

## NICTサイエンスクラウドを用いた磁気圏グローバルMHDシミュレーションデータとLANL衛星粒子データの比較

# 山本 和憲 [1]; 長妻 努 [2]; 久保田 康文 [2]; 村田 健史 [1]; 亘 慎一 [1]  
[1] 情報通信研究機構; [2] NICT

### Comparison between particle environment around GEO from global MHD simulation and that from LANL based on the NICT Science Cloud

# Kazunori Yamamoto[1]; Tsutomu Nagatsuma[2]; Yasubumi Kubota[2]; Ken T. Murata[1]; Shinichi Watari[1]  
[1] NICT; [2] NICT

<http://www.seg.nict.go.jp/>

Substorm injection is one of the important element of magnetospheric substorm, like auroral break up. Studying substorm injection is important to understand the physics of substorms. Also, substorm injection temporarily changes the particle environment around satellites at GEO. And dynamical variations of particle environment around GEO is one of the causes of satellite anomaly due to surface charging. We try to evaluate our magnetospheric global MHD simulation code by comparing output from global MHD code and LANL satellite particle data.

We have built a large-scale environment for the statistical analysis where the data obtained through satellites observations and computer simulation are used to construct a uniform, integrated dataset. In this environment, plenty of data are integrated in the following manner: (1)Archiving large quantities of data files, (2)Resampling time series and convert coordinates, (3)Extracting parameters from simulation data, and (4)Merging both data into one file.

**(1)Archiving large quantities of data files:** Using the STARS (Solar-Terrestrial data Analysis and Reference System) meta-database that provides meta-information of observation data files managed at distributed observation data sites over the internet, users download data files without knowing where the data files are managed. On the other hand, simulation data is saved from supercomputer to petabyte-scale distributed storage that is connected to 10GbE (JGN-X).

**(2)Resampling time series and convert coordinates:** We developed an original data class (SEDOC class) to support our reading of data files and converting them into a common data format. The data class defines schemata for several types of data. Since this class encapsulates data files, users easily read any data files without paying attention to their data formats. The SEDOC class supports a function of resampling time series through linear interpolations and converting them into major coordinates systems.

**(3)Extracting parameter from simulation data:** We have developed a 3-D visualization system that visualizes both of these data simultaneously and extract parameters from simulation data in the arbitrary coordinate value.

**(4)Merging both data into a single file:** Time scale and coordinate are regularized over data files.

We found a practical problem of the system, especially in case of long durational data analyses. It is the problem of the computational load on the processes two to four. It is necessary to solve this problem in order to achieve data-intensive processing for plenty of data files with non-negligible file I/O and CPU utilization.

To overcome this problem, we developed a parallel and distributed data analysis system using the Gfarm and Pwrake based on the NICT Science Cloud. The Gfarm shares both computational resources and perform parallel distributed processing. In addition, the Gfarm provides the Gfarm file system which behaves as a virtual directory tree among nodes. The Pwrake throws a job for each Gfarm node that has a target data file in the local disk. It utilizes local disk I/O to achieve effective load balance.

Detailed comparison between simulation and observation will be shown in our presentation.

サブストームインジェクションはオーロラブレイクアップ等と同様、磁気圏サブストームの典型的な現象の一つであり、これを研究することはサブストームの物理を理解する上で重要である。同時に、サブストームインジェクションは静止軌道の粒子環境を急激に変化させることから、人工衛星の表面帯電等のリスク要因にもつながる。

我々は、過去に蓄積した膨大な磁気圏グローバルMHDシミュレーションの計算結果と、LANL衛星の粒子データを比較することで、シミュレーションの計算結果の評価を試みると共に、シミュレーションの計算結果を用いてサブストームインジェクションによる静止軌道の粒子フラックス増大の推定を試みる。

本研究では両データの比較を行うために、NICTサイエンスクラウド(以下、サイエンスクラウド)が提供している計算機リソースおよびデータ処理システムをマッシュアップし、両データのデータフォーマットを統一したデータセットを作成する環境を構築した。データセットを作成するプロセスは、データアーカイブ、時間軸・座標系の統一、物理量の抽出、両データのマージの4つからなる。

データアーカイブ

分散管理された衛星観測データの収集には、NiCTy+Download Agentを利用することにより、分散管理情報をメタデータベース化したSTARSDBを参照して定期的にデータを収集できる。また、サイエンスクラウドではスパコンをハウジングできる環境を提供しており、スパコンから出力された大規模データは、基幹ネットワークが10GbEで接続された分散ストレージ(Gfarm)に出力することができる。

### 時間軸・座標系の統一

衛星観測データの時間分解能と座標系をシミュレーションデータに合わせる。サイエンスクラウドが提供する SEDOC を利用することで、衛星観測データのデータフォーマットの差異や時間単位で出力されたファイル群を意識することなく、指定した時間分解能にサンプリングして配列に格納される。また、衛星軌道データについては、主な座標系に変換することができる。

### 物理量の抽出

座標系が統一された衛星軌道の任意の座標値に該当する、シミュレーションデータの物理量を抽出する。サイエンスクラウドが提供する V\_Aurora は、衛星観測データとシミュレーションデータの融合表示および指定した座標値の物理量の抽出が可能である。

### 両データのマージ

時間軸・座標系が統一された両データを1つのファイルにマージする。

データアーカイブされた大量のデータファイルに対して、上記の一連の処理を行った場合、データ読み込み時のディスク I/O およびデータ補間時の CPU の処理性能がボトルネックになる。本研究では、この問題を解決するため Pwrake によるワークフローを作成した。Pwrake は分散配置されたファイルがあるノード上でプロセスを起動するようにジョブ管理するため、ローカルディスク I/O を活用したデータ処理が行える。

本発表では、過去に蓄積した LANL シリーズの衛星データと Global MHD シミュレーションデータを用いて比較を行った結果について紹介する。

