

ERGサイエンスセンターでのデータ アーカイブ・公開およびデータ引用に 関する取り組み

堀 智昭、三好由純、寺本万里子、津川靖基、T.-F. Chang、瀬川知紀、梅村宜生、小路真史、栗田 怜 (名古屋大ISEE)、松田昇也 (宇宙研)、桂華邦裕 (東大)、宮下幸長 (韓国天文研究院)、関華奈子 (東大)、田中良昌 (ROIS-DS-PEDSC/極地研)、西谷 望 (名古屋大ISEE)、浅村和史、高島 健、篠原 育 (宇宙研)



話の内容

- ▶ ERGプロジェクト、あらせ(ERG)衛星について
- ▶ ERGサイエンスセンターの役割
 - ▶ 科学データアーカイブ
 - ▶ 統合解析ツール
 - ▶ ERGでのデータ引用practice
- ▶ コメント・雑感 (まとめに代えて)

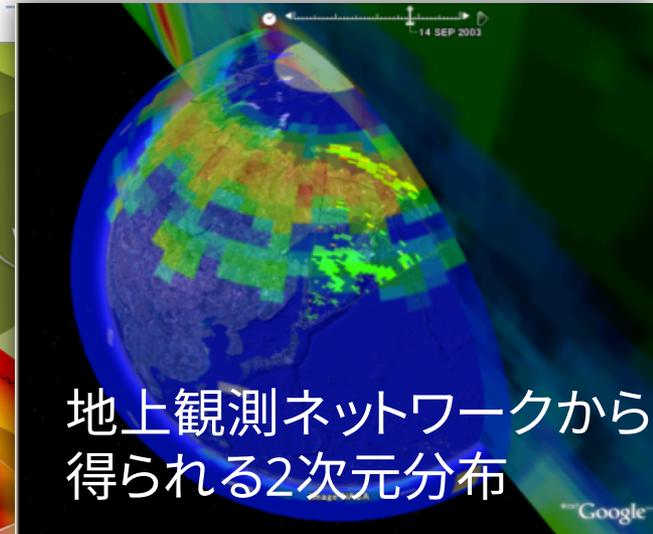
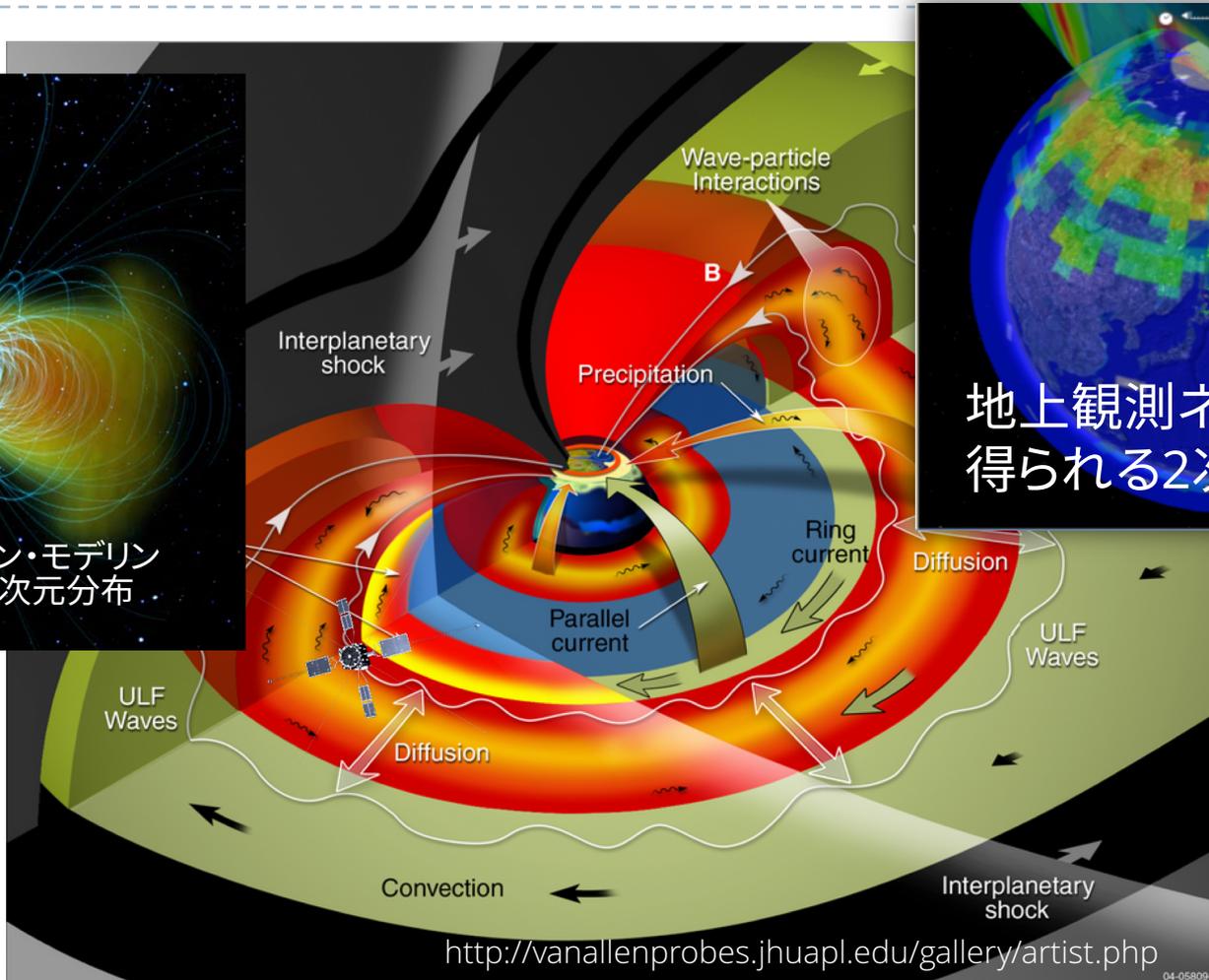
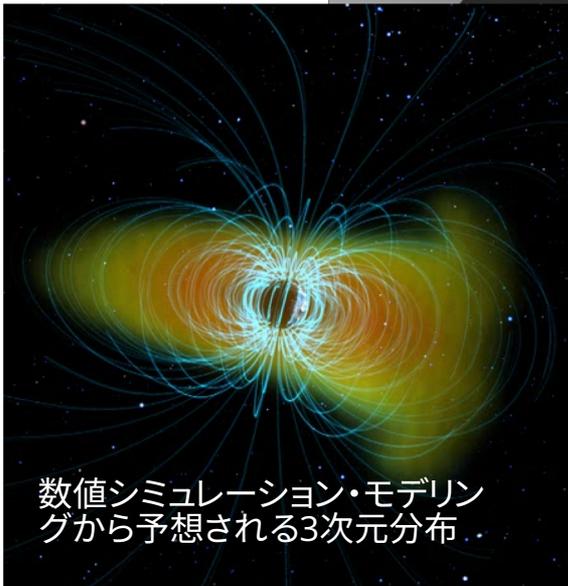
Exploration of energization and Radiation in Geospace ジオスペース探査(ERG)プロジェクト



▶ ERG projectのムービー

[Miyoshi+, EPS, 2018a]

ジオスペース研究と総合データ解析



- ▶ 衛星データ、地上観測データ、シミュレーション・モデリングデータを組み合わせた「統合解析」が必須

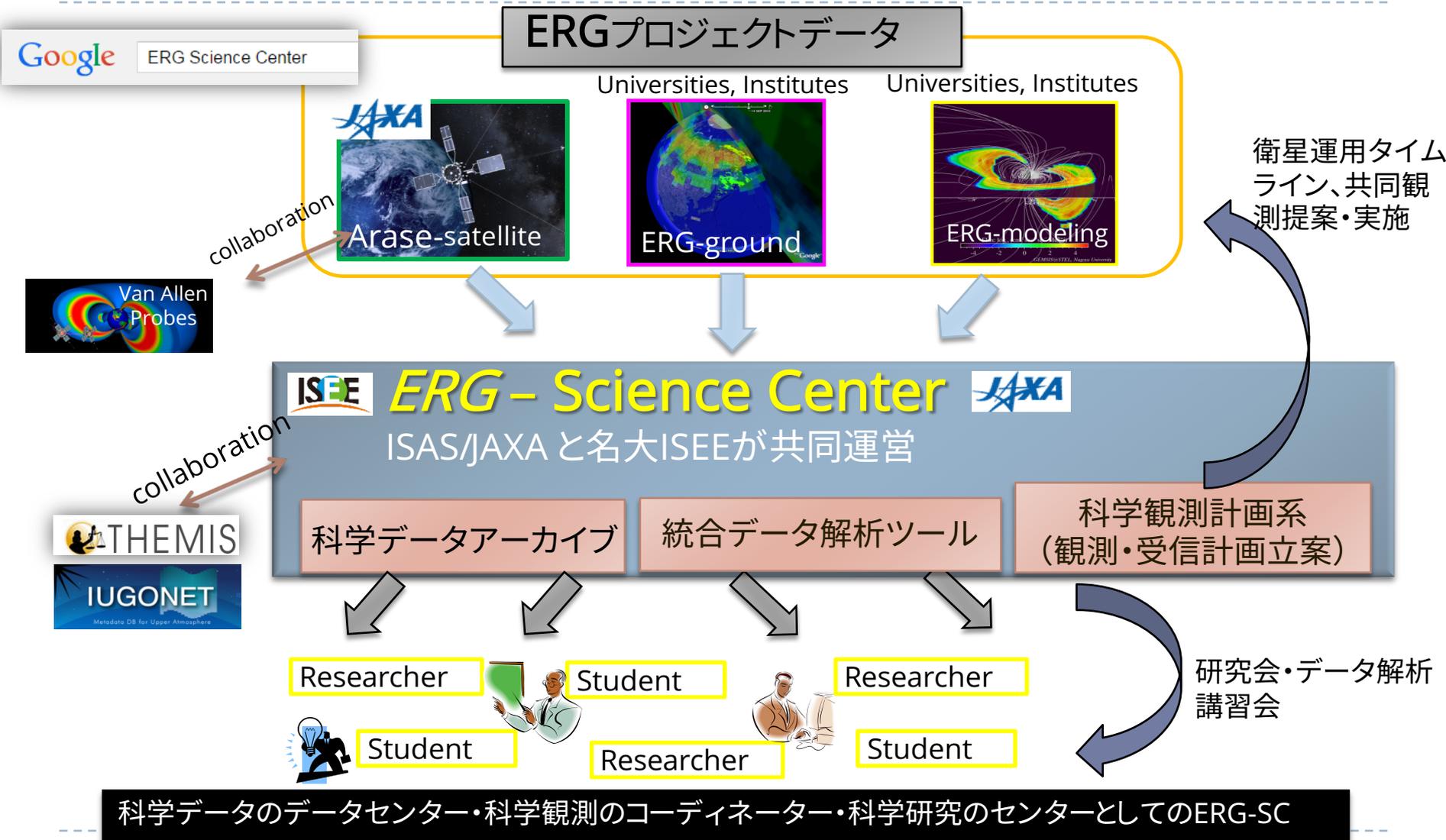
ERGプロジェクト =

Arase衛星 + 連携地上観測 + シミュレーション・モデリング





宇宙科学連携拠点ERGサイエンスセンター





文字で書くと...

ERGサイエンスセンターの役割

- ▶ 衛星及び地上・シミュレーション/モデリングデータの設計、アーカイブ、公開
- ▶ 統合データ解析ツールの開発・公開や講習会の開催
- ▶ あらせ衛星観測計画の立案および必要なツールの開発、関連プロジェクトとの共同観測のコーディネート

これらを通して、「**ERGプロジェクト**」の**科学成果創出の基盤**を担っている。



ERGプロジェクトの科学成果

Geospace Exploration by the ERG mission

The ERG project (Exploration of energization and Radiation in Geospace) is a geospace exploration project to elucidate acceleration, transportation and loss of energetic electrons in the radiation belts and dynamics of geospace storms. The projects consist of JAXA's satellite mission, ground-based observations and simulation and modeling studies and provide a comprehensive dynamical evolutions of the inner magnetosphere. The ERG project consists of a fleet of geospace satellites as well as ground-based network observation observations in geospace. The ERG satellite will be launched in FY 2016 and data will arrive after the launch.

This special issue focuses on all science instruments of the ERG satellite mission. This issue includes the papers on the satellite systems and also focuses on ground-based observations, simulation studies, and science center activities of the ERG mission.

Edited by: Tsugunobu Nagai, Barry Mauk, Ondrej Santolik, Takashi Kubota

Full Paper

The extremely high-energy electron experiment (XEP) onboard the Arase (ERG) satellite

The extremely high-energy electron experiment (XEP) onboard the Arase (ERG) satellite is designed to study energetic electrons in the Earth's radiation belts. The XEP was developed by taking advantage of the high-energy electron experiment (HEE) onboard the Arase (ERG) satellite.

Nana Higashio, Takeshi Takashima, Iku Shinohara and Haruhisa Matsumoto

Earth, Planets and Space 2018 70:134

Published on: 16 August 2018

[Full Text](#) [PDF](#)



The project and instrument papers published in Earth, Planets and Space (21 papers)

GRL special issue for the first scientific results of Arase (still many papers under rev.)

Geophysical Research Letters

AN AGU JOURNAL

Geophysical Research Letters publishes high-impact, innovative, and timely communications-length articles on major advances spanning all of the major geoscience disciplines. Papers should have broad and immediate implications meriting rapid decisions and high visibility.

HOME

BROWSE

HIGHLIGHTS

COLLECTIONS

ABOUT

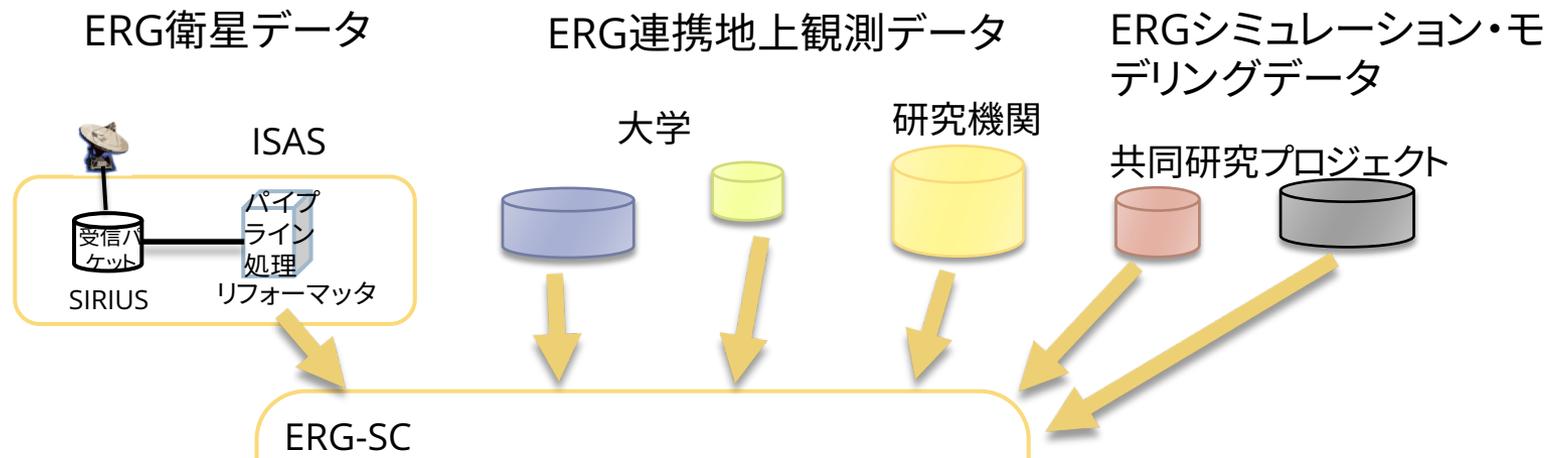
Initial results of the ERG (Arase) project and multi-point observations in geospace

Space Sciences | First published: Dec 18, 2017 | Last updated: 28 August 2018

The Arase (ERG) mission has been exploring the Earth's inner magnetosphere, in particular, the Van Allen radiation belts with six sets of onboard particle analyzers and two sets of onboard electromagnetic field measurements since its launch on December 20, 2016. The major purpose of the Arase mission is to study acceleration, transport, and loss processes of radiation belt particles and dynamics of the inner magnetosphere. One major characteristic of the Arase's observation strategy is that various kinds of ground-based coordinated observations have been carefully prepared. Coordinated multi-point observations with satellites in space and ground-based observatories must contribute comprehensive understandings of physical mechanism of dynamically varying geospace environment. Indeed, Arase and its coordinated measurements in geospace has successfully observed several severe geospace storm events. This special issue expands upon new findings from these observations of Arase and ground-based networks as well as interplay among different missions. Results from related observations and modeling/simulation studies on the inner magnetosphere are also welcome.



ERGプロジェクト科学データ統合アーカイブの構築



データアーカイブ設計・開発での問題意識:

- 地球物理観測データなので特に秘密性はなく、できるなら全部公開したい。しかし...
- データセットが多種多様、校正具合や生データファイル形式はバラバラ
 - データファイルを標準化して処理・解析効率を上げる
- 短期アーカイブフェーズである
 - データのバージョンは頻繁に上がる、走りながら考えることが多い
- マンパワーは限られる(専任は特任教員4名、非ITC専門家)
- なるべく業界内のデファクトスタンダードに合わせて色々楽したい
 - 技術的にはプリミティブ、将来伸びそうな統合解析ツールに乗せる



科学データファイルリポジトリ

Index of /data/ergsc/satellite/erg/mepe/l2/3dflux/2017/04

Name	Name	Last modified	Size	Description
Parent Directory	Parent Directory		-	
att/	erg_mepe_l2_3dflux_20170401_v01_00.cdf	2018-04-09 22:49	497M	
hep/	erg_mepe_l2_3dflux_20170401_v01_01.cdf	2018-06-18 22:11	497M	
lepe/	erg_mepe_l2_3dflux_20170403_v01_00.cdf	2018-04-09 23:07	391M	
lepi/	erg_mepe_l2_3dflux_20170403_v01_01.cdf	2018-06-18 22:38	391M	
mepe/	erg_mepe_l2_3dflux_20170404_v01_01.cdf	2018-06-18 22:53	424M	
mepi/	erg_mepe_l2_3dflux_20170405_v01_01.cdf	2018-06-18 23:12	371M	
mgf/	erg_mepe_l2_3dflux_20170406_v01_01.cdf	2018-06-18 23:27	402M	
orb/	erg_mepe_l2_3dflux_20170407_v01_01.cdf	2018-06-18 23:45	453M	
pwe/	erg_mepe_l2_3dflux_20170408_v01_01.cdf	2018-06-18 23:59	480M	
xep/	erg_mepe_l2_3dflux_20170409_v01_01.cdf	2018-06-19 00:14	524M	
	erg_mepe_l2_3dflux_20170410_v01_01.cdf	2018-06-19 00:29	356M	
	erg_mepe_l2_3dflux_20170411_v01_01.cdf	2018-06-19 00:49	443M	
	erg_mepe_l2_3dflux_20170412_v01_01.cdf	2018-06-19 01:09	396M	
	erg_mepe_l2_3dflux_20170413_v01_01.cdf	2018-06-19 01:28	362M	
	erg_mepe_l2_3dflux_20170414_v01_01.cdf	2018-06-19 01:45	476M	

データ取得から1年経ったら科学データ(Level-2データとそれ以降の高次データ)は自動で公開されていく



データファイルの追跡可能性のために

```
G_ATTRIBUTES = STRUCT = --(35 Tags/768 Bytes)-->
PROJECT = STRING = 'ERG>Exploration of Energization and Radiation in Geospace'
DISCIPLINE = STRING = 'Space Physics>Magnetospheric Science'
SOURCE_NAME = STRING = 'ERG>Inner Magnetosphere'
DATA_TYPE = STRING = 'hep_l2_omniflux>HEP Level-2 omni flux'
DESCRIPTOR = STRING = 'HEP>High-energy electron experiments'
DATA_VERSION = STRING = 'v01_01'
TITLE = STRING = 'Level-2 omni flux data obtained by the'
TEXT = STRING = ''
GENERATED_BY = STRING = 'ERG Science Center, Institute for Space and Astronautical Sciences'
GENERATION_DATE = STRING = '20180616'
MODS = STRING = 'Created 06/2018'
ADID_REF = STRING = ''
LOGICAL_FILE_ID = STRING = 'erg_hep_l2_omniflux_20170327_v01_01'
LOGICAL_SOURCE = STRING = 'erg_hep_l2_omniflux'
LOGICAL_SOURCE_DESCRIPTION = STRING = 'Exploration of Energization and Radiation in Geospace'
PI_NAME = STRING = 'Takefumi Mitani'
PI_AFFILIATION = STRING = 'ISAS, JAXA'
MISSION_GROUP = STRING = 'ERG'
INSTRUMENT_TYPE = STRING = 'Particles (space)'
TEXT_SUPPLEMENT = STRING = ''
RULES_OF_USE = STRING[14] = [ ... ]
LINK_TEXT = STRING = 'For more information, see'
LINK_TITLE = STRING = 'the ERG Science Center website'
HTTP_LINK = STRING = 'https://ergsc.isee.nagoya-u.ac.jp'
TIME_RESOLUTION = STRING = '8 s'
START_TI = STRING = '534099126'
END_TI = STRING = '539628142'
DATA_START_TIME = STRING = '20170327 000001.715358'
```

メタデータ部に以下の情報を自動で付加

▶ DATA_VERSION

- ▶ データファイルバージョン番号(ファイル名にも)

▶ GENERATION_DATE

- ▶ データファイルの処理日時

▶ SOURCE/ANCILLARY_FILE

- ▶ 元になった生データファイル・ancillaryファイル名

▶ GENERATION_CODE

- ▶ 処理プログラム名とそのバージョン(SVNでバージョン管理)

▶ RULES_OF_USE

- ▶ データポリシー、使用上の注意など

```
DATA_AVERAGING_TYPE = STRING = '8 s average/start'
SOURCE_FILE = STRING = 'HEP_1b8_DL_REP_20170327_L_hist_v01_00.l1bin HEP_1b8'
ANCILLARY_FILE = STRING = 'HEP_L_energy_step_v003.dat HEP_H_energy_step_v003'
GENERATION_CODE = STRING = 'makecdf_erg_hep_l2_omniflux.pro(rev.1321), hep11'
CALIBRATION_HISTORY = STRING = '201805 Initial check'
```



あらせデータのデータ引用？

- ▶ Acknowledgementsに、解析に使用した全データのバージョン番号を明記している。
 - ▶ ERG-SC siteにデータファイルが置いてあるので、誰でも元データを使って解析することで追試できる。
 - ▶ データバージョンはどんどん新しくなっていくが、**どの時点でのデータを論文に使ったか追跡可能**。
 - ▶ データファイルに格納されているメタデータにより、サイエンスセンター内で**生データやデータ処理ソフトのレベルまで簡単に追跡可能**。
- ▶ いわゆるデータ出版したものをReferencesに書いてデータ引用しているわけではないが...

Acknowledgments

The EMMA magnetometer data were provided by M. Vellante and B. Heilig, the PIs of the EMMA. We thank the institutes who maintain EMMA stations used for this study: the Finnish Meteorological Institute (Finland), Sodankylä Geophysical Observatory of the University of Oulu (Finland). Science data of the Arase (ERG) satellite were obtained from the ERG Science Center operated by ISAS/JAXA and ISEE/Nagoya University (<https://ergsc.isee.nagoya-u.ac.jp/index.shtml.en>). The present study analyzed the MGF v01.01 data and the MEP-i v01.01 data. The AL index was provided by the World Data Center for Geomagnetism, Kyoto. The

[Yamamoto+, GRL, 2018]

統合解析ツール Space Physics Environment Data Analysis Software (SPEDAS)



過去・現役のSTP分野の衛星

データリポジトリ CDF

THEMIS

CDF

Van Allen Probes

CDF

自動データファイルダウンロード機能を持つAPIをデータ毎に開発・リリース

CDF

IUGONET
Metadata DB for Upper Atmosphere

Ascii CDF

THEMIS-GBO ASI 2011-09-20/06:58:00

- ▶ 米国THEMIS衛星・地上観測データの解析ツールとしてUCB, UCLAで開発・リリース(当時はTDAS)。IDLスク립トの集合体。
- ▶ 昨年くらいからSTP分野のコミュニティツールを目指して、THEMISデータ部をモジュール化。他データを取り込みやすいようにAPI等を整理 → SPEDAS
- ▶ データ読込プログラムは自動でデータファイルをダウンロード・保存
- ▶ 異なる時系列データのプロットを並べたり時系列解析することが容易。配列数・構造を意識せずにプログラミング可能。



CDF あらせ衛星データ

CDF ERG連携地上観測データ

CDF? HDF? シミュレーション・モデリングデータ

ERG-SCでSPEDAS用のAPI及びデータリポジトリを開発・公開、SPEDASという国際標準化にも参画



▶ 良かった点

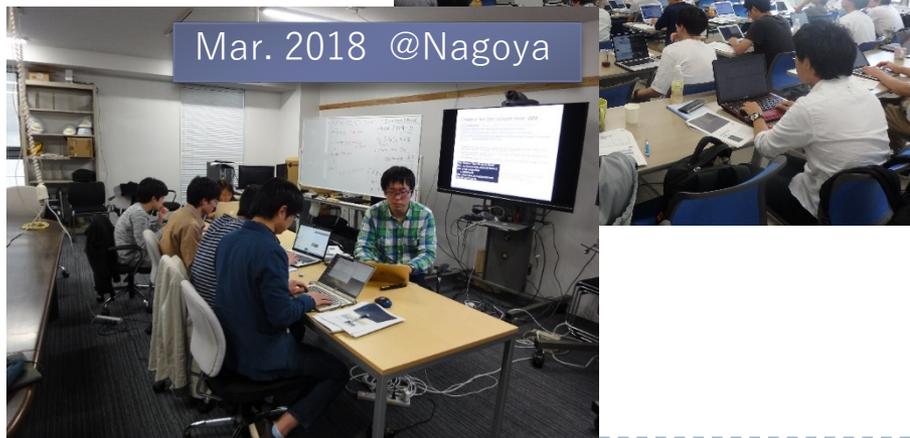
- ▶ データ配列の数や次元を意識せずに、抽象的にデータ処理・解析コードを書ける。
- ▶ 同じ日時・データ種でも、**最新のデータファイルを自動で認識してダウンロード**してくれる。
 - ▶ 実質データファイル更新をプッシュ通知してくれるわけで、地味に便利。
- ▶ 既に解析可能となっていた**多種多様な衛星・地上観測データとあらせデータとを合わせて、シームレスに解析**できる。

▶ 若干困る点

- ▶ コードのコア部分はSPEDAS開発チームの管理下(メリットもあるが)。
- ▶ 有償の商用ソフトであるIDLがベース。

あらせデータ解析講習会

- ▶ あらせデータを解析するための統合データ解析ツールに関する実践的な講習会を、年2回程度開催している。



来週、9月19-20日に東北大で行われるERG粒子データ解析ワークショップでも講習会を開催予定。

まとめに代えて:

ERGサイエンスセンターでの経験より、科学データアーカイブ
についてのコメント・雑感 (続き)



▶ 研究者(分野専門家)がデータアーカイブを担当する意義

- ▶ **データ設計は解析ツールと表裏一体**、解析ツールの趨勢・仕様・使い勝手に最適化できるのは、そのデータ・ツールのユーザーである研究者。
 - ▶ 10年前にSPEDASを選択しておいて、結果としてよかった。
- ▶ 研究者がアーカイブ作業を行うことのインセンティブをどう考えるか?の議論も必要。
- ▶ ただしICTのプロではないので、最新のICTは使えない。
- ▶ **プロジェクト初期の、成果を量産するフェーズ**では、高度な最新ICTが必ずしも必要でないなら、やはり**その道の科学研究者がデータアーカイブ開発を担当する方がよい**。そうでないと成果創出の瞬発力が出ない。

まとめに代えて:

ERGサイエンスセンターでの経験より、科学データアーカイブ
についてのコメント・雑感



▶ データ引用と追跡・追試可能性

- ▶ 「短期アーカイブ」フェーズでは、当然データや解析ツールのバージョンが頻繁に上がるので、それを踏まえたシステムが必要。
- ▶ 単にデータセンター・データベースの利用を追跡するだけなら、データセットのlanding pageを作ってそこへの引用を論文等にDOI付きで記述すればよい。
- ▶ しかし、バージョンがどんどん変わり得る段階のデータへも追跡・追試可能性を担保しようとするなら、上記では不十分。データのバージョン(ひょっとして解析ソフトのバージョンも?)を特定できるようなデータ引用のシステムを構築する必要がある。
 - ▶ DOI付与はMUSTで、DOI付き引用でDC・DB追跡、かつ論文の本文中にデータのバージョンを記載、あたりが現実的な線か。
 - ▶ 幾つか前の世代のデータファイルも一緒にアーカイブ、を意味する。

背景:

- 科学データ全般に追跡・追試可能性が求められるようになった。
- 米地球物理学会(AGU)誌ではregulationが厳格化し、データを開示しないと投稿を受け付けてもらえないし、出版時にデータファイルにアクセスできることが必要条件とされつつある。



(真の)まとめ

- ▶ ERG-SCは、ERGプロジェクトの科学データアーカイブを含めた「統合データ解析環境」を構築して、ERGプロジェクトの科学成果創出の基盤を担っている。
- ▶ Miyoshi, Y., Hori, T., Shoji, M., Teramoto, M., Chang, T. F., Segawa, T., et al. (2018). The ERG Science Center. *Earth Planets Space*, 70(1), 96.
<https://doi.org/10.1186/s40623-018-0867-8>