OPLACE: common observation preprocessing for LACE

Gergely Bölöni, László Szabó, Alena Trojáková

March 6, 2009

Contents

1 Introduction

As a part of the LACE data assimilation project (Bölöni, 2008), a common observation preprocessing system has been built up (OPLACE) at the Hungarian Meteorological Service (HMS), with the goal of providing all the available European observations in an appropriate format for assimilation purposes to all LACE members. The available observations are so far all data arriving to HMS through GTS and EUMETCAST. The appropriate format should be understood as any of the 3 acceptable formats (ASCII OBSOUL, GRIB and BUFR) for the BATOR program, which is the most common ARPEGE/ALADIN/AROME (AAA) software writing the observations into ODB format (Kertész, 2007). It is important to notice that a higher level of preprocessing (i.e. distribution of data in ODB format) is not possible due to the dependence of the ODB format on the IFS/AAA software cycle, which may vary among LACE members.

2 General description of OPLACE

The idea is to provide observational data in hourly time-slots with a quasi-continuous update. One time-slot is defined as the +/- 30 min. time interval of the given hour. On the top of the separation by time-slots, the data are also separated by observation types and by data format in case of existing multiple formats. Table 1 summarizes the presently available data for a given time-slot.

Observation ID	Observation type	Subtype	Variable	Format
1	SYNOP	LAND, SHIP	Ps, u, v, T, RH	OBSOUL
2	Aircraft	E-AMDAR	u, v	OBSOUL
3	AMV	GEOWIND (MSG-2)	u, v	OBSOUL
5	TEMP	LAND, SHIP	P, u, v, T, q, RH	OBSOUL
6	Wind Profiler	EUROPROFILE	u, v	OBSOUL
7	Satellite	ATOVS AMSU-A, B, HIRS (NOAA)	radiance	OBSOUL
7	Satellite	SEVIRI (MSG-2)	radiance	GRIB

Table 1: Available observation types and formats on OPLACE

The continuous update means that the data are recreated at every hour for the last few time-slots (max. 8). This update is necessary as most of the raw data arrives with a non negligible telecommunication delay. The maximum delay for each observation type was estimated based on statistics of a few days on the incoming raw data at HMS (Trojáková, 2008). According to this estimation, data are updated with the time delays shown in Table 2. The separation by time-slots and observation type gives a flexibility to the users to compose an appropriate observation window for their assimilation cycles either for surface CANARI OI, or for atmospheric 3DVAR, FGAT or 4DVAR in the future. It should b e mentioned that in the future it is expected to use more and

Observation ID	Observation type	Time delay
1	SYNOP	8 hours
2	Aircraft	8 hours
3	AMV	2 hours
5	TEMP	4 hours
6	Wind Profiler	8 hours
7	Satellite ATOVS	2 hours
7	Satellite SEVIRI	0 hours

Table 2: Time delays for updating the different observation types on OPLACE

more BUFR instead of the present OBSOUL formats. It is a already possible to read GEOWIND and ATOVS data directly in BUFR format in BATOR, however further tests are required to reproduce what we get with the present solution (Trojáková, 2008). The preprocessed observations are available via ftp for the members (see further details later on).

3 User's guide

The preprocessed observations are placed on ftp.met.hu, which is a server outside the HMS firewall. LACE members can login and download data for any time-slot and observation type from this server in the knowledge of the following details.

3.1 How to login to the server?

Members can reach ftp.met.hu via ftp (or ncftp) with a user name and password available on request from Gergely Bölöni (boloni.g@met.hu) or László Szabó (szabola@met.hu).

As an example we login from pc2039:

```
boloni@pc2039:~> ftp ftp.met.hu
Connected to ftp.met.hu.
220 FTPd (el)
Name (ftp.met.hu:boloni): lacepp
```

```
331 Please specify the password.
Password:
230 Login successful.
Remote system type is UNIX.
Using binary mode to transfer files.
ftp>
```

3.2   What are the data structure and naming conventions?

The data are placed on the server under the following directory structure:

```
{time_slot}/{observation_type}/{format}/{source}
```

where

- `time_slot` stands for the hours of the day in UTC with values `00,02,...,23`

- `observation_type` stands for the observation types as detailed earlier, with values `1,2,3,5,6,7`

- `format` stands for the data format with values `obsoul`, `grib`, `bufr`

- `source` stands for the source of the data with values `current`, `backup`. The directory `current` contains data from the most recent preprocessing run, while `backup` contains data from the run of the previous hour. As a consequence data under `backup` are less complete, which usually turns out from the file sizes also.

Under the above described directory structure the files are placed with the following naming convention:

```
{format}_{observation_type}_{subtype1}_{subtype2}_{yyyymmddhh}.gz
```

where

- `format` and `observation_type` are as described above

- `subtype1` and `subtype2` stands for the observation subtype. Presently `subtype1` is filled with characters `xxxxxx` except for the SEVIRI radiances where we fill these characters with `seviri`. So far, `subtype2` is filled with characters `xx` for all observation types and is kept for later distinctions if needed

- `yyyymmddhh` stands for the actual year, month, day, and hour

As an example, we login to ftp.met.hu and list the aircraft (E-AMDAR) and satellite (SEVIRI) data files for the date 2008.12.18 00 UTC:

```
boloni@pc2039:~> ftp ftp.met.hu
Connected to ftp.met.hu.
220 FTPd (el)
Name (ftp.met.hu:boloni): lacepp
331 Please specify the password.
Password:
230 Login successful.
Remote system type is UNIX.
Using binary mode to transfer files.
ftp>
ftp> ls 00/2/obsoul/current/*20081218*
229 Entering Extended Passive Mode (|||17302|)
```

3

```
150 Here comes the directory listing.
-rw-r--r--    1 ftp      ftp          16764 Dec 18 08:13 obsoul_2_xxxxxx_xx_2008121800.gz
226 Directory send OK.
ftp>
ftp> ls 00/7/grib/current/*20081218*
229 Entering Extended Passive Mode (|||20907|)
150 Here comes the directory listing.
-rw-r--r--    1 ftp      ftp          7049260 Dec 18 00:08 grib_7_seviri_xx_2008121800.gz
226 Directory send OK.
ftp>
```

3.3　When the data are updated and how long are they there?

OPLACE is run presently at 5 minutes past every hour and is running about 7-9 minutes. As a consequence it is recommended to always look for the fresh data under `current` about 15 minutes past every hour. In case the data did not appear under `current` on time for your assimilation run, it is recommended to fetch the data from `backup`. It is expected that in the future, when BUFR reading will appear for some data types, the runtime of OPLACE will be reduced significantly.

Under `current`, data are kept for 4 days, enabling to restart the assimilation cycle even if it was not running for a week-end, for instance. Under `backup`, indeed only the data prepared by the previous OPLACE run are kept.

The temporary storage of files on ftp.met.hu implies the very important fact that all members have to archive the preprocessed data locally for themselves.

3.4　Which data do I need?

The question, more exactly, is which observation types and for which time-slots do i have to download in order to run a meaningful data assimilation cycle? With other words again, with how large observation window (around the analysis time) should I use the different observation types? This will depend also on the actual data assimilation system I would like to use (surface CANARI, atmospheric 3DVAR/FGAT/4DVAR). Here we set up a recommendation for surface CANARI (Table 3.) and atmospheric 3DVAR (Table 4.) in case of 6-hourly assimilation cycle following the operational settings of ALADIN/HU. Using these settings the data amount to be downloaded at each analysis is about 35-40 Mbytes.

Observation ID	Observation type	Observation window
1	SYNOP	+/- 0 hours
5	TEMP	+/- 0 hours

Table 3: Recommended observation use for CANARI

Observation ID	Observation type	Observation window
1	SYNOP	+/- 0 hours
2	Aircraft	+/- 1.5 hours
3	AMV	+/- 0 hours
5	TEMP	+/- 0 hours
6	Wind Profiler	+/- 0.5 hours
7	Satellite ATOVS	+/- 3 hours
7	Satellite SEVIRI	+/- 0 hours

Table 4: Recommended observation use for 3DVAR

It is important to notice that at this stage one does not have the choice on the observed variables she or he wants to use in the assimilation. The selection of the used variables can be performed later on, on the level

of blacklisting, which is done while running BATOR or the so called screening procedure (Kertész, 2007). One can add, that in general in 3DVAR, the use of an observation window > 0 is meaningful in case of moving platforms only, such as aircrafts, AMV or polar satellites (ATOVS). Observations with fix locations will provide a time average over the observation window, which is probably less accurate than the observation taken at the analysis time exactly. For FGAT, experimentations are needed to chose optimal observation windows for each observation type. In 4DVAR, all hourly times-lots will be treated separately.

3.5   Merging time-slots and observation types

Downloaded data files should be un-zipped first. In order to feed the BATOR program the easiest way, OB-SOUL files for different time-slots and observation types should be merged into one large OBSOUL. In order to merge the OBSOUL files a tool has been placed on the LACE forum (obsoul_merge.pl), which is to be used as follows:

```
obsoul_merge.pl -obsoul {merged_obsoul_file} -files {listfile}
```

where `listfile` contains the list of OBSOUL files to be merged. As an example, we login to ftp.met.hu, get the aircraft (E-AMDAR) and TEMP data files for the date 2008.12.18 00 UTC and finally merge them into one large OBSOUL file:

```
boloni@pc2039:~> mkdir merge
boloni@pc2039:~> cd merge
boloni@pc2039:~/merge> ftp ftp.met.hu
Connected to ftp.met.hu.
220 FTPd (el)
Name (ftp.met.hu:boloni): lacepp
331 Please specify the password.
Password:
230 Login successful.
Remote system type is UNIX.
Using binary mode to transfer files.
ftp> get 00/2/obsoul/current/obsoul_2_xxxxxx_xx_2008121800.gz obsoul_2.gz
local: obsoul_2.gz remote: 00/2/obsoul/current/obsoul_2_xxxxxx_xx_2008121800.gz
229 Entering Extended Passive Mode (|||15953|)
150 Opening BINARY mode data connection for
00/2/obsoul/current/obsoul_2_xxxxxx_xx_2008121800.gz (16764 bytes).
100% |********************************| 16764         9.53 MB/s     00:00 ETA
226 File send OK.
16764 bytes received in 00:00 (430.60 KB/s)
ftp> get 00/5/obsoul/current/obsoul_5_xxxxxx_xx_2008121800.gz obsoul_5.gz
local: obsoul_5.gz remote: 00/5/obsoul/current/obsoul_5_xxxxxx_xx_2008121800.gz
229 Entering Extended Passive Mode (|||15958|)
150 Opening BINARY mode data connection for
00/5/obsoul/current/obsoul_5_xxxxxx_xx_2008121800.gz (154793 bytes).
100% |********************************|   151 KB   10.64 MB/s     00:00 ETA
226 File send OK.
154793 bytes received in 00:00 (3.94 MB/s)
ftp> by
221 Goodbye.
boloni@pc2039:~/merge> gunzip obsoul_*.gz
boloni@pc2039:~/merge> ls obsoul_2 obsoul_5 >LISTFILE
boloni@pc2039:~/merge> obsoul_merge.pl -obsoul OBSOUL -files LISTFILE
====================================================
Merge OBSOUL files
----------------------------------------------------
  -----> OBSOUL file will be OBSOUL
  -----> OBSOUL file is ready...
----------------------------------------------------
  Running time : 0 secs
boloni@pc2039:~/merge> ls -lrt
```

```
total 4455
-rw-r--r-- 1 boloni users 1309789 2008-12-18 04:12 obsoul_5
-rw-r--r-- 1 boloni users  203097 2008-12-18 08:13 obsoul_2
-rwxr-xr-x 1 boloni users    4135 2008-12-18 18:03 obsoul_merge.pl
-rw-r--r-- 1 boloni users      18 2008-12-18 18:14 LISTFILE
-rw-r--r-- 1 boloni users 1512861 2008-12-18 18:14 OBSOUL
boloni@pc2039:~/merge>
```

The large OBSOUL file provided by the merge will be an appropriate input file for BATOR (see the documentation by Kertész, 2007). Note, that the merge tool will work also for merging OBSOUL files for different time-slots, which is in fact essential in order to run the BATOR program. Also it should be mentioned that the merging of observation types is not obligatory. The BATOR program can be run in a loop for different observation types, but in this case the provided ODB will contain the different observation types as different ODB sub-bases (Kertész, 2007 or Bölöni, 2004).

4   References

Bölöni G., 2004: Technical Note on observation handling in cy28t0, RC LACE report.
(http://www.rclace.eu/File/Data_Assimilation/2004/GB_T_N_obs_handling_AL_cy28t0_2004.ps)

Bölöni G., 2008: LACE work plan on data assimilation 2008-2010
(http://www.rclace.eu/File/Data_Assimilation/plans/wp_da_2008_2010.7.doc)

Kertész S., 2007: Overview of the observation usage in the ALADIN variational data assimilation system, RC LACE report. (http://www.rclace.eu/File/Data_Assimilation/2007/lace_obspp.pdf)

Trojáková A., 2008: Report of work on common preprocessing, RC LACE report.
(http://www.rclace.eu/File/Data_Assimilation/2008/at_pplace_draft.pdf)