



# 宇宙科学・超高層大気科学分野データのデータ駆動型研究への利用促進のためのPythonベースのデータ取得・解析ツール開発

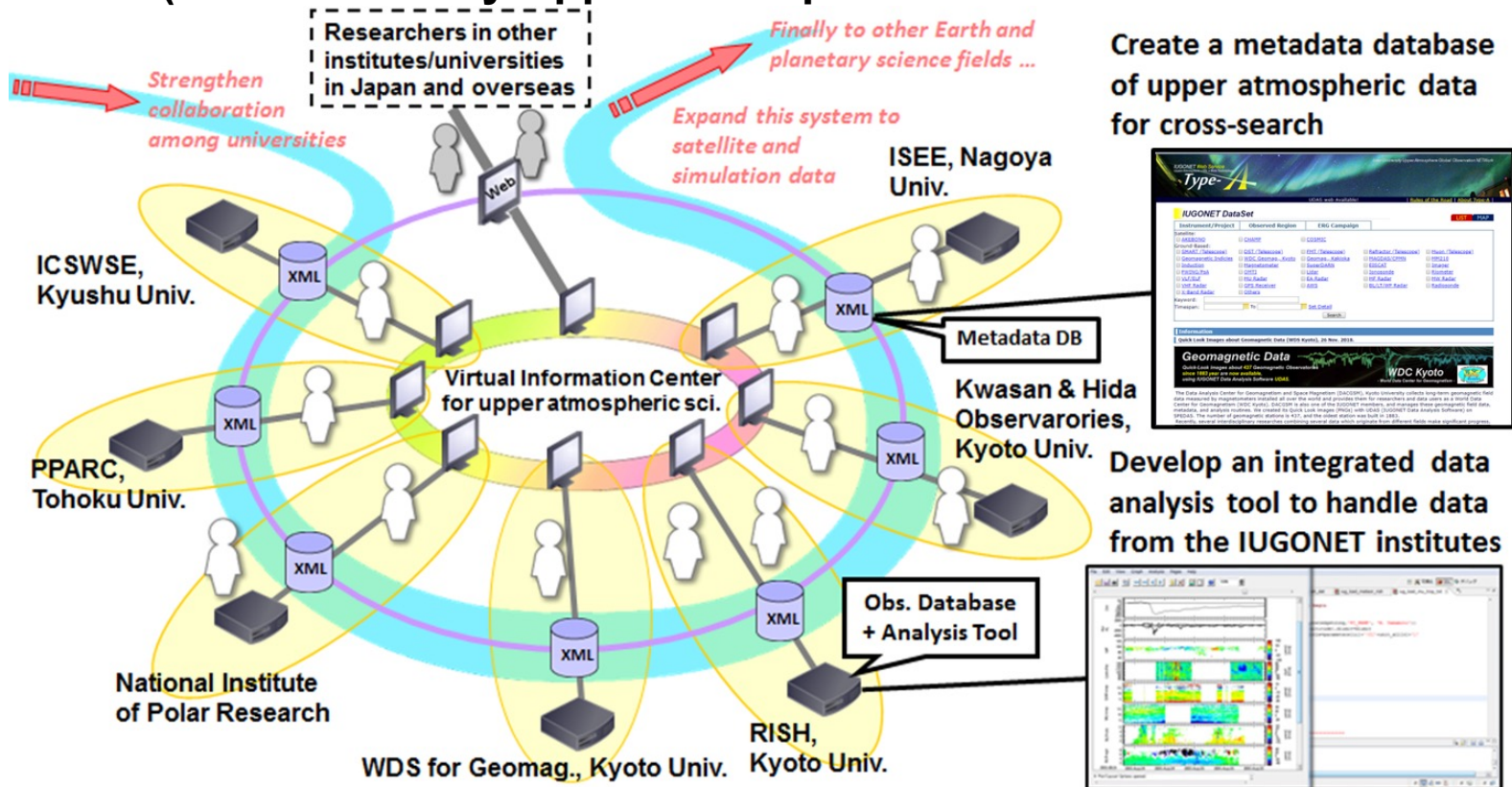
今城 峻<sup>1\*</sup>, 田中 良昌<sup>2</sup>, 阿部 修司<sup>3</sup>,  
新堀 淳樹<sup>4</sup>, 土屋 史紀<sup>5</sup>, 能勢 正仁<sup>6</sup>, 上野 悟<sup>7</sup>

- <sup>1</sup> 京都大学 理学研究科附属地磁気世界資料解析センター
- <sup>2</sup> 情報・システム研究機構 データサイエンス共同利用基盤施設
- <sup>3</sup> 九州大学 国際宇宙惑星環境研究センター
- <sup>4</sup> 名古屋大学 宇宙地球環境研究所
- <sup>5</sup> 東北大学 理学研究科附属惑星プラズマ・大気研究センター
- <sup>6</sup> 名古屋市立大学 データサイエンス学部
- <sup>7</sup> 京都大学 飛騨天文台



# IUGONET project

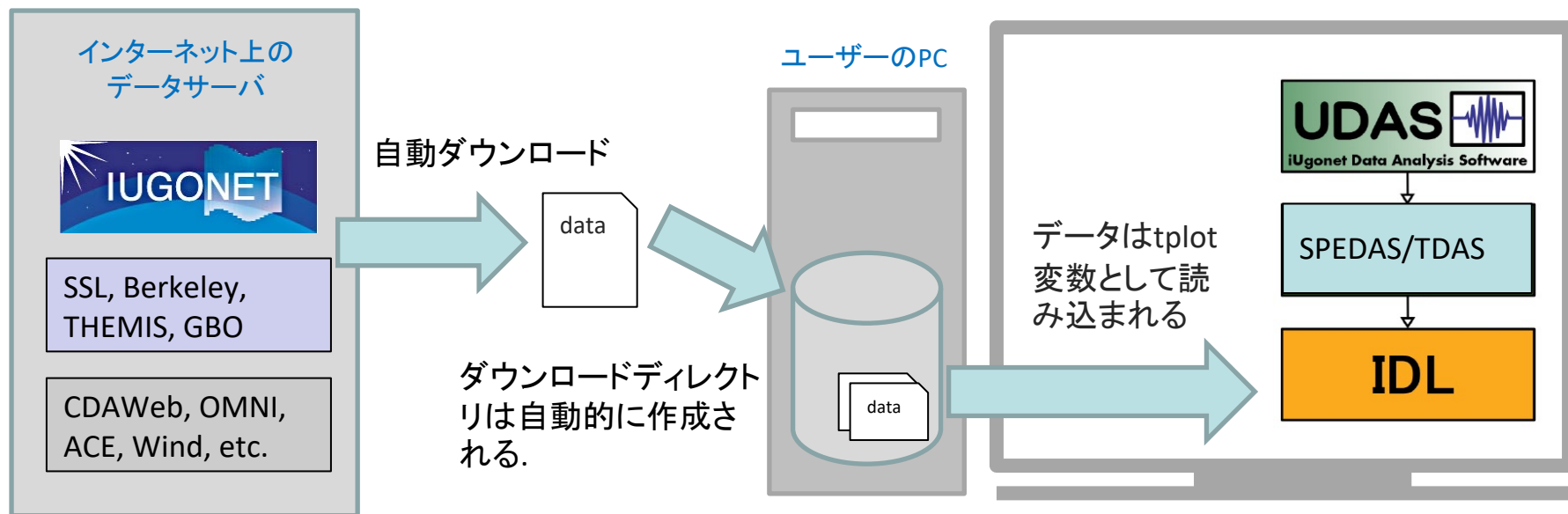
**IUGONET (Inter-university Upper atmosphere Global Observation NETwork)**



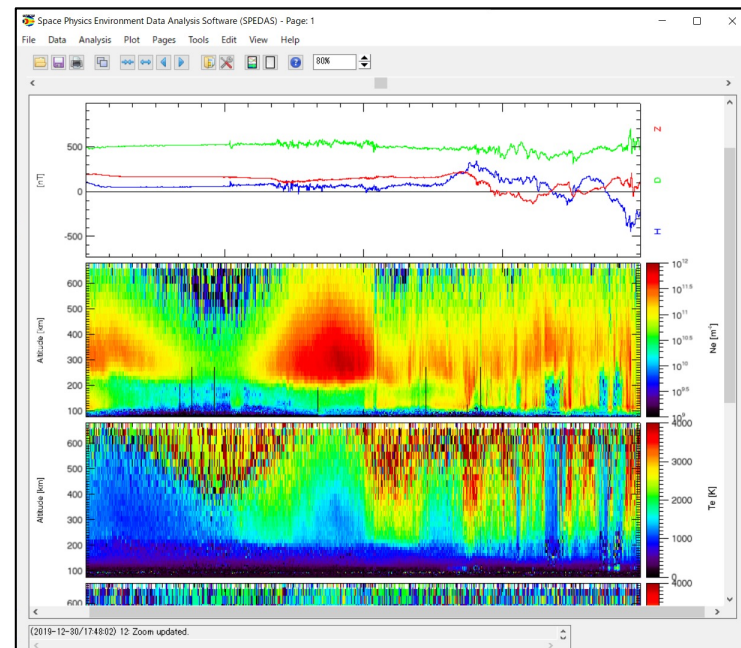
- 日本全国の大学・研究機関に分散する多様な上層大気データの公開を支援
- 効率的な解析のための研究基盤（メタデータデータベース、解析ツール）の整備
- 若手研究者育成のためのデータ解析ワークショップの開催



# UDASプラグインの開発



- 上層大気分野で主流言語であるIDL/SPEDASベースのデータ読み込みプラグイン「UDAS」を開発。
- 3つの基本コマンドでデータ読み込みと描画が可能。
  - `timespan, 'yyyy-mm-dd'`
  - `iug_load ***`
  - `tplot, ++`





# IUGONETによるデータ駆動型研究の促進

- 日本の多様な上層大気データに先端的なデータサイエンス手法を用いたデータ駆動型研究を促進したい。
- 読み込みプログラムがIDLベースであることの問題点。
  - 高額なライセンス料
  - 宇宙・上層大気分野以外ではほとんど普及していない
  - AI、機械学習系ライブラリが少ない。
- SPEDASのPython版「PySPEDAS」が米国で開発が進んでおり、広まる動きがある。
- Pythonは、
  - ライセンス料不要
  - データサイエンス系科学で幅広く普及
  - AI、機械学習系ライブラリが豊富
- 一方、国内のデータの読み込みプログラムはまだ十分用意されておらず、AI、機械学習系ライブラリの適用のハードルになっていた。
- 本プロジェクトでは、PySPEDASを拡張するプラグイン「PyUDAS」を開発
  - ・ 公開することで、国内の様々なミッションのデータをPythonで取得・可視化・解析することを可能にする。
- これにより特に機械学習の応用に関わるデータ駆動型研究の発展を目指す。



# PySPEDAS

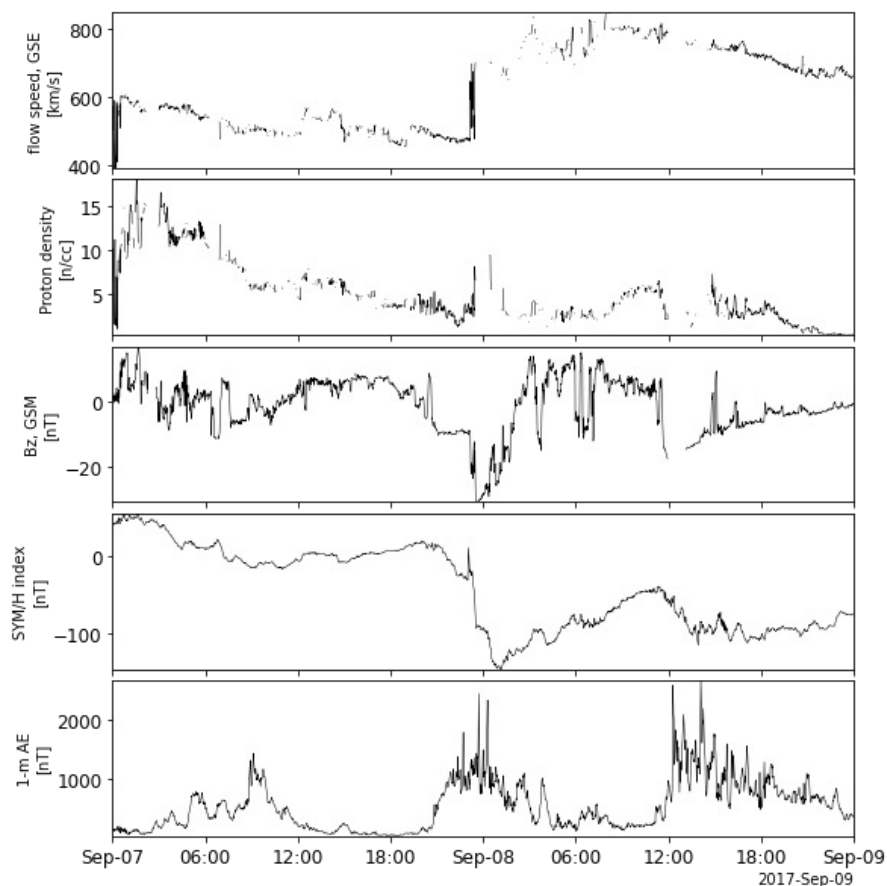
- `> import pyspedas`
  - `> import pyplot`
- モジュールのインポート

- `> pyspedas.omni.data(trange=['2017-09-07', '2017-09-09'])`

時間を指定してデータをダウンロード・  
変数に格納

- `> pyplot.tplot(['flow_speed',  
                  'proton_density', 'BZ_GSM',  
                  'SYM_H', 'AE_INDEX'])`

時系列データを可視化。



- 米国UCLAのチームが中心となって開発
- 基本的にIDL版SPEDASと同様の仕組みで動作
- まだ開発途上であり、SPEDASと全く同じデータの読み込み・解析ができる訳ではない





# 開発するロード関数

## CDFファイル

- asi\_nipr.py
- ask\_nipr.py
- eiscat.py
- eiscat\_vief.py
- elf\_hokudai.py
- gmag\_nipr.py
- gmag\_nipr\_induction.py
- hf\_tohokuu.py
- irio\_nipr.py
- kyushugcm.py
- lfrto.py

## NetCDFファイル

- ear.py
- gaia\_cpl\_nc.py
- gaia\_gcm\_nc.py
- gps\_isee.py
- gps\_ro\_rish.py
- ltr\_rish.py
- meteor\_rish.py
- mf\_rish.py
- mu.py
- radiosonde\_rish.py
- netcdf\_to\_tplot.py (ファイル読み込みモジュール)

## ASCIIファイル・その他

- aws\_rish.py
- blr\_rish.py
- gmag\_icswse\_iaga.py
- gmag\_wdc.py (複数のサブルーチンから構成)
- ionosonde\_rish.py
- iprt.py
- iprt\_highres.py
- ltr\_rish.py

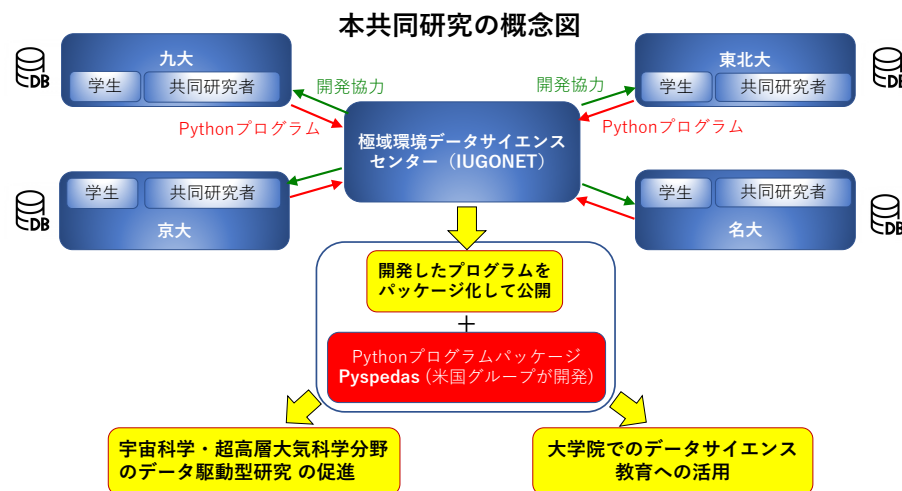
- CDF、netCDF、ASCIIなどのフォーマットのデータを読み込む約31の関数が開発の対象。



# 開発体制

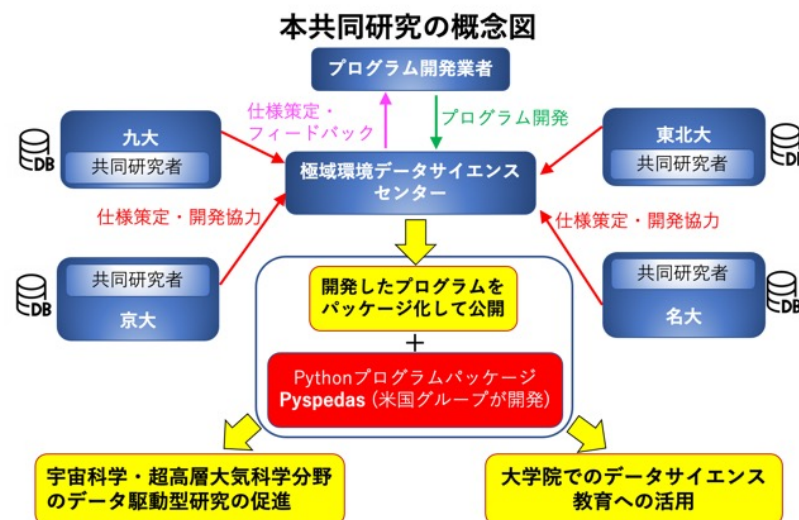
## • ~2022年度

- 各データ提供機関の学生をアルバイト雇用し、開発を進めた。
- 開発と、参加学生へのデータ教育両方の効果を狙った。
- 開発手順書・関数テンプレートを用意。



## • 2023年度

- ソフトウェア開発業者に発注し、開発。
- 各関数の仕様書を各データ提供機関の研究者で分担して作成。





# 学生を中心とした開発

- 京大地磁気センターで、学生を中心としたPythonロード関数開発を先行して行っていた。
- 九大、京大、名大、東北大で各数名の学生をROIS-DS-JOINTの研究費で雇用。
- Githubを用いた共同開発体制。
- 各学生に担当関数を割り振り、定期報告会を実施。
- 進捗に学生間でかなりの差があり、全関数の完成にはいかなかった。

template.py の使い方

9-1. template.py を複製し、ファイル名、及び、関数名「template」を変更する。

例： ファイル名を gmag\_nipr.py に、関数名を gmag\_mm210 に変更する。

```
1 import numpy as np
2
3 from pyspedas.utilities.time_double import time_double
4 from pyplot import get_data, store_data, options, clip, ylim, cdf_to_tplot
5 from ..load import load
6
7 def template(                                     gmag_mm210
8     trange=['2020-01-01', '2020-01-02'],
9     site='all',
10    datatype='all',
11    parameter='',
12    no_update=False,
13    downloadonly=False,
14    uname=None,
15    passwd=None,
16    suffix='',
17    get_support_data=False,
18    varformat=None,
19    varnames=[],
20    notplot=False,
21    time_clip=False,
22    version=None,
23    ror=True
24 ):
```

9-2. 「Set parameters (1)」の箇所を変更する。

```
26 #==== Set parameters (1) =====#
27 file_format = 'cdf'
28 remote_data_dir = 'http://iugonet0.nipr.ac.jp/data/'
29 local_path = '/nipr/'
30 prefix = 'nipr_'
31 file_res = 3600. * 24
32 site_list = ['']
33 datatype_list = ['']
34 parameter_list = ['']
35 time_netcdf=''
36 #=====
```

- file\_format は、現在は 'cdf' のみ。今後、netcdf、fits を追加予定。
- remote\_data\_dir は、データの所在 URL のベース部分を記入。remote\_data\_dir + pathformat がデータの URL を表す。ここで、pathformat は、「Set parameters (2)」で設定する。
- local\_path は、データ保存先。各機関の名称の略語を入れる。

作成した説明書・テンプレート



- ・ 各関数の仕様書テンプレートを作成し、各大学の研究者で分担して仕様書を作成。
- ・ データファイル読み込みに関わる一部モジュールはIUGONET側で作成。
- ・ ROIS-DS-JOINTの研究費でSPEDAS関連プログラムの開発で経験と実績のある業者に依頼。
- ・ 現在も開発中。

N o.	イベント・ 担当	2023/10	11	12	2024/1		2			3	
1	イベント	▲ 10月末 業者打ち合 わせ	▲ 11月末 発注		▲ 1月中旬 中間報告					▲ 3/20 最終締切	
2	IUG担当	各プログラム仕様書開発 仕様書テンプレート開発					テスト				
3	業者担当	製造・内部テスト									



# 開発の成果

## CDFファイル

- asi\_nipr.py
- ask\_nipr.py
- eiscat.py
- eiscat\_vief.py
- elf\_hokudai.py
- gmag\_nipr.py
- gmag\_nipr\_induction.py
- hf\_tohokuu.py
- irio\_nipr.py (データ側問題)
- kyushugcm.py
- lfrto.py (データ側問題)

## NetCDFファイル

- ear.py
- gaia\_cpl\_nc.py
- gaia\_gcm\_nc.py
- gps\_isee.py
- gps\_ro\_rish.py
- ltr\_rish.py
- meteor\_rish.py
- mf\_rish.py
- mu.py
- radiosonde\_rish.py
- netcdf\_to\_tplot.py (ファイル読み込みモジュール)

## ASCIIファイル・その他

- aws\_rish.py
- blr\_rish.py
- gmag\_icswse\_iaga.py
- gmag\_wdc.py (複数のサブルーチンから構成)
- ionosonde\_rish.py
- iprt.py
- iprt\_highres.py
- ltr\_rish.py

開発済み

途中まで作成済み or

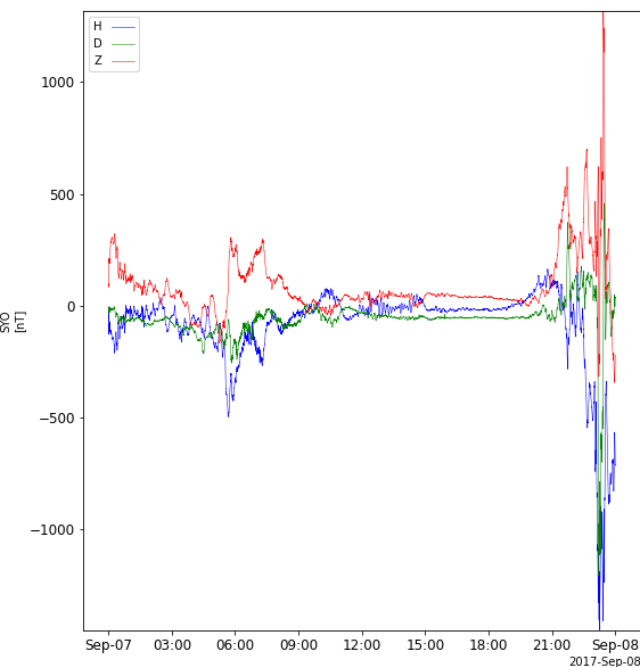
仕様書作成中

業者開発中

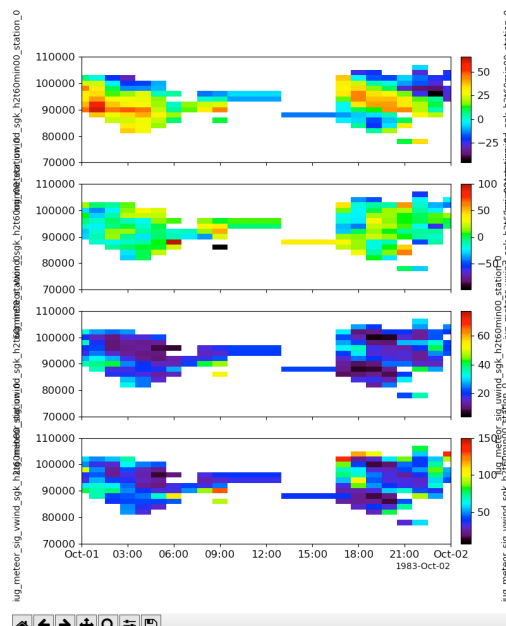
- 約1/3の数の関数が作成済み。
- 今回の業者発注でCDF/NetCDFのロード関数はほとんど完成する見込み。
- プログラムは<https://github.com/iugonet/pyudas>で公開されている。



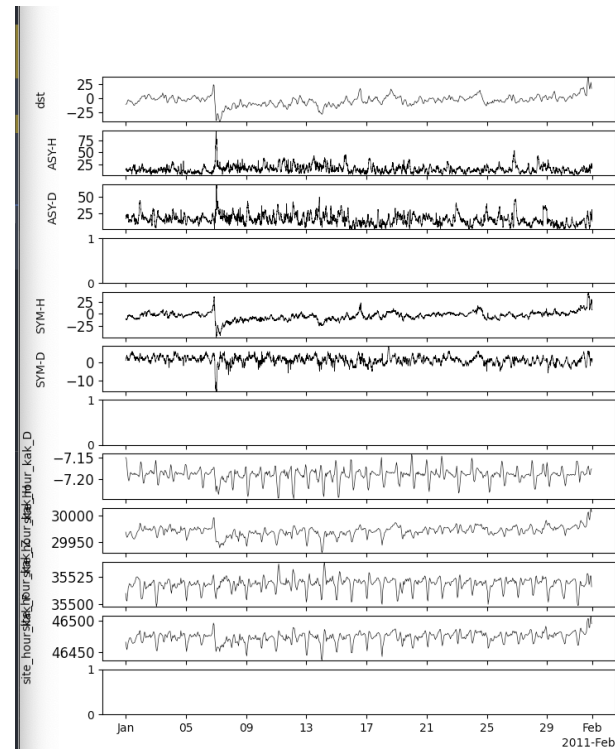
# 使用例



極地研 地磁気データ



京大rish 流星レーダー



WDC Kyoto 地磁気データ・指数

```
> from iugonet import gmag_nipr
    ロード関数のインポート。

> gmag_nipr(trange=['2017-09-07', '2017-09-08'], site='syo')

> tplot(['nipr_mag_syo_1sec'])
```



# 講習会を通した普及活動

- 2022、2023年度ISEE研究集会「太陽地球系物理学分野のデータ解析手法、ツールの理解と応用」において、PySPEDAS及び開発版PyUDASプラグインを用いたデータ解析講習会を行った。
- 現地・オンラインで約30名程度の学生・研究者が参加。(工学系大学・学部 of 学生も多数)
- 多様な太陽地球物理学分野のデータにpythonの最新の解析ツールを適用できる環境の構築をサポート。

The screenshot displays a Zoom meeting interface. On the left, a file explorer shows the contents of a Jupyter Notebook. The main area shows a Jupyter Notebook with Python code using PySPEDAS. The code includes imports, data loading, and plotting commands. On the right, a list of participants is visible, including S.UeNo (Hida Obs), Atsuki Shinbori, Shun Imajo, and others. The bottom of the screen shows the Zoom controls with buttons for '招待' (Invite) and 'ミュートを解除します' (Unmute).

```

[ ]: tplot(['flow_speed', 'proton_density', 'BZ_GSM', 'SYM_H', 'AE_INDEX'])
[ ]: pyspedas.omni.data?
[ ]: help(pyspedas.omni.data)
[ ]: pyspedas.thisis.gmag(trange=['2017-09-07', '2017-09-09'], sites='tro')
[ ]: pyplot.tplot_names()
[ ]: tplot(['thg_mag_tro'])
[ ]: from iugonet import gmag_nipr
[ ]: import site
[ ]: site.getsitepackages()
[ ]: gmag_nipr(trange=['2017-09-07', '2017-09-09'], site='syo')
  
```



## 今後の計画・課題

- 全ロード関数の完成
  - 一部関数はかなり構造が複雑で作業が難航。
  - ASCIIフォーマットのデータは個々のデータ特有の構造を理解する必要があり、作業が難航。
- PySPEDAS講習会の定期的開催による普及活動
  - 特に、情報工学系ではPythonは普及しており、同分野の学生・研究者への普及が、データ駆動型研究促進への鍵となる。
  - その意味では地球科学分野の集会以外での宣伝も必要かもしれない。
  - ライセンスが不要であるため、データ解析の入門者を対象とした教育でも幅広く利用が期待出来る。





## まとめ

- 宇宙科学・上層大気科学分野データの解析において、データの読み込みがハードルとなり、先端的なデータサイエンス手法の適用が難しい環境にあった。
- 本プロジェクトでは、Pythonベースの解析ツールPySPEDASを拡張するプラグイン「PyUDAS」を開発・公開することで、国内の様々なミッションのデータをPythonで取得・可視化・解析することを可能し、特に機械学習の応用に関わるデータ駆動型研究の発展を狙った。
- IUGONETチームが主導し、昨年度までは各大学の学生を主体に開発を行い、今年度は業者発注で開発を行っている。
- 約1/3のプログラムが開発済み、CDF/NetCDFのロード関数のほとんどは業者開発で近々完成する見込み。
- 研究集会において、pyspedas/pyudasを用いたデータ解析例の講習会を行い、特に情報工学系の学生・研究者への普及を狙っている。