MGI37-06 Room:413 Time:April 29 10:15-10:30

Comparison of grid data formats in meteorology: the reason for indexed sequential access method (ISAM) used in JMA

TOYODA, Eizi[1]*

[1]NPD, Japan Meteorological Agency

Many format standards are used for grid data of simulation such as numerical weather prediction (NWP). Wright and Gao (2008) argued there are direct-access and sequential formats, and a choice is needed between fastness of partial read and compactness of data files. However, a JMA's local standard NuSDaS (Toyoda, 2001) uses the third category, ISAM (indexed sequential access method) which achieves both fastness and compactness. These three types of file formats are comared (Table 1).

In the operational NWP (1) data structure is often sparse, (2) each file is written by a single process, and (3) many subsequent processes read a part of data tile. In this situation the weakness (cost of indexing) is not outstanding and the strength (fastness and compactness) are enjoyed.

It is also noted that the weakness of sequential access (full scan for partial data) will be aggravated in the future computing with larger data size.

References
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
TOYODA Eizi, 2001: NuSDaS: Numerical Prediction Standard Data-set System. JpGU presentation A2-011 http://goo.gl/JE0a3M

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

Types of File Structure

	Sequential Access	Direct Access	Indexed Sequential Access Metind (ISAM)
Partial Write	Simple: append to EOF (end of file)	Simple: seek and write	Complex: append to EOF, then index location
File Size	Most Compact	Sparse array Bloated	Compact
Partial Read	Slow: all records must be scanned	Fastest: only seek and read	Fast: look up index, seek and read
Other Strength	No definition needed when creating file	Parallel write to single file	
Meteorological Examples	GRIB GTOOL3	GrADS Binary netCDF	NuSDaS