

# Overview of the operational ALADIN variational data assimilation system

Sándor Kertész  
RC LACE Data Manager  
Hungarian Meteorological Service  
[kertesz.s@met.hu](mailto:kertesz.s@met.hu)

17th January, 2007

## 1 Preface

This overview is based on the ALADIN 30 model version and basically reflects the experience with the operational ALADIN data assimilation system at the Hungarian Meteorological Service (HMS).

## 2 General overview

The recent ALADIN data assimilation system comprises three main parts: **observation pre-processing**, **variational analysis** for the upper air fields (3D-VAR) and **optimal interpolation analysis** for the surface fields (CANARI, **conf. 701**). The 3D-VAR step consists of two further steps: at first the observation screening (quality control) takes place (**conf. 002**) and it is followed by the minimization step (**conf. 131**). The CANARI step is omitted in the operational practice (at Météo France and HMS) and replaced by simply taking the surface fields from the ARPEGE analysis.

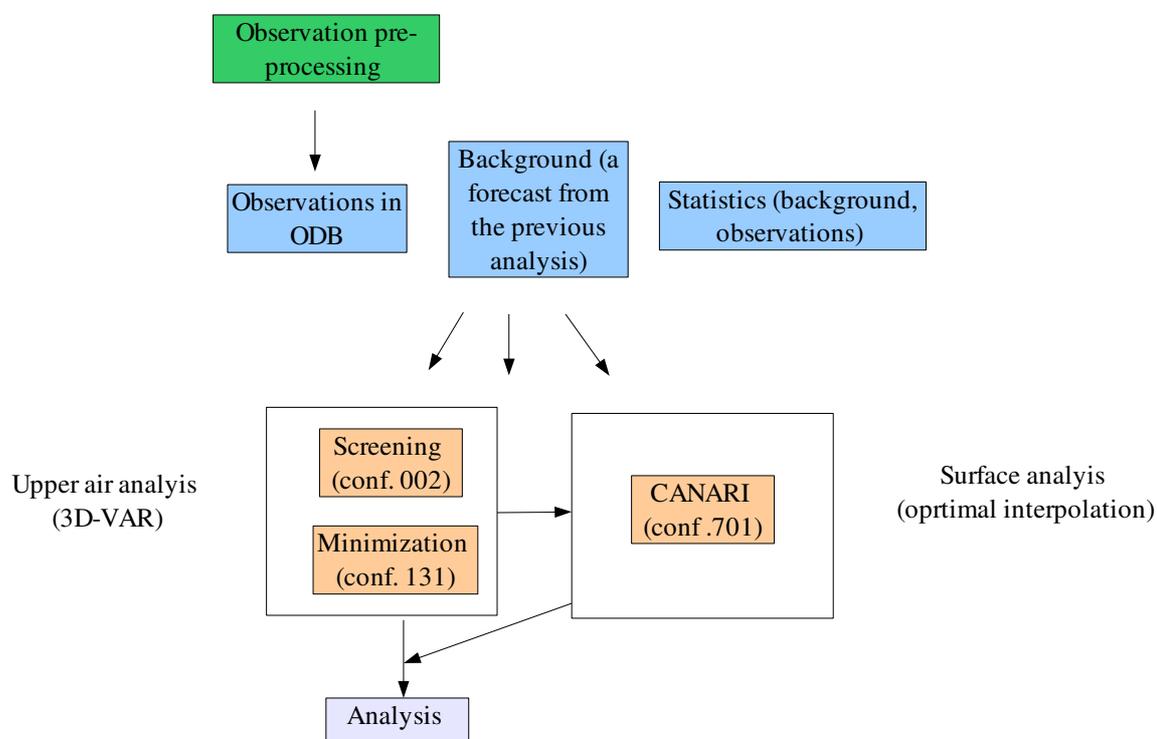


Figure 1: General overview of the ALADIN variational data assimilation system (the CANARI step is omitted in the present operational practice)

### 3 Observation pre-processing

The main goal of the observation pre-processing is to provide the analysis procedure with the observations in **ODB** format. At present all the 3 involved model configurations (002, 131, 701) require observations in ODB format. Two ODB formats are relevant to the ALADIN data assimilation: **ECMA** and **CCMA**. The CCMA format is a compressed version of ECMA. Screening and CANARI require an ECMA ODB while the minimization requires a CCMA one. Strictly speaking the observation pre-processing has to produce input ODBs only for screening and CANARI, because the input ODB for minimization is produced by the processing of the output ODB of screening.

The ECMA ODBs are produced by programme **BATOR**. BATOR also assigns the observation errors to the data and prepares blacklisting. At present, BATOR takes three kinds of input observation formats:

- **ASCII**: this is the so-called **OBSOUL** file produced by programme **OULAN** developed at MF. The original format of OULAN reads data from the local database of MF so it is not suitable for use at other sites. In order to use OULAN outside MF it has to be customized to the local observation format. At HMS we use OULAN to convert SYNOP, AIREP, SATOB (GEOWIND), TEMP, Windprofiler and NOAA AMSU-A and AMSU-B radiances to the OBSOUL format. The result of OULAN is one giant ASCII file with the above mentioned observations in the proper format.
- **BUFR**: BATOR can process AIRS, AMSU-B, QSCAT and GEOWIND data in BUFR format. This solution has never been tested at HMS so we have no experience with it.
- **GRIB**: The **SEVIRI** radiances are not put into the OBSOUL file, instead a separate pre-processing is applied that produces a set of **GRIB** files.

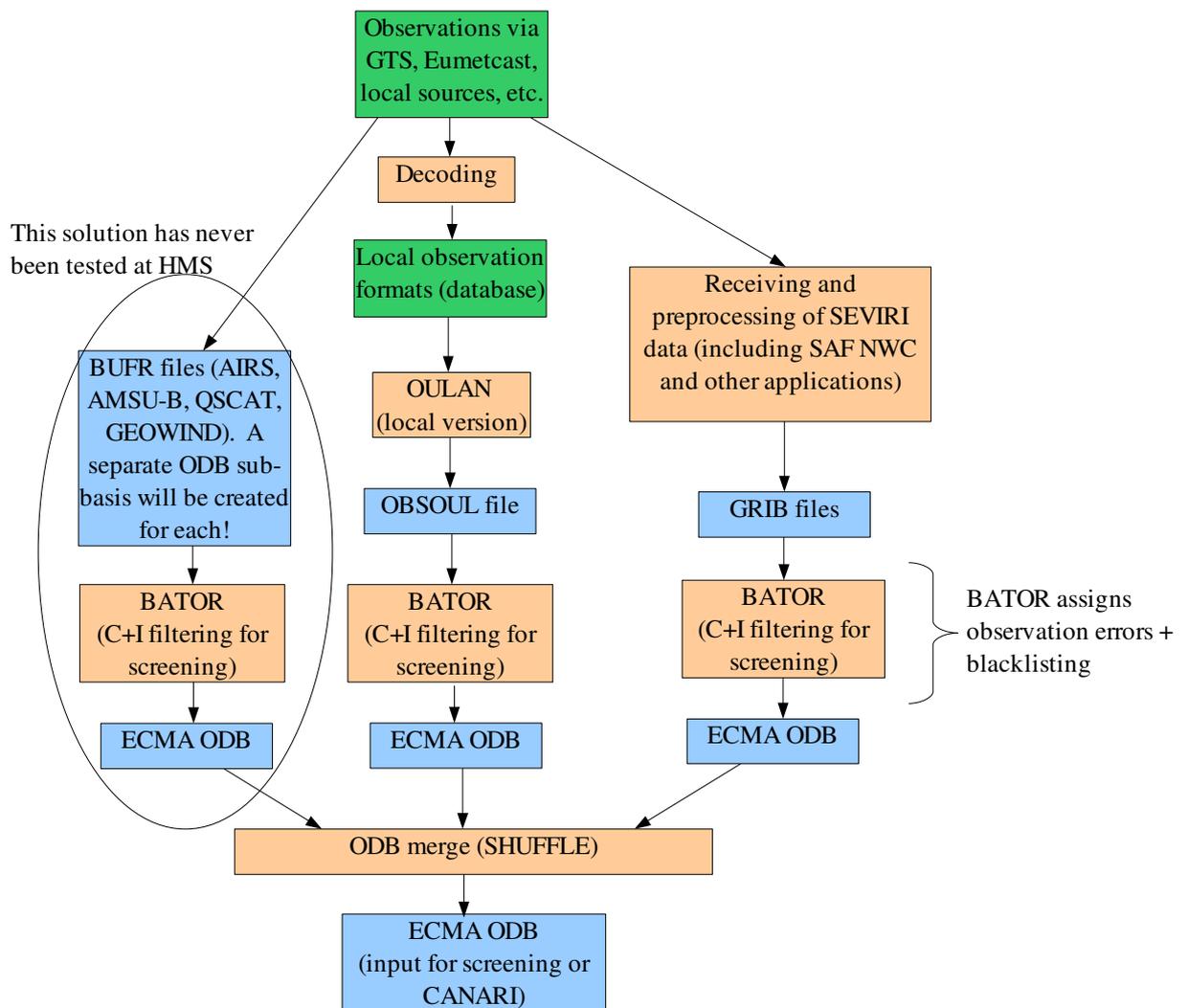


Figure 2: Data flow of the observation preprocessing

BATOR should be run for the OBSOUL file, SEVIRI GRIBs and BUFRs separately. The result is a so-called ECMA ODB sub-basis for each BATOR run. These sub-bases must be merged into full a ECMA ODB by programme SHUFFLE. The resulting ODB is now suitable for use in screening or in CANARI. The preparation of the ODB for screening requires a special care since screening is working only with an ODB that contains observation in the C+I zone. This kind of filtering can be activated in BATOR via an environmental variable and controlled through the namelist (however, this solution was not tested at HMS and we still use external programme LAMFLAG to perform the filtering). In practice it means that different ECMA ODBs are produced for screening and for CANARI.

## 4 Variational analysis

The first step in the variational analysis is the observation screening. The screening performs the quality control of observations and assigns status flags (active, passive, rejected, blacklisted) to reports and parameters. Screening takes as input:

- an ECMA ODB containing the observations only in the C+I zone
- the background (usually a 6h ALADIN forecast starting from the previous analysis)
- different statistics

The screening decisions, flags, computed departures are put into the input ODB. So there is no new ODB created by screening but the input ODB is updated. The resulting ODB is then converted to a CCMA ODB that will be the input of the minimization. This ODB conversion is done by programme SHUFFLE. During the conversion only the active data and the necessary ODB tables are kept, so the CCMA ODB is usually much smaller than the ECMA one.

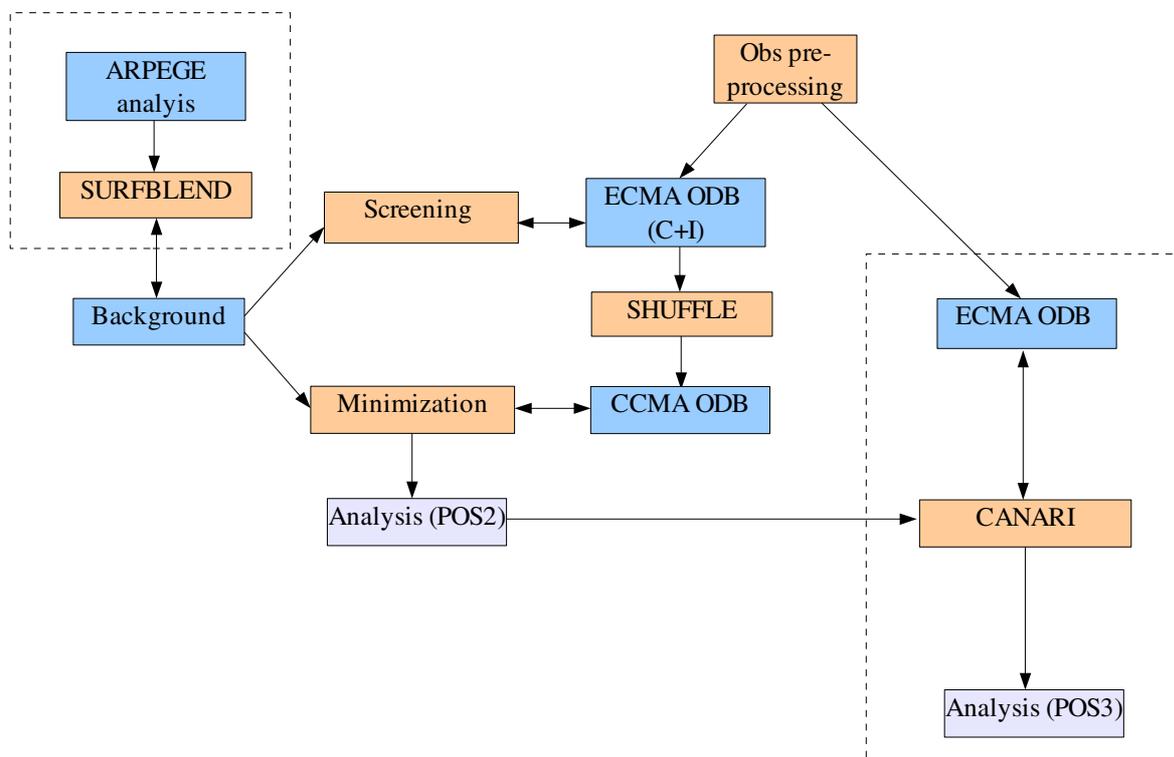


Figure 3: Data flow of the analysis process. The processes bounded by dashed line are optional: one of them should be applied

The minimization step takes the following inputs:

- CCMA ODB containing only the active data
- the background (the same as for screening)
- the B-matrix files
- different statistics

All the computed departures, analysis increments and other information are put into the input ODB. So there is no new ODB created by the minimization but the input ODB is updated. The real result of the minimization step is the so called POS2 file that now contains the analysis for the upper air fields.

If CANARI is not used this is usually the point where the replacement of the surface analysis is done by simply copying the selected surface fields from the ARPEGE analysis file into the **POS2** file. It is done by a modified version of surface blending that we call **SURFBLEND**.

## 5 Surface analysis by CANARI

The surface CANARI analysis takes the following inputs:

- an ECMA ODB containing the observations
- the first guess (it is the POS2 file produced by 3D-VAR analysis)
- different statistics

The quality control decisions (CANARI has its own quality control), flags, computed departures, surface increments etc. are put into the input ODB. So there is no new ODB created by CANARI but the input ODB is updated. The real result of CANARI is the so called **POS3** file that is now contains both the upper air analysis (by 3D-VAR) and the surface analysis (by CANARI).

## 6 Analysis cycling

The analysis procedure is usually organized into a cycle. Assuming a 6h cycle it means that the background for the next analysis is provided by a 6 hour forecast starting from the present analysis. None the less it seems to be easy it can be a rather difficult process in practice. The reason is that in an operational application when we want to run a forecast from the resulting analysis we have to consider that at the time when the analysis has to be performed:

- not all the observations are available (**cut-off time**)
- not the best possible LBCs are available for the 6h cycling forecast (cut-off time again).

It means that the analysis used for the operational (e.g. 48 h) forecast is usually based on short cut-off observations and LBCs. Then later if all the observations and the required LBCs are available the analysis is performed again with the long (or delayed) cut-off data to ensure that a more complete information could propagate towards in the analysis though cycling.

Many different cycling schemes can be applied starting from the very simple solutions to highly sophisticated ones. The question is always the scientific meaning and local computing resources should be also considered. A possible solution for the cycling arrangement can be seen in Figure 4.

## 7 Post-processing and archiving

The maintenance of a data assimilation system requires some post-processing tasks, such as observation monitoring, computation of statistics, computation of bias correction coefficients for satellite radiances etc. These tasks are usually based on the information stored in the resulting ODBs. Thus, having finished the analysis process. the ECMA and CCMA ODBs should be archived.

The input observations (in the raw format before ODB) should be also archived. It is needed because the ODB format can change with the model cycles and it is not guaranteed at all that the next model version will be able to read the present ODB format. So instead of archiving the input ODBs it is safer to keep the row observation data.

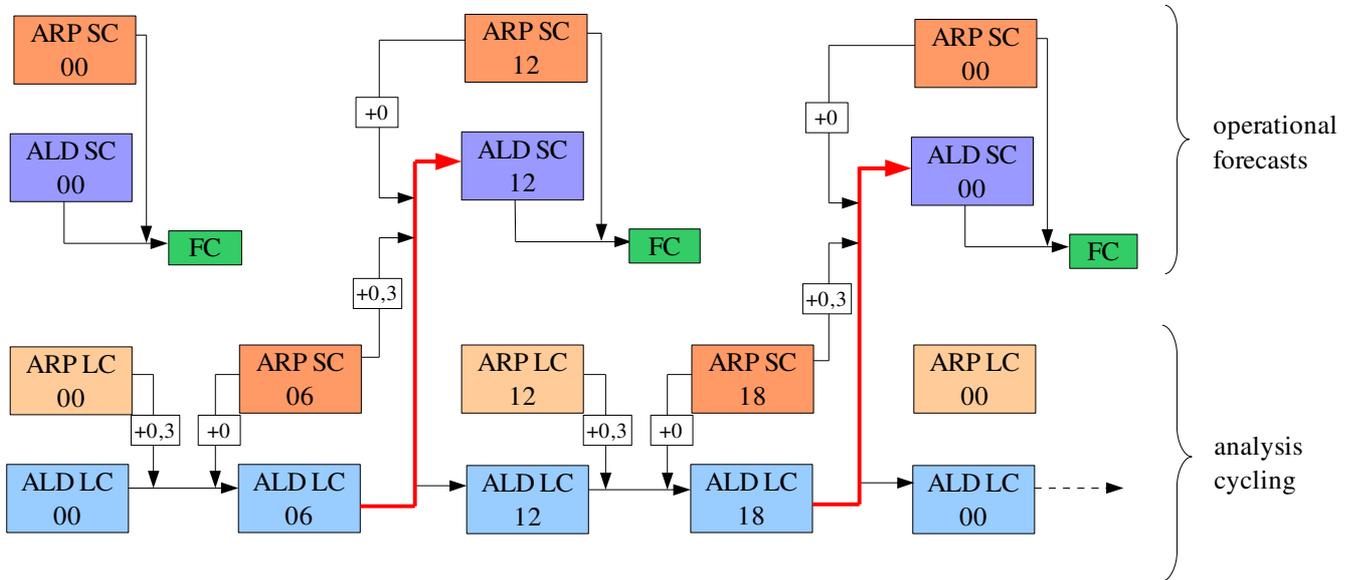


Figure 4: The 6h analysis cycling used at HMS formerly. The blue and purple rectangles denote the ALADIN analyses. The orange rectangles denote the ARPEGE LBC files used in cycling and in the production of the operational forecasts. The green rectangle denotes the 48 h operational forecasts performed two times a day (at 00 and 12 UTC). SC refers to the short cut-off and LC refers to the long cut-off ARPEGE or ALADIN analyses.