

数値予報標準データセットシステム NuSDaS

NuSDaS: Numerical Prediction Standard Data-set System

豊田 英司[1]

Eizi TOYODA[1]

[1] 気象庁 数値

[1] JMA NPD

0. 概要

2001年3月から気象庁の数値予報ルーチンで用いられているデータセット形式 NuSDaS について紹介する。

1. はじめに

気象庁の数値解析予報システム (NAPS) は 2001 年 3 月に更新され、スーパーコンピュータの OS が汎用機系 OS (VOS3) から UNIX に変わった。これにあわせて、NAPS で運用される数値予報ルーチンで格子点データ (GPV) を格納するファイル形式も UNIX に対応することが必要となった。このため従来用いられていた GVD1, GVS1 にかわって NuSDaS が開発され、現在運用されている。

2. データに関する想定

NuSDaS は数値予報ルーチンを効率よく組織的に運用することを目的として設計されたものであるため、NetCDF, HDF, XML などの汎用データ形式と比べれば制約の多いものであるが、いいかえれば数値予報ルーチンで用いられるデータに関しては無理なくコンパクトなプログラム記述で操作できるということでもある。

データ入出力の単位は整数型または浮動小数点型の 2 次元配列である。この 2 次元配列は空間に張られることを想定しており、これを「種別、基準時刻、識別符、対象時刻、面、物理量」を用いて指定して入出力を行う。つまり、GPV は時空間に張られた格子上に存在することが想定されている。プログラムの可読性のため、同定情報は時刻を除いてテキストである。GRIB データをそのまま格納することを想定して、物理量名は GRIB の物理量符号と対応付けて定義されている。

3. ファイル形式

NuSDaS がファイル形式でなくデータセット形式と称しているのは、データが格納されるファイルだけでなく一定のディレクトリ構造を保持することを規定しているからである。

最低レベルであるファイルは複数のレコードから構成される。レコードは先頭と末尾に 4 バイト整数でレコード長が記述されているもので、これは多くの UNIX 上の Fortran 処理系で書式なし順番参照ファイルとして用いられている形式である。

レコードの内容の先頭 8 バイト目から 4 バイトの文字列がレコードの種別をあらわす。レコードの種別は識別部 (書誌情報に相当)、管理部 (格子の張る空間)、アドレス部 (データ部オフセット)、補助管理部 (地形に沿う鉛直座標、レーダー運用情報など)、データ部 (2 次元配列本体)、

情報部(ユーザ自由記述)、終了部の7種類である。
データの利用に必要なメタデータは主に管理部、
補助管理部のフィールドとして規定されている。

データ部の2次元配列はオフセット・係数圧縮、ランレングス圧縮、
または GRIB 形式をサポートしている。精度を維持するため、
オフセット・係数圧縮のオフセットと係数はデータ内容により
2次元配列ごとに自動的に選択される。

4. データセット形式

ファイル形式は充分自己記述的であり、同定情報からファイルを
検索することは可能であるが、
運用効率を向上させるため、ファイルを一定の構造をもつ
ディレクトリツリーに置く。
ディレクトリツリーの頂点を NuSDaS ルートと呼ぶ。
アプリケーションはカレントディレクトリに NUSDAS<数字>
という形の名前のディレクトリを探して
この中にある定義ファイルと呼ばれるテキストファイルを解釈して
データファイルパスを決定する。

5. 制約、あるいは将来展望

時空間ではない多次元データなど、データの枠組みが異なる
応用には NuSDaS は向いていない。
しかしながら、気象用途におけるデータの実際的同定に必要なデータスキーマの
最小限を示すという意味においては地球惑星科学全体にとって
きわめて確実な意義をもっているといえよう。

また今後の大規模分散メモリ並列計算に対応していくためには、
並列計算時に2次元配列の部分入出力を可能にする仕組みが必要となるだろう。

6. 参照

数値予報課報告 別冊 (調べなきゃ!)