すなわち, 数値計算ソフトウェアのソースコードがその元となる物理モデルをどのように体现しているかを, かならずしもソフトウェアの専門家ではないユーザーが確認できるように, プログラム構造を設計することが必要である。

### 3. 誰でも使えるモデル・プログラム

`dcmodel` プロジェクトのモデル群はソースコードから全てインターネット上で公開されており, 自由に利用・改変・再配布が可能である。このことは単に利用者の利便を図っているだけでなく, 数値モデルを用いた科学研究において必須の事項だからである。他の研究者の結果を誰でも追試できることが科学研究を進める上で重要な点である。しかしながら, 近年の数値実験やシミュレーションによる研究を追試することは困難である。そもそも数値実験の追試を行うためには, 実験に用いられたモデルが誰でも使えるようになっていなければならない。

### 4. 様々な規模の計算機で実行できるモデル

パソコンによる低解像度計算からスーパーコンピュータによる大規模計算まで同じプログラムで使用できるようにモデルを設計・構築することをめざしている。学生が学部・大学院教育で用いた教材を, そのまま最先端の研究として用いることができる。大学で学習したことを社会人になっても生かすことができる。

## 5. 解説文書の整備されたモデル

dcmodel プロジェクトのモデル群では、初心者向けのチュートリアル文書、中級者向けの参考書および開発者向けのモデル解説書をモデル開発と共通の形式で整備することをめざしている。モデルのユーザーだけでなくプログラムの維持管理にとってもモデルの解説文書は重要である。すなわち、数値計算ソフトウェアのソースコードが記述しようとしていた元の物理モデルを確認する必要があるからである。

## 2.3 dcmodel プロジェクトのモデル群の構成要素

### 2.3.1 spmodel

spmodel<sup>6)</sup> は比較的簡単な地球流体数値実験用のスペクトルモデル集である。内部で Fortran77 のスペクトル変換ライブラリである ISPACK<sup>5)</sup> を用いている。1 次元 KdV 方程式から 2 次元順圧・浅水モデル・ブシネスク流体モデル、3 次元回転球殻磁気流体ダイナモモデルまでそろえている。図 1 に spmodel による流体計算の出力例を示している。

spmodel では内部で用いている独特の書法のお陰で、支配方程式から容易に想像できる表現をもったプログラムソースコードを構築することができている。例えば 1 次元 KdV 方程式、

$$\frac{\partial \zeta}{\partial t} = -\zeta \frac{\partial \zeta}{\partial x} - \frac{\partial^3 \zeta}{\partial x^3}$$

を spmodel 書法でプログラムをかくと

```
e_DZetaDt = -e_g(g_e(e_Zeta)*g_e(e_Dx_e(e_Zeta))) &
            -e_Dx_e(e_Dx_e(e_Dx_e(e_Zeta)))
```

といった具合である。ここで e\_Zeta は  $\zeta$  をスペクトル変換したデータ、e\_g, g\_e はスペクトルデータと格子点データの間の変換関数、e\_Dx\_e は空間微分演算を行う関数である。

### 2.3.2 deepconv

deepconv<sup>7)</sup> は地球を含めた惑星大気的应用を想定した、非静力学流体モデルである。順圧縮系を採用し、複数の大気構成分子の化学変化と相変化を導入しており、惑星大気の雲対流をシミュレートすることができる。計算設定の変更によって 2 次元計算と 3 次元計算を容易に切り替えることができるようになっている。表 1 に deepconv の主な仕様を、図 2 に deepconv による木星の雲対流の計算例を示している。

表 1 雲解像度モデル deepconv の主な仕様

力学過程	準圧縮系 (Klemp and Wilhelmson, 1978) 水平：等間隔格子点 鉛直：等間隔格子点
物理過程	
乱流過程	1.5 次のクロージャ (Klemp and Wilhelmson, 1978)
湿潤過程	複数凝結成分、主成分凝結を考慮 Kessler (1969) のパラメタリゼーション 雲粒の拡散成長
放射過程	一様冷却 / 加熱 地球用放射モデル
地表面フラックス	バルク法 拡散



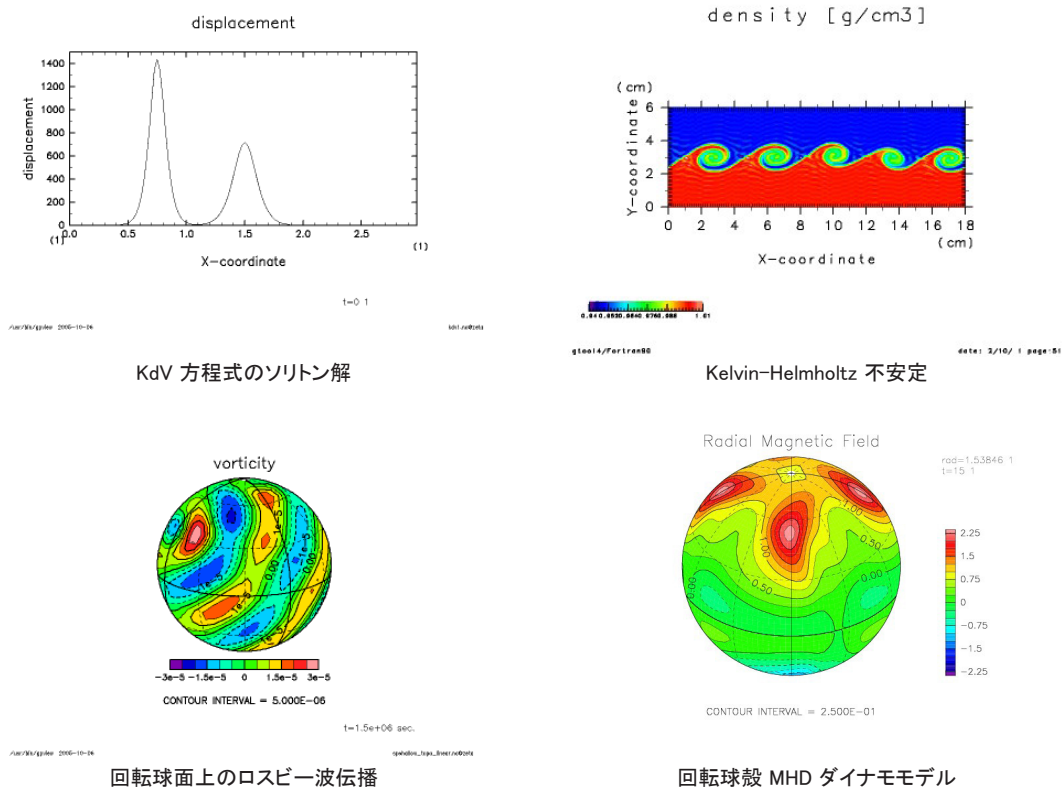


図 1 spmodel の計算例

### 2.3.3 dcpam

dcpam<sup>8)</sup> は汎用惑星大気大循環モデル (GCM) である。プリミティブ方程式に基づいており、力学部分は ISPACK を用いたスペクトル計算で行われる。放射・乱流・降水過程ならびに地表面過程が導入されている。設定によって、地球・火星および簡略化された金星および木星大気の大気大循環を計算できるようになっている。また、鉛直 1 次元計算と 2 次元軸対称計算も可能である。表 2 に dcpam の主な仕様を、図 3 に dcpam による地球側惑星大気大循環の計算例を示している。

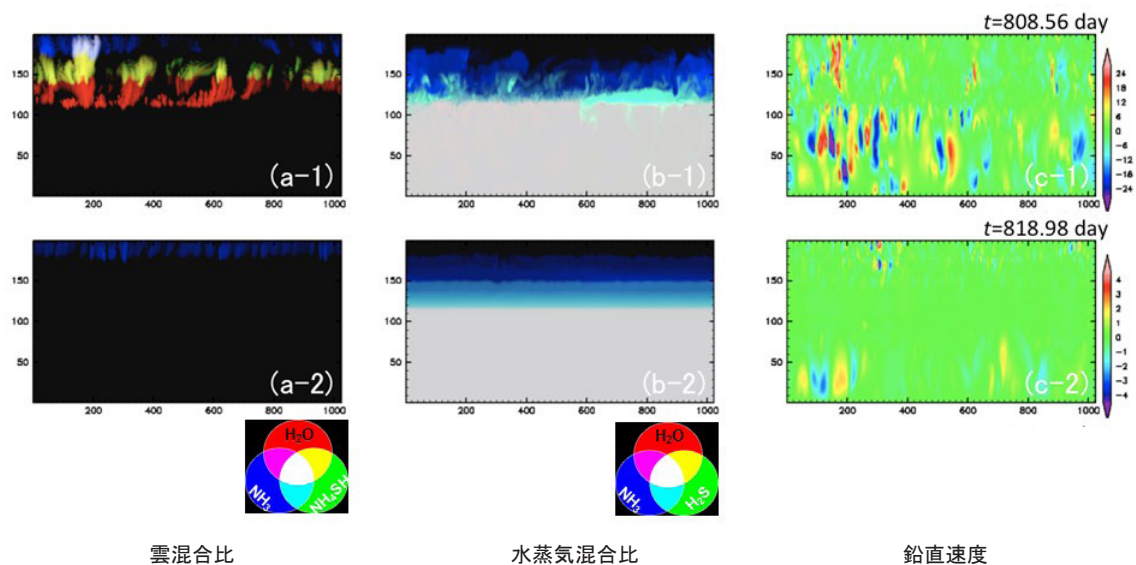


図 2 deepconv の計算例。上段が対流活動の活発期、下段が静穏期。

表 2 汎用惑星大気循環モデル (GCM) dcpam の主な仕様

力学過程	プリミティブ方程式系 水平：スペクトル変換法 (ISPACK 使用) 鉛直：Arakawa and Suarez (1983) の鉛直差分法
物理過程	
放射過程	地球大気放射モデル：CO <sub>2</sub> , H <sub>2</sub> O, O <sub>3</sub> , 雲 (Chou et al, 2001) 火星大気放射モデル：CO <sub>2</sub> , ダスト 灰色放射モデル
乱流混合	乱流：Mellor and Yamada (1974, 1982) 惑星表面フラックス：Louis et al. (1982)
凝結過程	積雲対流：Relaxed Arakawa-Schubert (Moorthi and Suarez, 1992) 大規模凝結：Manabe et al. (1965) CO <sub>2</sub> 凝結過程
陸面モデル	バケツモデル (Manabe, 1969) 土壌熱伝導モデル
雲モデル	簡単な予報方程式により雲水密度を予報 雲氷を無視 積雲対流と大規模凝結による生成, 定数寿命による消失, 乱流混合を考慮, サブグリッドスケールの部分雲は考慮しない (雲量 1)

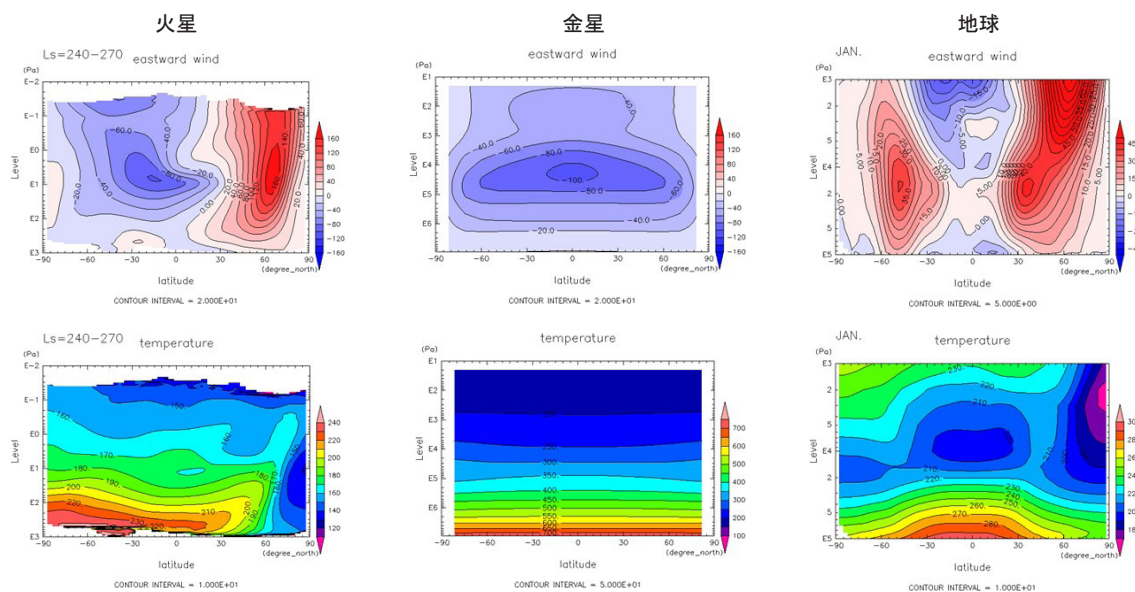


図 3 dcpam の計算例 . 上段が東西風 , 下段が温度

### 3. まとめ

惑星大気探査に基づいた研究を遂行するには、いまや数値シミュレーションモデルが必須である時代となった。惑星探査観測と有機的に結合でき、かつ、惑星科学分野で専有できる、地球をも含めた惑星科学の知見集積としてのシミュレーションモデルが必要である。一方で、用いられるシミュレーションモデルはソースコードが巨大化し、個人で開発維持することが困難になってきている。したがって、数値モデルの開発維持運用に専従するスタッフをそろえた中核となるグループが必要である。既存の地球環境科学 (天気予報, 気候予測) とは独立した組織で惑星大気科学のためのシミュレーション/データ解析のソフトウェアチームを編成し、惑星科学分野へサービスを提供するにしなければならない。こ

のような傾向は惑星大気研究に限らず惑星科学のあらゆる分野で顕著となり、それぞれの分野で洗練されたシミュレーションモデルとその開発維持体制が必要となるだろう。

### 参考文献

- [1] Lewis, S. R., Collins, M., Read, P. L., Forget, F., Hourdin, F., Fournier, R., Hourdin, C., Talagrand, O. Huot, J.-P., 1999 : A climate database for Mars. *J. Geophys. Res.*, 104, 24177–24194.
- [2] Montabone, L., Lewis, S. R., Read, P. L., 2005: Interannual variability of Martian dust storms in assimilation of several years of Mars global surveyor observations. *Advances in Space Research*, 36, 2146–2155.
- [3] Held, I. M.: The gap between simulation and understanding in climate modeling. *B. Am. Meteorol. Soc.*, 86, 1609–1614, 2005.
- [4] 地球流体電脳倶楽部 dcmmodel プロジェクト . <http://www.gfd-dennou.org/library/dcmodel/>, 地球流体電脳倶楽部 .
- [5] 石岡圭一 , ispack-1.0.0, <http://www.gfd-dennou.org/arch/ispack/>, (2013), 地球流体電脳倶楽部 .
- [6] Takehiro, S., Odaka, M., Ishioka, K., Ishiwatari, M., Hayashi, Y.-Y., and SPMODEL Development Group: SPMODEL: A series of hierarchical spectral models for geophysical fluid dynamics, Nagare Multimedia, available at: <http://www2.nagare.or.jp/mm/2006/spmodel/index.htm>, (2006).
- [7] 杉山耕一朗, 小高正嗣, 山下達也, 中島健介, 林祥介, deepconv 開発グループ, 非静力学モデル deepconv, <http://www.gfd-dennou.org/library/deepconv/>, (2010), 地球流体電脳倶楽部 .
- [8] 高橋 芳幸, 竹広 真一, 石渡 正樹, 納多 哲史, 小高 正嗣, 堀之内 武, 森川 靖大, 林 祥介, DCPAM 開発グループ, 惑星大気モデル DCPAM, <http://www.gfd-dennou.org/library/dcpam/>, (2011), 地球流体電脳倶楽部 .

# Effective system for simulating dust continuum observations on distributed computing resources

Kazutaka MOTOYAMA<sup>\*1</sup>, Yoshikazu TANAKA<sup>\*2</sup>, Kento AIDA<sup>\*1</sup>, Eisaku SAKANE<sup>\*1</sup>  
and Kenichi MIURA<sup>\*1</sup>

## Abstract

We present an effective system for simulating dust continuum observations by radiative transfer simulations. By using workflow management system RENKEI-WFT, we utilized distributed computing resources and automated a sequence of computational tasks required for radiative transfer modeling, namely, main radiative transfer simulations, pre-/post-processes, and data transfer between computing resources. Our system simultaneously executes a lot of radiative transfer simulations with different input parameters on distributed computing resources. This capability of our system enables us to conduct effective research by radiative transfer simulation. As a demonstration of our system, we simulated dust continuum observations of protoplanetary disk. We performed hydrodynamic simulation modeling photoevaporating protoplanetary disk irradiated by ultra violet radiation from nearby massive stars. Results of this hydrodynamic simulation were used as input data for radiative transfer simulations. Expected spectral energy distributions and intensity maps were obtained by our system.

**Keywords:** distributed computing, radiative transfer, scientific workflow

## 1. Introduction

Predicting radiative properties of astronomical objects is essential to interpretation of photometric and spectroscopic observations. Spectral line radiation from molecular gas and continuum radiation from dust grain provide useful information on physical and chemical properties of observed objects. Recent increasing of computational power allows us to simulate observations of astronomical objects by radiative transfer simulation. Comparison between observations and radiative transfer model helps understanding of physical and chemical properties of observed objects. Thanks to its high sensitivity and resolution, future powerful facilities such as the Atacama Millimeter Array (ALMA) and the Thirty Meter Telescope (TMT) will provide huge observational data at radio/infrared/optical wavelength. It is necessary to establish effective method for radiative transfer modeling to make comparisons with these huge observational data.

Automation of computational tasks for radiative transfer modeling, such as main radiative transfer simulations and pre-/post-processes, is needed to improve efficiency of radiative transfer modeling. Since radiative transfer modeling for simulating observations involves observing direction and observing frequency as

parameter, we need to run a lot of radiative transfer simulations with different parameters. It makes manual operation of computational tasks inefficient.

When all computational tasks for simulating observations run on single computing resource, shell script provides us an easy way to automate these computational tasks. However, it is very common situation that main simulations and pre-/post-processes run on different computing resources. In such case, we encounter difficulty to automate computational tasks with shell script. We propose using a workflow management system RENKEI-WFT to automate computational tasks. RENKEI-WFT was developed for REsources liNKage for E-science (RENKEI) project<sup>1</sup> launched by National Institute of Informatics (NII) and partner institutes and universities. RENKEI-WFT coordinates multiple job submission and data transfer on distributed computing resources integrated with grid middleware such as NAREGI<sup>2</sup>. One of the advantages of using RENKEI-WFT is that it supports parallel execution of a lot of independent jobs with different input parameters. Therefore, RENKEI-WFT is suitable for executing parameter survey application such as radiative transfer simulation.

Other advantage of using RENKEI-WFT is that it ensures secure communications between computing resources. Security is one of key issues in distributed computing environment, particularly when utilizing computing resources distributed over multiple organi-

<sup>\*1</sup> National Institute of Informatics, 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan

<sup>\*2</sup> Meisei University, 2-1-1 Hodokubo Hino, Ome, Tokyo 191-8506, Japan

<sup>1</sup><http://www.e-sciren.org/>

<sup>2</sup><http://www.naregi.org/>

zations. RENKEI-WFT uses the Grid Security Infrastructure (GSI) as an authentication mechanism. GSI is designed based on the X.509 public key infrastructure, where a certificate authority digitally signs a user's public key and issues a digital certificate. The certificate is used for authentication between computing resources.

This paper aims to propose method for automation of simulating observations on distributed computing resources with RENKEI-WFT. The layout of this paper is as follows. Section 2 describes our radiative transfer simulation. Section 3 presents architecture of RENKEI-WFT. Section 4 describes design and implementation of our system and demonstrates how it works. Finally, section 5 summarizes our work presented in this paper.

## 2. Radiative Transfer Simulation

This section briefly describes radiative transfer simulation that is main component of our system. Now we define the specific intensity,  $I_\nu$ , as the amount of radiative energy at frequency  $\nu$  crossing a surface of unit area per unit time, per unit solid angle, per unit frequency interval. As described in Rybicki and Lightman (1979),<sup>[1]</sup> the change in specific intensity traveling a distance  $s$  in a specific direction is given by radiative transfer equation:

$$\frac{dI_\nu}{ds} = -\alpha_\nu I_\nu + j_\nu, \quad (1)$$

where  $\alpha_\nu$  and  $j_\nu$  are the absorption coefficient and the spontaneous emission coefficient, respectively. Although both of molecules and dust can be sources of emission and absorption, we focus only on transfer of thermal continuum radiation from dust in radiative equilibrium with local radiation field in this paper. Since the equilibrium temperature of dust is set by balance between emitted and absorbed radiation, it is written as

$$j_\nu = \alpha_\nu B_\nu(T_{dust}), \quad (2)$$

where  $B_\nu$  is the Planck function at the dust temperature  $T_{dust}$ . The absorption coefficient is related to the dust opacity per unit mass,  $\kappa_\nu$ , by

$$\alpha_\nu = \kappa_\nu \rho_{dust}, \quad (3)$$

where  $\rho_{dust}$  is the mass density of dust. The dust opacity depends on the shape, size distribution, and constitution of the dust grain.

Radiative transfer code RADMC-3D<sup>3</sup> is used for radiative transfer simulations in our system. Simulation of thermal dust continuum radiation consists of two steps. At first step, equilibrium dust temperature  $T_{dust}$  satisfying equation (2) is computed by Monte Carlo

method (Lucy 1999;<sup>[2]</sup> Bjorkman & Wood 2001<sup>[3]</sup>). The positions, spectra and luminosities of individual stars inside the observed system, and spatial distribution of dust density are given as input data. At second step, expected images and spectra are computed by ray tracing. The radiative transfer equation given by equation (1) is solved along direction pointing to the observer.

## 3. RENKEI-WFT

We assume the typical application scenario in which the user utilizes computing resources in both a laboratory/department grid (or a local site) and a national grid infrastructure to run the application. Hereafter in this paper, we refer to the former resources as a Laboratory Level System (LLS) and the latter resources as a National Infrastructure System (NIS). The workflow application consists of jobs for pre-/post-processes and the main simulation. The user would want to run jobs for the pre-/post-processes on the LLS and the job for the main simulation on the NIS.

RENKEI-WFT enables submission of jobs in a single workflow to both the LLS and the NIS even if they are operated with different grid middleware. Figure 1 illustrates an example of an LLS and a NIS setting using RENKEI-WFT. The LLS consists of an LLS-Portal server and computing resources, e.g., PC cluster(s). Here, we assume that two PC clusters in the LLS are operated with different grid middleware, GridSAM<sup>4</sup> and WS GRAM<sup>5</sup>. Both types of middleware provide lightweight services for job submission and monitoring of submitted jobs. The NIS consists of supercomputers operated through NAREGI. NAREGI includes the information service (IS) and the brokering service (Super Scheduler), which performs matchmaking and dispatches jobs to suitable computing resources in the NIS.

RENKEI-WFT, which consists of the workflow tool (WFT) and the workflow engine (WF engine), is installed on a portal system in the LLS (LLS-Portal). The WFT works as a user interface for editing a workflow, submitting jobs in the workflow, and monitoring the submitted jobs. In addition, the WFT offers its service on the portal so that the user does not have to install and configure software for RENKEI-WFT in the local machine. The WF engine dispatches jobs through its job submission interfaces. The user can save edited workflows in the application repository, Application Hosting Service (AHS),<sup>[4]</sup> and share the workflows with other users.

RENKEI-WFT is designed to cooperate with the grid file system, Gfarm,<sup>[5]</sup> to enable file sharing among distributed computing resources. Jobs running in the

<sup>3</sup><http://www.ita.uni-heidelberg.de/~dullemond/software/radmc-3d/>

<sup>4</sup><http://www.omii.ac.uk/wiki/GridSAM/>

<sup>5</sup><http://www.globus.org/toolkit/docs/4.0/execution/wsgram/>



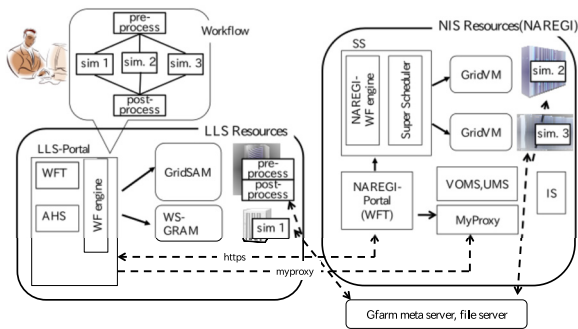


Figure 1: Example of an LLS and a NIS setting

LLS and the NIS are able to share files through Gfarm, which enables the stage-in/stage-out of files between the LLS and the NIS. The interface to access Gfarm is implemented by the GridFTP protocol<sup>[6]</sup> or a proprietary protocol.

Resources in the NIS consist of supercomputers in production service. They are usually operated with rigid firewall policies. Administrators of the NIS may not be able to open special TCP ports for submitting jobs from the LLS. We designed RENKEI-WFT to use limited protocols from the LLS to the NIS. Access from the LLS to the NIS is conducted by using the https request between RENKEI-WFT and NAREGI-Portal.

The detailed scenario presented in Figure 1 is as follows. The user accesses the LLS-Portal and launches RENKEI-WFT. Then, the user edits the workflow, which consists of five jobs, **pre-process**, **sim1**, **sim2**, **sim3** and **post-process**. The user is able to submit the workflow, or jobs, and through the WFT, the user can monitor the submitted jobs that are running in the LLS-Portal. The WFT shows statuses, “queued”, “running”, or “done” of the submitted jobs on the GUI. The WF engine in the LLS-Portal submits jobs to the LLS directly and to the NIS via the NAREGI-Portal using the https protocol. In this example, **pre-process**, **post-process**, and **sim1** run on the LLS, and **sim2** and **sim3** run on the NIS. Through the WFT, the user is able to indicate resources (LLS or NIS) to which the jobs are assigned. The rest of this section presents the details of RENKEI-WFT.

### 3.1. Workflow Tool

The Workflow Tool (WFT) is a core software module in RENKEI-WFT. It enables the user to edit the workflow, submit the workflow, and query the status of the workflow (or jobs) running on the LLS and the NIS through the GUI. A workflow consists of programs to be executed once, programs to be executed a specified number of times, data, and sub-workflows. These com-

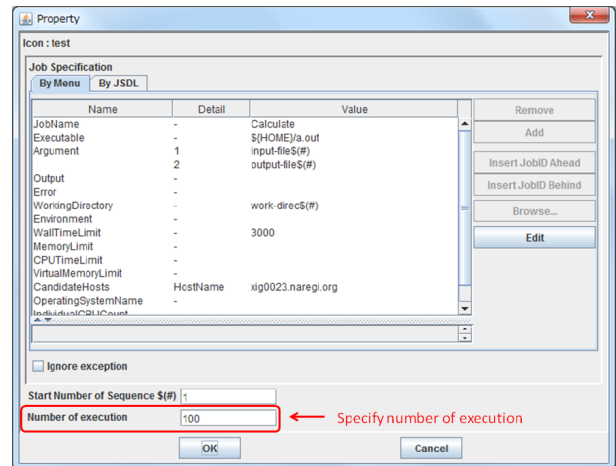


Figure 2: GUI to define the property of a calculate icon

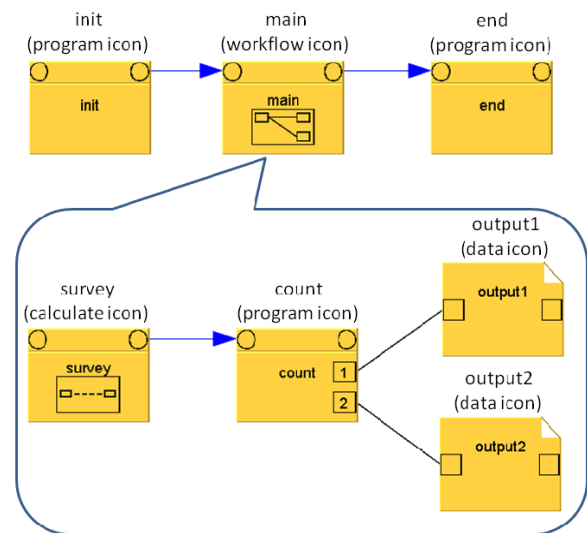


Figure 3: An example of a workflow

ponents are presented as program icon, calculate icon, data icon, and workflow icon in the GUI, respectively. The user defines the property for each icon. For example, a data icon has a property indicating a file path. The property of a program icon includes the path of a program executable, environment variables, and the computing resource to run the program. The property of a calculate icon includes a number to specify how many times a program should be executed. Working directories and input/output files can be named with a sequential number. It is useful for parameter survey application. Figure 2 shows a screenshot of the GUI for defining the property of a calculate icon. A sub-workflow contains other workflows, such that the WFT is able to create a hierarchical workflow structure. Complex workflow descriptions such as loops and conditional branches are also supported.

Figure 3 illustrates an example of a workflow with a two-level structure. The workflow at the outermost level is composed of two program icons, **init** and **end**,

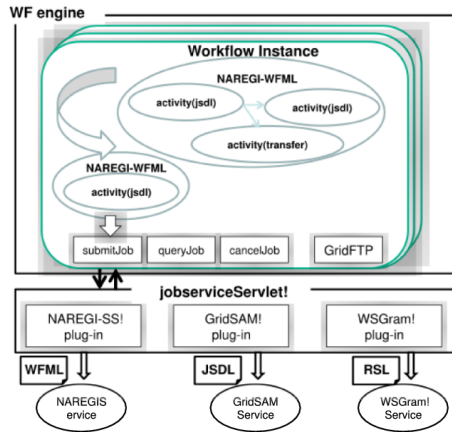


Figure 4: Workflow Engine

and a workflow icon, **main**. The workflow, **main**, consists of a calculate icon, **survey**, and a program icon, **count**, and two data icons, **output1** and **output2**. The lines and the lines with an arrow denote the data flow and the control flow between two icons, respectively. In this example, **init** runs first, then **main** runs after **init** finishes. In the execution of **main**, **survey** runs first, then **count** runs after **survey** finishes and writes data into **output1** and **output2**. Finally, **end** runs after **main** finishes.

### 3.2 Internal Representation of the Workflow

RENKEI-WFT enables users to submit jobs to computing resources operated by different middleware. Thus, an internal representation of the workflow in RENKEI-WFT should be architecture independent among the various middleware used by the computing resources. We designed the intermediate representation of the workflow, the NAREGI-Workflow Markup Language (NAREGI-WFML). NAREGI-WFML is extended from the Grid Service Flow Language (GSFL)<sup>[7]</sup> by adding a sub-workflow representation. The grammar of NAREGI-WFML is architecture independent, and NAREGI-WFML presents a job (or a program) in the Job Submission Description Language (JSDL).<sup>[8]</sup> JSDL is standardized in the Open Grid Forum (OGF)<sup>6</sup>, and many grid middleware services support JSDL to improve interoperability.

Figure 4 illustrates the structure of the Workflow Engine (WF engine) and the servlet to submit jobs (jobServiceServlet). The WFT creates a workflow instance per each workflow and puts the workflow into the threads pool, that is, multiple workflows are executed in parallel using multithreading. We use the ExecutorService interface in Java for thread management. For each workflow, the WF Engine decomposes the

<sup>6</sup><http://www.ogf.org/>

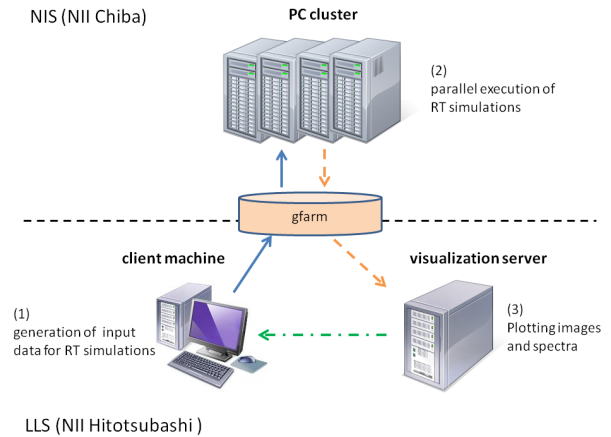


Figure 5: Schematic diagram of system configuration. Arrows indicate directions of data transfer.

workflow represented by NAREGI-WFML into sub-workflows consisting of computing jobs and file transfer jobs. Then, it submits jobs in the sub-workflows to computing resources in the LLS or the NIS through the jobServiceServlet. The WF Engine creates a sub-workflow of one activity (or one job) for a job submitted to GridSAM or WS GRAM. The GridSAM plug-in module extracts a JSDL document from the NAREGI-WFML document and submits the job to the GridSAM Service in the LLS. The WS GRAM plug-in module extracts a JSDL document, translates it to a Resource Specification Language (RSL)<sup>7</sup> document and submits it to the WS GRAM Service in the LLS. NAREGI supports NAREGI-WFML. The WF Engine creates sub-workflows of multiple activities, and the NAREGI plug-in module submits the sub-workflow in NAREGI-WFML to the NAREGI service. File transfer jobs in the LLS are executed directly using the GridFTP protocol. The WF engine monitors the status of submitted jobs by using the polling mechanism.

### 3.3 Authentication

Authentication is an important issue when utilizing computing resources distributed over multiple organizations. RENKEI-WFT uses the Grid Security Infrastructure (GSI) as an authentication mechanism for submitting jobs to computing resources in the LLS and the NIS. We use the User Management Server (UMS) in NAREGI to share user certificates between the LLS and the NIS, where user certificates are archived in a certificate repository of the UMS.

When starting to use RENKEI-WFT, the user first creates a proxy credential from the user's certificate archived in the UMS. The proxy credential is stored in the MyProxy server<sup>8</sup>. As long as the proxy credential is

<sup>7</sup>[http://www.globus.org/toolkit/docs/2.4/gram/rs1\\_spec1.html](http://www.globus.org/toolkit/docs/2.4/gram/rs1_spec1.html)

<sup>8</sup><http://grid.ncsa.illinois.edu/myproxy/>

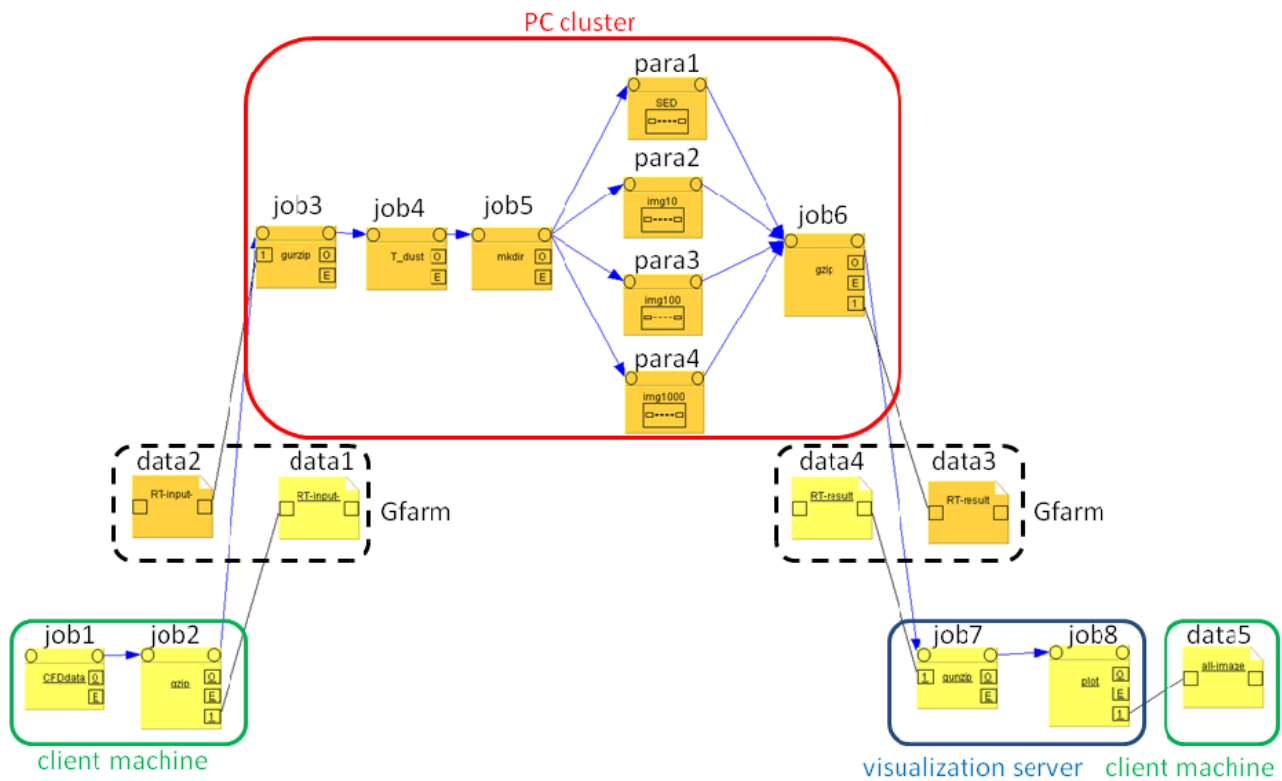


Figure 6: Workflow of our system. Program icons, calculate icons, and data icons are labeled with "job", "para", and "data", respectively.

valid, the user is able to sign on to the LLS-Portal. The LLS-Portal and the NAREGI-Portal are able to obtain the proxy credential from the MyProxy server. When the user submits a job, the proxy credential is delegated to the computing resource that runs the user's job.

The credential includes not only the basic authentication information but also the virtual organization (VO) information issued from the Virtual Organization Membership Service (VOMS).<sup>[9]</sup> The VO information is mainly used in the NIS operated by NAREGI.

## 4 Design and Implementation

Figure 5 shows a schematic diagram of system configuration. The system is composed of distributed computing resources that are located at NII Chiba site and NII Hitotsubashi site. Following application scenario described in section 3, we refer to computing resources in NII Chiba site and those in NII Hitotsubashi site as NIS and LLS, respectively. Input data for radiative transfer simulations are generated on client machine in LLS. Radiative transfer simulations run on PC cluster in NIS. Expected spectra and images are plotted with visualization software VisIt<sup>9</sup> on visualization server in LLS, and generated image files are sent to client machine. Gfarm file system is used to share files between

NIS and LLS.

Figure 6 shows workflow of our system. The workflow consists of data icons, program icons, and calculate icons. These icons are labeled with "data", "job", and "para", respectively. Workflow in Figure 6 is executed in the following steps:

1. Input data for radiative transfer simulations are generated on client machine in LLS (job1).
2. These input data are transferred to PC cluster in NIS via Gfarm. Client machine in LLS compresses data files by tar command with gzip option before transfer (job2), and write these files on gfarm (data1). Then, PC cluster in NIS reads these files on gfarm (data2) and decompressed them (job3).
3. Equilibrium dust temperature is computed by Monte Carlo method. RADMC-3D is executed with option "mctherm" (job4).
4. Ray tracing calculations are performed on PC cluster. Working directories for each job are constructed (job5). Spectral energy distributions (para1), intensity maps at 10  $\mu\text{m}$  (para2), 100  $\mu\text{m}$  (para3), and 1000  $\mu\text{m}$  (para4) with different viewing angles are computed in parallel. RADMC-3D is executed with option "sed" and "image".

<sup>9</sup><https://wci.llnl.gov/codes/visit/>

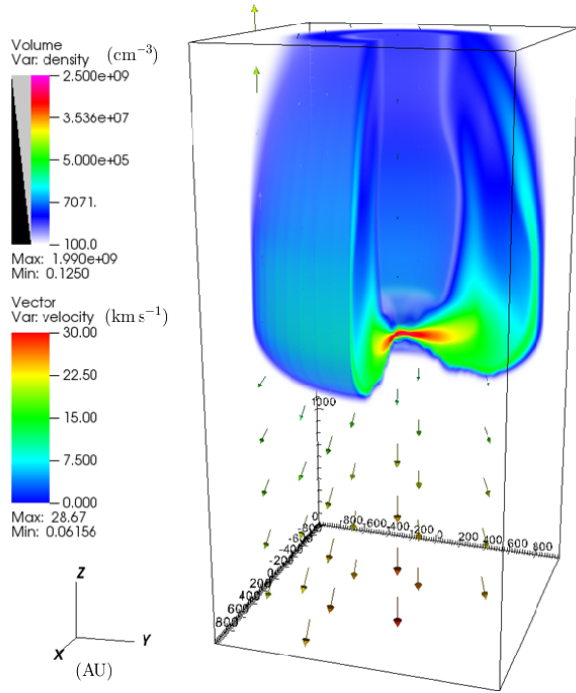


Figure 7: Volume rendering map of neutral gas and velocity field (arrows) at  $t = 4.5$  kyr.

5. Results of radiative transfer simulations are transferred to visualization server in LLS in the same manner as step 2 (job6, data3, data4, and job7).
6. Using results of radiative transfer simulations, spectra and images are plotted on visualization server (job8).
7. All image files generated on visualization server are directly transferred to client machine using the GridFTP protocol (data5).

For demonstration of our system, we simulated observations of photoevaporating protoplanetary disk using results of hydrodynamic simulation. A number of young stellar objects surrounded by ionized envelope have been detected in H II regions. These cometary-shaped objects are called "proplyd" and supposed to be photoevaporating protoplanetary disks irradiated by strong ultra violet radiation from nearby massive stars. We performed hydrodynamic simulation modeling photoevaporating protoplanetary disk. We assumed that disk is 1 pc away from exciting star whose mass and surface temperature are  $18 M_{\odot}$  and 30000 K, respectively. Mass of central star and initial mass of disk were set to be  $1 M_{\odot}$  and  $0.01 M_{\odot}$ , respectively. Figure 7 shows number density and velocity field at  $t = 4.5$  kyr. Ultra violet radiation enters computational domain from bottom boundary and propagates along the z-axis. We calculated density distribution of dust assuming gas-to-dust mass ratio of 0.01, and used it as input data for radiative transfer simulations.

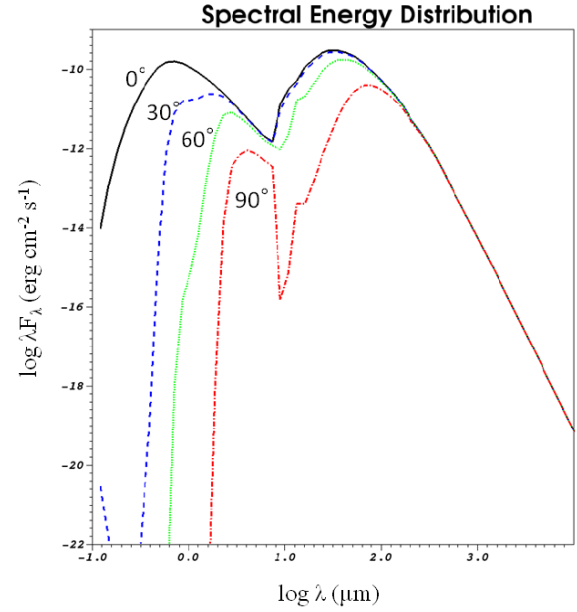


Figure 8: Spectral energy distributions calculated with viewing angle of  $0^{\circ}$  (solid),  $30^{\circ}$  (dashed),  $60^{\circ}$  (dotted), and  $90^{\circ}$  (dash-dotted).

Figure 8 shows some selected spectra obtained by our system. Viewing angle is defined as angle between direction to observer and the positive z-axis. Therefore, spectra with viewing angle of  $0^{\circ}$  are that expected when disk is observed face on. As shown in this figure, increasing of viewing angle decreases observed energy flux at short wavelength. This is because emission from hot inner region of the disk is obscured by cold outer region of the disk when observing edge on disk. Figure 9 shows some selected intensity maps at  $1000 \mu\text{m}$  obtained by our system. Increasing of viewing angle increases observed specific intensity, because dust continuum at this wavelength reflects column density along observed direction.

## 5 Summary

In this paper, we proposed the system for simulating dust continuum observations on distributed computing resources. Computational tasks and data transfer are automated with RENKEI-WFT in our system. Our work is important contribution because capability of parallel job execution of our system enables us to conduct effective research by radiative transfer simulation.

Although we focused on simulations of dust continuum observations in this paper, simulations of spectral line observations can be implemented in our system with a similar manner. Next step would be extension of our system to simulations of spectral line observations.

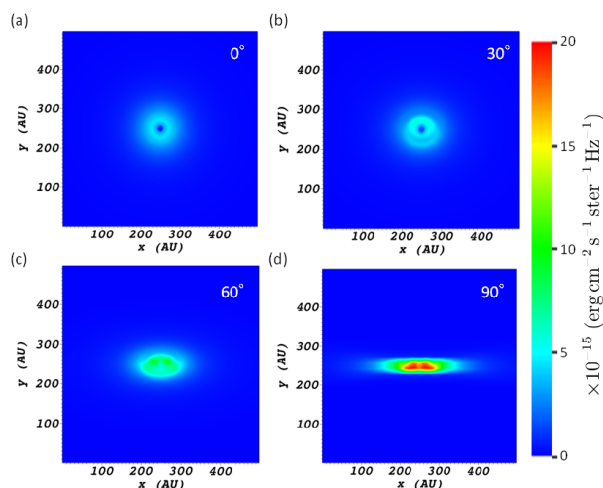


Figure 9: Contour maps of dust continuum intensity at  $1000\ \mu\text{m}$  with viewing angle of  $0^\circ$  (a),  $30^\circ$  (b),  $60^\circ$  (c), and  $90^\circ$  (d).

## Acknowledgment

This work is partially supported by the Ministry of Education, Culture, Sports, Science and Technology in Japan.

## References

- [1] G. B. Rybicki and A. P. Lightman. *Radiative processes in astrophysics*. Wiley-Interscience, 1979.
- [2] L. B. Lucy. Computing radiative equilibria with Monte Carlo techniques. *Astronomy and Astrophysics*, 344:282–288, April 1999.
- [3] J. E. Bjorkman and K. Wood. Radiative Equilibrium and Temperature Correction in Monte Carlo Radiation Transfer. *The Astrophysical Journal*, 554:615–623, June 2001.
- [4] H. Kanazawa, N. Onishi, Y. Mizusawa, T. Tsunekawa, and H. Usami. Application Hosting Services for Research Community on Multiple Grid Environments. *Journal of Convergence Information Technology*, 2010.
- [5] O. Tatebe, N. Soda, Y. Morita, S. Matsuoka, and S. Sekiguchi. Gfarm v2: A Grid File System That Supports High-performance Distributed and Parallel Data Computing. In *Proc. of the 2004 Computing in High Energy and Nuclear Physics (CHEP04)*, 2004.
- [6] W. Allcock. GridFTP: Protocol Extensions to FTP for the Grid. In *OGF Document CFD-R.20*, 2003.
- [7] S. Krishnan, P. Wagstrom, and G. von Laszewski. GFSL: A Workflow Framework for Grid Services, July 2002. draft.
- [8] A. Anjomshoa, F. Brisard, M. Drescher, D. Fellows, A. Ly, S. McGough, D. Pulsipher, and A. Savva. Job Submission Description Language (JSDL) Specification, Version 1.0. In *OGF Document CFD-R.136*, 2008.
- [9] R. Alfieri, R. Cecchini, V. Ciaschini, L. dell’Agnello, Á. Frohner, A. Gianoli, K. Lorente, and F. Spataro. VOMS, an Authorization System for Virtual Organizations. In *Proc. of the First European Across Grids Conference*, 2004.



