

ASM-2001 discussion on Observation handling

Heikki Järvinen, Kristian Mogensen and Sami Saarinen

1 Foreword

One of the Working groups at the ASM-2001 was devoted to “Observation handling”. This note consists of the discussion paper prepared for the working group, added with the outcome of the discussions, and a synthesis of the opinions and contribution of many of our Hirlam colleagues not able to attend the ASM this year.

2 Introduction

Volumes of the remote sensing data files are on increase. ECMWF has developed the Observational DataBase (ODB) software for enabling them to cope with the next generation satellite instrument data. ECMWF operations has been using ODB since end of June 2000 in 4D-Var (4D-Var/12h since September 2000). Software is also heavily used by the ERA-40 re-analysis project for resolving duplicate data problem in the multiple (conventional) data supplier context. Their database currently contains over 100GB online ODB-data split into 6-hourly time windows, since 1943. Météo-France will use ODB soon in operations through the IFS/ARPÉGÉ collaboration with ECMWF. Also DWD has shown great interest toward ODB and is currently performing remote familiarization with the software. UKMO has also been interested, and may come back in about one year time. ODB has also been introduced to the Canadian Meteorological Centre.

ECMWF is a global system with global data assimilation and with a very long data cut-off. Hirlam, on the other hand, is a limited area system. Data is “local” and the cut-off is very short. Data volumes in Hirlam will never be very large with the ECMWF standards.

Hirlam 3D/4D-VAR aims at exploiting a comprehensive set of remote sensing data. Computing facilities of Hirlam institutes will always remain modest, as compared with

those of ECMWF. Bottle-necks in data handling and accessing are therefore to be expected, even with the planned data volumes.

Currently the Hirlam reference system (Optimum Interpolation) supports MAOF and AOF/COF observation data file format. HIRVDA on the other hand makes use of OBSPROC and CMA file format. Input data format for observation processing is BUFR both in MAOF and in OBSPROC. AOF/COF file format will disappear as soon as OI disappears. OI may however survive longer than perhaps expected a year or two ago. Input data format for the surface analysis is recently changed to BUFR. ODB can be considered as a sort of observation file format as well, although it provides a number of additional functionalities. Conversion to ODB would mean abandoning CMA file format in HIRVDA.

The conversion from CMA format to ODB implicates at ECMWF, that a large part of OBSPROC code, roughly 1/3, will be included into IFS (Integrated Forecasting System of ECMWF), and thus reducing BUFR-handling to the BUFR-to-ODB conversions.

3 Observational DataBase

ODB stands for Observational DataBase. It is a home-grown relational database software developed at ECMWF. It comprises of Fortran90 interface for data access, flexible data definition layout- and fast data extraction (query) language catered for observation handling. It has also embedded data partitioning, where MPI-interface can be used for parallel data passing between processors.

ODB software has been coupled with IFS via a purpose built interface layer. In practice, IFS requests observational data through this layer: data queries are executed and the end results – one or more matrices containing observational data – are propagated back to IFS. Data queries can either be parallel or local, the latter being more favourable due to obvious performance reasons.

Most of the ODB software is in fact written in C-code. Even data queries are currently transformed into well vectorizable C-code, where applicable. Speed of ODB comes from many factors, like: only the necessary operations needed in observational processing are implemented in the software (lacking the typical transaction processing locks for instance); near optimal and implicitly parallel I/O is used; data is kept cached in memory as much as possible to avoid extra I/O; vectorized data packing is employed leading to loss-less, but yet compact data files.

ODB will solve the satellite volume problem by overlaying satellite data from different instruments or even data partitions into the memory in a transparent manner. That is to say, not all data will be kept in memory at any given time, because so called (semi-)automatic “virtual memory” will rotate the data. This part of the software is still under serious development.

4 A summary of items to be considered

A list of items *in favour* of replacing OBSPROC/CMA with ODB.

1. Next generation satellite information cannot be fully exploited without a proper observation processing and accessing system.
2. Hirlam would not be alone with ODB, but accompanied by ECMWF, Météo-France and possibly DWD.
3. Time consuming encoding/decoding of BUFR messages can be avoided as the observation data stream to HIRVDA using ODB need not to be in BUFR.
4. ODB is a suitable archiving format, but is less compact in size than BUFR.
5. ECMWF is developing BUFR-to-ODB and ODB-to-BUFR conversions, MARS online support for ODB (archiving still using BUFR), Metview/Magics graphics interface for ODB, and a proper ODB documentation and User’s Guide
6. ODB is well suited for data monitoring due to its data query language and Fortran90 connection. It can also be used in time-series studies, since data partitioning makes sure that scanning over longer periods always fits into the memory.

A list of items *supporting of delaying* the replacement of OBSPROC/CMA with ODB.

1. HIRVDA development aims with priority to an operational 3D/4D-VAR system. Achieving this goal requires concentrating the limited resources. (We do not want a mess with file formats right now, to say it frankly!)
2. Conversion from CMA to ODB is a very technical task. Experiences from ECMWF and Météo-France can be applied to HIRVDA, and thus shorten the conversion, and making the “wait”-option less hazardous.
3. OBSPROC/CMA fully supports 3D-VAR and 4D-VAR with observations sorted according to the time slots, and from this point-of-view, Hirlam 4D-VAR can go ahead.

4. ODB is a non-trivial piece of software mostly written in C-language, containing Fortran90 interface to the outside world. However, it removes most of the burden from the top level observation handling software development, since data queries retrieve the needed attributes on demand. So there is very little need to worry about explicit I/O, data packing/unpacking or even data sorting, since all this is managed “behind the scenes” for the user.
5. Operational requirements and facilities are highly variable at different Hirlam member institutes.
6. KNMI uses BUFR for pre-processing of SeaWinds scatterometer data with BUFR in- and BUFR output, and BUFR format is not a bottle-neck here since many observations are packed in one BUFR message.

5 The recommendation of the working group

The working group recommends

1. to delay the final decision of the conversion by one year, and thus ensure operational implementation of 3D-VAR, and in the meantime,
2. to gain more detailed information about the ODB, and
3. to reduce the internal dependency of HIRVDA code on the CMA file structure.