Towards a unified observations monitoring system

M. Dahoui and L. Isaksen, ECMWF

Introduction

The quality and the availability of the meteorological observing system are key elements of successful data assimilation. Not even the most advanced data assimilation algorithms works without good quality controlled observations. Therefore, the monitoring of the quality and the availability of observations is a crucial component of the data assimilation diagnostic system. Data monitoring tools are designed to produce statistics over large data samples. Statistics are generally computed for parameters that are used to evaluate the data usage (first guess and analysis departures, bias correction, observation values and data counts). Monitoring statistics are generated according to various data selection criteria reflecting the different data usage possibilities (like data assimilation flags, land surface mask, and data provider quality indicators). Monitoring results are generally exchanged between NWP centres for inter-comparison purposes. Monitoring results are also very valuable for data providers. They are using this information to improve their processing algorithms and also to adjust their strategies regarding future missions.

With the substantial increase of satellite data it is becoming necessary to have a generic monitoring tool able to process all data types and produce common monitoring results.

Towards a unified data monitoring system

Several tools are being used at ECMWF to monitor the observing system. These tools are not necessarily complementary and require considerable maintenance and support effort. To optimise the data monitoring activity it has been decided, in the context of the ECMWF's Data Handling Project, to merge features from the existing monitoring tools to end-up with a generic monitoring system able to fulfil almost monitoring needs for research and operations. Such system should:

- Optimize the maintenance and support
- Ensure the handling of all existing data types (conventional and satellite observations from all Data Assimilation systems),
- Provide an easy mechanism for the introduction of new data types,
- Provide tools for an easy handling and use of monitoring statistics,
- Contain a built-in Alarm System,
- Provide tools for observations co-location.

It has been decided to use OBSTAT software (among the other monitoring tools) to evolve as the main monitoring tool. OBSTAT will be the core of an integrated system taking advantage of the ECMWF reliable archiving and plotting packages MARS[1], Magics++ and Metview. OBSTAT will ingest data available online (for operations) or archived in MARS and will produce gridded and scatter statistics. Gridded data will be encoded in GRIB-2 format and will be archived in MARS. Scatter statistics will be stored in the ECMWF's storage system ECFS. OBSTAT outputs are intended to feed several monitoring applications. The main ones are the plotting package and the Alarm System. The plotting package will produce generic products with many possibilities for plots customization. The Alarm System will triggers warnings in case of some goes wrong with the quality and the availability of satellite and conventional observations. Although OBSTAT is being designed to be a self-contained package (useable outside ECMWF environment), it will be fully integrated within Metview. Indeed, Metview will allow the browsing and visualisation of OBSTAT outputs and will also offer the possibility to trigger the statistics calculation.

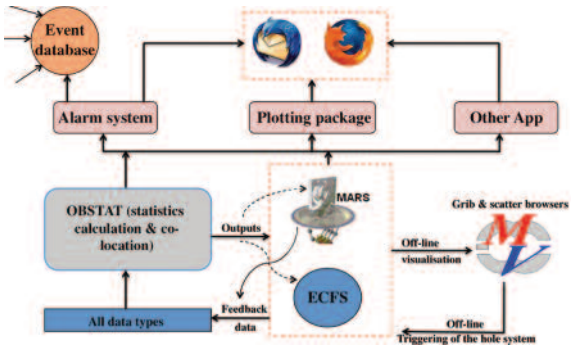

Fig. 1 Schema describing the unified monitoring system

[1]ECMWF's Meteorological Archive and Retrieval System

OBSTAT architecture

OBSTAT is a modular software generating statistics for almost all observations presented to the IFS[2] data assimilation system. OBSTAT is based on Fortran 90 and scripting languages. Almost all data specifications and plotting definitions are supplied to OBSTAT through configuration files in ASCII format, which facilitates the implementation of new data types. OBSTAT can ingest data from ODB[3] databases and from BUFR feedback files. Recently OBSTAT has been upgraded to ingest data from user-defined formats. For this, users need to customize an interface file, in ASCII format, to specify to OBSTAT the meaning of the file content. This option is particularly useful for external users. OBSTAT is able to produce statistics in four optional formats: gridded statistics in high and low resolutions encoded in GRIB-2, scatter statistics and the OBSTAT classical ASCII statistics (like profiles and bar charts of departures).

OBSTAT outputs

To be able to fulfil almost all monitoring needs efficiently by producing gridded statistics in two different resolutions: higher and coarser. High-resolution gridded statistics will be used to produce geographical maps of statistics and will be of great help for special and detailed investigations related to the use of satellite data. Low-resolution gridded statistics are to be used for the production of all kind of time series statistics. Scatter data (and associated plots) are very useful to easily investigate the statistical relationship between variables and diagnostic quantities. For many years, OBSTAT classical outputs have been of great help in the evaluation of the observing system usage.

GRIB-2 format

OBSTAT gridded statistics are currently being encoded in GRIB-2 format. The choice of this format was based on the following considerations:

- GRIB-2 is a standard WMO format which benefit from a long-term commitment.
- Offers a good compression
- Supported by ECMWF Archiving system (MARS): Archiving with indexation & retrieval with filtering. A new GRIB-2 template is being defined and MARS being extended to cater with the archiving of monitoring statistics
- Well handled by ECMWF graphical tools (Magics, Magics++ and METVIEW) and almost all other international meteorological graphical packages,
- Suitable for plotting on demand (if data are quickly accessible from highly available disks)
- Offers good possibilities for statistics exchange and inter-comparison between NWP centres and data providers
- Easily convertible to other formats (e.g. NetCDF, ASCII)

OBSTAT high-resolution gridded statistics

OBSTAT high-resolution gridded statistics are designed mainly to produce geographical maps of statistics. Statistics are produced for a large number of diagnostics and for various data selection criteria. The production of high-resolution statistics is characterized by:

- The production of the high resolution GRIB files is optional (default yes)
- Three grid types supported: REGULAR, GAUSSIAN and REDUCED GAUSSIAN
- For REGULAR GRID, Users are free to specify the resolution (default 1°x1°)
- For GAUSSIAN and REDUCED GAUSSIAN the available and supported resolutions are: N16, N24, N32, N48, N64, N80, N108, N128, N160, N200, N256, N320, N400, N512, N576, N640 and N1024,
- A bitmap is used to handle missing values and consequently reduce disk space requirements,
- Possibility to normalize statistics over a period (one GRIB produced) or keep statistics per analysis cycle

The OBSTAT plotting package is using the high-resolution to produce geographical maps of statistics. If requested, OBSTAT can produce maps of statistics differences between two data sets (generally the same data from two different experiments)

High-resolution gridded statistics can be also be used to produce data coverage charts. Finally, GRIB files can be browsed and plotted via Metview.

[2]Integrated Forecast System

[3]ECMWF Observations DataBase

OBSTAT low-resolution statistics

OBSTAT low-resolution gridded statistics are designed to produce time series of large area means. The production of the low resolution GRIBs is optional (default yes). For low-resolution gridded statistics only REGULAR grid is supported with a free specification of the resolution (default 2.5°x30°). OBSTAT offers the possibility to produce statistics according to the surface type: land, sea, sea-ice and all surface types. The products that can be derived from the low-resolution gridded statistics are:

- Large areas time series. The plot contains several panels contain several diagnostics each. The number, content and order of panels. This product is generally produced for selected areas and several data selection criteria,
- Overview plots containing large area time series for several channels (maximum 20),
- Time-latitude Hovmöller diagrams showing the time evolution of statistics according to latitude bands (by 2.5°),
- Channels-time Hovmöller diagrams showing the time evolution of statistics associated to a number of channels/levels (up to 100),
- Channels-latitude Hovmöller diagrams showing time averaged statistics of channels/levels according to latitude bands (by 2.5°).

OBSTAT scatter plots

OBSTAT has been upgraded to allow the computation and plotting of scatter statistics. The production facility of scatter statistics is highly flexible and available for all data types. The OBSTAT scatter production facility has the following characteristics:

- The production of scatter data is optional (default no)
- Up to 18 pairs of variables can be defined for comparison
- For each diagnostics, the users need to define the binning to be used for statistics gathering,
- OBSTAT offers the possibility to produce scatter statistics for different areas
- Scatter data can be produced, in one go, for different channels/levels/ layers/areas and for different data selection criteria
- Population statistics are stored in a self-documented ASCII file.

OBSTAT classical products

Since its first release, OBSTAT has been used to produce several observational statistics. Statistics generating these products are in ASCII format.

- Classical OBSTAT plots are: statistics RMS/ST.DEV plots (vertical profiles of statistics), Histograms and Usage charts
- OBSTAT offers the possibility for experiment statistics Superimposition,
- OBSTAT offers the possibility to customize the final product by gathering individual plots

OBSTAT main features

Several new functionalities have been implemented in OBSTAT. The summary of these changes is:

- OBSTAT can produce statistics for a large number of pre-defined quantities. Users have the possibility to define their own diagnostics without any change in the OBSTAT code. A configuration file (in ASCII) is available for that
- OBSTAT can produce statistics according to several pre-defined data selection criteria. Statistics related to several data criteria can be produced in one go. Users have the possibility to define their own data selection criteria without any change in the OBSTAT code (if only one channel/level/layer is involved). A configuration file (in ASCII) is available for that
- OBSTAT can handle all existing vertical coordinate: Channels, Pressure levels and pressure layers
- OBSTAT offers several possibilities for pressure layers binning:
- Standard pressure binning is defined in OBSTAT (through configuration files)
- For each data type, the user can define a different pressure binning (which has the priority over the standard one)

- OBSTAT offers the possibility to define pressure layers which are overlapped (or embedded within each other)
- OBSTAT offers the possibility to define the time binning of data (from 1 hour to Data Assimilation window length). The default time binning is the DA window length (12h for DCDA and 6 for DA)
- OBSTAT offers a slow option to browse the input database and produce generic statistics for all present data,
- The plotting is separated from the statistics calculation that require heavy data transfers. The users are advised to request the maximum statistics to be produced in one OBSTAT run. The plotting package can be applied separately to a subset of statistics.

Road map

The OBSTAT software will be developed further to ensure more flexibility, versatility, efficiency and robustness. The planned development actions are:

- Implement the archiving of the monitoring statistics in MARS
- Adapt and Plug-in the Automatic Alarm System
- Adapt and implement the existing co-location tools
- Implement OBSTAT as the main ECMWF operational monitoring software
- Implementation, by the Graphics section, of an interface to OBSTAT within METVIEW

Conclusion

A significant work has been performed towards the implementation of a comprehensive observation monitoring system based on the OBSTAT software. Further enhancements are ongoing to allow more flexibility, versatility, efficiency and portability of OBSTAT software.