

気象用格子データ形式の比較：気象庁ではなぜ索引付き順番探索を使うのか Comparison of grid data formats in meteorology: the reason for indexed sequential access method (ISAM) used in JMA

豊田 英司^{1*}
 TOYODA, Eizi^{1*}

¹ 気象庁予報部数値予報課

¹NPD, Japan Meteorological Agency

数値天気予報などのシミュレーションで作成される格子点データにはWMO（世界気象機関）標準のGRIB、OGC標準のnetCDF、気象庁内で使われるNuSDaSなど多数の形式がある。Wright and Gao (2008)はデータ形式をファイル構造により順番探索と直接探索に二分し、部分読み込みの高速性とファイルサイズを選択と論じた。しかし気象庁で用いられるNuSDaS（豊田、2001）は索引付き順番探索ファイルであり、高速性とサイズが両立する。これら3種類の性能上の長短を比較した（表1）。

現業数値予報でのデータアクセスには、しばしばデータ構造が疎な配列となる、ファイル書き出しは一度に行われて追加が行われない、ファイルの一部を読みだすことが多く性能が問われるという特徴がある。索引付き順番探索ファイルの短所（索引作成）が目立たず、読み書き速度とコンパクトなファイルサイズの両立という利点が享受しやすい。

また、将来データサイズが巨大化してゆくと、単純な順番探索は不利になってゆくことも注意される。

引用文献

Bruce Wright and Feng Gao, 2008: GRIB vs NetCDF: Evaluation of the Technical Aspects. WMO ET-ADRS Doc.2.3(1) <http://goo.gl/AFrsls>

豊田英司, 2001: 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の紹介. 合同大会講演A2-001 <http://goo.gl/JE0a3M>

キーワード: GRIB, netCDF, NuSDaS, 格子データ, 索引付き順番探索

Keywords: GRIB, netCDF, NuSDaS, grid data, indexed sequential access method

ファイル構造とその特性

	順番探索	直接探索	索引付き順番探索
一部書き出し処理	単純: ただ先着順	単純: 決まったところにシークするだけ	複雑: 先着順、その位置を別途索引に保存
ファイルサイズ	コンパクト	疎配列が膨張	比較的コンパクト
一部読み込み処理	遅い: 先頭から読まねばならない	速い: 決まったところにシークするだけ	比較的早い: 索引で決まる場所にシーク
他の利点	ファイル作成時に内容未確定でもよい	単一ファイルへの並列書き出し	
気象格子データの例	GRIB2 GTOOL3	GrADS Binary netCDF	NuSDaS