



ISSN 2433-2216(Online)
JAXA-RR-17-009

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 第七号

2018年3月

宇宙航空研究開発機構

Japan Aerospace Exploration Agency

目 次

1. 再現と共有を可能とするデータ可視化方法について
今井 弘二, 村山 泰啓, 海老沢 研, 池田 大輔, 瀬口 大介 1
2. DOI を用いた天文学研究資料の横断的な整理・公開に向けた取り組み
根本 しおみ, 高田 良宏, 堀井 洋, 堀井 美里, 飯野 孝浩, 林 正治 9
3. ビッグデータ太陽系天文学の創成のための ALMA キャリブレーション観測データの
悉皆的解析システムの検討と構築
飯野 孝浩, 並木 美太郎, 山田 崇貴 19
4. 時空間変動データからのホットスポット自動抽出・要約システムの開発
本田 理恵, 林 諒 33
5. 惑星の位置関係に基づく深発月震分類のための特徴量の検討
加藤 広大, 山田 竜平, 山本 幸生, 廣田 雅春, 横山 昌平, 石川 博 43
6. 「はやぶさ 2」地球スイングバイに係る見かけの等級の変化推定
三浦 昭, 山本 幸生, 吉川 真 53

再現と共有を可能とするデータ可視化方法について

今井 弘二^{*1}, 村山 泰啓^{*1}, 海老沢 研^{*2}, 池田 大輔^{*3}, 瀬口 大介^{*3}

Reproducible and shareable data visualization method

Koji Imai^{*1}, Yasuhiro Murayama^{*1}, Ken Ebisawa^{*2}, Daisuke Ikeda^{*3}, Daisuke Seguchi^{*3}

Abstract

Data management technology is becoming more and more important to promote scientific development in the society brimming with data. We improved the data visualization web service of earth, planetary and space sciences (Cross-Cutting Comparisons; C3) as the system controlled by the human-understandable query string (QS) to make reproducible and shareable charts. By including information of data handling procedures in the QS in an orderly manner, the chart is easy to understand, remake and share via text-based communication tools.

Keywords: reproducibility, data visualization, geoscience, cross-cutting research

概要

データが希少であった時代から溢れる時代となり、データマネジメント技術がますます重要視されるようになってきている。我々は、地球惑星科学におけるウェブサービス (Cross-Cutting Comparisons ; C3) を人間可読なクエリ文字列でコントロールするシステムとすることで、図やグラフなどのチャートを再現・共有できる新たな方法を実現した。クエリ文字列には、データの選択や処理手順などが明記されているため、チャートの理解が容易であり、また、同じクエリ文字列をテキストベースのコミュニケーションツールを利用すれば、同じチャートを再現することも可能である。

キーワード データマネジメント, データの再現性, 可視化, 地球惑星科学, C3

doi: 10.20637/JAXA-RR-17-009/0001

* 平成 30 年 1 月 18 日受付 (Received January 18, 2018)

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

^{*2} 宇宙航空研究開発機構 宇宙科学研究所
(Japan Aerospace Exploration Agency, Institute of Space and Astronautical Science)

^{*3} 九州大学大学院システム情報科学 研究院・学府
(Graduate School of Information Science and Electrical Engineering, Kyushu University)

1 はじめに

科学技術の進歩によって、観測機器やセンサーネットワーク、計算機が小型化・高精度化・低コスト化され、膨大な量のデータが収集できるようになり、これにもとづく新しい研究や科学の発展が求められている。そして、それらのデータ資源の公開や共有などの取り組みとして、「オープンデータ」、あるいは「オープンサイエンス」が叫ばれるようになってきた¹⁾。しかし、それぞれの科学技術分野におけるデータを、国境や分野間の隔たりがなく、シームレスに活用するための社会を実現するためには、技術的、政策的、文化的、人材育成面といった乗り越えるべき壁が幾重にもあるため、研究機関や学協会などによるボトムアップな努力に加え、国際組織（Research Data Alliance, ICSU-WDS, OECD, G8/G7等）がトップダウンの政策を視野に入れて中長期的な課題解決を目指して取り組んでいる^{2, 3)}。

これらの国際組織による議論では、一拠点にデータを保有・管理する集中型から、各国の各機関がデータを保有・管理しながら相互に利用する、分散協調型のグローバルなデータマネジメントの体制を確立することが重要になると考えられている。科学技術分野では、これまでに個別のデータごとに対して、処理や可視化をするシステムが開発されてきた傾向があるが、分散協調型のグローバルなデータマネジメントの体制においては、分散配置された異分野のデータを統一的に取り扱うシステムがこれまで以上に求められると考えられる。また、近年では、第三者が再現できない論文が指摘されており、科学技術分野の発展の大きな妨げになっている^{4, 5)}。したがって、次世代のデータマネジメントシステムにはデータやチャート（ここでは、図やグラフなどの情報の視覚表現をまとめてチャートとよぶ）が、どのように処理され、どのように可視化されたかなどの作業過程を確認できる仕組みが必要である。

我々は遠隔のサーバからデータを取り込み、ブラウザ上にチャートを生成するシステムを開発した。チャートにはデータの選択や処理手順などを明示的に指定するクエリ文字列が付与されており、チャートの管理・共有が従来と比べて容易になると期待される。本稿では、第2章にこの新しい仕組みを実現した地球惑星科学のデータ可視化サービス C3（Cross-Cutting Comparisons）^{6, 7, 8)}の通信処理の流れや活用例について、第3章に本研究のまとめを記述する。

2 C3 について

C3 は太陽から人類生存圏までの空間における様々な現象が、どのように影響を及ぼしているのかを総合的に理解するために、地球惑星科学における分野横断型の複合研究を支援するウェブサービスである。多種多様なデータ形式を取り扱っており、データマネジメントシステムを実践・評価するテストベッドとして最適である。この章では、データやチャートの再現や共有を可能としたシステムの通信処理の流れや、活用例を中心に記述する。

2.1 通信処理の流れ

C3 のシステムは、ウェブサーバとデータサーバで構成している (図 1)。ユーザは、C3 のウェブページにアクセスし、インタラクティブな入力フォームに従って必要な情報を入力/選択することで、ブラウザ上にチャートを作成することができる。

この時、ユーザのリクエストはクライアント側で確認され (図 1 の (1))、適切なクエリ文字列が動的に生成される (クエリ文字列の文法については、2.2 節を参照)。そして、生成されたクエリ文字列はウェブサーバに送信され (図 1 の (2))、CGI プログラムによってユーザのリクエストに対するデータ処理・解析が実行される (図 1 の (3))。

その後、ウェブサーバからのレスポンスとして、データとチャートを作成するためのウェブコンポーネント (例えば、HTML や JavaScript のプログラム、CSS スタイルシートなど) がクライアントに送信され (図 1 の (4))、ブラウザ上にチャートが作成される (図 1 の (5))；C3 が提供するチャートの例については、付録の図 A を参照)。このように、ウェブサーバはクライアントとデータサーバの仲介役として位置づけられており、ユーザの反応に応じて、その都度、ウェブサーバがレスポンスを返すシステムではなく、クライアントとウェブサーバ間の通信を、1つのリクエストに対して1回のレスポンスで対応することによって、ウェブサーバの負荷を低減させるとともに、ネットワークの通信速度の遅延も避けている。

データサーバは、内部用と外部用に分けている。第 1 章に記述したように、各機関の膨大なデータを一つのデータサーバで一括に管理することは非現実的であるため、比較的容量の小さいデータセットは内部のデータサーバで保有し、容量の大きなデータセットは外部のデータサーバから、ユーザのリクエストに応じて取得している。

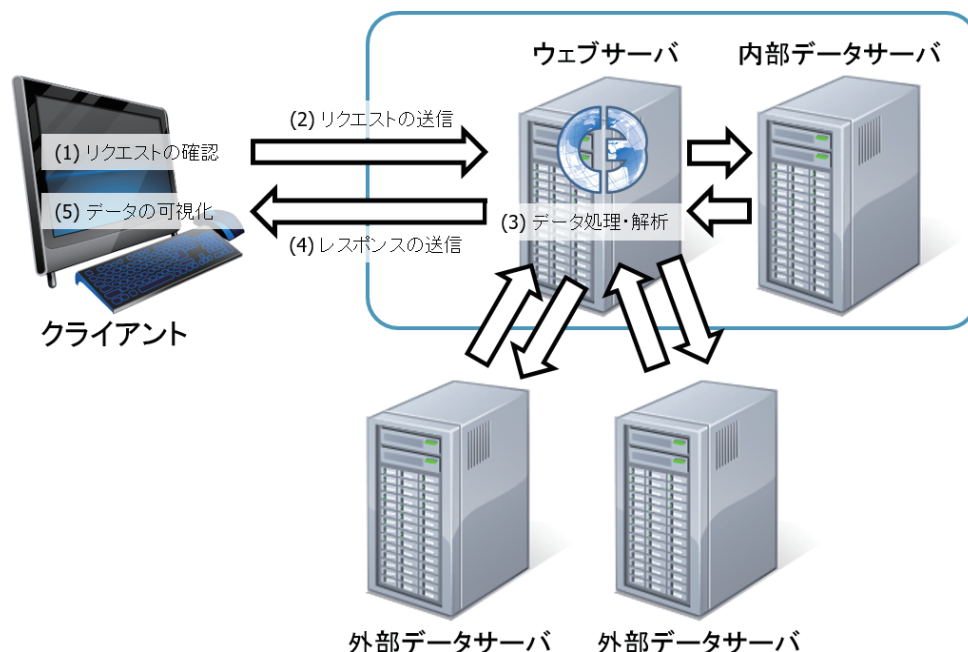


図 1 通信処理の流れ

図中の通し番号は、ユーザがチャートを作成するまでの一連の順序である。

2.2 クエリ文字列の取り扱い

地球環境における長期変動や突発的な現象の特徴を捉えるためには、データを時空間的に柔軟に表現して調査する必要がある。C3 は、HTML と JavaScript のライブラリを利用し、地球惑星科学における多種多様なデータ形式をブラウザ上に柔軟に表現している (図 A)。

ハイパーテキストシステムである WWW (World Wide Web) では、クエリ文字列は URL (Uniform Resource Locator) の一部であり、通常はクエスチョンマーク「?」に続けて記述される。そして、C3 の場合は、チャートを作成するための情報をクエリ文字列として次のように定義している。

```
C3_XXXX.cgi?en+SORCE_TSI_17_timeSeries+20100101_20100201
```

C3_(システムのバージョン).cgi?**ヘッダ部**+**データ選択部**+**データ抽出部**+オプション部

上の例は、SORCE (Solar Radiation and Climate Experiment) ⁸⁾ が捉えた全太陽放射照度 (Total Solar Irradiance ; TSI) のバージョンが 17 のデータを時系列図で、2010 年 1 月 1 日から 2010 年 2 月 1 日まで表示するためのクエリ文字列である (チャートは図 A (a) を参照)。クエリ文字列は、プラス「+」で分けられた 4 つの部分で構成している。先頭はヘッダ部であり、その後にデータ選択部、データ抽出部、そして必要であればオプション部が続く。

ヘッダ部には、チャートの表示言語などの情報が、データ選択部には、データのメタ情報と表示するチャートの種類が記載されている。そして、データ抽出部に選択したデータの期間や緯度、経度、高度などの抽出条件が記載されている。最後のオプション部には、チャートのタイトルやサブタイトル、データの表示形式の切替情報などが付加されている。このように、クエリ文字列にデータの選択や抽出条件などの情報を入れて、人間可読な形式とすることで、作成されるチャートの理解が容易となり、ユーザによる編集も可能となる。

C3 では、システムのバージョンによって、データの処理と可視化を、データのバージョンによって、データサーバにあるデータを制御しているため、同じクエリ文字列を含む URL (以下、URL クエリ文字列と呼ぶ) を用いれば、同じチャートが再現される。ただし、外部にあるデータは管理していないため、同じチャートを保証するわけではない。

2.3 活用例

C3 が提供する URL クエリ文字列は、選択するだけでブラウザ上にチャートを再現することができるため、URL クエリ文字列を管理しておけば、画像ファイルを保有する必要がなく、チャートを多用する研究でも、ユーザが管理するデータサイズを抑えることができる。さらに、テキストベースのコミュニケーションツール (例えば、E メールやツイッターなど) を使って同じ URL クエリ文字列を第三者にも共有することができるため、ユーザ間の通信

量を低減する新たなチャートの共有方法としても期待される (図 2)。

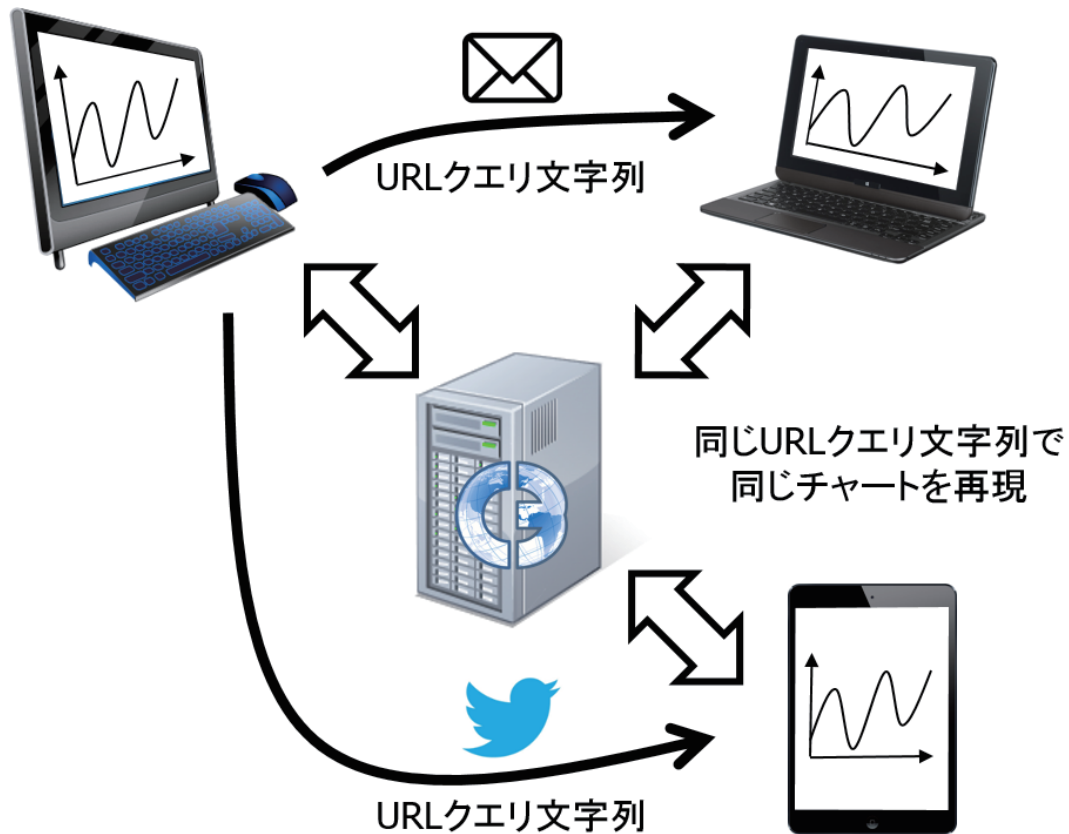


図 2 URL クエリ文字列を用いたチャートの共有

3 まとめ

データがあふれる現代となり、信頼のおけるデータを共有できるデータマネジメントを確立することが重要視されている。我々は、地球惑星科学における多種多様なデータを扱うウェブサービス C3 を、人間可読なクエリ文字列でコントロールするシステムに改良した。これによって、クエリ文字列からデータの選択や処理手順などが理解でき、ユーザによる編集も可能となった。また、テキストベースのコミュニケーションツールを利用してクエリ文字列を共有すれば、同じチャートが再現・共有できるだけでなく、ユーザ間の通信量も低減されるため、新たなチャートの共有方法として期待される。

謝辞

本研究は宇宙航空研究開発機構・宇宙科学研究所，科学衛星運用・データ利用センター (C-SODA)，JSPS 科研費 15H02787 の助成を受けたものです。ここに感謝の意を表します。

参考文献

- 1) Huijboom, Noor, and Tijs Van den Broek. "Open data: an international comparison of strategies." *European journal of ePractice* 12.1 (2011): 4-16.
- 2) 渡邊堯. "ICSU 世界データシステム (WDS) について." *学術の動向* 17.6 (2012): 6_11-6_15.
- 3) ICSU World Data System, <https://www.icsu-wds.org> (参照日 2017.11.20).
- 4) Baker, Monya. "1,500 scientists lift the lid on reproducibility." *Nature* 533.7604 (2016): 452-454.
- 5) Iqbal, Shareen A., et al. "Reproducible research practices and transparency across the biomedical literature." *PLoS Biol* 14.1 (2016): e1002333.
- 6) Imai, Koji, et al. "Quick look service for geoscience." *Journal of Space Science Informatics Japan*, Vol. 5 (2016): 93-109.
- 7) 今井弘二, 村山泰啓, 松下幸市朗. "地球科学における対話形式のデジタル教材の開発とその評価の試み.", *教育情報研究*, 33 巻 1 号.
- 8) Cross-Cutting Comparisons, <https://darts.isas.jaxa.jp/C3/> (参照日 2017.11.20).
- 9) Rottman, Gary. "The SORCE mission." *Solar Physics* 230.1-2 (2005): 7-25.
- 10) Bernath, Peter F., et al. "Atmospheric chemistry experiment (ACE): mission overview." *Geophysical Research Letters* 32.15 (2005).
- 11) Kikuchi, Ken-ichi, et al. "Overview and early results of the Superconducting Submillimeter - Wave Limb - Emission Sounder (SMILES)." *Journal of Geophysical Research: Atmospheres* 115.D23 (2010).
- 12) Imai, Koji, et al. "Validation of ozone data from the Superconducting Submillimeter - Wave Limb - Emission Sounder (SMILES)." *Journal of Geophysical Research: Atmospheres* 118.11 (2013): 5750-5769.
- 13) U. S. Geological Survey, <https://www.usgs.gov/> (参照日 2017.11.20).

付録 A

図 A は C3 が提供するチャートの例である。図 A (a) は $SORCE^8$ の全太陽放射照度の時系列図であり、データ点にマウスオーバーして（マウスカーソルを重ねて）、ポップアップで詳細な情報を表示している様子を示している。図 A (b) は ACE-FTS (Atmospheric Chemistry Experiment-Fourier Transform Spectrometer) ⁹⁾ の温度の高度分布図であり、ドラッグして拡大表示している様子を示している。図 A (c) は SMILES (Superconducting Submillimeter-Wave Limb-Emission Sounder) ^{10, 11)} のオゾンの全球分布図である。そして、図 A (d) は USGS (United States Geological Survey) ¹²⁾ の震源図である。データ点にマウスオーバーして詳細な情報を表示している。チャート自体をドラッグして回転させることも可能である。このように、C3 は点、線、面、そして疑似的な 3 次元まで表現し、自在なスケール調整やデータの詳細表示など、多機能なチャートを提供している。

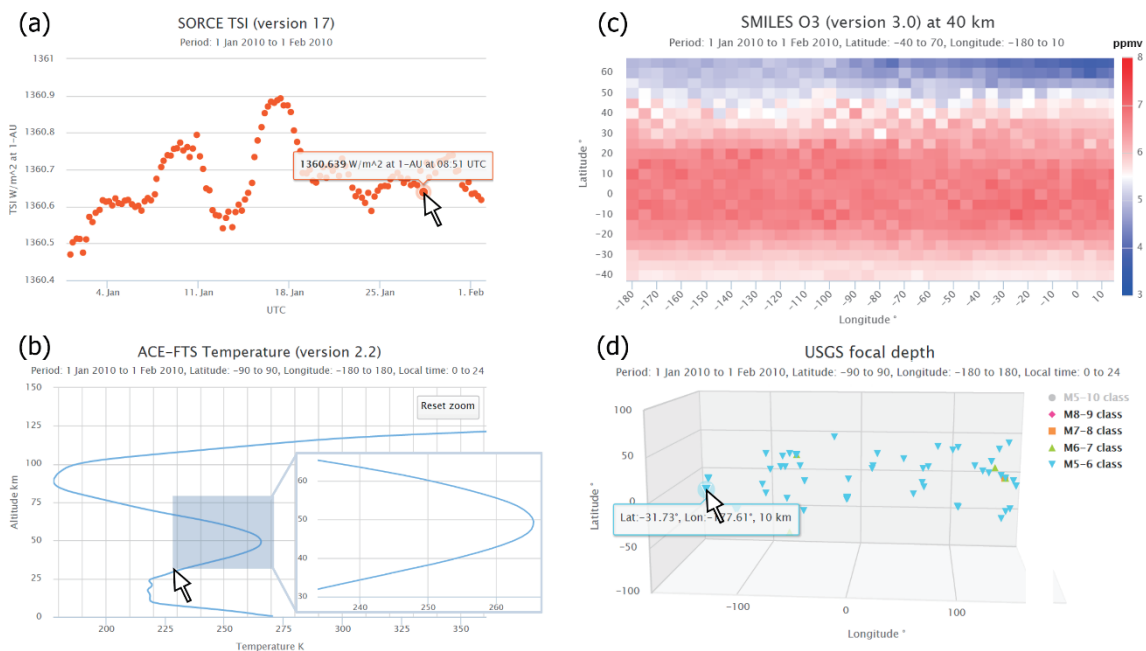


図 A 2 C3 が提供するチャートの例

- (a) 全太陽放射照度の時系列図, (b) 温度の高度分布,
(c) オゾンの緯度経度断面図, (d) 震源図

DOI を用いた天文学研究資料の横断的な整理・公開に向けた取り組み —岡山天体物理観測所を一例として—

根本しおみ^{*1}, 高田良宏^{*2}, 堀井洋^{*3}, 堀井美里^{*3}, 飯野孝浩^{*4}, 林正治^{*5}

An attempt for cross-cutting organization and publication of astronomical research materials using DOI: A case study of Okayama Astrophysical Observatory

Shiomi NEMOTO^{*1}, Yoshihiro TAKATA^{*2}, Hiroshi HORII^{*3}, Misato HORII^{*3},
Takahiro IINO^{*4} and Masaharu Hayashi^{*5}

Abstract

Astronomical research materials such as observing instruments, photographic plates and blueprints are important materials that records the history of the improvement of not only astronomy but also modern natural science. These materials are stored in many universities and research institutes with their manners. For the cross-cutting materials organization, we addressed to construct a “subject repository” that uses Digital Object Identifier(DOI). For the first attempt, we tried to arrange materials stored in Okayama Astronomical Observatory. In this paper, first results of the arrangement of astronomical research materials and meta data inherence using Dublin Core method.

概要

天文学研究資料は機器や乾板、設計図等多岐にわたっており、我が国の天文学や自然科学の発展を記録した重要な資料である。また、これら資料群は国内の大学や研究機関に広く保管・収蔵されている。資料の種類や機関を横断した資料情報の整理の手法として、我々はデジタルオブジェクト識別子 (DOI) を用いた「サブジェクトリポジトリ」の構築を目指している。その嚆矢として、国立天文台岡山天体物理観測所の資料群を対象に、天文学研究資料の分類を試みた。本稿では天文学研究資料の分類や、Dublin Core を用いたメタデータ付与について報告する。

Keywords: 天文学研究資料リポジトリ, サブジェクトリポジトリ, デジタルオブジェクト識別子 (DOI), 岡山天体物理観測所

doi: 10.20637/JAXA-RR-17-009/0002

* 平成 30 年 1 月 18 日受付 (Received January 18, 2018)

^{*1} 国立天文台 (National Astronomical Observatory of Japan)

^{*2} 金沢大学 (Kanazawa University)

^{*3} 合同会社 AMANE (AMANE.LLC)

^{*4} 東京農工大学 (Tokyo University of Agriculture and Technology)

^{*5} 国立情報学研究所 (National Institute of Information Science)

1. はじめに：天文学研究資料の永続的な保存の意味

国内各地の天文関連施設・学術機関には、過去の観測や研究に使用された観測機器や乾板・フィルムなどの天文学研究資料が多数現存している。それらは、本国の天文学の発展に資する重要な学術資源であり、確実な保存と次世代への継承が求められている。天文学研究資料を保存・継承することの意義は、①過去に実施された観測・研究の妥当性・有意性を検証するための根拠、②過去の成果を最新の技術を用いて再度分析・検証することによる、新たな学術的知見の獲得、③教育や公開を通じた、科学技術について広く社会へ普及・啓蒙、などが挙げられる。天文学分野では、常に最新の科学技術が研究・観測に適用され、かつ観測プロジェクト単位で観測機器を独自に製作・運用することが多いことから、科学技術史的な観点から、観測データと併せて観測機器やマニュアル等の関連資料についても記録・保存することが求められている。

関連する近年の取り組みとして、宇宙科学研究所（ISAS）が所有する写真や各種資料の整理・公開の取り組み「宇宙科学資料データベース」が先端的事例として挙げられる（川上ら，2016）。同所の保有する80万枚にも及ぶネガ・デジタル写真や実験計画書等、60年間の研究活動に際して産生された資料群にデジタル化とメタデータ付与を施し、同所のデジタルアーカイブとして公開を行っている。さらに、天文学研究においては、写真・文書とともに観測装置等の機器資料も重要である。国立天文台では保有する機器群を整理するアーカイブ室を天文情報センター内に2008年に設立（現在は閉室）、各種観測・測定装置の発掘・復元や、行方不明の装置の捜索・回収、そして写真資料等のデジタル化を行った。また、研究過程の記録として重要なオーラルヒストリー記録の取り組みも行われつつある（高橋ら，2017）

さらに、多くの先行研究では資料群は機関毎に整理・情報公開がなされてきたが、整理対象とすべき同様の貴重な資料群は国内の大学・研究機関等に多数存在していると考えられる。著者らは、これまでに自治体や大学等と連携して、地域に現存する歴史資料の調査・デジタル化・公開に取り組んできた（堀井ら，2012）。2014年に一般社団法人学術資源リポジトリ協議会¹（以下、本協議会）を設立し、歴史資料など“非文献資料”を含む学術資料全般を対象にした組織・分野横断的な情報共有を目指している。その一環として、明治・大正期の日本において、教育・研究目的で使用された科学実験機器および教育掛図資料の資料情報に対して、デジタルオブジェクト識別子(DOI: Digital Object Identifier)²を付与する試みを実施してきた（堀井ら，2012）。個々の資料情報に対して、固有識別子を付与す

¹ <http://www.repon.org> , 一般社団法人学術資源リポジトリ協議会（2017年6月30日確認）

² <https://www.doi.org>, the International DOI Foundation（2017年6月30日確認）

ることにより、資料の保存・整理の観点からは資料の存在肯定の手段が確立され、活用の観点からは活用成果と学術情報を結びつける効果・役割が期待される。

科学実験資料を含む学術資料を“情報資源化”し、それらを広く社会に対して公開・利用促進するオープンサイエンスの推進は、世界の潮流となりつつある。国内においても、最終成果である論文等成果に加え、その根拠となる研究資料、さらには、研究過程、もしくは、研究環境を示す資料も併せてのリポジトリ化・オープン化が急速に進んでいる。しかしながら、天文学分野では、常に最新の科学技術が研究・観測に適用され、かつ観測プロジェクト単位で観測機器を独自に製作・運用する、いわゆる一品物の観測機器等が多いことから、観測データを再利用するには併せて観測機器やその図面、マニュアル等の関連資料についてもリポジトリ化・オープン化することが求められる。

2. 固有識別子 DOI を介した資料間の関連付けと資料情報の公開

本研究では、天文学資料およびその関連資料のデジタルデータ（画像データ等）とそのメタデータのセット（コンテンツ）に対して、固有識別子である DOI を付与する。前述した通り、天文学資料には、観測データや論文等の学術成果物と併せて、観測機器や関連資料・関係者へのインタビューなど、研究活動の全貌を記録・検証するために不可欠な多種多様な資料が含まれる。図 1 に DOI を介した資料情報の関連付けに関する概要を示す。

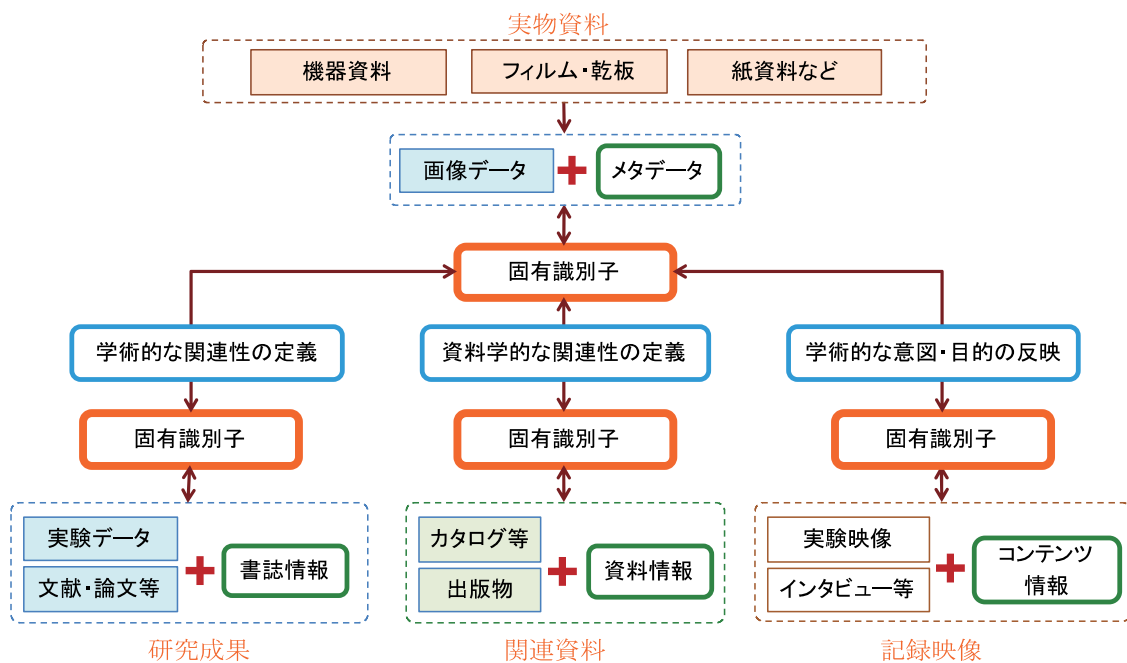


図 1: 固有識別子を介した資料関連付けの概要

天文学研究資料に対してDOIを付与する意義は、まず第1に多様な資料・文献間における参照関係の明確化である。論文や著書などの文献上で、天文学研究資料などの資料を参照する場合には、これまで所蔵機関名および資料名を記述することが一般的であった。しかし、参照された資料を特定し、資料情報への参照を行う場合には、所蔵機関へ問い合わせるか、過去に刊行された目録を閲覧する必要があった。これに対して、資料情報にDOIを付与することによって、物・紙・電子データなど多様な資料間の関係性を電子的に定義することが可能となり、複数の資料からなる学術的な意味を有した“群”の構築と公開が実現される。第2に、DOIを固有IDとして採用したことで、論文など文献資料と同様に横断検索の対象となることである。これにより、組織横断的な資料比較研究や、文献資料と資料情報を一体的・総合的に扱う新たな研究環境の構築が実現される。現在、学術機関リポジトリデータベース（IRDB）³は、国内の機関リポジトリからメタデータを収集し、論文検索サービスCiNii（<http://ci.nii.ac.jp/>）や国立国会図書館との連携を実施しているが、それらを利用した隣接・他分野との情報連携についても可能となる。

第3に、天文学研究資料の存在肯定手段の確立である。現在、認知されている天文学研究資料の大部分が大学や天文台の所蔵機関に所蔵されているが、目録や管理台帳等に資料情報が記載されていない、所謂“未整理資料”が少なからず存在することは事実である。これまでは、資料台帳や目録などの紙媒体への資料情報の記載が資料の存在肯定における代表的な手段であったが、これに加えて、資料情報へのDOI付与とその電子的な蓄積・公開が資料の存在を社会的に認知する手段となることが期待される。

本研究では、天文学研究資料情報を“天文学資料リポジトリ”として公開することを目指している。次章において、岡山天体物理観測所を事例として、具体的な資料の概要について述べる。

3. 岡山天体物理観測所における天文学研究資料調査の概要

ここからは、本研究の要点である天文学研究資料調査の一環として実施した、国立天文台岡山天体物理学観測所における資料整理について記す。同観測所は1960年に東京大学東京天文台の施設として発足し、現在は188 cm(74インチ)、91 cm、50 cm反射望遠鏡、65 cmクーデ型太陽望遠鏡が設置されている。中でも、発足当時に建設された口径188 cm反射望遠鏡は当時国内最大（世界7位）の口径を誇り、さまざまに測器を載せ替えながら、現在に至るまで光赤外の観測的天文学研究を推進してきた。

³ https://www.nii.ac.jp/irp/archive/system/irdb_harvest.html, IRDB ハーベストについて(2017年6月30日確認)

188 cm 鏡にはこれまで多様な測器が開発・搭載されている。天体の観測には主に撮像、分光、偏光、測光などの手法が用いられるが、これらの手法の発展に合わせ、同鏡には多種多様な検出器が開発・搭載されてきている。撮像観測を例にとると、現在は CCD を始めとした電子的手法が用いられるが、開所から 1980 年代までは写真乾板が用いられていた。また、天文学研究における重要な手法である分光観測に用いられる分光器は、1960 年代から 20 年にわたって用いられた Glass 分光器にはじまり、1999 年から現在まで用いられる高分散分光器 HIDES など、10 種を越える機器が搭載されてきた。観測対象の天体も変遷があり、初期には恒星の分光観測が広く行われていたが、現在は系外惑星へと広がりを見せている。同所は国内において現役で利用されている大学共同利用機関の中では有数の歴史を持っており、特に同一の測器を使いながら第一線級の成果を挙げ続けてきていることに特色がある。同所が所有する多様な資料のリポジトリ化は、2 章に示した天文学研究資料の保存・継承のみならず、国内における他の自然科学分野のリポジトリ化に際しても参考事例として重要になると考えられる。

同所には開所当時からの観測機器、観測野帳、マニュアル類、写真乾板、図面などが所蔵されており、本研究ではこれらの天文学研究資料を対象に、調査と資料情報の作成を実施した。各資料の概要を表 1 に示す。さらに、図 2 から図 5 に各資料の外観を示す。

表 1：調査対象資料の概要

| 資料種別 | 点数 | データ形式 | 内容 |
|-------------------|--------------------|-------|----------------------|
| 観測野帳 | 1 点 (107 ページ) | PDF | 観測原簿 vol.69 (1975 年) |
| 機器マニュアル (望遠鏡・分光器) | 3 点 (計 336 ページ) | PDF | 74 インチ望遠鏡マニュアル他 |
| 図面 (望遠鏡・設備) | 41 点 | JPEG | 望遠鏡本体・動設備・ドーム設計図面他 |
| ガラス乾板 | 837 点 | fits | 岡山理科大加藤研究室 |
| インタビュー | 2 点 | MOV | 前原英夫氏, 中桐正夫氏 |

観測ログを記載している観測野帳は、当時の観測対象、観測手法だけでなく、観測手順や戦略についても窺い知ることのできる貴重な一次資料である。観測野帳は望遠鏡ごとに存在し、そこには観測者の氏名、観測に使われた機器、観測対象の天体名、撮像の露出時間やその間の天候の状態などが記載されており、観測が行われた過程や環境を知る上で

貴重な資料である。本研究では、多数現存している観測野帳から事例として1点を抽出し、デジタル化を実施した。

機器マニュアルについては、188 cm 反射望遠鏡が岡山に導入された当時のものを含む3点について、デジタル化を実施した。前述したとおり、天文観測機器では、分光器などの一部を最新式に更新しつつ、長期間使用することがあるため、導入当時の機器に関するマニュアル資料を把握することには、大きな意味がある。観測野帳および機器マニュアルについては、冊子体のため接写台上でデジタル撮影を行い、PDF形式で公開を行う。

また、図面資料については、建設当時およびその後の改築時に作成されたものについて、デジタル化を実施した。デジタル化に際しては、資料のサイズが大型かつ表記が細かいため大判カメラを用いて撮影し、JPEG形式の画像データを生成した。

ガラス乾板については岡山理科大学の加藤研究室によってデジタル化したデータを提供していただいた。加藤研究室では、市販のスキナー（EPSON GT-X980）を用いて岡山天体物理観測所のガラス乾板のデジタル化を行なっている。スキャン密度は2400dpi、.fits形式で画像データを保存している。

インタビューは、元岡山天体物理観測所所長の前原英夫氏と、元岡山天体物理観測所職員の中桐正夫氏について行なった。編集後、1時間から1時間半程度のビデオ画像とした。

なお、岡山天体物理観測所は、2018年度より組織改編が行われ、規模が縮小される。現在、同観測所で保管されている写真乾板や野帳、古い機器のマニュアル類は、三鷹へ移動し、保管する予定である。

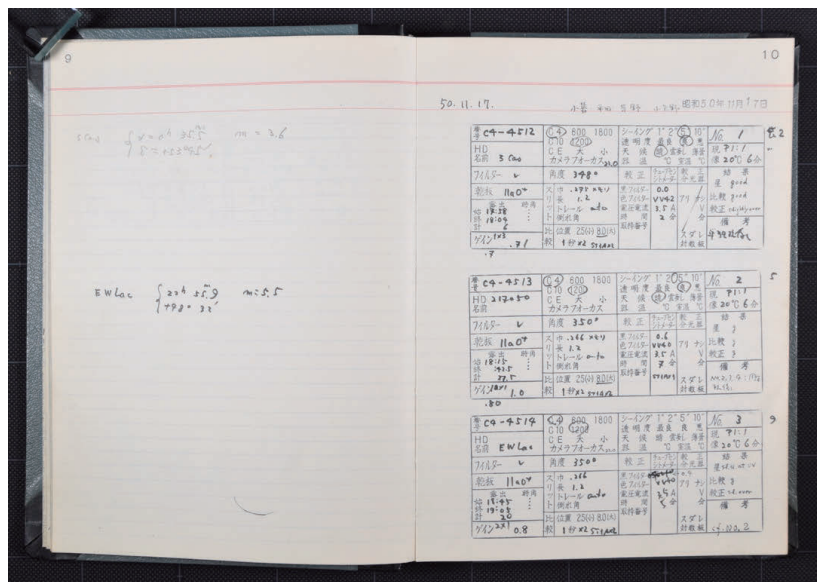


図2：観測野帳資料の例（資料名：観測原簿 vol.69）

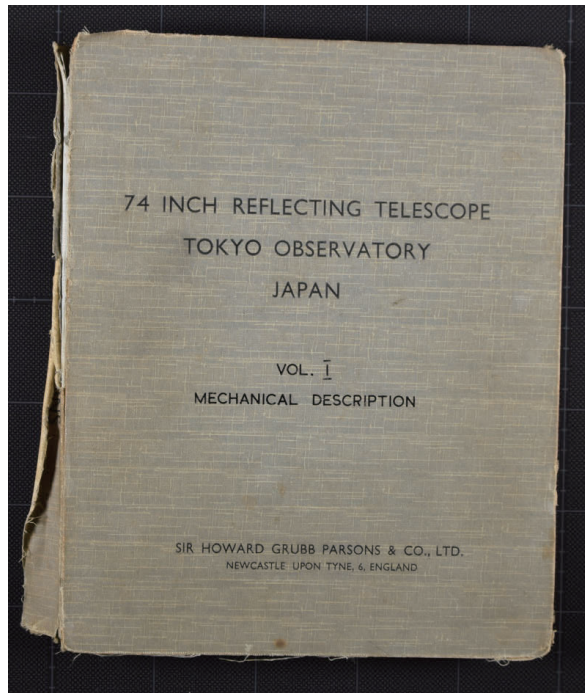


図 3 : 文書資料の例. 188cm 望遠鏡を製造したイギリスの企業によるマニュアル
(資料名 : 74INCH REFLECTING TELESCOPE TOKYO OBSERVATORY JAPAN
vol. I MECHANICAL DESCRIPTION)

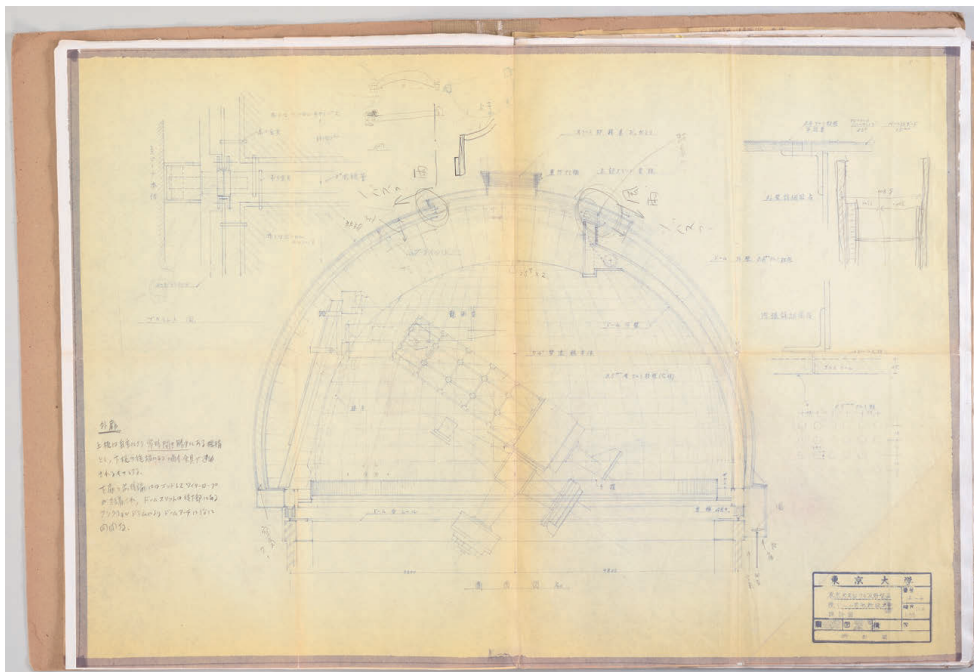


図 4 : 図面資料の例
(資料名 : 東京天文台 74"反射望遠鏡ドーム其他新設工事設計図 14-6 背面図)

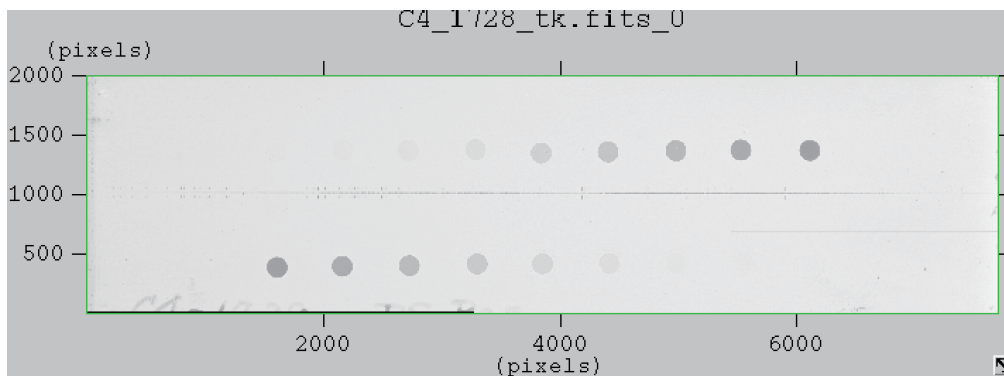


図 5：ガラス乾板資料の例（資料名：DS Peg. のスペクトル）

4. Dublin Core を用いた天文学研究資料へのメタデータ付与

本研究では天文学研究資料情報の記述のために、ネットワーク上の文献等の情報記述を目的に定義された Dublin Core メタデータ（以下、DC）⁴ へ対応させることを試みた。その理由は以下の 3 点である。

1. DC はネットワーク上のリソース記述のための汎用メタデータフォーマットとして設計されており、共通メタデータ形式として広く認知されている。また、IRDB が収集するメタデータ形式 `junii2` (<https://www.nii.ac.jp/irp/archive/system/junii2.html>) は DC に準拠している。
2. 独自拡張語彙の定義など、非文献資料情報の特性を反映した拡張が可能である。
3. DC に対応した既存の文献資料リポジトリシステムの改修や将来的な文献資料リポジトリとの統合が期待できる。

表 2 に天文学研究資料に関する記述例の一部を示す。非文献資料メタデータの DC への対応については、これまで複数の研究が行われており、博物資料などの非文献資料資料に対して DC の項目が満足に適合するかが問題である。本研究では、`title` および `description` を中心として拡張語彙を定義し、特に `description` については、非文献資料に含まれる意味的な情報を注記として記述し、性質が異なる資料の情報の差異を吸収させた。

⁴ <http://dublincore.org/documents/dces/>, Dublin Core Metadata Element Set, Version 1.1: Reference Description (2017 年 11 月 19 日確認)

表 2 : 天文学研究資料に関するメタデータ記述例 (一部)

| DC | 項目名 | 資料 1 | 資料 2 | 資料 3 |
|---------------------------|-----------------|---|--|---------------------------------|
| title | タイトル | 74INCH REFLECTING TELESCOPE TOKYO OBSERVATORY JAPAN vol. I MECHANICAL DESCRIPTION | 観測原簿 vol. 69 | 動力配線系統図 74 インチドーム T88-601 |
| title | タイトル | 74 インチ望遠鏡マニユ アル 1 | 野帳 | |
| title.collection | 統一資料名 (大分 類) | 国立天文台岡山天体物 理観測所資料 | 国立天文台岡山天 体物理観測所資料 | 国立天文台岡山天 体物理観測所資料 |
| description.author | 作成者 | SIR HOWARD GRUBB PARSONS & CO., LTD. NEWCASTLE UPON TYNE, 6, ENGLAND | 観測者 : 石田, 小 暮, 福田, 岩崎, 西村, 兼古, 寿岳, 山下, 高窪, 近藤, 磯部 | 株式会社スズキ麵 工 |
| coverage.spatial | 場所についての情 報 | ENGLAND | 日本 | 日本 |
| coverage.temporal | 時代についての情 報 | | 観測年代 : 昭和 50 年 (1975 年) 11 月 16 日-昭和 50 年 (1975 年) 12 月 27 日 | 作成年 : 昭和 63 年 (1988 年) 6 月 |
| subject | トピック | 冊子 | 冊子 | 図面 |
| description.source-format | 形態 | 冊子 (バインダ) | 冊子 | 一枚 |
| language | 言語 | en | ja | ja |
| rights.rightsholder | 所蔵 | 国立天文台岡山天体物 理観測所 | 国立天文台岡山天 体物理観測所 | 国立天文台岡山天 体物理観測所 |

5. まとめ

本稿では, 天文学研究資料を対象にした調査および情報公開について, 岡山天体物理観測所所蔵資料を事例とした取り組みの概要について述べた. 多様な資料間の関係性を定

義・記述するための資料に対する固有識別子 DOI を付与や、公開のための資料画像のデジタル化およびメタデータの記述について実際の資料を対象に検証を行った。今後は、現在公開に向けて最終的な構築段階であるリポジトリの完成を目指す。完成したリポジトリは一般社団法人学術資源リポジトリ協議会の web サイト上 (<http://www.repon.org>) で公開する予定であるが、詳細については国立天文台と調整中である。さらに、岡山天体物理観測所以外の天文関連施設についても天文学研究資料に関する調査を行うとともに、オープンサイエンスを推進させるための天文学研究資料の活用などについても、より具体的な提案と検証を実施する予定である。

謝辞

本研究は、岡山天体物理観測所の戸田博之氏、柳澤顕史氏、岡山理科大学の加藤賢一氏、国立天文台の中桐正夫氏、元岡山天体物理観測所長の前原英夫氏の方々のご協力を受けて遂行されました。また、以下の研究の一部として実施されました。関係各位に感謝致します。 科研費基盤研究(B):24300310 科研費挑戦的萌芽研究:25560140 平成 28 年度国立情報学研究所共同研究「DOI 付与に基づいた横断的な博物資料情報共有モデルの検討」

参考文献

1. 川上修司ら, 宇宙科学資料データベースの開発, 宇宙科学情報解析論文誌, 第 6 号, pp 89 - 98, 2017 DOI : 10.20637/JAXA-RR-16-007/0008
2. 高橋慶太郎ら, 川口市郎氏ロングインタビュー, 天文月報 110 巻 7 号, pp. 488-498, 2017
3. 堀井洋ら, 学術資源リポジトリ構築に向けた現状と課題 - 明治期以降の科学実験機器資料のリポジトリ化を事例として-, 人文科学とコンピュータシンポジウム論文集, Vol. 2012. No. 7 pp. 17-22, 2012
4. 堀井洋ら 博物資料情報に対する DOI 付与の意義と展望, 情報知識学会誌, Vol. 26 (2016) No. 2 p. 217-220

ビッグデータ太陽系天文学の創成のための ALMA キャリブレーション観測データの悉皆的解析システムの検討と構築

飯野孝浩^{*1}, 並木美太郎^{*2}, 山田崇貴^{*3}

A feasibility study of exhaustive analysis of ALMA calibration data for the creation of big-data driven solar system astronomy

Takahiro Iino^{*1}, Mitaro Namiki^{*2} and Takayoshi Yamada^{*3}

概要

2011 年から科学運用を開始した ALMA は、ミリ波・サブミリ波帯域においてこれまでにない高感度・高空間分解能観測を実現する測器であり、太陽系天体のリモートセンシングにおいてもブレイクスルーをもたらすと考えられる。観測毎に蓄積されるキャリブレーション観測データを科学観測用途に転用することで、周波数・時間・天体方向に対して巨大な太陽系天文学観測データの入手が可能となる。一方で、そのデータ容量は膨大であり、最終段階までデータ処理を行わなければデータ品質の見極めが困難であり、大量のデータ処理を可能にするソフトウェア・ハードウェアを必要とすることが分かった。我々は ALMA キャリブレーション観測データの科学観測への転用を目指し、データ処理の歩留まり向上とバッチ処理化、そして専用システムの構築に取り組んだ。キャリブレーション観測データを生データから fits ファイルおよびクイックルック用画像ファイルまで処理するシステムを構築し、海王星の Band6 以上のデータについては公開済み全ファイルの処理を完了した。大容量データに対応するスケーラビリティと高速性を持つ SAS ストレージを導入し、広く用いられている 1 Gbps のネットワーク接続ストレージと比べて、最も処理時間のかかるプロセスであったキャリブレーション処理時間を 19-36% に短縮させることに成功した。さらに、同プロセスの並列化により、処理時間を約 13% に短縮した。これにより、研究室で保有できる程度の計算機システムにより、現実的な時間で ALMA キャリブレーション観測データの悉皆的解析を行うための道筋をつけることができた。

Abstract

The ALMA that started its operation in 2011 has significantly high sensitivity and spatial resolution in millimeter and sub-millimeter waveband and expected to introduce a breakthrough on solar system remote sensing. Use of calibration data accumulated during every observations enables us to obtain the vast amount of solar system astronomical data. In turn, the total amount of data is quite large and

doi: 10.20637/JAXA-RR-17-009/0003

* 平成 30 年 1 月 18 日受付 (Received January 18, 2018)

^{*1} 東京農工大学科学博物館

(Nature and Science Museum, Tokyo University of Agriculture and Technology, E-mail: iino@nagoya-u.jp)

^{*2} 東京農工大学工学部情報工学科

(Department of Computer and Information Sciences, Tokyo University of Agriculture and Technology)

^{*3} 東京工業大学大学院総合理工学研究科

(Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology)

all of the data need to be processed to see if the data is suitable for the scientific analysis. Therefore, we attempted to construct an automated ALMA data processing system and a new computer system specified for the conversion of ALMA calibration data into the scientific data. As a result, we succeeded to construct the automated data processing system that produces fits cube file and quick-look data of spectrum and 2-D brightness map, and to process the whole data of Neptune above Band 6. In addition, introducing a new SAS storage that enables both capacity, scalability and high-speed access succeeded to reduce the time required for calibration process to $\sim 19\text{--}36\%$ comparing with a standard NAS storage installed on 1 Gbps LAN network. A parallel data calibration system also decreased the processing time up to $\sim 13\%$. In summary, we succeeded to draw a path for the comprehensive analysis of ALMA calibration data in a realistic machine-time.

Keywords: Radio astronomy, ALMA, Planetary atmosphere

1. はじめに

干渉計型望遠鏡は、複数の望遠鏡で同時に観測された天体からの信号に相関処理を施し、高い空間分解能を実現する測器である。中でも、2011年に運用が始まったアタカマ大型ミリ波サブミリ波干渉計 (Atacama Large Millimeter/submillimeter Array; ALMA) は、66台の12 m および 7 m 望遠鏡からなる電波干渉計であり、31 GHz から 950 GHz という広い周波数帯において連続波および分光観測を可能とする。実現される空間分解能は、2017年現在 (Cycle-5 シーズン) に実用化されている観測バンドにおいて、最高の空間分解能を達成できる最長基線モードで 0.042 秒 (Band 3, 100 GHz)–0.024 秒 (Band 10, 870 GHz)^{*4} であり、また、最大 66 台のアンテナの同時稼働により、アンテナ 1 台による単一鏡型望遠鏡に比べて高感度での観測も期待できる。

高感度・高空間分解能観測という ALMA の特性は太陽系天文学においても重要である。太陽系内天体のうち、単一鏡型望遠鏡で空間分解可能な天体は太陽・金星・火星・木星・土星にとどまる。しかし、ALMA は視直径 2.2–2.3 秒の海王星や、0.6–0.7 秒の土星衛星タイタンなど、多くの惑星・衛星の空間分解観測を実現する。探査機に匹敵する高空間分解能での分光・連続波観測は、大気組成や大気ダイナミクス、熱分布の導出を可能とし、太陽系天文学にブレイクスルーをもたらすと期待されている¹⁾。しかし、ALMA を用いた科学観測の機会は限定されており、サイクル 0 からサイクル 3 までの観測プロポーザル採択倍率は 8.2, 5.8, 3.9, 3.9 倍で推移してきた。また、通常のプロポーザルにおいて申請できる 1 回あたりの観測時間は、12 m アレイを用いる場合で 50 時間までと制限されている^{*4}。現状では観測機が限定されているのみならず、長大な観測時間を投下したラインサーベイ観測や、分子・熱分布の時空間変動取得を目指した長期の観測も困難な状況である。

そこで、我々は ALMA がキャリブレーション目的で行っている惑星観測データの科学観測データへの転用に着目した。同様のアプローチにより、これまで土星の衛星タイタンについて、ALMA の空間分解能を活かしたタイタン大気中同位体比や窒素化合物検出が報告されている^{2),3)}。図 6 に ALMA の典型的な観測プロシージャ (Project code: 2011.0.00628.S) を示す。R CrA IRS7B が科学観測天体であり、それ以外に海王

^{*4} Cycle-5 ALMA Proposer's guide による (<https://almascience.nao.ac.jp/proposing/documents-and-tools/cycle5/alma-proposers-guide>)

星およびクエーサー J1924-292 がそれぞれ Flux および Phase/Gain キャリブレーションとして観測されている。惑星は視直径や温度がある程度分かっているため、Flux キャリブレーションとして頻りに用いられており、各観測セットごとに1回ずつ観測される。科学観測と同じアンテナ台数が使用されており、観測時間は短いものの、高い観測性能を持つと期待できる。キャリブレーションとして用いられている太陽系内天体は、火星、木星、天王星、海王星といった惑星だけでなく、ガニメデ、カリスト、タイタンといった衛星、そして小惑星であるセレスと幅広い。

キャリブレーション観測データは、積分時間が短いためにプロポーザルをベースとした科学観測に比べて感度で劣るものの、1. 天体、2. 周波数、3. 時間軸、の3軸に対して巨大なデータ群になっていることが重要である。これらの特質により、それぞれ1. さまざまな対象天体に対し、2. 特定の瞬間における多様な分子観測および大気・地表温度構造の取得、さらにその3. 時間変動の取得、という観測手法の展開が可能になる。これにより、大気化学、大気動力学、惑星形成といった、多方面からのサイエンスの展開が期待される。一方で、太陽系内天体は絶対強度のキャリブレーションに用いられており、確からしい Spectral Energy Distribution モデルを再現するように、観測された天体のフラックスをスケールアップするのに用いられる。すなわち、各天体表面においてグローバルな連続波強度の変動を捉えることは原理上不可能であることに留意する必要がある。

本研究では、キャリブレーション観測データを太陽系天文学ビッグデータととらえ、その悉皆的な解析を通じて、上記のサイエンスにおけるブレイクスルーを目指す。その前段階として、本稿では、キャリブレーション観測データの個数や観測周波数・対象天体の特性等を明らかにし、期待されるサイエンスを整理する(2章)。また、悉皆的な解析において重要なキャリブレーション観測データのバッチ処理を可能にするとともに(3章)、専用の環境を構築し、将来の悉皆的な解析において必要とされるマシンタイムの見積もりを行った(4章)。

2. ALMA キャリブレーション観測データの概要と、期待されるサイエンス

2.1 ALMA キャリブレーション惑星データ概要

キャリブレーション観測データのフィジビリティ検討のため、主要な太陽系内天体(火星、木星、ガニメデ、カリスト、タイタン、天王星、海王星、セレス)について、観測日時、観測周波数およびデータ容量の調査(タイタンのみ)を行った。これら天体群は大気の分光観測が期待できる群(タイタン、海王星、天王星、火

| Date | Timerange (UTC) | Scan | FldId | FieldName |
|-------------|-------------------------|------|-------|--------------------------------|
| 04-May-2012 | 11:04:04.0 - 11:09:21.2 | 4 | 0 | J1924-292 Phase, Gainキャリブレーション |
| | 11:14:32.7 - 11:19:27.9 | 7 | 2 | Neptune Flux キャリブレーション: 惑星 |
| | 11:23:38.5 - 11:24:08.8 | 10 | 0 | J1924-292 |
| | 11:26:32.5 - 11:34:33.7 | 12 | 3 | R CrA IRS7B 科学観測天体 |
| | 11:36:17.6 - 11:36:47.9 | 14 | 0 | J1924-292 |
| | 11:38:34.2 - 11:46:35.4 | 16 | 3 | R CrA IRS7B |
| | 11:48:19.0 - 11:48:49.2 | 18 | 0 | J1924-292 |
| | 11:50:35.9 - 11:58:37.1 | 20 | 3 | R CrA IRS7B |
| | 12:00:32.6 - 12:01:02.8 | 22 | 0 | J1924-292 |
| | 12:03:01.1 - 12:11:02.2 | 24 | 3 | R CrA IRS7B |
| | 12:12:45.9 - 12:13:16.1 | 26 | 0 | J1924-292 |
| | 12:15:01.9 - 12:23:03.1 | 28 | 3 | R CrA IRS7B |
| | 12:24:46.6 - 12:25:16.8 | 30 | 0 | J1924-292 |
| | 12:27:02.6 - 12:33:41.8 | 32 | 3 | R CrA IRS7B |
| | 12:35:37.1 - 12:36:07.3 | 34 | 0 | J1924-292 |

図1 海王星を含む観測プロシージャの例。観測天体として、1つの科学観測天体と2つのキャリブレーション天体が含まれる。

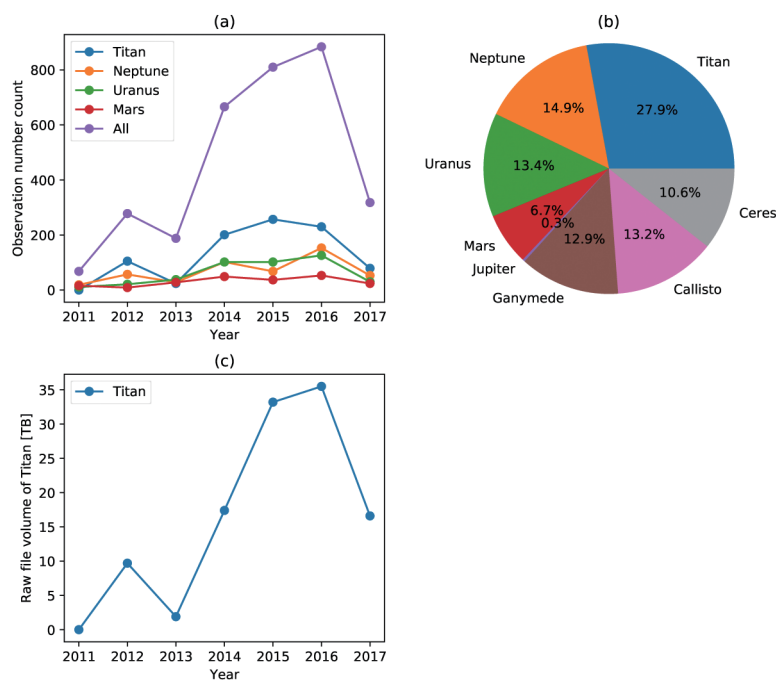


図2 (a) 年毎の太陽系天体全体及び主要天体の総観測数の推移。2017年は5月までの数値である。2014年以降は総数が600回を越えており、今後も同程度で推移すると期待される。(b) 天体ごとの観測数の割合。(c) タイタンの観測年ごとのデータ容量の推移

星)と、固体天体群(ガニメデ、カリスト、セレス)に二分できる。ALMAアーカイブより上記全天体の公開データ(観測日、観測周波数、積分時間等)のクエリを行い、テキストデータとしてダウンロードし、調査を行っている。データの個数、天体ごとの内訳、データ容量(タイタンのみ)を図2に示す。2017年は5月までの値である。(a)に示した総観測数はほぼ右肩上がりであり、2016年には800件超となっている。一部の天体(タイタン、海王星、天王星、火星)の観測数推移も併せてプロットしているが、同様に増加を示している。(b)にはキャリアレーション天体のデータ数の内訳を示した。タイタンが最多で27.9%を占める。他の天体は10%台であるが、国内外に研究者の多い火星は6.7%と少ない。(c)はタイタンの総データ容量の推移を示す。2015–2016年は30TBを越える。後述するように、タイタンは複雑な大気組成を持つため、その大気組成や同位体比の導出のためには、可能な限りのデータを用いたラインサーベイ観測が重要である。トータルで100TB超という大容量のデータを高速で処理するために、本研究では大容量・高速ストレージが必要となること分かる。アーカイブでは他に空間分解能も公開されているが、楕円形となるビームの短辺・長辺の平均を示しており、ビーム形状の詳細についてはデータをダウンロードし、CLEAN処理を施すまで不明である。

2.2 ALMA キャリブレーション観測データ利活用の意義

2.2.1 周波数方向

惑星大気由来分子スペクトルの輝線幅は圧力に比例する「圧力広がり(Pressure broadening)」が支配的であり、数MHz/torr程度が一般的である。気圧が100mbar以下と低い成層圏に偏在する分子の線幅は典型的

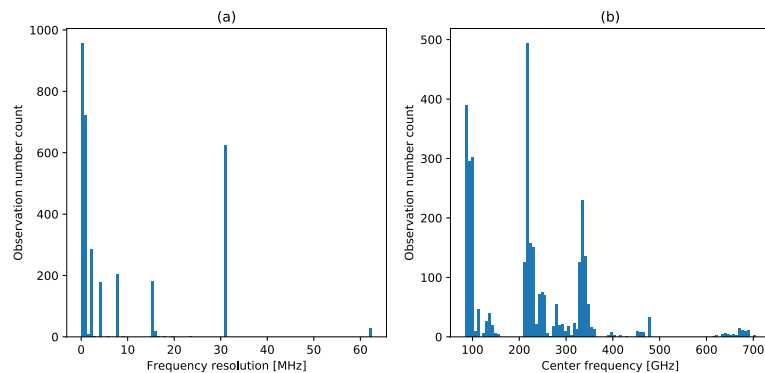


図3 (a) キャリブレーション観測データの周波数分解能のヒストグラム, (b) キャリブレーション観測データの中心周波数ヒストグラム. Band 8 以上の高周波数帯が乏しく, ミリ波・サブミリ波天文学で実績のある 100 – 300 GHz に集中している.

に数十 MHz 程度であり, スペクトル形状を十分に分解するには高い周波数分解能が必要となる. 図 3(a) からは, キャリブレーション用データにおいて多数を占めるのは, 1 MHz 以下という高い周波数分解能を持つデータであることが分かる.

周波数方向に広く (無バイアスに) 分光観測が行われていることは, 観測的天文学のメソッドとして重要である. 図 3(b) に示した中心周波数ヒストグラムからは, 観測データが Band 3, 6, 7 の 100~300 GHz 帯に集中していることが分かる. この波長域における観測数は少ないものの, 単一鏡型望遠鏡においても一般的ではなかった Band 8, 9 のデータが取得可能になっている意義は大きい. 図 4 にサイクル 0 におけるタイタンのキャリブレーション観測データのスペクトルを示す. HC_3N 分子の回転および振動遷移, そして同位体がとらえられている. このような分光観測を周波数方向に多数結合することで, 大気中の微量分子や同位体比をもれなく観測することが可能になる. この手法は, 濃密な大気を持ち, 分光観測による効果が見込める火星, タイタン, 天王星, 海王星について重要である. 特に, N_2 を主成分とし, HCN , HC_3N , $\text{C}_2\text{H}_5\text{CN}$ 等の窒化物を豊富に含むタイタン大気においては, 複雑な大気化学プロセスの解明が重要な課題であり, その観測的制約のために本アプローチが果たす役割は大きい. また, 分子の分布の導出も分光観測から可能になる. 特定の大気層から放射される分子輝線形状は, 主に自転によるドップラーシフトおよび, 圧力広がりによって決定される. 視線方向に分割された各大気層はそれぞれが異なる温度・圧力・分子存在量を持っており, 輝線形状を考慮して十分に周波数方向に分解した輻射輸送計算を行い, 得られた輝線形状と観測された形状を比較することで, 分子の鉛直方向の分布を決定することが可能である (=大気フォワードモデル). さらに高空間分解能データを用いれば, 分子 3 次元分布の導出も可能であり, 実際に一部の窒化物分子群の極域への局在が報告されている³⁾. 分子の存在量と併せて, 分布を観測的に制約することが大気化学の理解においては重要である. これはタイタン以外の天体においても重要であり, 例えば海王星においては成層圏に偏在している HCN 分子の空間分布や, 小天体の降り込みによって生成される分子群の存在量の観測的導出から, 成層圏微量分子の起源の制約が見込める. これまで, 単一鏡型望遠鏡を用いた系統的な観測により, 海王星成層圏は硫化物に乏しいことが明らかになっている⁴⁾. キャリブレーション観測データを用いれば, より広範な微量分子の探索や, 検出済み分子の 3 次元構造の導出が可能になる.

2.2.2 天体方向

ミリ波・サブミリ波を用いた観測的研究は, キャリブレーション観測対象天体の中では, 海王星及びタイタ

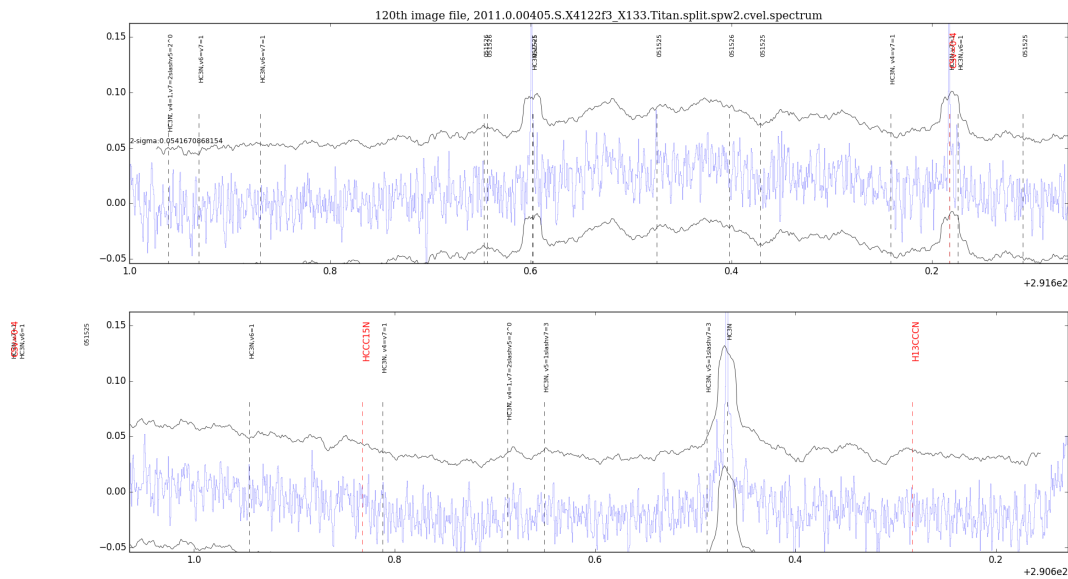


図4 本研究で生成されたタイタンのBand 6におけるスペクトルの例。HC₃N分子の回転、振動遷移が検出されている。黒色の線は $\pm 2\sigma$ を、縦の点線は検出可能性のある分子の位置を示す。

ンにおいて活発に展開されてきた。一方、ガニメデ、カリスト、セレスの4天体の大気は非常に希薄であり、分光観測のフィジビリティが低い天体とみなされてきた。だが、特に氷天体であるガニメデとカリストは表面の氷が昇華することで大気が形成されると考えられ、実際に、氷天体である海王星衛星トリトンや冥王星大気分光観測から、CH₄やCOといった氷由来分子の輝線・吸収線が観測され^{5),6)}、またその時間変動も捉えられている⁷⁾。ALMAの科学観測により、サブミリ波帯域での冥王星大気中CO、HCN分子検出も報告されている⁸⁾。氷天体においても同様の分光観測の成果が期待される。また、固体天体の温度分布マップからは地表の熱慣性や内部熱源といった物理量の制約が可能になると考えられる。衛星表面の中間赤外線による温度マップ取得はガリレオ探査機によりなされたのみであり、全球を継続観測できるキャリブレーション観測データの活用は重要である。

2.2.3 時間方向

太陽系内天体においては、大気構造やダイナミクス、太陽風相関イベントなど、日・月単位の短いタイムスケールで変動するイベントが存在し、その素過程の理解には観測的制約が重要である。国立天文台のミリ波およびサブミリ波測器である野辺山45-m望遠鏡やASTE望遠鏡は、大気条件の問題から観測時期が年の半分程度に制限されているが、ALMAはほぼ通年で観測を実行しており、時間変動のあるイベントの観測に適している。

大気プローブとして重要な一酸化炭素(CO)分子輝線の観測は太陽系外天文学においても重要であることから、アーカイブ中にも多数含まれている。火星大気の主成分はCO₂分子であり、その光解離で生成されるCO分子はミリ波・サブミリ波帯で最も強力な吸収線である。CO分子をプローブとし、太陽活動と連動したCO分子3次元分布や、ドップラースhiftを用いた大気ダイナミクスの測定からは、火星気象の理解において重要な観測的制約を行うことが可能である。

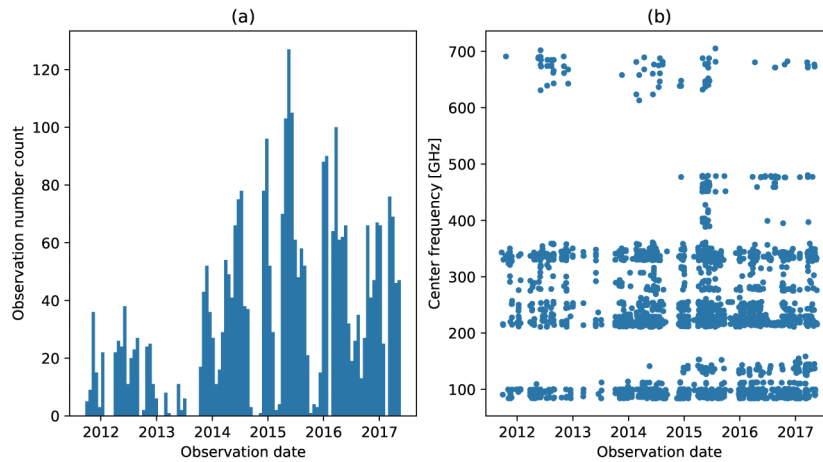


図5 (a) 調査天体に対する観測回数の推移. 2016年以降は絶え間なく観測が行われている. (b) 観測周波数の推移. これを周波数ヒストグラムにしたものが図3(b)である. 受信機が整備される毎に観測周波数が拡大している. 2017年からは, 200 GHz 周辺の Band 5 が共同利用に供される.

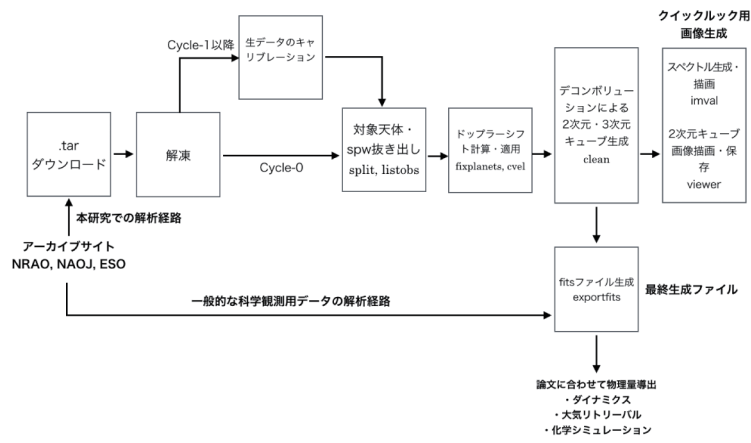


図6 本研究におけるALMAデータ処理のプロセス

3. データ処理プロセスとバッチ処理化

本研究では, 膨大なキャリブレーション観測データの悉皆的解析のために, データ処理プロセスのバッチ処理化と処理歩留まりの向上を検討した. ALMAにおけるデータ処理のプロセスを図6に示した. 科学観測の場合には, ALMAスタッフにより解析, 生成されたfitsキューブファイルが提供される. いっぽう, キャリブレーションデータの場合には生データ(ASDM)ファイルからfitsファイルの生成までを全てユーザーが行う必要がある. ASDMファイルは最終プロダクトであるfitsファイルと比べて容量も大きい. 例として, 火星をキャリブレーションデータとして用いているデータのうち, IDの先頭が2013となるデータからASDMとfitsの含まれるtarファイルの容量・個数を調べると, それぞれ2.1TB(439個)および222GB(116個)であり, 容量でほぼ10:1であった. 解析用スクリプトはfitsファイルに同梱されているため, 本研究ではASDM, fits双方をダウンロードする必要がある. 本研究は対象のデータも多いため, 通常の科学観測データを用いたアプ

ローチに比べてはるかに大容量のストレージを必要とすると言える。データ処理には電波天文学データの汎用処理ソフトウェア CASA(Common Astronomy Software Application) を用いている。CASA は Linux もしくは Mac 上でのみ動作し、Windows はサポートされない。本研究ではデータ処理に Linux 計算機システムを用いている(表 1)。

全プロセス中で、最もマシンタイムを要するのはキャリブレーションプロセスである。キャリブレーションには観測データ毎に専用の Python スクリプト群が用意されており、CASA 上で呼び出す (`execfile` タスク) ことで動作する。処理すべきデータのスクリプトを順に実行することでバッチ処理が可能である。前処理として、事前に関連するキャリブレーションスクリプト群に対し、キャリブレーション後に生成されるファイル容量を抑制するオプションの記述や、ASDM とスクリプトのファイル数の不整合や予期せぬライブラリ読み込みによるエラーを回避するための処理を行っている。キャリブレーションスクリプトの実行歩留まりは低く、さまざまなエラーで CASA 自体が停止するため、失敗しても処理全体が停止しないよう、外部スクリプトからキャリブレーションスクリプトを呼び出している。

キャリブレーションスクリプト実行時間や生成前後のディレクトリの容量を表示させることで、実行効率の検証を行っている。実行時間とファイル容量の関係については後述する。

生成されたキャリブレーション済データ (`.ms.split.cal` 形式) は天体および、異なる観測周波数を持つ分光計のチャンネル群である Spectral window(`spw`) に分割する (`split` タスク)。強い分子輝線はマスクされている場合があるため、マスクを外す処理 (`unflag`, `plotms` タスク) を行う。観測周波数は科学観測天体のドップラーシフトに合わせて設定されているため、キャリブレーション天体のドップラーシフトの補正を行う (`cvel` タスク)。特に天体の分子輝線の同定において重要なプロセスであるが、本プロセスは本稿執筆時点では正常動作の歩留まりが低く、今後の重要な課題である。CLEAN プロセスにより、2 次元・3 次元輝度分布キューブを生成 (イメージング) する。同プロセスは輝度が集中すると予想される任意の領域を指定する必要があるが、太陽系内天体の場合には既に視直径が分かっているため、これもバッチ処理が可能である。イメージングにより生成されたキューブファイルの可視性を高めるため、クイックルック用に連続波マップ及びスペクトルを生成する。png 形式の連続波マップは、CASA のデータ可視化タスク (`viewer` タスク) を用いてバッチ生成する。スペクトル画像の生成時には、いったん天体の任意領域を指定して周波数-強度 (Jy/beam) の 2 カラムのテキストデータを生成したのち (`imval` タスク)、Python の可視化ライブラリ `matplotlib` で png ファイルを生成している。スペクトル図の例を図 4 に示す。同時に主要な分子輝線をオーバーレイすることで輝線同定に役立てている。

キューブファイルは CASA のみで読み込み可能であるため、`ds9` や `aplpy` といった汎用の天文学データ処理ソフトウェア・ライブラリでの利用のために、キューブファイルを `fits` 形式に変換する (`exportfits` タスク)。fits ファイルは最終アウトプットとして別ストレージにも格納し、冗長性を高める。

ALMA よりダウンロード・解凍されたデータの基本構造を以下に示す。Project code の下に、Science goal, Group, Member の各 ID 毎に階層が形成される。ひとつの Project code 中に複数の science goal が設定される際には、それぞれの science goal 以下に独自の ID を持つ階層が形成される。

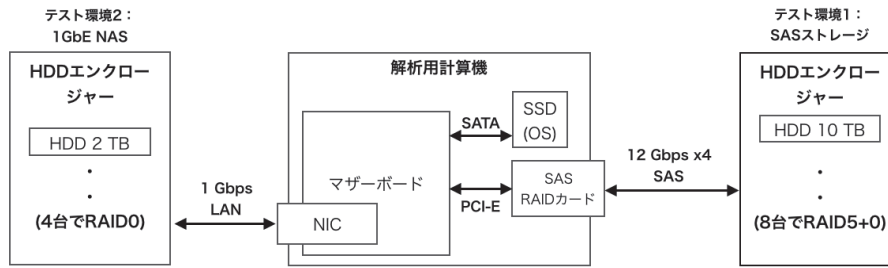
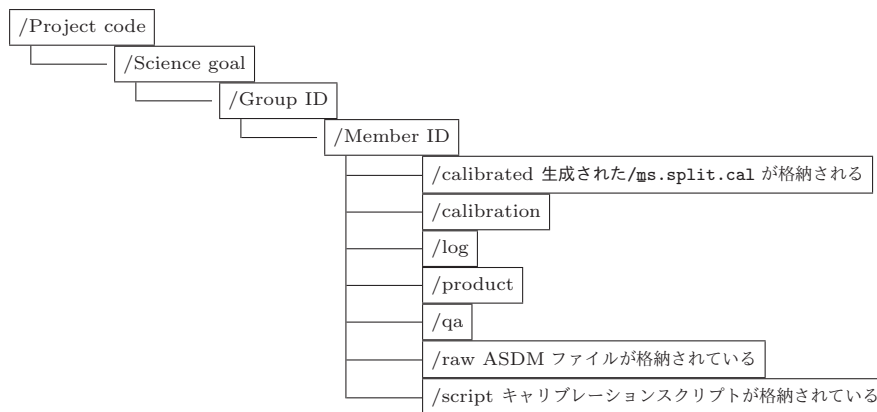


図7 本研究で用いているシステムのブロックダイアグラム.



本研究では、'Project code.Member ID.天体名.spw. 処理段階' という命名規則のもとに、処理後のファイル、ディレクトリを生成している。処理段階は図6の split 処理以降に対応する。

4. 生成データの概要と、解析処理高速化の取り組み

4.1 悉皆的解析に必要な新解析システムの要件と構築

表1 本研究に用いている計算機およびストレージの仕様

| 機能 | 仕様 |
|--------------|--|
| 型名 | Precision T7500 (Dell 社製) |
| CPU | Xeon 5690 (6 コア 12 スレッド, 最大動作周波数 3.73 GHz) 2 基搭載 |
| チップセット | Intel 5520 |
| RAM | DDR3 1333MHz 駆動 ECC Registered 96 GB (4 GB × 8 + 16 GB × 4) |
| システムディスク | 250 GB ソリッドステートドライブ, SATA 6 Gbps 接続 |
| OS | Ubuntu Linux 16.04 LTS |
| SAS RAID カード | ASR8885Q(Adaptec 社製), 内外 8 ポート, PCI-Exp. 3.0 ×8 接続 |
| HDD ケース | ARC-4038(areca 社製), SATA 8 ベイ, 48 Gbps SAS 接続, SAS エキスパンダー搭載 |

本研究では、ガスを持つ天体のラインサーベイ観測および、固体天体・ガス惑星輝度マップ観測、の2テーマが大きな観測ゴールである。そのためには、最大の観測データ数を持つタイタンを含む、大容量のアーカイ

ブデータの悉皆的な解析が必要である。そのため、大容量のストレージがシステムに求められる要件のトップに挙げられる。さらに、データの増加に対応できるスケーラビリティ、耐障害性、計算機の入れ替えへの対応、そして高速性が併せて求められる。

大容量ストレージとして 2017 年現在で一般的であるのは Network Attached Storage(NAS) である。2-8 台の内部ディスクに加え、USB や eSATA を用いた拡張にも対応可能である。また、ネットワークを介して複数の計算機からの同時アクセスが可能である。しかし一方で転送速度が難点であり、一般的には 1 Gbps に留まる。10 Gbps Ethernet を用いる NAS は普及途上であり、ネットワーク全体を再構成するコストが生じる。内蔵 SATA ポートを用いてソフトウェア RAID を組む場合には、各 HDD と 6 Gbps で接続できるため速度では優位性があるが、SATA ポートと HDD ベイの数が容量の上限を決めてしまうこと、計算機リプレースに対応しないことが問題となる。ハードウェア RAID 機能を搭載したハードディスクケースは計算機リプレースに対応できるが、接続速度とスケーラビリティにおいて劣る。Thunderbolt 3 規格に対応したハードディスクケースは速度面で優位であり、デジチェーンによる拡張性も併せ持つが、Linux 環境で十分にテストされた製品は見つからなかった。これら競合する規格との比較検討のうえ、本研究では Serial Attached SCSI (SAS) 接続によるストレージの構築を計画した。SAS はケーブル 1 本あたり 12 Gbps の高速接続速度（さらに接続本数で通信できる）と、デジチェーン・パラレル接続による拡張性を併せ持つ規格である。ハードウェア RAID 機能を持った拡張カードを用いることで、計算機のリプレースにも対応する。いっぽう、複数の計算機からの高速接続には対応しない点がデメリットである。

本研究で用いているシステムのブロックダイアグラムを図 7 に示す。本研究では、前段階として 1 Gbps 接続の NAS 類似ネットワークストレージ（図 7 中のテスト環境 1）を複数構築し、データ処理プロセスの構築やアーカイブデータのポテンシャルの確認を行い、図 4 のような高感度データが多く含まれることを確認した。その中で、キャリブレーション観測データの容量が大きく、複数のストレージにデータを分散させる不便が大きいこと、そしてデータ処理中は常に NAS と計算機間のデータ転送容量が 1 Gbps の上限値に張り付いているため、外部記憶との転送速度が処理のボトルネックになっていると見られたことから、SAS を用いた高速・大容量ストレージの構築（図 7 中の新ストレージ）を検討・実施した。本稿執筆時点では、10 TB のエンタープライズ向けハードディスク 8 台で RAID5+0(RAID5 のストライピング) を構築、48 Gbps で計算機と接続している。実効容量は 58.4 TB である。本ストレージは SAS ストレージの性能確認用として導入したものであり、近い将来に 24 ベイのエンクロージャを導入予定である。

SAS ストレージの転送速度の試験として、`hdparm` コマンドを用い、3 秒間のシーケンシャルリードで転送した容量を 10 回計測し、平均値と標準偏差を計測した。結果は 1309.6 ± 7.7 MB/s であり、1 Gbps LAN の理論上の転送速度の 10 倍程度を実現できた。転送帯域を使い切っていない状況であり、今後は RAID5 を構成するディスク数の増加による速度の向上も期待できる。

4.2 実行効率とビーム形状

2017 年 5 月～6 月に、海王星の Band 6 以上の公開済み全データのキャリブレーション処理を完了させることができた。図 9 に、海王星を対象としたキャリブレーションの所要時間 (a) とキャリブレーションの実行効率 (ASDM データ容量/実行時間) (b) を示す。所要時間は大まかに ASDM 容量と比例していることが分かる。(b) の実行効率はデータによる差異が大きく、複数のピークを持つが、おおまかに 2-9 MB/s の範囲に収まる。ここから、将来のより巨大なデータの処理に必要な時間を見積もる事が可能である。タイタンは提供されているデータ全体の処理によりラインサーベイ解析が可能であり、提供済みの総データ容量は 114.3 TB である。うち 90% を ASDM ファイルと、実行効率を海王星と同等の ~ 5 MB/s と仮定すると、キャリブレーション

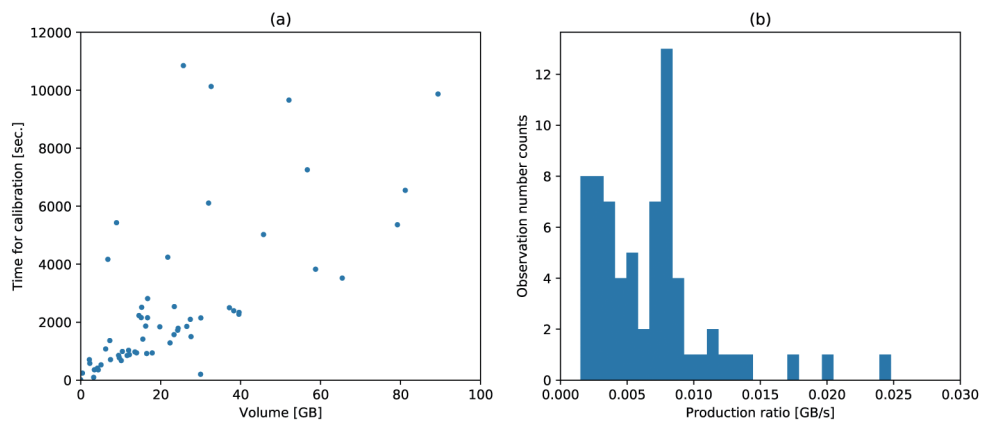


図8 (a) キャリブレーション所要時間 (縦軸) と ASDM ファイル容量 (横軸) の関係 (b) データ処理効率のヒストグラム. 2-9 MB/s に値が集中している.

ション処理に要するマシンタイムは、現状のシステムで ~265 日となる。

海王星のサイクル 1 以降のキャリブレーション後のデータとサイクル 0 のデータのうち、spw=0 のみに split, CLEAN 処理を行い、ビーム形状を調べた結果を図 9 に示す。(a) はビーム長辺の長さを示す。海王星の視直径は 2.2-2.3 秒程度であるが、データの多くは視直径以下のビーム径を持つことから、空間分解観測が可能なポテンシャルを持つデータが多いことがわかる。(b) は扁平率 (1-短辺/長辺) であり、値が大きいほどビームが扁平であることを表す。マッピング観測時にはビームが円形に近いことが求められるが、短辺:長辺=1:2 となる 0.5 を越える扁平率を持つデータも多い。本稿で扱ったデータ処理終了後には、輻射輸送計算により分子の存在量や気温分布といった物理量を導出することになるが、輻射輸送コードには楕円ビーム形状の処理の実装が重要である。これは円形ビームを仮定できる単一鏡型望遠鏡にはないデメリットである。(c)(d) はそれぞれ海王星視直径とビーム長辺比、視直径とビーム長辺の時間推移をプロットした。(c) は 1 を越えると海王星の空間分解がなされているデータとなる。最高で 4 を越えるデータがあるが、これはサイクル 0 シーズンに行われたものであり、現在論文化に取り組んでいる。干渉計に特有の問題として、一定以上に広がった輝度分布を取得できない Resolving out という現象が挙げられる。特に惑星は見かけの形状が円形であり、温度分布の観測時にこの影響を受けることが予想される。ALMA の場合には、空間分解能と Resolving out する角度の比は各バンドにおいて典型的に 1:10 程度であり^{*4}、海王星においては 0.2 秒程度が最大の空間分解能となるが、現状では該当する高空間分解能のデータは含まれていない。(d) からは、定常的に高空間分解能データが提供され続けていることが分かり、温度分布の時間変動が観測の可能性を示している。火星や氷天体群は季節変化のタイムスケールが短いため、同様のデータが提供されていれば温度や組成分布の季節変動をとらえられる可能性がある。

4.3 高速ストレージ導入によるキャリブレーション処理高速化の検証

SAS ストレージの導入による高速化の検証のため、旧ストレージ (図 7 中「テスト環境 2」) と SAS ストレージでのキャリブレーション処理の実実行時間の比較を行った。キャリブレーション観測データとして、火星のデータを 5 個用い、両環境で 1 回ずつキャリブレーション処理を実行している。実実行時間の積み上げグラフを図 10 に示す。実実行時間は各データにおいてテスト環境 2 と比べて 19-36% に、全体として 32% に短縮した。キャリブレーション観測データは一つ一つが数十 GB と巨大であり、またキャリブレーション後の出

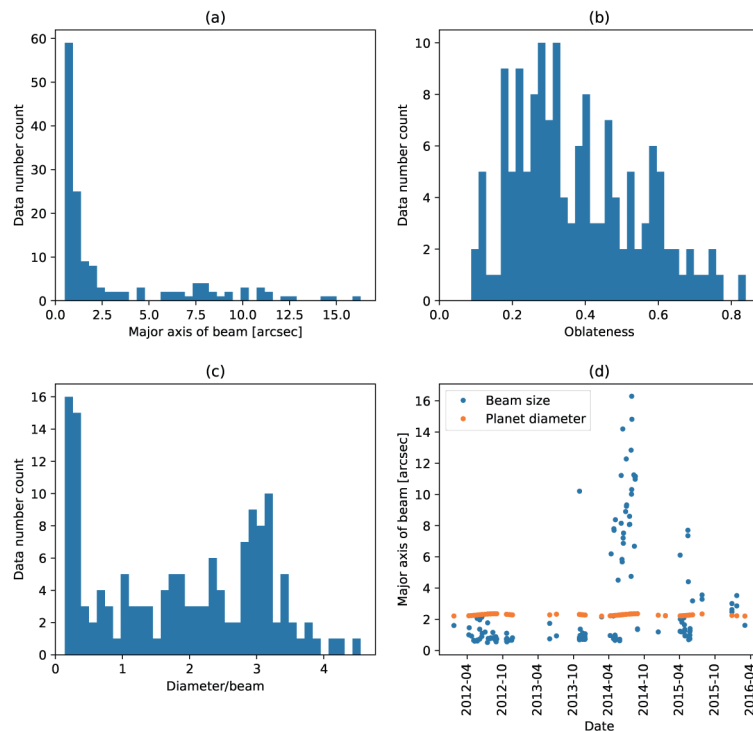


図9 イメージング後の海王星データ群から取得した、ビーム形状と海王星視直径の関係。(a) ビームの長辺の FWHM (b) ビームの扁平率 (c) 海王星視直径とビーム長辺の比 (d) ビーム長辺の大きさと海王星視直径の時間推移

カデータは元データより大きいいため、外部記憶の転送速度がボトルネックになっていたことが示された。一方、処理速度の短縮はシーケンシャルリード速度の比ほど劇的ではないため、現状では CPU やメモリ速度がボトルネックと考えられる。

4.4 並列実行によるキャリブレーション処理高速化

前節のストレージ高速化に引き続き、複数のキャリブレーション処理のプロセス並列による高速化に取り組んだ。並列処理には Python の `multiprocessing` モジュールを使い、キャリブレーションスクリプトをサブプロセスとして実行することにより実装した*⁵。並列実行するプロセス数は、12 個の物理 CPU コア数に対して 10 プロセスとした。実行速度は単純に 10 倍にはならず、メインメモリ・キャッシュメモリの帯域や容量、外部記憶のオーバーヘッド等により実行効率が低下すると予想される。そこで、キャリブレーションスクリプトの実行時間から並列化の効果測定を実施した。シングルプロセス実行時は、前節と同様に複数のキャリブレーションスクリプトを連続実行し、12 のスクリプトの実行時間を計測した。マルチプロセス実行時は、常に 10 のプロセスが実行されるようにしながら、それぞれの実行時間を計測した。シングル、マルチプロセスそれぞれの総実行時間は 45146 および 61559 秒であり、マルチプロセス時には全てのスクリプトで実行時間が長くなった。双方での各スクリプト実行時間の積み上げグラフを図 10 に示す。各スクリプトでの実行時間の比 (マルチプロセス/シングルプロセス) は 1.34 であり、標準偏差は 0.16 であった。10 のプロセスを同時実行する今回の計測設定では、キャリブレーションの実行効率を 7.4 倍 (13.4%) に高速化することができた。

*⁵ <https://docs.python.jp/3/library/multiprocessing.html>

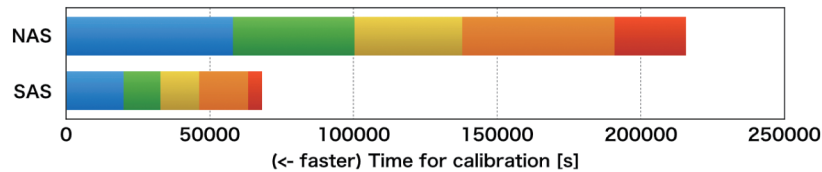


図 10 (a) 1 GbE-NAS・SAS 両システムでの実キャリブレーション時間の比較。SAS システムでは NAS システムに比べて処理時間が 19–36% に短縮している。(b) キャリブレーションスクリプトのシングルプロセス実行時とマルチプロセス (10 プロセス) 実行時の処理時間の比較。各種オーバーヘッドによりマルチプロセスでは各スクリプトの処理時間がシングルプロセス時に対して長くなっているが、その比は 1.34 ± 0.16 倍に留まっており、トータルでは 7.4 倍の高速化となっている。

4.2 節で推算したタイタンの総計算時間は 265 日であったが、これが 35 日まで短縮されることになる。

4.5 今後の展開

本研究で得られる多様なデータの論文化のため、複数の研究グループの立ち上げ、クイックルックデータの共有が重要である。国内外では火星、タイタン、固体天体、ガス惑星と天体ごとに研究コミュニティが存在するが、観測データを通じてこれらコミュニティとの有機的な結合をはかることで、ALMA とキャリブレーション観測データの特徴を活かしたビッグデータ太陽系天文学時代の創成を目指していく。

謝辞

本研究は、2017 年度電気通信普及財団研究助成、日本学術振興会科研費 (17K14420) および大学共同利用機関法人自然科学研究機構アストロバイオロジーセンター (AB291014) のサポートを受けて行われています。本テーマで発表を行った平成 28 年度 JAXA 宇宙情報解析シンポジウムでは、フロアから多くの有用なコメントをいただきました。また、解析手順等について慶応大学の田中邦彦助教および京都産業大学の佐川英夫准教授から、そして ALMA データ処理全般について国立天文台チリ観測所の西合一矢助教、中西康一郎准教授から多くのコメント、ご意見、サポートをいただきました。名古屋大学環境学研究科卒業生の仲本悟さんには、在学時に ALMA データ処理について多くの有益なトライアンドエラーの知見を提供いただきました。匿名の査読者の方には、専門的な視点からの的確なコメントを頂戴いたしました。ここに感謝申し上げます。This paper makes use of the following ALMA data: ADS/JAO.ALMA#2011.0.00405.S. ALMA is a partnership of ESO (representing its member states), NSF (USA) and NINS (Japan), together with NRC (Canada), MOST and ASIAA (Taiwan), and KASI (Republic of Korea), in cooperation with the Republic of Chile. The Joint ALMA Observatory is operated by ESO, AUI/NRAO and NAOJ.

参考文献

- [1] Hideo Sagawa. アルマ望遠鏡による太陽系研究. 天文月報, 110(5):317–324, 2017.
- [2] Joseph Serigano, Conor A. Nixon, Martin A. Cordiner, Patrick G. J. Irwin, Nicholas A. Teanby, Steven B. Charnley, and Johan E. Lindberg. ISOTOPIC RATIOS OF CARBON AND OXYGEN IN TITAN'S CO USING ALMA. *The Astrophysical Journal*, 821(1):L8, apr 2016.
- [3] M. a. Cordiner, M. Y. Palmer, C. a. Nixon, P. G. J. Irwin, N. a. Teanby, S. B. Charnley, M. J. Mumma, Z. Kisiel, J. Serigano, Y.-J. Kuan, Y.-L. Chuang, and K.-S. Wang. Ethyl Cyanide on Titan: Spectroscopic Detection and Mapping Using Alma. *The Astrophysical Journal*, 800(1):L14, 2015.
- [4] Takahiro Iino, Akira Mizuno, Taku Nakajima, Takehiro Hidemori, Takashi Tsukagoshi, and Chihaya Kato.

- Search for sulfur-bearing species as remnant of cometary impact on Neptune. *Planetary and Space Science*, 104:211–215, dec 2014.
- [5] E Lellouch, C De Bergh, and B Sicardy. Exploring the spatial, temporal, and vertical distribution of methane in Pluto’s atmosphere. *Icarus*, 2014.
- [6] E. Lellouch, C. de Bergh, B. Sicardy, S. Ferron, and H.-U. Käufel. Detection of CO in Triton’s atmosphere and the nature of surface-atmosphere interactions. *Astronomy and Astrophysics*, 512:L8, apr 2010.
- [7] Takahiro Iino, Yasuhiro Hirahara, Takehiro Hidemori, Takashi Tsukagoshi, Taku Nakajima, Satoru Nakamoto, and Chihaya Kato. Observational constraint on Pluto’s atmospheric CO with ASTE. *Publications of the Astronomical Society of Japan*, 68(1):L1, feb 2016.
- [8] E. Lellouch, M. Gurwell, B. Butler, T. Fouchet, P. Lavvas, D.F. Strobel, B. Sicardy, A. Moullet, R. Moreno, D. Bockelee-Morvan, N. Biver, L. Young, D. Lis, J. Stansberry, A. Stern, H. Weaver, E. Young, X. Zhu, and J. Boissier. Detection of CO and HCN in Pluto’s atmosphere with ALMA. *Icarus*, 286:289–307, apr 2017.

時空間変動データからの ホットスポット自動抽出・要約システムの開発

本田 理恵^{*1}, 林 諒^{*1}

Development of Automatic Extraction and Summarization System for Hot spots in Spatio-Temporal Data

Rie HONDA^{*1} and Ryo HAYASHI^{*1}

Abstract

"The hot spots" which have the values unlike ones in the neighborhood often appear in the spatio-temporal data set in the field of space and planetary sciences. Method of automatic extraction of such hotspots as objects from the spatio-temporal data by using mixture distribution of multivariate normal distributions are described, which aims to understand changing patterns such as co-occurrence or temporal rules among the hot spots and to predict them. This paper also introduces the example system in which hotspots are extracted from meteorological satellite imagery. Plans for future work are also discussed based on the problem that became clear through this study example.

Keywords: spatio-temporal data, hot spot, object extraction, pattern discovery, mixture distribution

概要

宇宙、惑星科学の分野で取得される時空間データには、周囲とは異なる値を持つ“ホットスポット”に着目すべき問題が現れることが多い。このホットスポット領域をオブジェクトとして自動的に抽出して記載し、その時空間変動パターンを抽出することができれば様々な問題に応用出来ると考えられる。本研究では時系列グリッドデータから多変量正規分布の混合分布でモデル化し、そのモデルパラメータを求めることによってホットスポットを自動抽出する手法と、これを使って気象衛星画像からインタラクティブ知識発見システムの構築例について紹介し、これらの研究例を通して明らかになった問題から今後の方向性について述べる。

1. はじめに

近年諸分野で大量の時空間データが蓄積されるようになってきている。日常生活ではセキュリティカメラの画像などが典型的な例であるが、地球観測衛星や宇宙惑星科学分野でもリアルタイムモニタリングされた様々なデータが蓄積され、時間とともに成長する巨大な時空間データのアーカイブを形成している。こうした時空間のビッグデータから変動パターンを抽出し、自然・社会現象の変動パターンのモデルを推定することができれば、現象の理解や予測に利用することができる。

このような時空間データには周囲と異なる値をもついわゆる”ホット(あるいはコールド)スポット”が含まれることが多い。気象衛星画像であれば雲の存在領域が、太陽観測衛星のデータにおいては太陽黒点などがこれにあたる。他にも地球観測衛星で得られる植生指標の異常値領域、地震学的なトモグラフィーで得られる地球内部の高温・低温領域もこれにあたる。流体力学のシミュレーションによって生成された様々な物理量(温度、渦度など)もホットスポットの典型例を含む。

このような時空間データは、空間的には2次元、すなわち時系列画像であることが多かったが、最近では3次元データの空間データも取得されるようになってきている。3次元シミュレーションの物理量の記録はこの典型例であるが、観測の分野でもフェーズドアレイ気象レーダによってリアルタイムの3次元の反射強度(降水コアの存在を示す)が取得されるようになってきている^[1]。

本論文では、このホットスポットに基づいて時空間データの要約と知識発見を行う問題について、気象衛星画像やフェーズドアレイ気象レーダデータに対して、筆者らが検討しているオブジェクト抽出システムの検討内容を汎用的な時空間データからのホットスポット抽出の観点からまとめて報告し、そこに見られる問題点を整理して、今後の開発の方向性を探ることを目的とする。

doi: 10.20637/JAXA-RR-17-009/0004

* 平成30年1月18日受付 (Received January 18, 2018)

^{*1} 高知大学 (Kochi University)

2. 手法

2.1 ターゲット，タスク定義

最初にターゲットとなるデータの性質について定義する．ターゲットは2次元，または3次元のグリッド状に配置されたフィールド値の時系列データとする．このフィールド値は気象画像では輝度，フェーズドアレイ気象レーダでは反射強度などの物理量を反映した値にあたる．実世界の様々な分野では，まばらにおかれたセンサや移動するセンサから時間とともに取得されるデータなど非グリッド状のデータも普遍的に存在するが，ここでは簡単化と，実際に地球観測衛星や宇宙惑星科学データで取得されるデータは最終的にはグリッドデータの形に変換されることが多いという理由から，対象をグリッドデータに限るものとする．

2.2 ホットスポットの定義

抽出対象であるホットスポットは以下のような特徴をもつものとする．

- (1) 周囲と異なる値をもつ塊状の領域として現れる．
- (2) 誕生から消滅に至る生存期間をもつ．
- (3) 2つ以上のホットスポットは重なって存在しうる．
- (4) ホットスポット同士は分裂，融合などの相互作用を行い得る．

これは気象画像での台風等の自然界で発生するホットスポットを想定したものとなっている．

こうしたホットスポットは“オブジェクト“の1種として捉えることができる．”オブジェクト“という言葉は画像認識では物体を表すものとして，またオブジェクト指向では属性と手段をもつ実体を意味するものとして用いられる．本研究ではホットスポットをオブジェクトとして表現することで，このような特徴を持つものとしてホットスポットを特徴付ける．例えばホットスポットの抽出後にはその細かい特徴(サイズ，分散，あるいはパターン内部のテクスチャ)はオブジェクトの属性として記載する．以降では一般論としてのオブジェクトとホットスポットを必要に応じて使い分ける．

2.3 多変量正規分布の混合モデルによるオブジェクト抽出手法(標準法)

まず図1のように塊状に分布する M 次元のデータ集合 D の塊をオブジェクトとして表現する事を考える．

$$D = \{d_i \in R^M \mid i = 1, 2, 3, \dots, n\} \quad (1)$$

ここでは重なって存在する不特定の形状を扱うため，オブジェクトを多変量正規分布で表現し，この混合分布で全体を表現するものとする．観測事例の分布を多変量正規分布の混合モデルで近似する手法は機械学習分野でのクラスタリングに一般的に用いられる手法^[2]でもある．この手法について以下に説明する．

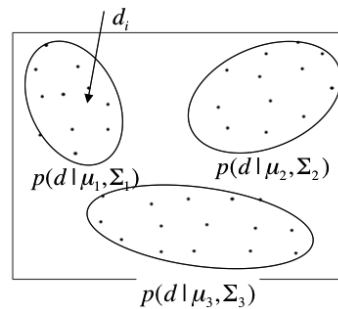


図1. オブジェクトの多変量正規分布の混合分布近似 ($M=2$ の場合)

図1の1点 d を観測する確率密度分布が下記の通り 1つのオブジェクトの存在を表す多変量正規分布でモデル化できるものとする．

$$p(d | \mu, \Sigma) = \frac{1}{(\sqrt{2\pi})^M \sqrt{|\Sigma|}} \exp\left\{-\frac{1}{2}(d - \mu)^T \Sigma^{-1}(d - \mu)\right\} \quad (2)$$

ここで $\mu \in R^M$ は多変量正規分布の中心ベクトル(平均値)， Σ は $M \times M$ の分散共分散行列とする． μ はオブジェクトの中心， Σ はオブジェクトの広がりを表す．複数のオブジェクトが存在しうる時，座標 d においてオブジェクトを観測する確率密度分布は複数成分からなる多変量正規分布の重み付きの平均で表すことができる．

$$P(\mathbf{d}) = \sum_{j=1}^K \omega_j p(\mathbf{d} | \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j) \quad (3)$$

ここで ω_j は重み係数($\sum_{j=1}^K \omega_j = 1$), K は成分数を表す. このパラメータをまとめて以下の通りとする.

$$\boldsymbol{\theta} = \{(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j, \omega_j) | j = 1, \dots, K\} \quad (4)$$

観測値 \mathbf{D} は, それぞれどの多変量正規分布から発生したものかを知ることができないいわゆる不完全データであり, 最尤推定などの手法で直接パラメータを求めることができない. このため, EM (Expectation and Maximization) アルゴリズム^[3]でパラメータを求める. EM アルゴリズムは, 仮に与えたパラメータ $\boldsymbol{\theta}$ からモデルを計算し, さらにモデルの対数尤度を増加させるような新しいパラメータ $\boldsymbol{\theta}'$ を推定するという操作を繰り返すことによって対数尤度 $L(\boldsymbol{\theta})$ を最大化するパラメータ $\boldsymbol{\theta}'$ を求めるものである.

$$L(\boldsymbol{\theta}) = \log\left(\prod_{i=1}^n P(d_i | \boldsymbol{\theta})\right) \quad (5)$$

$$\boldsymbol{\theta}' = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} L(\boldsymbol{\theta}) \quad (6)$$

まず標準的な方式での EM アルゴリズムによるパラメータの計算手順^[4]を示す.

<EM アルゴリズムによるパラメータ推定>

E ステップで寄与率を推定し, これをもとに M ステップで対数尤度を最大化するパラメータを推定し, この値を用いて寄与率を推定するというプロセスを収束するまで続ける. ここで寄与率 $z(i, j)$ は, 観測値 \mathbf{d}_i が j 番目の多変量正規分布 (オブジェクト) から発生したことを示す推定量である.

1. $\boldsymbol{\theta} = \{(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j, \omega_j) | j = 1, 2, \dots, K\}$ に初期値を与える.
2. **E-Step:** 寄与率 $z(i, j)$ を計算する. このとき変化 $z(i, j) - z'(i, j)$ が小さければ終了. そうでない場合 3 を実行する.

$$z'(i, j) = \frac{\omega_j p(\mathbf{d}_i | \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}{\sum_{k=1}^K \omega_k p(\mathbf{d}_i | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)} \quad (7)$$

3. **M-Step:** 対数尤度を最大化させるようモデルパラメータを更新し, $z(i, j) \leftarrow z'(i, j)$ として2に戻る.

$$\omega_j' = \frac{\sum_{i=1}^n z'(i, j)}{n} \quad (8)$$

$$\boldsymbol{\mu}_j' = \frac{\sum_{i=1}^n z'(i, j) \mathbf{d}_i}{\sum_{i=1}^n z'(i, j)} \quad (9)$$

$$\boldsymbol{\Sigma}_j' = \frac{\sum_{i=1}^n z'(i, j) (\mathbf{d}_i - \boldsymbol{\mu}_j) (\mathbf{d}_i - \boldsymbol{\mu}_j)^T}{\sum_{i=1}^n z'(i, j)} \quad (10)$$

なお成分数 K については総当たりで上記の手法でもとめたモデルパラメータに対してバイズの情報量基準 (BIC)^[5]を用いて, 最適な(すなわちこれを最小化する) K を求める. F は自由度である.

$$\text{BIC} = -2nL(\boldsymbol{\theta}) + F \log n \quad (11)$$

2.4 フィールド値をもつグリッドデータからのホットスポット抽出への応用

2.3 の手法をグリッドデータに適用するためには, 単純には観測値が閾値をこえた座標をサンプリングするアプローチをとることができる^{[6][7][8]}. M 次元のグリッドの座標 \mathbf{d}_i に存在するフィールドデータ $I(\mathbf{d}_i)$ に対してある閾値 I_{th} を超えた座標 \mathbf{d}_i をホットスポットの存在する点の座標集合 \mathbf{D} としてサンプリングして使用する.

$$D = \{\mathbf{d}_i | I(\mathbf{d}_i) > I_{th}, i = 1, 2, 3, \dots, n\} \quad (12)$$

ただしこの手法では1座標(グリッド点)に対して観測点が1点となってしまう、生成されたデータは平坦な確率密度分布を示す事になり、モデルとした多変量正規分布からの乖離が大きくなってしまう。気象レーダの反射強度などでは塊の中心にむかってその値が増加し、フィールド値自身が多変量正規分布に類似した分布をしている場合も多いため、問題によっては観測値そのものを有効活用したほうが良い。

以下では気象レーダデータのようにフィールド値自身が多変量正規分布に調和的であり、中心ほどその値が高いというケースを考える。この場合、観測値の大きさ自体が観測頻度に比例すると仮定することによって、フィールド値の分布により適合したモデルパラメータを求めることができる。あらかじめバックグラウンドノイズやオフセットの影響を取り除くため、フィールド値に対して、閾値 I_{th} を差し引くオフセット処理を行う。

$$\delta I_i = \max\{I(\mathbf{d}_i) - I_{th}\} \quad (13)$$

この項を利用して、2.3の標準法(8),(9),(10)式の $\sum_{i=1}^n 1$ を $\sum_{i=1}^n \delta I_i$ で置き換えることによって、フィールド値が確率密度を反映するという効果を簡易に取り込むことができる。以下にその変更部分のみを示す。

<EM アルゴリズムによるパラメータ推定。フィールド値を活用する場合の3. M-Step 変更部分のみ抜粋>

$$\omega_j' = \frac{\sum_{i=1}^n z'(i,j) \delta I_i}{\sum_{i=1}^n \delta I_i} \quad (14)$$

$$\mu_j' = \frac{\sum_{i=1}^n z'(i,j) \mathbf{d}_i \delta I_i}{\sum_{i=1}^n z'(i,j) \delta I_i} \quad (15)$$

$$\Sigma_j' = \frac{\sum_{i=1}^n z'(i,j) (\mathbf{d}_i - \mu_j') (\mathbf{d}_i - \mu_j')^T \delta I_i}{\sum_{i=1}^n z'(i,j) \delta I_i} \quad (16)$$

多変量正規分布の混合分布から生成した3次元の模擬データを用いた実験^[9]では、この変更を施すことによって、標準手法では縮退や分裂を起こして正しい成分数を求める事ができなかったケースでも、正しい成分数の解が求められるようになり、モデルとデータの一致度が大きく改善したことを確認している。

なお、この手法はそれぞれデータの性質によって選択して使用されるべきものである。2.6で述べる気象画像からの雲塊の抽出などの問題では、中心ほど信号が高いわけではなく、閾値処理によって抽出されたサンプリング点の分布自体がモデルパラメータの決定において重要である。このような場合は従来のフィールド値を使わない標準手法が適当である。一方、気象レーダの反射強度など中心付近にむかって信号が大きくなる性質のある問題においては、フィールド値が確率密度分布モデルの値に比例することを仮定する今回の手法(2.4節)を用いる事が適当であると考えられる。

2.5 ホットスポットの消滅、融合、生成

ここまで述べた手法で、ある時間断面で抽出されたホットスポットは、一定の期間存続し、やがて消滅する。また生存期間のうちに他のホットスポットと融合したり、分裂してあらたなホットスポットを生成したりする。この過程を表現するために、パラメータの抽出過程において前の時間の解を次の時間に継承させる手法を提案している^{[7][8]}。具体的には前の時間の解をラベルとともに次の時間の解の初期値に継承させるという単純な手続きによって、時間的な連続性をもってホットスポットの追跡を行うことを可能としている。一方、消滅や分裂が発生する可能性も考慮して、解の継承の際に下記のような複数の初期値を与えて、並列(下記の場合 $2q+1$ ケース)にモデリングし、BICに従って最良の解を選択する手法をとっている。

- (1) 前の時間の解そのものを初期値として与える(成分数 K , 1ケース)
- (2) 前の時間の解から重み係数の低い順に解を q 個まで消去して初期値として与える(成分数 $K-1, \dots, K-q$, q ケース)
- (3) 前の時間の解から重み係数の大きい順に解を q 個までそれぞれ2つに多重化して初期値として与える(成分数 $K+1, \dots, K+q$, q ケース)

実際に気象画像や気象レーダデータに適用した例では、時間ステップで分布が大きく様変わりすることなければ、 q は2-3で適当なことが多い。また、ここでは新たに別の箇所にホットスポットが発生するケースは陽には扱っていないが、実際には(3)の過程で分裂用に生成された初期解が、この新規発生成分へ移動して

発見されていることが多い(2.6, 図2参照)が、この弊害として”新規発生”と”分裂”が混同されて扱われてしまうことになる。このような影響についても、事前、事後に判定するプロセスを付加する等、今後より詳細な検討を進めていく予定である。

2.6 ホットスポット抽出の実例

本章で述べたホットスポット抽出手法の実データへの適用例として、気象衛星ひまわり IR 画像 (MTSAT-7, 高知大気象情報頁^[9]) からの雲塊抽出の例を紹介する。ターゲットデータは2次元画像であり、画像の輝度が高い部分に雲が存在する。これがホットスポットであるとみなして、この手法を適用する。この場合は雲塊の中心ほど輝度が高いということはないため、閾値によるサンプリングのみ用いる標準的な手法(2.3参照)を使用する。

図2に6時間おきの画像に対して、求められたモデルの多変量正規分布のカラーコンターとそのラベルで抽出結果を示す。図2下の得られたモデルの各成分に付与されているIDは2.5でのべた手順でつけられた同一性を示すラベルである。このうち図2下中央の7-1, 7-2のような番号はその前の時間で1つであった7番のホットスポットから分裂したことを示している。全般的には、発見的に雲の塊を個数も含めて決定できていることがわかる。また、時間を追って同一のホットスポットを追跡できていることもわかる。一方で最初の時間から次の時間に移行する過程で7番目の成分が2つに分裂したとみなされているが、この1つは実は新規発生であって分裂を示している訳ではない。これは2.5で述べた通り現在のフレームワークには新規発生が含まれていないためであり、今後の検討が必要である。

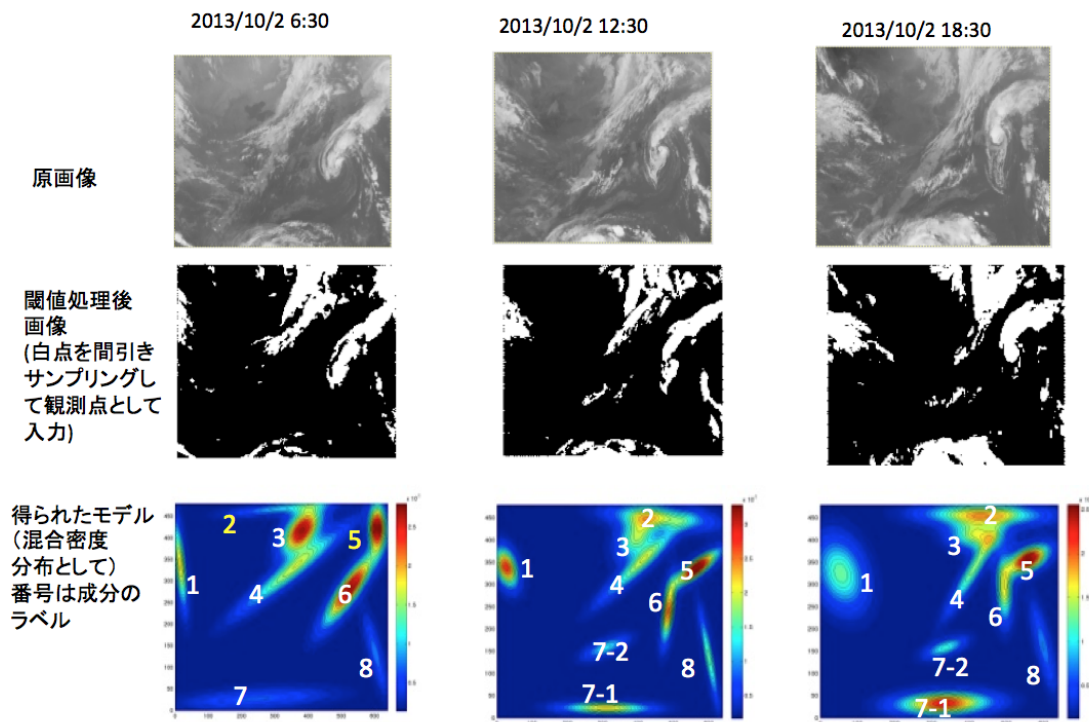


図2. 気象衛星画像(2013年10月2日)からのホットスポット抽出の例。

次に3次元データへの適用例として、フェーズドアレイ気象レーダで観測した反射強度データからの降水コア抽出^[1]への適用例を示す。図3の例では赤みがかかった部分ほどレーダの反射強度が大きく、この塊状の領域が濃い雨雲のある降水コアにあたる。降水コアの中心に向かって反射強度の値が大きくなっていることと、この降水コアの上昇や下降が豪雨の発展過程と関わっていると考えられることから、2-4の強度の重みを使用する手法によるモデリングの精度の向上が有効と考えられる。

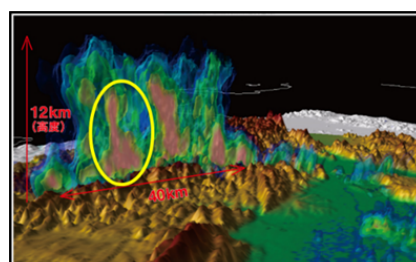


図3 フェーズドアレイ気象レーダデータに現れた雨雲とホットスポットの例^[1]

図4に実際に2012年7月16日大阪大学吹田キャンパス付近で発生した豪雨をフェーズドアレイ気象レーダで観測した反射強度データに対して本手法を適用した例を示す^[10]。ここでは閾値を30dBZとして座標のみを取り出して2.3の標準的な手法を使用したケース(図4左)と、閾値処理後フィールド値(反射強度)を重みとして2.4の改良手法を使用したケース(図4右)の比較を示す。簡単のため各観測点において寄与率が最大の成分毎に同じ色で示すことで、抽出された成分の分布を示している。座標のみを取り出して入力値をしたケース(図4左)では、特に右上の大きな塊が過剰に細かい成分に分裂しているが、フィールド値を用いた手法(図4右)では、この分裂が抑制され、自然な分布に近づいていることが観察できる。よって予備的には中心に向かってフィールド値が大きくなるようなグリッドデータからホットスポットをオブジェクトとして抽出する問題においては、フィールド値を使用することが効果的であることが示される。

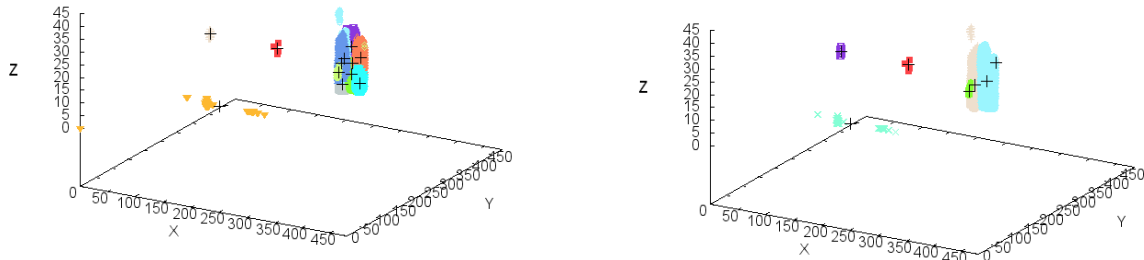


図4. 2012年7月16日17時00分の大阪大学吹田キャンパスのフェーズドアレイ気象レーダ^[1]で観測した反射強度のモデリングの結果得られた多変量正規分布の成分の分布。左は敷居値処理(閾値30dBZ)によって座標のみを入力値として使用した結果で、右は敷居値処理後、強度による重み付けを行った結果^[10]。各点を寄与率が最大の成分毎に同じ色で表現している。また黒い十字は各成分の重心である。

3 時空間ホットスポット抽出、要約システム

3.1 システムの概要

ここまで述べてきたホットスポットの抽出手法をコアプロセスとして、汎用的な時空間ホットスポット抽出、要約システム的设计について述べる。満たすべき要件は以下の4つになる。

- (1) ホットスポット=オブジェクトを自動的に時系列2次元, 3次元グリッドデータから抽出できること
- (2) ホットスポットの特徴に応じて意味的ラベル付けができること
- (3) これらの情報がデータベースに格納され、簡易に検索できること
- (4) これらの情報を活用して、ホットスポットの相互作用や変遷を時間、空間の両面で要約、可視化できること

抽出されたホットスポットには2で述べられた手法を適用することにより、中心位置、広がり、存続期間、親子関係のラベルなどが自然に付帯するが、そのホットスポットが何を表しているのかという意味的なラベル(2)がさらに付与されると、次の段階での知識発見における有用性がさらに増す。この意味的ラベル付けにはパターン認識、機械学習のアルゴリズムを用いる。先行研究^{[7][8]}ではKohonenの自己組織化マップ^[11]による教師なし学習によるクラスタリングに基づくラベル付けを行っているが、教師データを積極的に利用して、Deep Learningなどのより精度の高い分類アルゴリズムを用いてもよい。

これを実現するシステムのイメージ図は図5のようになる。意味付けの箇所は必要に応じて様々なアルゴリズムで置き換えることが想定される。またユーザとのインターフェースは実際にはweb severを通じて行うことを想定する。

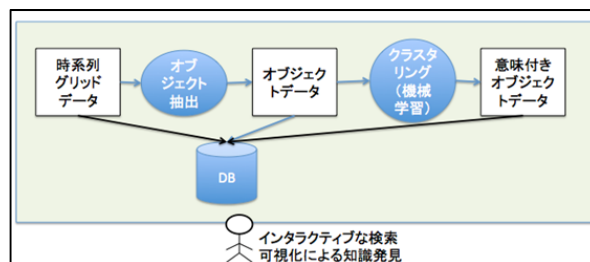


図5 時空間ホットスポット抽出、要約システムの概念図

3.2 データモデル

このようなシステムを構築するにあたっては、オブジェクトについてのデータモデルを整理しておく必要がある。図6に松永ほか(2016)^[8]の検討結果を示す。ここに示されているように、ここまで抽出対象としてきた

オブジェクトは”スナップショット的な画像内のみかけのオブジェクト”であり、その背後で時間的に変化しながら存続し続ける”真のオブジェクト”の時間断面にすぎない。またオブジェクト間には分裂によって生じる親子関係としてファミリーに相当するものが存在する。実際にはこの3つは図6下の図の概念図のように階層的な構造で表現可能である。なおこの概念図ではオブジェクトの融合についてはまだ含まれていない。

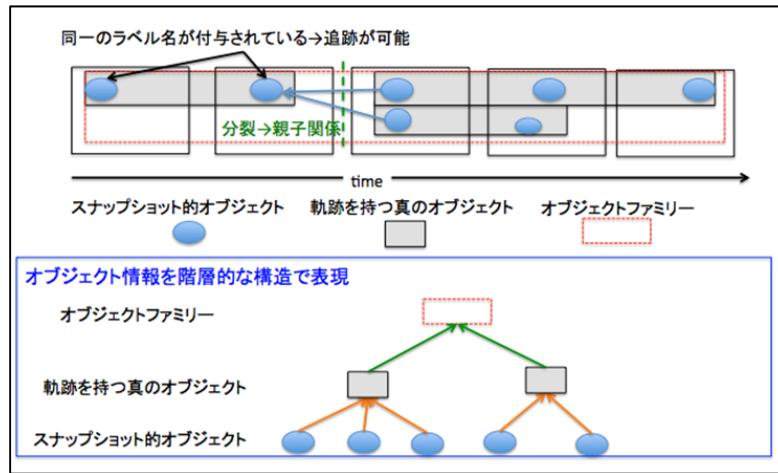


図6 オブジェクトの階層構造([8]に基づく)

図7にはさらにこの3つの階層のオブジェクトのモデルをUML (Unified Modeling Language) のクラス図で示す。”真のオブジェクト”は”スナップショットオブジェクト”の集約として、また”オブジェクトファミリー”はこの真のオブジェクトの集約として定義される。真のオブジェクト、オブジェクトファミリーの情報は、2で述べたオブジェクト抽出手法で自然に得られるスナップショット的オブジェクトの情報から再構成され、データベースに格納されるものとする。

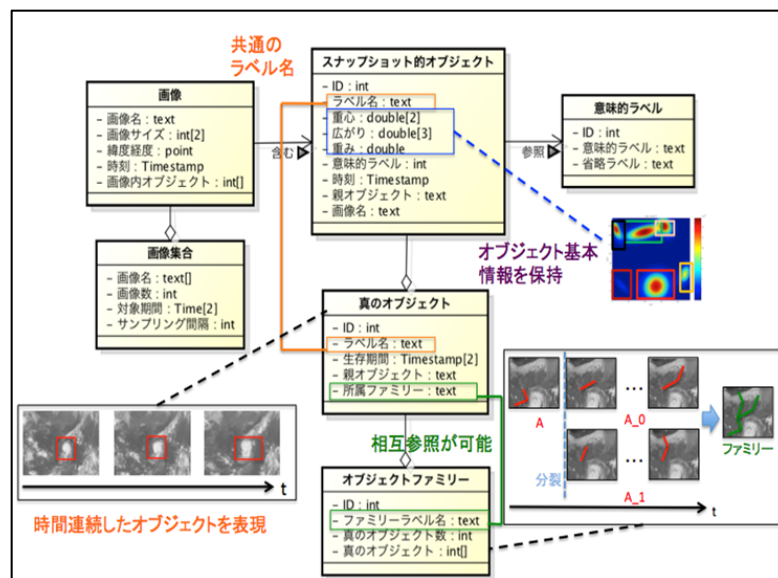


図7 オブジェクト (ホットスポット) のデータモデル([8]に基づく)

3.3 実装例 -気象画像-

実装例としてここでは 2006 年 1 年間の日本付近のひまわり画像 (MTSAT-7, 高知大気象情報頁^[9]) に対して構築したシステムの例を紹介する^[4]。3 時間おきにサンプリングし、640 x 480 pixel に整形された日本付近の 2920 枚の画像に対してホットスポット抽出をおこなった。ここでは先行研究との用語の統一のために、ホットスポットをオブジェクトという言葉で表現するものとする。ホットスポット (オブジェクト) の意味的ラベリングにはオブジェクトを切り出した画像の FFT パワースペクトルを特徴量として自己組織化マップ^[11]を使用してクラスタリングさせ、さらに人間が当時の天気図をもとにしてクラスタごとに意味付けすることによって行った。

実験の結果、抽出されたオブジェクト数は 14058、所要時間は Intel core i5 3.2GHz (iMac) で 7094sec

であった。なお学習に要した時間は同じ環境で 4239sec であった。クラスタリングでは初期に 36 ユニットに分類された結果について、類似のクラスタをスーパークラスタとしてまとめることによって、最終的に 7 種のグループにまとめ、意味的ラベル(台風, 台風の一部, 寒冷前線, 梅雨前線, 温暖前線, 停滞前線, それ以外)を付与した。意味的ラベルの確からしさをクラスタごとの適合オブジェクトの割合 (精度: Precision) で評価するとその平均値は 61 % 程度であった。この値はパターン認識の精度(最近の畳み込みニューラルネットワーク等による画像分類のコンペティション ILSVRC2016 では上位では 95%以上達成^[12], 実用のためには 80%以上程度は必要と考えられる)としては低いが, これは教師なしのアルゴリズムで発見的にクラスタリングを行っていることによるもので, 教師学習アルゴリズムを使用することで性能を向上させられると考えられる。

図 8 に構築されたシステムのトップページを示す。月ごと, もしくは年ごとにどのような種類のホットスポットがどれだけ抽出されたのか, その頻度分布が表示されている。さらに時間指定をすることによって, 図 9 のようにスナップショット画像の上に抽出したオブジェクトの領域を表示し, その意味的ラベルを色で表示するビューを開くことができる。図 9 の対象画像には 2 つの台風が含まれているが 1 つは台風としての認識に成功し, もう 1 つは失敗していることがわかる。さらにそれぞれのオブジェクトの軌跡を表示することもできる。

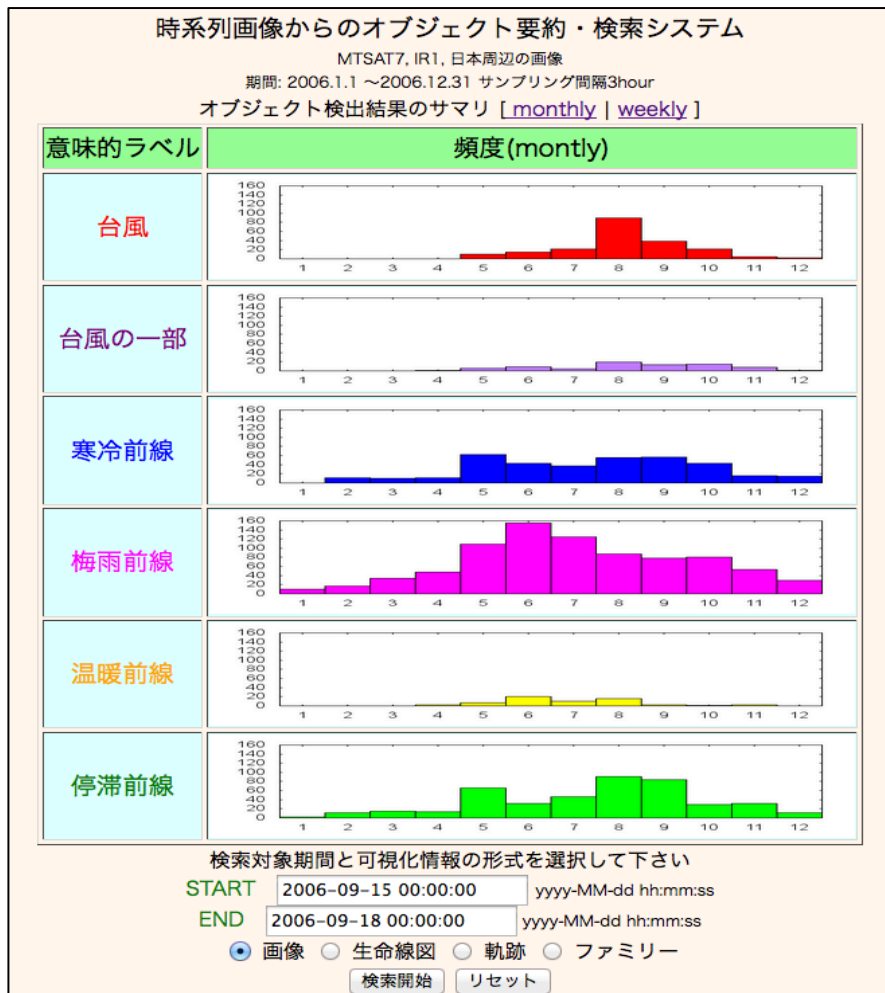


図 8 構築されたシステムのサマリー頁([8]に基づく)

また図 10 のように, 横軸を時間, 縦軸をオブジェクト ID として, UML の生命線図に類似した形で, オブジェクトの生存期間や分裂, 相互作用, 共起関係の視覚的な発見に用いることのできる画面を表示することができる。これらは高次の知識発見の前に, ユーザが試行錯誤をしながらインタラクティブにパターン発見を実施し, 仮説を検証するのに活用できると考えられる。図 10 を詳細に見ると, 同じオブジェクトでも時間によって異なる意味的ラベルがつけられていることがわかる (グラフ上部の赤枠内で台風が頻繁にその他にラベル付けされているなど)。これは, 現象としてはホットスポットの性質が時間とともに代わりうると考えられることから自然なことではあるが, 本来はある安定度をもって同じラベルが付与されるべきもので頻繁に入れ替わるべき物ではない。これは意味的ラベル付けの精度がまだ良くないことと, 各時間で独立にラベル付けをしてしまっていることの影響をうけている。意味的ラベル付けの精度向上とその時間的連続性の考慮は, 今後検討すべき課題である。

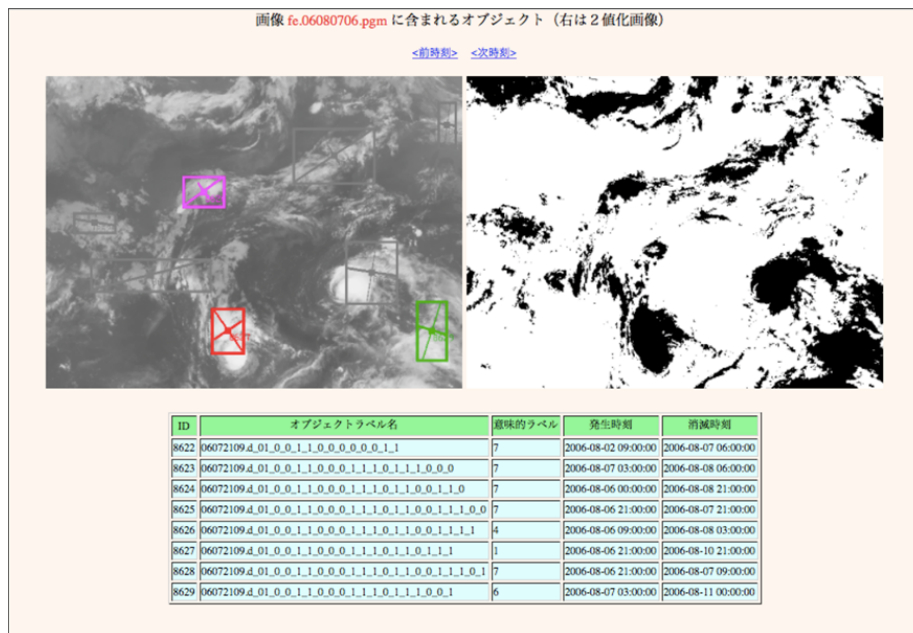


図9. スナップショットで抽出されたホットスポットの表示. オブジェクトの色は図8の意味的ラベル毎の表示と同様(たとえば, 台風は赤). 灰色はその他のオブジェクトに分類されたことを示す. ([8]に基づく)

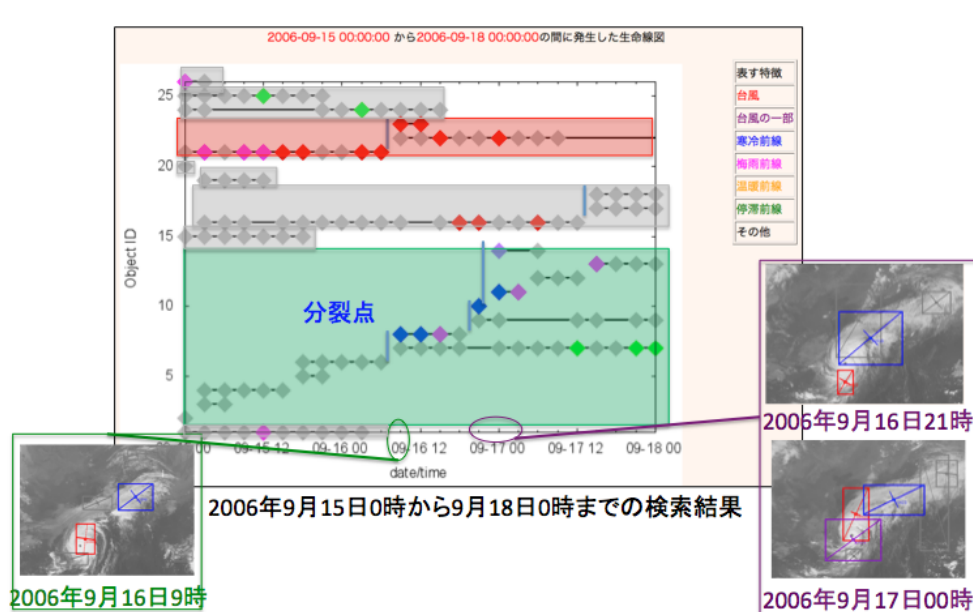


図10. オブジェクトの生命線図. オブジェクトの意味的ラベル毎の色は図8, 9の表示と同様. ([8]に基づく)

4. 議論と今後の検討課題

以上の内容から汎用的な時空間グリッドデータからのホットスポット抽出, 要約システムの構築の検討課題についてまとめる. 気象画像やフェーズドアレイ気象レーダへの適用結果から, フィールド値の値そのものがホットスポットの中心や多変量正規分布の形状を規定している場合とそうでない場合を識別し, その両者に適したアルゴリズムを示した. 特に現在検討を始めているフェーズドアレイ気象レーダデータからの降水コアの抽出^[10]に関してはフィールド値を活用することによってより精密なモデリングが可能になる可能性が予備的に示された. またスナップショットオブジェクト, 真のオブジェクト, オブジェクトファミリーの3つでデータモデルを整理することによって, インタラクティブな知識発見システムを構築することが可能になった.

これらによって, ホットスポットの自動抽出, 要約のための基本的な要件が押さえられたと考えられるが, (a) 時間的連続性についての単純な扱い, (b) 意味的ラベルの精度の不足などの問題が残っている. これらの問題は互いにかかり合っているが, 意味的ラベルの精度不足については, 教師有り学習の活用がその解決策としてあげられる. 一方, 時間的連続性の取り扱いが単純過ぎることで, 新規発生と分裂が混同される, 意味的ラベルが頻繁に入れ替わるといった問題が起こっていることに関しては, 事後の検証で修正する手法

と、時間変化を考慮したモデルに拡張することが考えられる。短期的な検討としてはまず現在の検討の自然な延長として事後の検証を取り入れることによる修正を検討する予定である。これらを検討しながらリアルデータでの実装システムの構築と、汎用システムの構築を目指した基礎的な検討を並行して進めていく予定である。

5. まとめ

本研究では、時系列からホットスポットをオブジェクトとして抽出、要約することによって時空間知識発見を行うためのシステムを構築するための基礎手法とモデルについて検討し、その結果を気象画像に適用した例を紹介した。現時点で、基礎的な手法と検討が出そろった状態であるが、一方で、(a)時間的連続性の扱い、(b)意味的ラベルの精度の不足などの問題が現れている。今後はこれらの問題についてリアルデータでの実装システムの構築と、汎用システムの構築を目指した基礎的な検討を並行して検討していく。

謝辞

気象衛星画像の解析を担当した松永知也氏、3次元に対応した手法の開発を分担した松岡愛美氏、フェーズドアレイ気象レーダデータを提供いただいたNICT佐藤晋介氏、村田健史氏に感謝します。また有益な意見をいただいた査読者に感謝します。

参考文献

- [1] 佐藤晋介, 牛尾知雄, 水谷文彦, フェーズドアレイ気象レーダの研究開発, NICT News, 2013.1, 2013, 3-5.
- [2] MacQueen, J., Some methods for classification and analysis of multivariate observations. In Proceedings of the fifth Berkeley symposium on mathematical statistics and probability (Vol. 1, No. 14), 1967, 281-297.
- [3] Dempster, A. P., Laird, N. M., and Rubin, D. B., Maximum likelihood from incomplete data via the EM algorithm. Journal of the royal statistical society. Series B (methodological), 1977, 1-38.
- [4] Bishop, C. M., Pattern Recognition and Machine Learning, Springer Verlag New York, LLC., 2006.
- [5] Schwarz, G., Estimating the dimension of a model, Annals of Statistics 6, 461-464.
- [6] Honda, R., Wang, S., Kikuchi, T. and Konishi, O., Mining of objects from time-series images and its application to satellite weather imagery, Journal of Intelligent Information Science, 19:1, 2002, 79-93.
- [7] 松永知也, 本田理恵, 時系列画像からのオブジェクトベースデータマイニング -オブジェクトの抽出とデータベース化-, DEIM フォーラム 2015, 2015, P2-4, pp.6.
- [8] 松永知也, 森啓太, 本田理恵, 時系列画像に含まれるオブジェクト特徴の変遷要約とその可視化, DEIM フォーラム 2016, 2016, P1-2,
- [9] 高知大学気象情報頁, <http://weather.is.kochi-u.ac.jp>, 2017.6.30 参照.
- [10] 林諒, 本田理恵, 佐藤晋介, 村田健史, 村永和哉, 鶴川健太郎, 佐々浩司, 村田文絵, 時系列 3 次元グリッドデータからのホットスポットの自動抽出・追跡法の開発 -フェーズドアレイ気象レーダデータによる局地的大雨解析への適用-, DEIM フォーラム 2017, 2017, P4-4, pp.7.
- [11] Kohonen, T., Self Organizing Maps, Springer, 3rd ed., 2000.
- [12] UNC vision lab, Large Scale Visual Recognition Challenge 2016 (ILSVRC2016), <http://image-net.org/challenges/LSVRC/2016/index>, 2017.

惑星の位置関係に基づく深発月震分類のための特徴量の検討

加藤広大^{*1}, 山田竜平^{*2}, 山本幸生^{*3}, 廣田雅春^{*4},
横山昌平^{*5}, 石川博^{*1}

Investigation of Orbit Parameters to Classify the Deep Moonquake Sources

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

Abstract

In this paper, we evaluate features based on time of occurrence of deep moonquakes different from the waveforms to classify seismic sources. Classification of the sources of the deep moonquakes is an important issue for analyzing the focal mechanisms and the lunar deep structures. Previous researches found that deep moonquakes which occur from the same source have similar waveforms. Some studies have been conducted to identify the source of deep moonquake using the waveform similarities. However, classifying some deep moonquakes using only the waveforms is difficult due to large noise and the small amplitude. If we could show that other features different from the waveforms are effective for classification of deep moonquakes, we can increase the number of classifiable moonquakes even if moonquakes include noise and small amplitude of the waveforms. Therefore, we use other features to classify deep moonquakes (position and velocity relative to the Earth, Sun, Jupiter, and Venus, as seen from the Moon). We apply these features to classify deep moonquakes that are not classified based on only waveforms, and it is useful to analyze the deep moonquake occurrence mechanisms. Our experiments showed that the position and velocity relation between the Moon and the Earth are effective for classification.

Key Words: Moonquake, Machine Learning, Random Forest

概要

本論文では、深発月震の発生時刻に基づいた特徴量から、機械学習を用いて、震源を分類するのに有効な特徴量の発見を試みる。月震の解析は、月の内部構造や月震の発生原因などを解析する上で非常に重要であり、現在も解析が行なわれている。月震解析の一つに、深発月震の震源分類がある。深発月震に着目すると、同一の震源から発生する深発月震は波形が類似していることが知られており、従来、深発月震の震源分類は、主に波形の相互相関から分類が行われてきた。しかし、深発月震は振幅が小さく、ノイズが含まれているものも多いため、波形の類似性のみでは震源が推定できないものが多数存在している。本論文では、震源分類のための特徴量として、月、地球、太陽、木星の位置関係、速度を特徴量として検証した。これらの特徴量は、波形では分類できなかった深発月震の震源分類にも適用可能である。また、震源ごとの発生要因の分析への利用が可能である。結果として、月-地球の位置関係、速度が震源分類に寄与し、また、分類する震源の組み合わせによって、寄与する特徴量が異なることを示した。

doi: 10.20637/JAXA-RR-17-009/0005

* 平成 30 年 1 月 18 日受付 (Received January 18, 2018)

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

^{*2} 会津大学コンピュータ理工学部 (Computer Science and Engineering, The University of Aizu)

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

^{*4} 岡山理科大学総合情報学部 (Department of Information Science, Okayama University of Science)

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

1 はじめに

NASA が行った月への有人宇宙飛行計画であるアポロミッションの目的の一つには、月への観測装置の設置があり、Apollo Lunar Surface Experiments Package (ALSEP) と呼ばれる観測装置群が月面に設置された。ALSEP に含まれる地震計を用いて、Passive Seismic Experiment (PSE) と呼ばれる月地震の観測が行われた。観測の結果、月にも地震のような揺れがあることがわかった（これらを月震と呼ぶ）観測されたデータはすべて Web 上で公開され、データの閲覧、取得が可能となっている [17]*1。

PSE によって取得されたデータは、月の研究において貴重なデータであり、アポロミッションから約 40 年が経過した現在でも解析が行なわれており、月震の発生メカニズムの解明や、月の内部構造の推定に利用されている [12, 9]。

これまでの解析から、月震は、地球の地震と大きく異なる性質を持つ事がわかっている。まず、地球での地震の発生要因となるプレート運動は月には存在しない。そして、月震は、発生要因や震源の深さから、深発月震、浅発月震、隕石衝突、熱月震などに分類されている。月震の約半数は深発月震である。これまでに、約 13,000 件のイベントが発見されており、深発月震は約 7,300 件、浅発月震は約 30 件、隕石衝突は約 1,700 件である。深発月震に着目すると、同一の震源から発生する深発月震は波形が類似していることが知られており [2, 8]、これらの類似性を利用した震源分類基準の作成と分類が行われた [11] (深発月震の震源は、A1, A6 のように Axx とラベリングされる)。本研究の目的は、未分類の深発月震を分類し、月の内部構造の解明を行うことである。

また、最近の研究では、月震波形の前処理や、検出手法を改良することで、Bulow ら [3]、Endrun ら [6] は、これまで発見されていなかった月震の検出、未分類であった月震の分類を行った。また、機械学習を用いて、波形の周波数成分に着目した震源の分析も行われている [5]。機械学習を用いる利点は、専門家が人手で行っていた分析を半自動化することで、人的コストの削減につながることである。

しかし、これまでの月震分類の手法は、すべて波形の特徴を用いた分類である。波形は震源分類に有効な特徴量のひとつであるが、ノイズや振幅が小さいために、波形によるこれまでの震源分類を適用できない月震は未だ多く存在している。また、これまでの研究は、波形情報のみで震源分類を行っているため、波形以外の様々な要因は考慮されていない。そこで、我々は波形以外の情報に着目した分析を行う。深発月震は、同じ震源から周期的に発生することがわかっており、潮汐力と関連させた分析が行なわれている [16, 7, 10]。本研究では、深発月震の発生時刻に着目し、時刻から推定できる特徴量から、震源の分類に有効な波形以外の特徴量を、機械学習を用いて検証する。波形以外の特徴量を考慮した基準を作成することで、ノイズ等の影響で分類できなかった深発月震の震源分類が可能になる。

本論文は、以下の構成に従う。2 章では、本研究に関連する研究について述べる。3 章では、深発月震の震源分類に寄与する特徴量の評価手法について述べる。4 章では、実験結果について述べる。5 章では、考察について述べる。6 章では、本論文のまとめを述べる。

2 関連研究

本章では、これまでの月震分類手法について述べる。

一般に深発月震の震源は、複数の観測点で観測された月震波が震源から観測点まで到達する時間である走時データを利用することで推定される。しかし、振幅の小ささやノイズの影響で、走時データを利用した震源

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

表1 SPICE を用いて算出する座標特徴量のパラメータ

| Target | Observer | 座標系 | 本論文における呼称 |
|-------------------------|-------------------------|-----------|--------------------|
| EARTH BARYCENTER | MOON | IAU_MOON | earth_from_moon |
| SOLAR SYSTEM BARYCENTER | MOON | IAU_MOON | sun_from_moon |
| JUPITER BARYCENTER | MOON | IAU_MOON | jupiter_from_moon |
| SOLAR SYSTEM BARYCENTER | EARTH BARYCENTER | IAU_EARTH | sun_from_earth |
| JUPITER BARYCENTER | EARTH BARYCENTER | IAU_EARTH | jupiter_from_earth |
| SUN | SOLAR SYSTEM BARYCENTER | IAU_EARTH | sun_perturbation |

推定ができない月震は、月震波形の類似性から震源位置の違いが判別されている。ALSEP によってデータが取得された 1970 年代当初は、同一の震源から発生する月震波形が類似することに着目し、特に目視で分類を行っていた [13].

その後、計算機の発達に伴い、Nakamura ら [11] による、月震波形の相互相関係数を用いた最短距離法に基づく階層的クラスタリングによる月震分類が行われた。これが、現在の震源分類基準となっている。最近の研究では、月震波形の前処理を改良することで、Bulow ら [3] は、新しい A1 震源の月震を多数発見している。また、Endrun ら [6] は、隠れマルコフモデルを用いた手法で、大規模な月震データベースから、これまでに発見されていなかった新たな月震を検出することに成功している。後藤ら [5] による月震の可視化システムでは、従来の月震波形の相互相関係数ではなく、周波数成分を特徴量として、教師なし学習のひとつである SOM による深発月震の分類を可視化した。SOM の結果では、複数の震源が混合したクラスタリング結果が得られている。また、ノイズによって波形の性質が埋もれている波形が集まるセルも見られた。菊池ら [18] は、機械学習を用いて深発月震を分類する際の手法比較や最適な波形の切り出し位置などを検討した。しかし、これらの研究は全て波形をもとにした分析を行っている。我々は、波形以外の情報に着目することで、これらの研究で行うことができなかった深発月震に対しても震源分類を行うことを目的としている。

3 提案手法

本論文では、機械学習における代表的な教師あり学習のひとつである Random Forest[1] を拡張した Balanced Random Forest[4] を適用し、特徴量の検証を行う。検証方法として、従来の震源ラベルを正解ラベルとし、月震から生成した特徴量に基づいて、従来の震源ラベルが、本論文の特徴量を用いた Balanced Random Forest によって再現可能かを確認する。

3.1 特徴量

本論文で用いる座標特徴量を表 1 に示す。本論文では、月、太陽、地球、木星それぞれの惑星における位置 (x, y, z) 、速度 (vx, vy, vz) 、距離 (lt) を、表 1 に基づき、月震発生時刻をもとに SPICE*3 を用いて計算し、特徴量とする。sun_perturbation は、太陽摂動のことである。IAU_MOON 座標系は、月を中心とする固定座標系である。z 軸は月の北極方向であり、x 軸は月の子午線方向、y 軸は x-z 平面に対して右方向である。IAU_EARTH 座標系は、地球を中心とする固定座標系である。z 軸は慣用国際原点の方向であり、x 軸は本初子午線の方向、y 軸は x-z 平面に対して右方向である。

また、earth_from_moon の距離における近地点の周期と、近地点の周期をもとにした周期、太陽摂動の x 座標及び y 座標の周期を算出する。これらの周期的特徴量と、これをもとにした位相角から \sin と \cos を特徴量

として抽出する。加えて、moon_from_earth と、sun_from_earth における位置の cos 類似度を恒星月の特徴量として算出する。これらの 11 種類の特徴量のすべての組み合わせのパターンである 55 特徴量を本論文における特徴量として実験を行う。

3.2 Balanced Random Forest

Random Forest[1] は多数の決定木を組み合わせるアンサンブル学習であり、データマイニングなどの分野で広く利用されており、特徴量の寄与率が算出可能という特徴がある。Random Forest の問題点として、学習するデータ数に大きな差が有る場合、分類器が数の多いクラスに偏って学習される問題がある。一般的に、不均衡データに対応した Random Forest は、データ数の少ないクラスに重みをつけることで対応していた。しかし、データ数の偏りが激しいと、少数のクラスのデータの重みが大きくなってしまい、少数のクラスへの過学習の原因になると考えられる。深発月震は、震源ごとにイベント数に大きな差があるため、不均衡データに対応した手法を適用する必要がある。

本論文では、各決定木を構築する際のサンプル数を、クラスごとに均等にする Balanced Random Forest[4] を適用することで、不均衡データに対応した分析を行う。Balanced Random Forest を構築する各決定木は、ジニ係数を基準に分割を行う。ジニ係数とは不純度さを表す指標であり、0 から 1 の間の値をとる。0 に近いほど純度が高く、データのバラつきが少ない。特徴量の寄与率は、特徴量ごとに、木の分岐におけるジニ係数の減少率を計算することで算出される。最終的な寄与率は、各決定木の寄与率の平均値である。

4 実験結果

本章では、震源分類に有効な特徴量を評価するための実験、及びその結果と考察について述べる。Balanced Random Forest の分類性能と、特徴量の寄与率をもとに、本論文で用いた特徴量における震源間の関係性を分析する。

本論文における特徴分析の概要を以下に示す。

- 月震発生時刻をもとに、特徴量を算出する
- Balanced Random Forest を、全ての震源のペアごとに学習する
- Balanced Random Forest の分類性能と、特徴量の寄与率を算出し、分析を行う

本論文では、one-vs-one 法で、データセットにおける 2 つの震源に関する分類器をすべてのペアに対して構築し、震源ごとの性質や、震源間の関係性を考慮した分析を行う。Random Forest は各分類器ごと、1,000 個の木を構築する。各決定木の構築に用いるサンプルは、各クラスごと 50 ずつ、bootstrap 法を用いて復元抽出する。また、Random Forest を構築する決定木の実装は scikit-learn[14] の Decision Tree を利用した。

また、本論文では、特徴量選択として、以下の 2 つについて分析を行う。

- 抽出した 55 特徴量すべてを学習した分類器を作成する
- 分散拡大係数 (VIF) を用いて、特徴量を削減した後に分類器を構築する

VIF とは、多重共線性の評価に使用される指標のひとつである。本論文では、各特徴量の VIF が 6 未満となるよう、特徴量を削減したサブセットで実験を行った。特徴量は、4.3 節の結果をもとに、寄与率が低い特徴量から順に VIF を計算し、VIF が 6 以上の特徴量を削除した。VIF の計算には statsmodels[15] を用いた。

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

表2 データセットの内訳

| 震源 | A1 | A5 | A6 | A7 | A8 | A9 | A10 | A14 | A18 | A20 | A23 | A25 | A35 | A44 | A204 | A218 |
|------|-----|----|-----|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|------|
| データ数 | 441 | 76 | 178 | 85 | 327 | 145 | 230 | 165 | 214 | 153 | 79 | 72 | 70 | 86 | 85 | 74 |

4.1 データセット

表2は、本論文におけるデータセットを示したものである。本論文では、観測されたイベント数が70以上の16震源の計2,480イベントをデータセットとする。

4.2 評価指標

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

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

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

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

適合率は、分類の正確性を測る指標であり、再現率は、分類の網羅性を測る指標である。適合率と再現率はトレードオフであるため、適合率と再現率の調和平均であるF値は、適合率と再現率のバランスを考慮した評価指標である。本論文における分類器のスコアは、分類器が対象としている2クラスのF値の平均値とする。

4.3 全特徴量を用いた実験結果

4.3.1 分類性能

図1は、震源ごとの分類器のF値の平均である。縦軸、横軸ともに震源であり、各値は分類器のF値のスコアである。図1において、最も高い分類性能は0.96であり、複数の震源のペアで観測された。また、最も分類性能が低いものは、A9とA25の分類器の0.54である。図1は、震源の組み合わせによって分類が可能であるものと分類が難しいものが存在することを示している。また、分類性能が0.9以上の分類器数は20個で全体の約17%、0.8以上0.9未満の分類器は60個で全体の50%である。分類性能が0.6を下回る分類器は1つで、多くの分類器では、高い分類性能が出ており、惑星の位置関係が深発月震の震源分類に有効であることを示した。

4.3.2 特徴量の寄与率

図2は、特徴量別の寄与率の平均値である。寄与率の上位はすべて月を原点としたときの地球の特徴量である。また、月を原点としたときの木星の特徴量の一部が、月を原点としたときの地球の特徴量の次に寄与率が高いことを示している。月を原点としたときと、地球を原点としたときの特徴量を比較すると、月を原点とした特徴量が、地球を原点とした特徴量より寄与率が高い。図2は、月と地球の関係性が最も分類に影響していることを示した。しかし、特徴量間の相関などが含まれている可能性があり、それぞれの特徴量について独立に、より分析する必要がある。そこで、4.4節では、特徴量間の相関を考慮し、VIFを用いた特徴量削減後の結果を述べる。

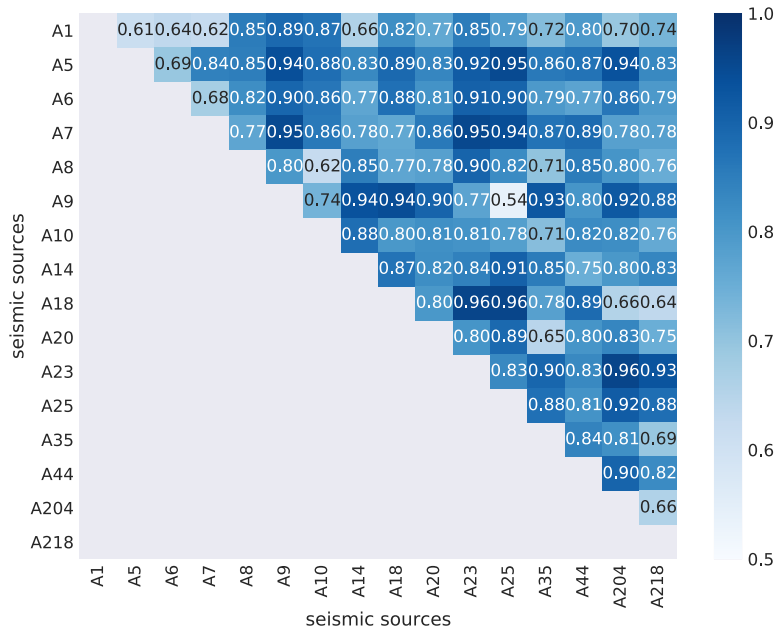


図1 震源別のF値

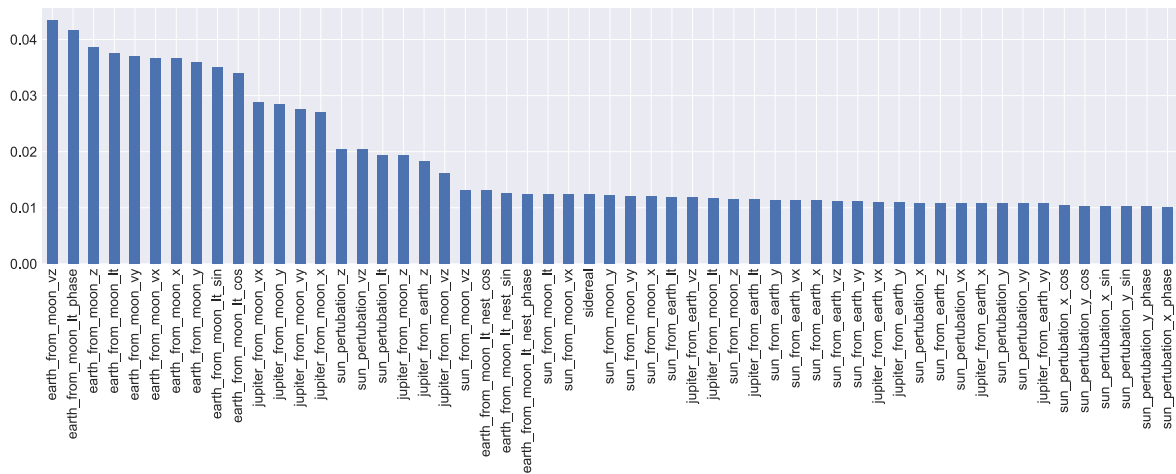


図2 特徴量別の寄与率の平均値

4.4 VIF を用いて特徴量削減を行った実験結果

4.4.1 分類性能

図3は、特徴量を削減した場合の分類器のF値平均である。図1と同様に、縦軸、横軸は震源であり、各値は分類器のF値のスコアである。また、分類性能が0.9以上の分類器数は26個で全体の約22%、0.8以上0.9未満の分類器は54個で全体の45%である。分類性能が0.6を下回る分類器は1つであった。これらは、図1と比較して、分類性能が大きく変わらないことを示している。

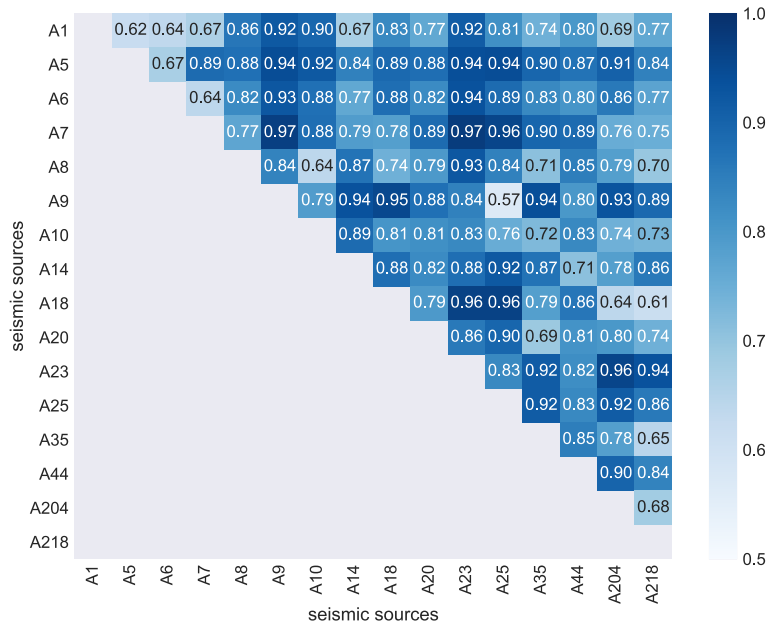


図3 特徴量削減後の震源別のF値の平均値

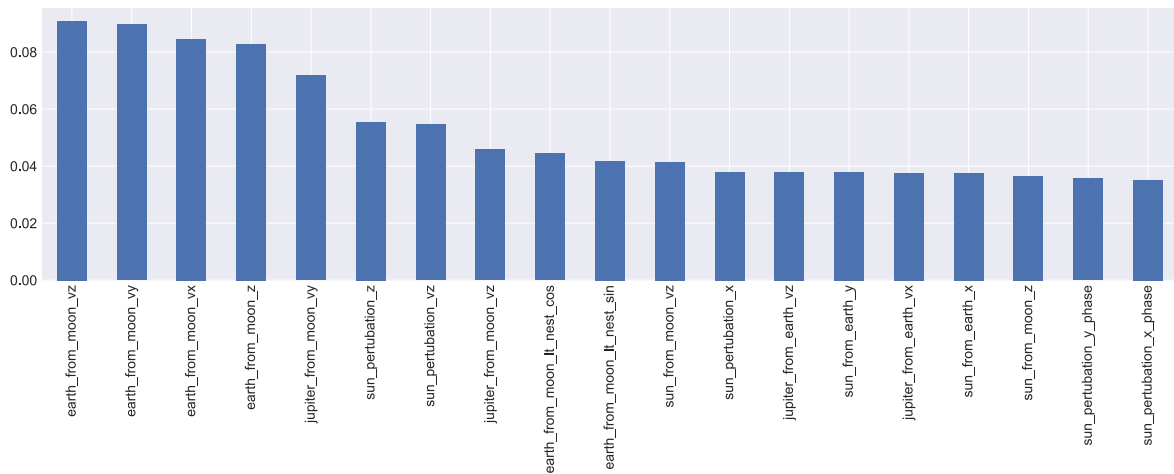


図4 特徴量削減後の特徴量別の寄与率の平均値

4.4.2 特徴量の寄与率

図4は、特徴量削減後の各震源の寄与率の平均値である。特徴量削減後は、月を原点としたときの地球の特徴量は、削減前の上位10特徴量のうち、4特徴量まで削減された。また、月を原点としたときの木星の特徴量は、図2の上位11位から14位の4特徴量のうち、3つが削減された。木星の他のパラメータは、他の特徴量の影響を受けていたと考えられる。本節での、特徴量削減後の特徴量のサブセットは、多重共線性の影響は小さいと考えられる。そのため、月を原点としたときと地球の特徴量と、一部の木星の特徴量が分類に有効である可能性が示された。

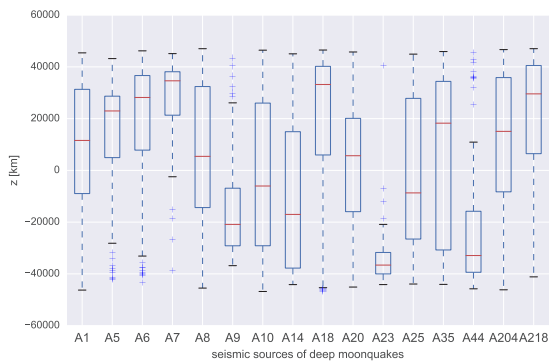


図5 earth_from_moon における z 軸の位置の震源別の箱ひげ図

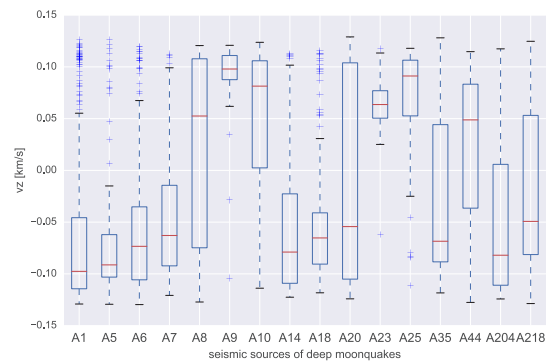


図6 earth_from_moon における z 軸の速度の震源別の箱ひげ図

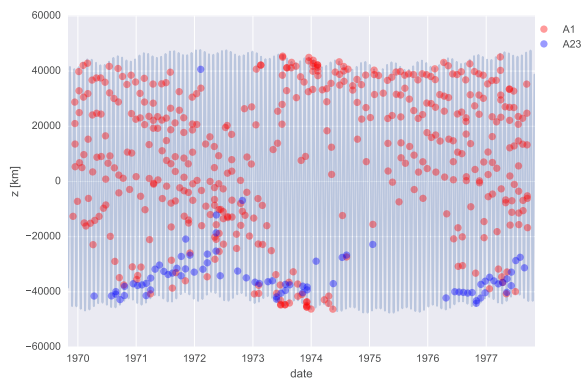


図7 earth_from_moon における z 軸の位置の時系列の変化

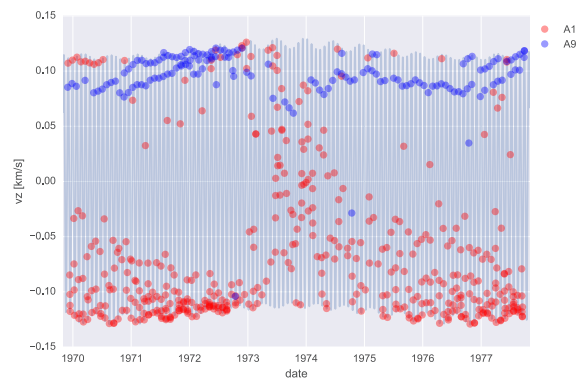


図8 earth_from_moon における z 軸の速度の時系列の変化

ここで、 z 軸における地球の位置の各震源ごとの箱ひげ図と、 z 軸における地球の速度の各震源ごとの箱ひげ図を、それぞれ図5、図6に示す。

図5の箱ひげ図は、A23の発生時刻における地球の z 座標の位置の分布が約-45,000kmから、約-20,000kmであり、A7の発生時刻における地球の z 座標の位置の分布は約0kmから約50,000kmである。A1やA8、A10などの震源の発生時刻における地球の z 座標の位置の分布は約-50,000kmから50,000kmであり、幅広い分布である。同様に、図6の箱ひげ図は、A9の発生時刻における地球の z 座標の速度の分布は約0.06km/sから約0.13km/sである。A5の発生時刻における地球の z 座標の速度の分布は約-0.13km/sから約-0.02km/sである。A8やA20、A35などの震源の発生時刻における地球の z 座標の速度の分布は約-0.13km/sから約0.13km/sであり、幅広い分布である。

図5、図6は、震源ごとに、特徴量の分布が異なることを示している。本論文の分析手法で寄与率が高い特徴量の傾向として、図5におけるA23や、図6におけるA9のように、分布が他の震源に比べて限定的である特徴量が、分類性能が高い震源であり、A1のように特徴量の分布が他の震源と差別化できない震源は、分類性能が低い震源と推察される。

ここで、A1とA23における z 軸における地球の位置の時間変動、A1とA9における z 軸における地球の速度の時間変動を、それぞれ図7、図8に示す。図7は、A23の発生時刻における地球の z 座標の位置が、観測期間を通して変動が小さく、1975年から1976年にかけて、発生回数が他の期間に比べて減少していること

を示している。また、A1は、1973年から1975年にかけて、特徴量の分布が変化している。同様に、図8は、A23の発生時刻における地球のz座標の速度が、観測期間を通して変動が小さく、A1において、1973年から1975年にかけて、特徴量の分布が変化していることを示している。

上記の結果より、深発月震の発生時刻における本論文における特徴量は、震源によって時期的変動があることを示している。A1の分類性能は、図7、図8のような時期変動が捉えられる特徴量を抽出することで、分類性能が向上すると推測される。

4.5 手法、特徴量について

Balanced Random Forest を利用することで、分類性能に加えて、特徴量の寄与率を容易に算出可能であるため、本論文のような特徴量分析に有用であった。しかし、本手法では、分類手法などのパラメータについて、パラメータ毎に震源の分類先の変化などの検討の余地がある。また、より高い分類性能を得るためには、多くの分類手法について検討する必要がある。さらに、波形情報も考慮した手法を適用する必要がある。加えて、本論文で得られた知見は、あくまで相関であり、深発月震の発生メカニズムの因果を直接に推定することは難しい。しかし、本論文の結果は、専門家の新たな分析や知見創出に有用であり、専門家の知識を用いることで、震源と宇宙空間との因果関係や、月震の発生原因の解明につながると考えられる。

5 まとめ

本論文では、Balanced Random Forest を用いて、惑星の時間的、空間的特徴量から、深発月震の震源分類に有効な特徴量の分析を行った。本論文での実験及び考察において得られた知見を以下に示す。

- 従来の波形特徴量を用いず、時間的、空間的特徴量を用いて分類が可能な震源が存在する
- 月を原点としたときの地球の特徴量が、本論文で検証した特徴量では最も分類に有効であった
- 時期変動を持たない特徴量が、本論文では寄与率が高い特徴量として抽出された

本論文で得られた知見は、専門家の新たな分析や知見創出に有用であり、深発月震の発生メカニズム解明に繋がる可能性がある。今後の課題は、以下に示すとおりである。

- 惑星間の orbit parameter 関係に相関や交絡を考慮した特徴量の検証
- 各特徴量の周期性や惑星間の相対的な関係性を考慮した分析
- 新たな特徴量の検証

物理的解釈につながる特徴量、知見創出に展開できる分析が今後の課題である。

参考文献

- [1] Leo Breiman. Random forests. *Machine learning*, Vol. 45, No. 1, pp. 5–32, 2001.
- [2] 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 (1991–2012)*, Vol. 112, No. E9, 2007.
- [3] R. C. Bulow, C. L. Johnson, and PM Shearer. New events discovered in the apollo lunar seismic data. *Journal of Geophysical Research: Planets (1991–2012)*, Vol. 110, No. E10, 2005.
- [4] Chao Chen, Andy Liaw, and Leo Breiman. Using random forest to learn imbalanced data. *University of California, Berkeley*, pp. 1–12, 2004.
- [5] Yasumichi Goto, Ryuhei Yamada, Yukio Yamamoto, Shohei Yokoyama, and Hiroshi Ishikawa. Som-based visualization for classifying large-scale sensing data of moonquakes. In *P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC), 2013 Eighth International Conference on*, pp. 630–634. IEEE, 2013.
- [6] B. Knapmeyer-Endrun and C. Hammer. Identification of new events in apollo 16 lunar seismic data by hidden markov model-based event detection and classification. *Journal of Geophysical Research: Planets*, Vol. 120, No. 10, pp. 1620–1645, 2015.
- [7] Junji Koyama and Yosio Nakamura. Focal mechanism of deep moonquakes. In *Lunar and Planetary Science Conference Proceedings*, Vol. 11, pp. 1855–1865, 1980.
- [8] David R Lammlein, Gary V Latham, James Dorman, Yosio Nakamura, and Maurice Ewing. Lunar seismicity, structure, and tectonics. *Reviews of Geophysics*, Vol. 12, No. 1, pp. 1–21, 1974.
- [9] P. Lognonné, J. Gagnepain-Beyneix, and H. Chenet. A new seismic model of the moon: implications for structure, thermal evolution and formation of the moon. *Earth and Planetary Science Letters*, Vol. 211, No. 1, pp. 27–44, 2003.
- [10] Y. Nakamura. A1 moonquakes-source distribution and mechanism. In *Lunar and Planetary Science Conference Proceedings*, Vol. 9, pp. 3589–3607, 1978.
- [11] Y. Nakamura. New identification of deep moonquakes in the apollo lunar seismic data. *Physics of the Earth and Planetary Interiors*, Vol. 139, No. 3, pp. 197–205, 2003.
- [12] Y. Nakamura, G. V. Latham, and H. J. Dorman. Apollo lunar seismic experiment—final summary. *Journal of Geophysical Research: Solid Earth (1978–2012)*, Vol. 87, No. S01, pp. A117–A123, 1982.
- [13] Y. Nakamura, G. V. Latham, H. J. Dorman, and J. Harris. Passive seismic experiment long-period event catalog. *Galveston Geophysics Laboratory Contribution*, Vol. 491, , 1981.
- [14] F. Pedregosa, et al. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, Vol. 12, pp. 2825–2830, 2011.
- [15] Skipper Seabold and Josef Perktold. Statsmodels: Econometric and statistical modeling with python. In *9th Python in Science Conference*, 2010.
- [16] RC Weber, BG Bills, and CL Johnson. Constraints on deep moonquake focal mechanisms through analyses of tidal stress. *Journal of Geophysical Research: Planets*, Vol. 114, No. E5, 2009.
- [17] Ryuhei Yamada, Yukio Yamamoto, Jun Kuwamura, and Yosio Nakamura. Development of an online retrieval system of apollo lunar seismic data. *Journal of Space Science Informatics Japan*, No. 1, pp. 121–131, 2012.
- [18] 菊池栗, 加藤広大, 山田竜平, 山本幸生, 廣田雅春, 横山昌平, 石川博. 深発月震源分類に適した機械学習法の研究. *宇宙科学情報解析論文誌*, 第 6 号, pp. 51–62, 2017.

「はやぶさ 2」地球スイングバイに係る見かけの等級の変化推定

三浦 昭^{*1}, 山本 幸生^{*1}, 吉川 真^{*1}

Estimation of the Variation of Apparent Magnitude during the Earth Swing-by of "Hayabusa2"

Akira MIURA^{*1}, Yukio Yamamoto^{*1} and Makoto YOSHIKAWA^{*1}

Abstract

On December 3, 2015, an Earth Swing-by operation of the asteroid explorer "Hayabusa2" was performed. Several observatories successfully measured the brightness of "Hayabusa2". The variation of the apparent magnitude of "Hayabusa2" measured during the swing-by is hard to explain using a simple hypothesis based on the distances of the observatories and "Hayabusa2".

In this paper, apparent magnitude of "Hayabusa2" is estimated using modeling and rendering techniques based on computer graphics. The method is found to be useful to roughly estimate the variation of the magnitude, while more consideration is required to perform precise estimation.

Keywords: Hayabusa2, Earth swing-by, visualization, magnitude

概要

2015年12月3日の小惑星探査機「はやぶさ 2」地球スイングバイにおいて、各地点で観測された「はやぶさ 2」の見かけの等級の変化は、各観測地点と「はやぶさ 2」との距離のみの簡単な仮定では説明の難しいものであった。

本稿においては、「はやぶさ 2」の見かけの等級を CG 技術に基づいたモデリングとレンダリングによって推定した。その手法は、光度変化のあらましを推定するには有用である一方で、精密な推定のためには、なお一層の検討が必要であることが明らかとなった。

キーワード: はやぶさ 2, 地球スイングバイ, 可視化, 等級

1. はじめに

1.1. 背景

2015年12月3日に、小惑星探査機「はやぶさ 2」が、小惑星 Ryugu (162173, 1999 JU₃) に向かう途上で、地球スイングバイを実施した。日本上空の

通過は18:00から19:00 [JST] 頃で、19:08 [JST] 頃が最接近の時刻であった。

これに際して、はやぶさ 2 地球スイングバイ観測キャンペーン²が展開され、多くの観測地点が「はやぶさ 2」の観測に成功した。その内、いくつかの観測地点は「はやぶさ 2」の見かけの等級を測定した。その概要を図 1^[1]に示す。以後、本稿における

doi: 10.20637/JAXA-RR-17-009/0006

* 平成30年1月18日受付 (Received January 18, 2018)

^{*1} 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)

² 日本惑星協会, はやぶさ 2 地球スイングバイ観測 Campaign Site, <http://www.planetary.jp/hayabusa2/hayabusa2swingby/> (2017年10月28日現在)

時刻表示は、別段の記載が無い限り、図 1 に準じて、1 日を単位とした世界時 (UT) で表すものとする。

図中の各観測地点及び観測に使用された装置類を以下に示す。

- NHAO: 兵庫県立大学西はりま天文台³
なゆた望遠鏡 / 可視多波長撮像装置 MINT
- KISO: 東京大学木曾観測所⁴
105cm シュミット望遠鏡 / Tomo-e Gozen
- BSGC: 美星スペースガードセンター⁵
1m 望遠鏡 / Volante
- かわさき宙と緑の科学館²
20cm 屈折型望遠鏡 / WAT-902H2 ULTIMATE

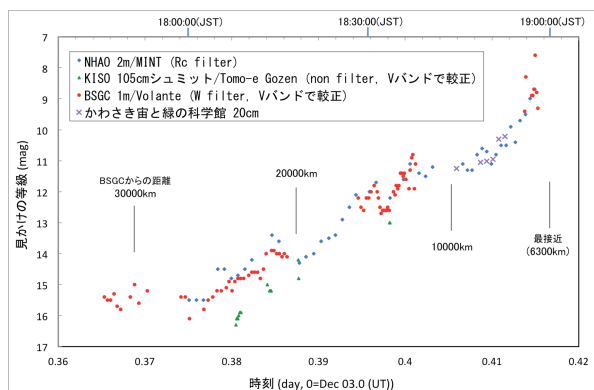


図 1 「はやぶさ 2」地球スイングバイ時の見かけの等級の変化^{[1]6}

観測結果の内、かわさき宙と緑の科学館で得られた見かけの等級については、同館の紀要において、 ± 0.5 等から ± 1.1 等のエラーが付された報告^[2]がなされている。同報告によると、ここで見積もられたエラーは、「測定誤差も含まれるが、はやぶさ 2 が非常に短時間に光度を変化させていた事実もあり、

³ 兵庫県立大学天文科学センター西はりま天文台、「ニュース： なゆた望遠鏡による地球スイングバイ中のはやぶさ 2 の撮影(まとめ)(2015/12/04)」, <http://www.nhao.jp/research/news/news151204.html>, (2017 年 10 月 28 日現在)

⁴ 東京大学木曾観測所、「はやぶさ 2 地球スイングバイ観測プロジェクト」, <http://www.ioa.s.u-tokyo.ac.jp/kisohp/NEWS/hayabusa2/hayabusa2.html>, (2017 年 10 月 28 日現在)

⁵ 日本スペースガード協会 はやぶさ 2 観測速報 (2015 年 12 月 3 日), <http://www.spaceguard.or.jp/RSGC/haya2.html>, (2017 年 10 月 28 日現在)

⁶ はやぶさ 2 プロジェクト、「はやぶさ 2」アウトリーチ活動についての天文学会での発表内容, <http://www.hayabusa2.jaxa.jp/topics/20160314/>, (2017 年 10 月 28 日現在)

この 2 つの影響が値に反映していることにより少々大きな値をとっている」ものである。

図 1 の大凡の傾向としては、「はやぶさ 2」が観測地点に接近するに連れて明るくなる様が見て取れる。しかしながらその変化は、観測地点と「はやぶさ 2」との距離のみを根拠としたのでは説明のつかないものであった。

1.2. 研究の方向性と、本稿のあらまし

筆者らは従来より「はやぶさ」の小惑星イトカワ探査等で CG 技術を用いた可視化^{[3][4]}等、係る探査分野における、リアリティの高い可視化手法についての研究を進めてきた。これはアウトリーチ分野でのリアリティ向上に寄与するのみならず、探査ミッションの分野で実用に耐えうる模擬手法の確立をも目指すものである。

例えば「はやぶさ 2」探査においては、分離型の装置の幾つかが搭載カメラの撮像対象に含まれることが想定されており、事前の運用訓練においても、各機器類の形状モデルを含めて、実際の光学観測を模擬することが求められている。そのためには形状モデルやその反射特性についての検討が課題となる。

このように、CG 技術を用いた模擬手法は、本稿に述べるような、地上観測における光度変化推定にとどまらず、探査機搭載の撮像系に係る計画や訓練等においても有用であると考えられる。

そのような模擬手法の確立を目指すにあたって、本稿においては、形状モデルに主眼を置いて、「はやぶさ 2」の地球スイングバイにおける見かけの等級の変化を 3D CG ソフトウェアにて模擬し、その手法の可能性と課題について検討する。

以下、まず 2 節において CG を用いた等級の計算手法について述べる。3 節では、その手法を単純な形状モデルに適用した場合の計算について述べる。4 節では観測されたような光度変化を生じうる形状について検討し、5 節で簡易の「はやぶさ 2」形状モデルを用いた光度変化の計算について述べる。6 節で、まとめを述べる。

2. 等級の計算手法

本節では、CG を用いた等級の計算手法について述べる。

2.1. 計算手法の概要

本稿における「はやぶさ 2」の見かけの等級の求め方を以下に示す。

- 「はやぶさ 2」地球スイングバイの各時点における「はやぶさ 2」や太陽、観測地点等の位置・姿

勢情報等に基づいて、それらの位置関係を CG ソフトウェアを用いて再現し、観測地点 (CG ソフトウェアにおける視点) から見た「はやぶさ 2」の形状モデルをレンダリングする。

- 各シーンにおける視線方向は、「はやぶさ 2」形状モデルの原点が視野中心となるように計算する。
- レンダリング結果の画像について、全画素の明るさを合計したものを、そのシーンにおける「はやぶさ 2」の明るさとする。
- 計算された明るさに基づいて、等級を計算する。

次節以降に、各要素の概要を示す。

2.2. CG ソフトウェア

CG ソフトウェアとしては POV-Ray⁷ を用いる。POV-Ray はスクリプト (テキスト) で記述されたシーンファイルに基づいて レイトレーシング を実行するソフトウェアである。

2.3. 各種パラメータ

本稿においては、3 次元空間上に観測地点 (CG における視点位置) と「はやぶさ 2」(次節に述べるような、種々の形状モデル) を配置したスクリプトを作成し、それぞれの時刻に観測地点から見た形状モデルをレンダリングする。

観測地点に関する値を表 1 に示す。本稿において、地球は球で近似し、その球上の緯度・経度に基づいて観測地点の座標を定める。本稿においては、係る緯度・経度は美星スペースガードセンター付近を採用する。

表 1 観測地点に関する値

| 項目 | 値 |
|-------|---|
| 観測地点 | 北緯 34° 40′ 19″ 東経 133° 32′ 35″ (美星) |
| 地球の半径 | 6,356.752314 [km] ⁸ |

各時刻における「はやぶさ 2」の位置・姿勢、太陽の位置、地球の自転は SPICE カーネルから求める。「はやぶさ 2」の位置・姿勢は Hayabusa2 Science Data Archives から公開されている SPICE カーネ

ル⁹を用いる。太陽・地球の位置や自転は、NASA の NAIF (Navigation and Ancillary Information Facility)¹⁰ から公開されている SPICE カーネルを用いる。それぞれの SPICE カーネルを表 2 に示す。

表 2 座標計算に用いる SPICE カーネル

| 項目 | SPICE カーネル |
|-------------|--|
| 「はやぶさ 2」の位置 | hyb2_20151123-20151213_000 1m_final_ver1.oem.bsp ⁹ |
| 「はやぶさ 2」の姿勢 | hyb2_aocsc_2015_v01.bc hyb2_hk_2015_v01.bc ⁹ |
| 太陽・地球の位置 | de432s.bsp ¹⁰ |
| 地球の自転 | earth_000101_151229_151007. bpc ¹⁰ |

2.4. 形状モデルの倍率

現実の「はやぶさ 2」のサイズと位置情報を用いると、本稿で用いたソフトウェアとスクリプトの組み合わせでは正常にレンダリングされなかったため、計算される軌道等に対して、レンダリングされる「はやぶさ 2」の縮尺は元の形状モデルの 10,000 倍とすることとした。この場合、最接近時の「はやぶさ 2」の見かけの大きさは、視野角にして 0.7 度程度となる。これは「はやぶさ 2」の形状をレンダリングするにあたって若干の誤差を生じるものであるが、次節以降に述べるように、本稿の手法はそれ以上にモデリングの誤差が見込まれるため、本稿における検証の範囲では 10,000 倍の拡大は支障のないものであると判断した。また本稿で計算対象とした時間帯 (時刻 0.36 から 0.42 の範囲) において、レンダリングされた「はやぶさ 2」の 10,000 倍の形状モデルが地球の形状モデルに隠蔽されないことを、事前に確認した。

視野角は、別途記載する場合を除いて、原則として 0.7 度の設定とする。これは計算対象とする時間帯に渡って「はやぶさ 2」全体が視野内に収まるようにするためである。

⁷ Persistence of Vision Raytracer Pty. Ltd., POV-Ray - The Persistence of Vision Raytracer, <http://www.povray.org/> (2017 年 10 月 28 日現在)

⁸ Wikipedia, 地球, <https://ja.wikipedia.org/wiki/地球> (2017 年 10 月 28 日現在)

⁹ 宇宙航空研究開発機構宇宙科学研究所, Hayabusa2 Science Data Archives, <https://www.darts.isas.jaxa.jp/planet/project/hayabusa2/> (2017 年 10 月 28 日現在)

¹⁰ NASA, SPICE An Observation Geometry System for Space Science Missions, <https://naif.jpl.nasa.gov/naif/> (2017 年 10 月 28 日現在)

2.5. 輝度・反射モデル

レンダリングにあたって、背景は黒色とする。

反射モデルは、「拡散反射」もしくは「鏡面反射」を用いる。また、反射モデルでの表現が困難と思われる場合や、比較の目的等では、陰影やハイライト等を伴わない単色のレンダリングを併用する。

本稿において、各反射モデルの実装方法は次のように定めた。これらの反射モデルや各属性値は経験的に定めたものであり、本稿で示すような初歩的な検討を対象としたものである。将来、厳密な等級計算をする段階に至った場合は、係る属性値のみでなく、各部品の素材に対していかなる反射モデルを採用するかを含めた個別検討が必要になると考えられる。

● 拡散反射

拡散反射の計算にあたっては、POV-Rayにおけるdiffuse属性を用いる。diffuse属性は、拡散反射モデルに基づいて明るさを計算するための属性である。diffuse属性の値は、計算結果の画像の各画素が飽和しない程度の値とする。

● 鏡面反射

鏡面反射で近似できるオブジェクトに光源が反射される様（ハイライト）を計算するため、POV-Rayにおけるspecular属性とroughness属性を用いる。この内、specular属性はハイライトの明るさを計算するための属性である。roughness属性は、その「鏡面」の粗さを指定する属性である。本稿においては、鏡面反射は原則として光源（太陽）の映り込みのみ考慮し、他のオブジェクトが映り込む様は計算対象外とする。

Specular属性の値は、計算結果の画像の各画素が飽和しない程度の値とし、roughness属性の値は、原則として0.003とする。

● 単色

陰影を考慮しない、もしくは考慮することが困難と思われるオブジェクトについては、単色で表現する。単色の計算にあたっては、POV-Rayにおけるemission属性を用いる。emission属性は、発光による明るさを計算するための属性であり、光源の影響を受けない。

2.6. 等級の計算

以上の条件でレンダリングしたCG画像について、画像内の全画素の明るさを合計した値をその画像の明るさとして等級に換算する。等級に換算するにあたっては、天文の等級表現と同様に、等級差が5のときに明るさが100倍となるように対数計算する。

等級換算の根拠となる明るさがCG画像由来であるため、結果として得られる等級の絶対値は実際に観測された見かけの等級と直接比較できるものではない。また各部品の反射特性を厳密にモデリングしたものではないため、本稿においては異なる部品同士の等級を絶対値で比較することも控えることとする。

以後、等級に関する図には、傾向を比較するために、同時刻に実際に観測された見かけの等級として、図1に示した観測結果を重ねて表示するが、これらは直接明るさを比較するためのものではない。

3. 単純な形状モデル

本節では、単純な形状モデルを用いた計算について述べる。

3.1. 形状モデル

等級の算出にあたっては、2節の手法に以下の形状モデルを適用する。

● 立方体（鏡面反射）

鏡面反射の立方体で近似したモデルである。立方体の各面は、直方体で近似される「はやぶさ2」筐体の各面に並行とする。

● 立方体（拡散反射）

拡散反射の立方体で近似したモデルである。立方体の各面は、直方体で近似される「はやぶさ2」筐体の各面に並行とする。

● 球体（単色）

単色の球体（レンダリング結果としては、単色の円）で近似したモデルである。

● 球体（拡散反射）

拡散反射の球体で近似したモデルである。

3.2. 計算結果

図2に、前述の単純な形状モデルを用いて等級の変化を計算した結果を示す。図示した等級は、実際に観測された等級と傾向を比較しやすいようにオフセットを加えてある。

以下、幾つかの時刻に分けて考察する。

実際に観測された等級の傾向としては、時刻0.375付近までは殆ど増光しない、もしくは観測誤差が小さいのであれば、減光傾向にあると読み取れる。この傾向を説明できる可能性があるのは球体（単色）もしくは球体（拡散反射）であるが、いずれも「はやぶさ2」が接近するにつれ増光する傾向にあり、仮に減光があったとしたら、それがどのような原因によるものか検証する必要があると考えられる。

時刻 0.375 付近から時刻 0.4 付近までは不規則な増光と減光を繰り返しており、そのゆらぎが観測誤差によるものなのか、探査機の形状に由来するものなのかは、観測誤差・形状モデル双方からの検討が必要であると考えられる。おおまかな傾斜は立方体（拡散反射）の傾斜と似ているが、本稿の対象とする時間帯においては、「はやぶさ 2」の+Z 面は太陽に対してほぼ垂直となっていたため、+Z 面が観測地点を向いていなかった、時刻 0.392 付近までは立方体（拡散反射）をほとんど視認できない状態となり、立方体（拡散反射）のみで現実の光度変化を説明するのは困難である。

実際に観測された等級では、時刻 0.405 から 0.412 付近の間は、再び立方体（拡散反射）の傾斜と同様の増光が見て取れる。0.413 以降の急激な増光は、立方体（鏡面反射）の類似で説明できる可能性がある。

仮に、これらの単純な形状モデルを組み合わせる実際に観測された光度変化を説明するとしたら、図 3 のような加算の例が考えられるが、細部については、さらに詳細な形状モデルの検討が必要であると考えられる。

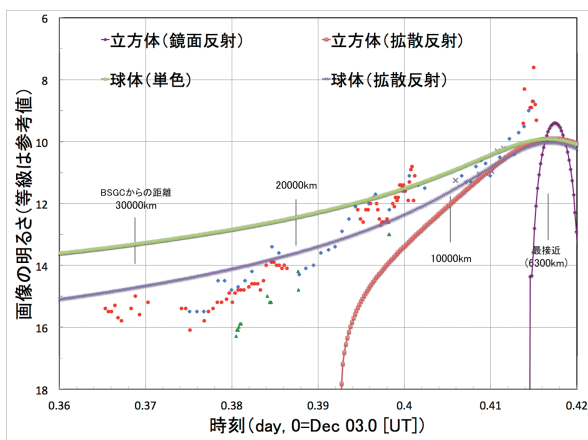


図 2 単純な形状モデル（単体）

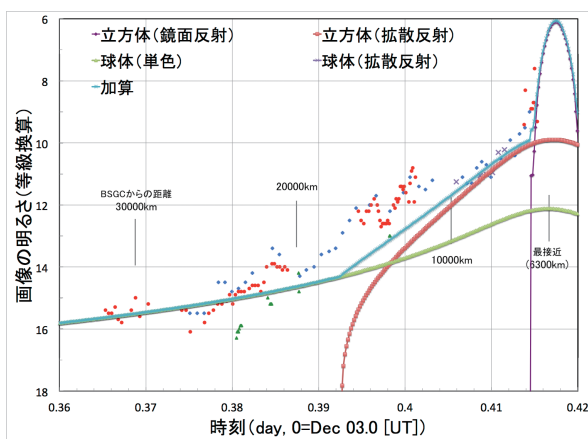


図 3 単純な形状モデル（加算）

4. 特定時間帯の増光

本節では、3 節に挙げたような単純な形状モデルの組み合わせでは説明困難と思われる、時刻 0.39 付近から 0.4 付近の光度変化について、別の形状モデルを用いた検討を行う。

4.1. 円錐を用いた形状モデル

「はやぶさ 2」の筐体、特に+Z 側の面には、様々な部品が取り付けられており、さらに各々の部品は多層断熱材 (Multi-Layer Insulation: MLI) 等の柔軟な素材で覆われて不定形の外觀となっている。このような形状を CG モデルで再現するのは容易ではないため、本稿においては、近似モデルを用いて、時刻 0.39 付近から 0.4 付近にかけて増光するようなパターンがありうるかの検討を行う。

以下に近似モデルの作成方法を示す。数値は「はやぶさ 2」形状モデルにおける 1m を単位とする。

- 構成要素は底面が-Z 方向を向いた円錐とし、その高さを 0.01 とする。
- 円錐の底面(円)の半径を r とすると、間隔 $\sqrt{2}r$ の正方格子の、各格子点上に円錐を配置する。
- 原点から半径 1 以下の範囲にある円錐は鏡面反射とし、原点から半径 1 を超えて 2 以下の範囲にある円錐は黒色（反射しない）とする。
- 他の格子点には円錐を配置しない。

以上の手順により、概ね半径 1、高さ 0.01 の円状で、+Z 方向（太陽方向）に対して一定角度の面を持つ鏡面反射オブジェクトを作成できる。

4.2. 計算結果

図 4 に、円錐の頂角を 170 度から 80 度まで 10 度刻みで変化させた場合の、係るオブジェクトの光度変化を示す。比較のために、平面（半径 1 の円）で構成された鏡面反射オブジェクトも示す。またオブジェクトの見え方の例を図 5 に示す。係る視野角は 0.7/3 度とした。

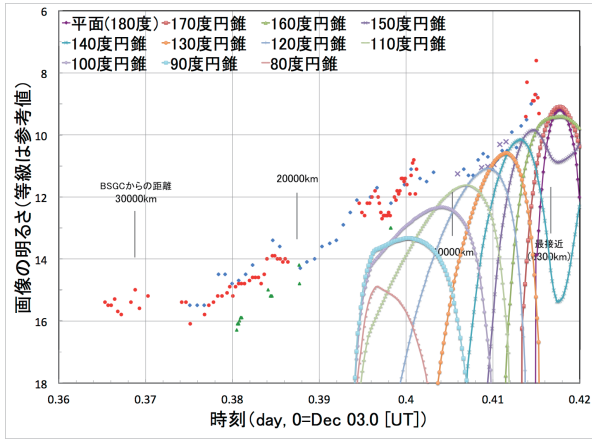


図 4 円錐を用いた形状モデル

この結果より、例えば頂角 90 度の円錐を構成要素とするオブジェクトは、0.395 から 0.45 付近で増光することが見て取れる。また他の、頂角を変えたオブジェクトも、特定の時間帯に増光する様子が見て取れる。

本節の結果から即断できるものではないが、「はやぶさ 2」の +Z 面に取り付けられた様々な構造物は、その取付角度（もしくは貼り付けられたシートがなす角度）に応じて、特定の時間帯に増光する要因となっている可能性があると考えられる。



頂角 90 度 時刻 0.395 頂角 90 度 時刻 0.400 頂角 160 度 時刻 0.420

図 5 オブジェクトの見え方の例

5. 「はやぶさ 2」形状モデル

本節では、「はやぶさ 2」の CG モデルを用いた光度計算について述べる。

5.1. 形状モデルの概要

図 6 に、使用した形状の概略を示す。図中の矢印は、「はやぶさ 2」形状モデルの座標系における +X 方向、+Y 方向、+Z 方向を示している。なお図中の色は各部品の形状や位置関係を視認しやすくするために着色したものであり、実際の計算に用いた反射モデルを反映したものではない。実際の計算に用いた反射モデルの使い分けは 5.2 に示す。

対象時刻における姿勢に基づいて計算した場合に太陽光が届かないと思われる部品は、原則として割愛している。現実には多重反射により太陽光が届く

範囲があると考えられるが、本稿の反射モデルでは多重反射を考慮していないため割愛した部品も存在する。形状モデルの源泉となったデータは CAD データであるため、筐体の多くを覆っている MLI の形状は再現されていない。+Z 面に置かれた種々の部品の多くは割愛し、4 節で述べたような円錐での近似を用いた。図 6 では円錐の配置が不明瞭であるため、図 7 に詳細を示す。

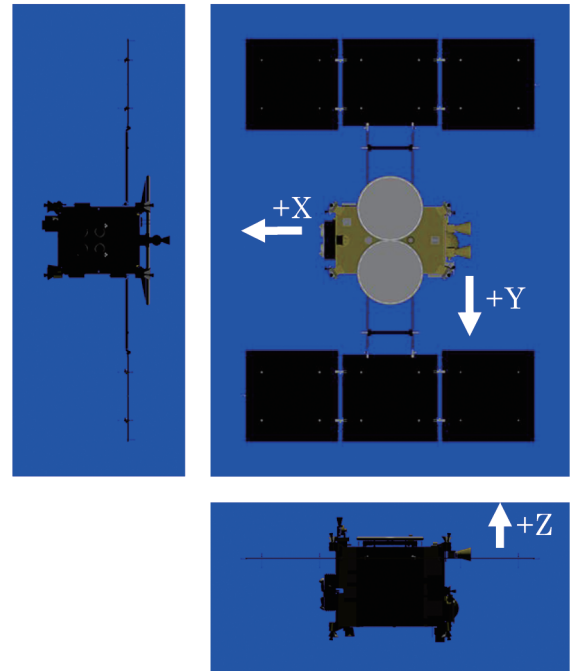


図 6 「はやぶさ 2」形状モデル

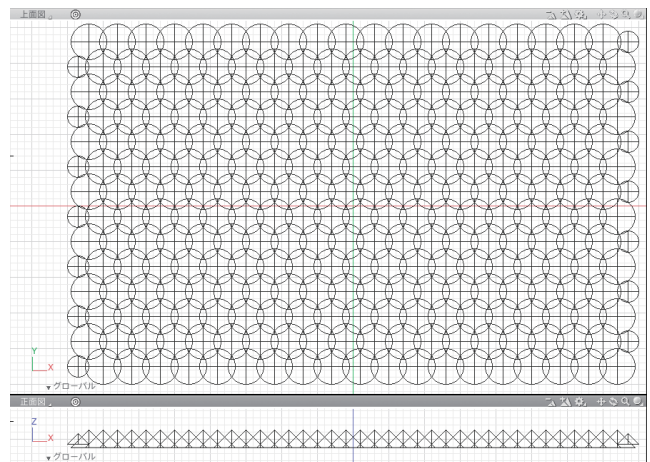


図 7 「はやぶさ 2」形状モデル+Z 面の円錐配置

5.2. 形状の構成要素(部品)と反射モデル

「はやぶさ 2」の形状は、以下の部品群に分けて光度計算を行う。部品名称の後に、各部品の光度計算に用いる反射モデルを記す。

- 鏡面反射 (+Z): 鏡面反射
+Z 側に位置する鏡面反射的要素の内、平面で近似

しうるものを対象とする。太陽電池パドル、高利得アンテナを覆うゲルマニウム蒸着カプトンシート等が想定される。高利得アンテナのシートは、地上で撮影された写真においては、シートの自重により下の構造が浮き出ているが、重力の影響が異なる宇宙空間での形状を表しているとは限らないため、本稿においては特別な凹凸は仮定しないこととする。

● 銀色（平面）：鏡面反射

他で特筆されていない、平面で構成される銀色の素材（露出部分）を対象とする。

● MLI（平面）：鏡面反射

MLI で覆われた部分の内、比較的平面で近似できると思われる範囲を対象とする。

● IES：鏡面反射・拡散反射（混在）

イオンエンジン（Ion Engine System: IES）関係の部品を対象とする。

● MLI (+Z 側)：鏡面反射

筐体の+Z 側を覆う MLI を角度 45 度の円錐の集合で模擬した場合の計算結果を示す。その形状を図 7 に示す。これは実際の MLI の形状を再現したものではないため、相応の誤差が見込まれる。

● HGA（縁）：鏡面反射

高利得アンテナ（High-Gain Antenna: HGA）の縁を覆うシートを模擬している。図 6 に示すように、不規則な形状をトーラス状のモデルで模擬しているため、MLI 同様に、相応の誤差が見込まれる。

● MLI (MGA)：鏡面反射

中利得アンテナ（Medium Gain Antenna: MGA）の MLI 部分の形状を対象とする。

● MGA Horn：単色

中利得アンテナの矩形ホーンを対象とし、最大限見えた場合の計算結果を示す。実際には、その多くが暗色のシートで覆われているため、レンダリングされる面積ほどの影響は無いと考えられる。

● SAP（拡散反射）：拡散反射

太陽電池パドル（Solar Array Panel: SAP）の拡散反射成分を対象とする。太陽電池パドルも暗色であるため、レンダリングされる面積ほどの影響は無いと考えられる。

● STT：鏡面反射

スタートラッカ（Star Tracker: STT）を対象とする。

● HGA (-Z 側)：単色

高利得アンテナの裏側（-Z 側）の構造物を対象とする。この部分は、本来であれば、筐体の+Z 側を

覆う MLI や高利得アンテナの裏側等で太陽光が複数回反射したものを計算する必要がある。しかしながら、いずれも複雑な形状となっており、詳細な計算は困難である。そのため本稿においては、高利得アンテナの背面が見える領域を面積計算し、明るさの代替とする。

● その他：鏡面反射・拡散反射（混在）

上記以外の諸々の部品を包含する。

5.3. レンダリング事例

図 8 に、時刻 0.4 付近におけるレンダリングの例を示す。

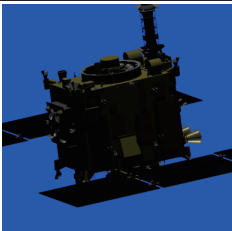
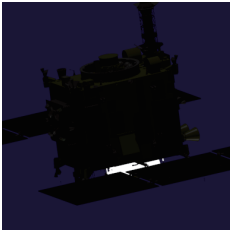
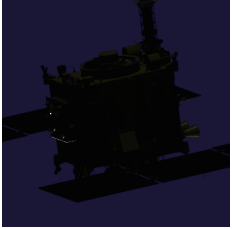
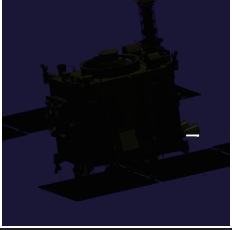
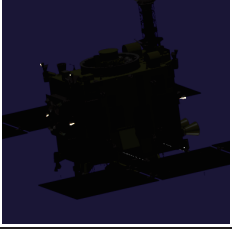
| レンダリング例 | 該当する位置 |
|---------------|---|
| 参考形状 |  |
| 高利得アンテナ (HGA) |  |
| イオンエンジン (IES) |  |
| スタートラッカ (STT) |  |
| その他 |  |

図 8 レンダリングの例（時刻 0.4 付近）

参考形状は、当該時刻に美星スペースガードセンターから観測したと仮定した場合の「はやぶさ 2」の概要を示している。その形状モデルやレンダリング

結果は図中で形状を確認するために調整したものであり、等級の計算に用いられたものとは異なる。それより下は、各部品毎のレンダリング例と、「はやぶさ2」の参考形状における該当箇所を示す。レンダリング例の視野範囲や明るさは、例示として印刷用に調整している。

5.4. 計算結果(部品毎)

図9, 図10, 図11に、各レンダリング結果から計算した等級を示す。

● 鏡面反射 (+Z), 銀色 (平面), MLI (平面), IES

図9に、鏡面反射 (+Z), 銀色 (平面), MLI (平面), IES の計算結果を示す。

鏡面反射 (+Z), 銀色 (平面), MLI (平面) の3者については、再接近付近で急激に明るくなる様子が見て取れる。実際に観測された見かけの等級も、再接近付近で急激に増光していると考えられ、上記のような部品が寄与していることが示唆される。

MLI (平面) で時刻0.41付近から増光しているのは、-X+Z側の筐体に取り付けられたスラスターの台座からの反射が計算されたものである。

IESについては、0.396から0.4付近でイオンエンジン関係の部品が太陽電池パドルに隠されたことにより、減光が計算されている。実際に観測された見かけの等級にも、同じタイミングで同じような傾向の減光が見て取れるが、イオンエンジン関係の部品は、計算された画像中では大きな面積を占めておらず、また暗色の部品も少なからぬ割合で存在するため、この結果のみで両者の因果関係を見極めることは困難である。

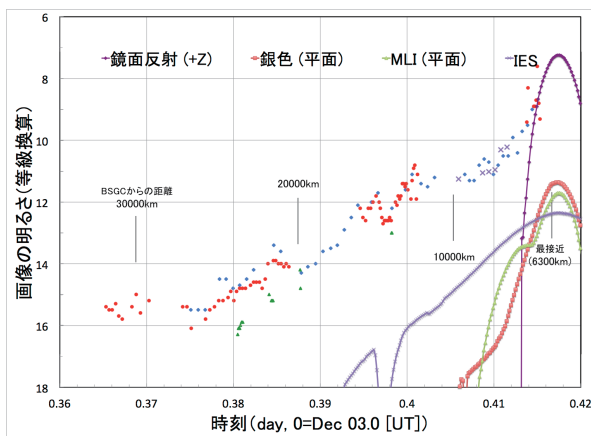


図9 鏡面反射 (+Z), 銀色 (平面), MLI (平面), IES

● MLI (+Z側), HGA (縁), MLI (MGA), MGA Horn

図10に、MLI (+Z側), HGA (縁), MLI (MGA), MGA Horn の計算結果を示す。

MLI (+Z側) の計算結果は、時刻0.39から0.4付近までの増光を説明できる可能性がある。前述の

ように、この部品は円錐による近似で形状モデルが構成されているため、具体的な明るさの変化を論じるためには、形状モデルの詳細な検討が必要となる。

HGA (縁), MLI (MGA), MGA (Horn) は、いずれも類似の傾向にあり、0.38付近で、それまで太陽電池パドルに遮られていたものが現われている。実際に観測された明るさが0.38付近で増光していることを説明できる可能性がある。

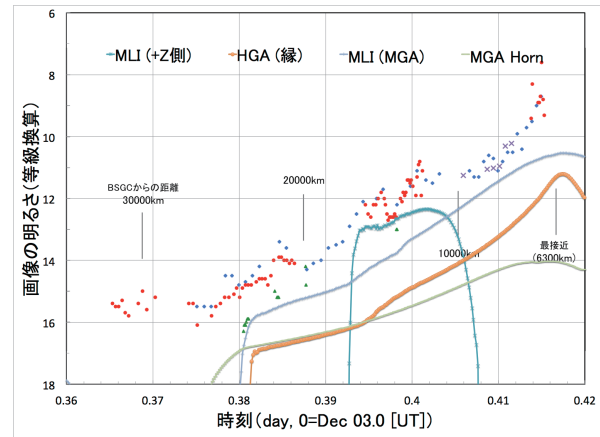


図10 MLI (+Z側), HGA (縁), MLI (MGA), MGA Horn

● SAP (拡散反射), STT, HGA (-Z側), その他

図11に、SAP (拡散反射), STT, HGA (-Z側), その他の計算結果を示す。

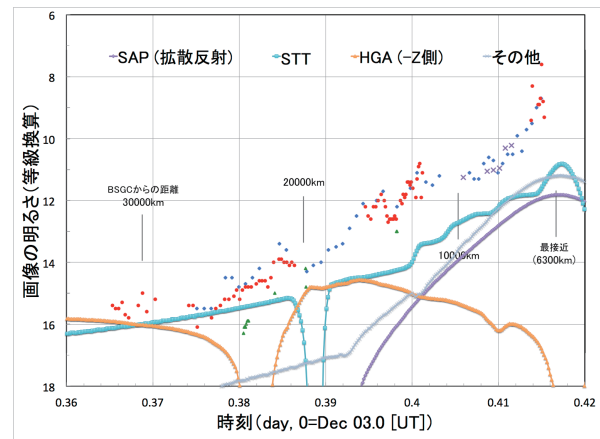


図11 SAP (拡散反射), STT, HGA (-Z側), その他

SAP (拡散反射) は、時刻0.41付近の明るさの傾斜を説明できる可能性があるが、太陽電池パドル自身は暗色であり、また他にも類似の傾向を示す部品があるため、寄与の程度は不明である。

STTは時刻0.39付近で太陽電池パドルに隠されることにより、急激な減光が計算されている。同時刻帯に実際の観測でも減光しており、その現象を説明できる可能性がある。また時刻0.4付近から階段状に増光しており、これは川崎市青少年科学館 (か

わさき宙と緑の科学館)における観測で、「はやぶさ2」が短時間に光度を変化させていた[2]との記述や同科学館で計測された等級の変化を説明できる可能性もある。しかしながら本稿のモデルで簡略化した、MLI に代表される複雑な鏡面反射が影響している可能性もあり、STT が光度変化の主因であると断じることにはできない。

HGA (-Z 側) の見え方の変化を図 12 に示す。図中の黄色で示された領域が、HGA (-Z 側) の見える領域である。この領域は、時刻 0.38 付近までの接近に伴い徐々に太陽電池パドルに隠蔽され、時刻 0.38 以降、暫くの間は殆ど見えなくなる。その影響が大きかったと仮定すると、この時間帯には接近に伴い「はやぶさ2」が減光した可能性が示唆される。これは、実際に観測された見かけの等級が、係る時間帯において増光しない、もしくは減光の傾向にあったことと矛盾しない。

その他については、0.39 を過ぎた辺りから増光している。これは、係る時刻から、太陽光を受けていた+Z 面が見え始めたことにより、関連する部品が増光したためと考えられる。

時刻 HGA (-Z 側) の見え方 (黄色部分)

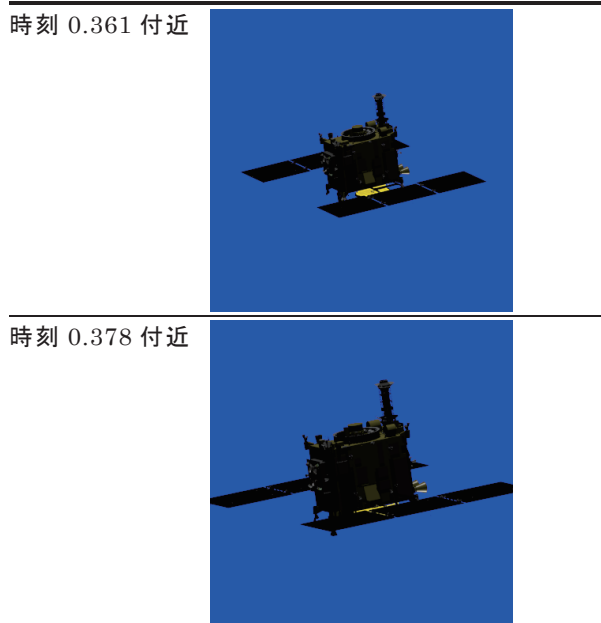


図 12 HGA (-Z 側) の見え方 (黄色部分)

5.5. 計算結果 (全体)

5.4 に述べた要素を加算した一例を図 13 に示す。前述のように、各部品が「はやぶさ2」全体の明るさの変化に対してどのように寄与したかは、この比較だけでは明らかではないので、本稿の範囲では、あくまでも例示となる。

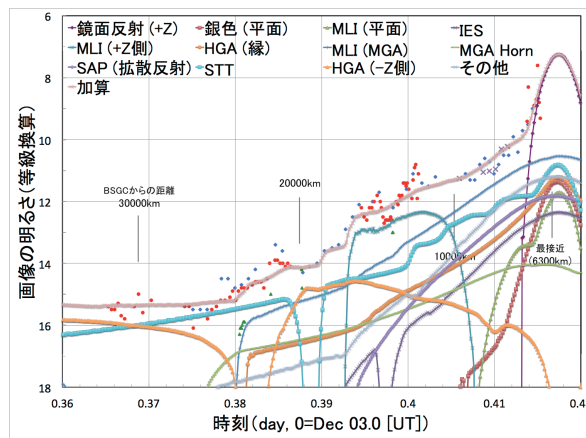


図 13 各部品の明るさの加算例

その結果として、本稿に述べた各部品によって大まかな変化が説明できると思われる時間帯がある一方で、時刻 0.385 前後の増光や時刻 0.395 から 0.4 前後の不規則な変化等は、観測誤差の詳細を吟味しつつ、その妥当性を検証する必要があると思われる。

6. おわりに

「はやぶさ2」の地球スイングバイにあたって、CG モデルを用いた、見かけの等級の変化推定を行った。その結果、「はやぶさ2」を構成する各部品の明るさの変化の計算結果から、実際に観測された見かけの等級の変化を大まかには説明しうることが示唆された。

その一方で、詳細な変化を説明するためには幾つかの課題が残された。それらの課題を以下に示す。

本稿で対象とした「はやぶさ2」地球スイングバイは、観測地点に対する「はやぶさ2」の相対的な姿勢は急激に変化していないにも関わらず、実際に観測された見かけの等級は短時間に変化していたことが見て取れる。また「非常に短時間に光度を変化させていた」[2]ことも報告されており、これを説明するためには、本稿で検証したような各部品のモデリングに加えて、複雑な鏡面反射を如何に考慮するかが課題となると考えられる。そのためには、詳細な部品配置の再現、MLI に代表される柔軟な構造物のモデリング、多重反射の考慮等、解決すべき要因が複数存在している。

また明るさの変化が類似している部品も存在しているため、全体の明るさに対する各部品の寄与の程度を示すためには、汎用の CG ソフトの反射モデルに依らず、個々の部品の表面材質に応じた反射モデルを構築する必要があると考えられる。

これらの課題を考慮しつつ、各観測地点における観測条件や、見かけの等級を計算した際の誤差の評

価等，総合的に検討することが，さらに詳細な推定につながると考えられる。

謝辞

本研究に際して、「はやぶさ2」の見かけの等級観測にご尽力いただいた，兵庫県立大学西はりま天文台，東京大学木曾観測所，美星スペースガードセンターおよび川崎市青少年科学館（かわさき宙と緑の科学館）の各位に感謝致します。

参考文献

1. 「はやぶさ2」地球スイングバイ観測キャンペーンデータの解析. 吉川真, ほか. 2016年, 日本天文学会 2016年春季年会, ページ: Y01a.
2. 川崎で捉えた小惑星探査機「はやぶさ2」のスイングバイの観測. 佐藤幹哉, 弘田澄人, 成瀬裕子. 2017年, 川崎市青少年科学館紀要 第27号.
3. 「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. 2014年, 宇宙科学情報解析論文誌 第3号, ページ: 7-15.
4. はやぶさの軌跡の可視化: タッチダウン時の位置推定. 三浦昭, 山本幸生, 吉川真. 4, 2015年, 宇宙科学情報解析論文誌 第4号, ページ: 173-183.

宇宙航空研究開発機構研究開発報告 JAXA-RR-17-009
JAXA Research and Development Report

宇宙科学情報解析論文誌 第七号

発行 国立研究開発法人 宇宙航空研究開発機構 (JAXA)
〒182-8522 東京都調布市深大寺東町7-44-1
URL: <http://www.jaxa.jp/>

発行日 平成30年3月9日
電子出版制作 松枝印刷株式会社

©2018 JAXA

※本書の一部または全部を無断複写・転載・電子媒体等に加工することを禁じます。
Unauthorized copying, replication and storage digital media of the contents of this publication, text and images are strictly prohibited. All Rights Reserved.

