Summary report on DAsKIT video-conference, 26 March 2021

Participants: *Mohand Ouali Ait Meziane (Algeria), Alex Deckmyn, Idir Dehmous, Lesley De Cruz (Belgium); Andrey Bogatchev, Milan Tsankov, Konstantin Mladenov (Bulgaria); Fatima Hdidou, Zahra Sahlaoui (Morocco), Malgorzata Szczech-Gajewska, Marcin Kolonko (Poland); Maria Monteiro (Portugal); Simona Tascu (Romania), Yelis Cengiz (Turkey); Alena Trojakova (OPLACE, Czech Republic); Benedikt Strajnar (LACE DA Area Leader, Slovenia), Roger Randriamampianina (ACCORD DA Area Leader, Norway). Excused: Boryana Tsenova(Bulgaria), Haythem Belghrissi, Wafa Khalfaoui (Tunisia) excused themselves.*

The main topics of this video-conference were:
1. Tour de Table: status & progress of local implementations
2. Recent news from the ACCORD DA Area Leader (Roger Randriamampianina)
3. On-going work: status on a common basic scripts set for combDA (Maria Monteiro & Alex Deckmyn)
4. DAsKIT wiki page (Yelis Cengiz)
5. AOB

Short status per country:

ALGERIA
> data acquisition:
> Conventional: SYNOP (3-hour GTS; 1-hour GTS automatic stations, ASCII/BUFR), SYNOP SHIP (BUFR), TEMP (5 stations/day, ASCII/BUFR), BUOY (BUFR), AMDAR (hourly GTS BUFR files);
> Non-conventional: ASCAT (MetopA, MetopB/2 times intervals (8-10, 19-22 UTC) GTS BUFR), GNSS ground (hourly BUFR files).
> data pre-processing:
> SYNOP (conversion to BUFR; duplications and amends are tackled); TEMP (filtered over ALADIN domain); AMDAR (filtered over ALADIN domain; selection of template 311010); ASCAT (filtered over ALADIN domain); pre-processing of GNSS data.
> monitoring tools:
> OBSMON installation is on-going; MANDALAY (CY40T1); odbsql; Metview Bundle (2018-12.0, ecCodes, Magics, Metview, OBD_api)
> verification tools:
> HARP not installed.
> surface DA:
> AROME OI_MAIN (CY40T1_bf07) is being cycled under test mode with GTS SYNOP data (first results available for a 20-day cycling period), but without blendsur.
> upper-air DA:
> BATOR (CY40T1_bf07): SYNOP, TEMP, AMDAR;
> BATOR (CY43T2_bf10): tested for SYNOP, TEMP, AMDAR, ASCAT, GPS ground;
> ALADIN 3D-Var (CY43T2, local HPC, tested in belenos) is under pre-operational mode (at 00,06,12,18UTC, 6-hours cycling; B-matrix for 6 km);
> AROME 3D-Var (CY43T2, local HPC, tested in belenos) is under pre-operational mode (AROME B-matrix has been computed from AEARP downscaling with an hybrid cycle (CY40T1_bf07 for BATOR and CT43T2_bf10 for 001)).
> Plans: build ALADIN 3D-Var and AROME 3D-Var in belenos.
> combined surface+ upper-air DA:
> Plans: building a pre-operational version of 3D-Var cycle, combined with OI_MAIN, for AROME at CY43T2.

BELGIUM

data acquisition:
Conventional: SYNOP, TEMP, AMDAR from GTS (BUFR);
Non-conventional: ground-based GNSS (E-GVAP samples from Royal Observatory of Brussels, ROB/GTS, BUFR or COST716, BUFR), MODE-S EHS (KNMI, BUFR), radar data (RMI, HDF5).

data pre-processing:
Python script (PoP-RMI) that deals with duplications and amends; SAPP (ECMWF) server is installed; GPS data is converted from COST format to GTS BUFR with an UK tool (met Office) from E-GVAP. In a first approach, statistics on the observation error had to be determined, in particular, mean bias and error deviation, which are used when varBC method is not in use. However, in this case a 'whitelist' has to be used to enable this observation type.
Plans: configuration of SAPP to local needs.

monitoring tools:
OBSMON is technically working (see 2020 LACE DAWD & DAsKIT WD for details); MANDALAY installation (CY43T2_bf10), odbsql.

verification tools:
HARP.

surface DA:
3-hour cycling of an ecflow suite for surface DA (OI_MAIN, AROME 1.3, CY43T2_bf10) was set in operational mode by switching off CANOPY;

upper-air DA:
computing B-matrix for ALARO 4.0 and AROME 1.3 by the EDA method.

combined surface+ upper-air DA:
3-hour cycling of 3D-Var+OI_MAIN (AROME1.3, CY43T2_bf10) is under test;
3-hour cycling of 3D-Var+CANARI (ALARO4.0 under test, ALARO1.3 on-going, CY43T2_bf10).
Plans: validation of MODE-S and OPERA-ODIM radar; increase cycling from 3-hours to 1-hour in order to build Rapid Update Refreshing for nowcasting purposes.

operational systems:
CY43T2 by surface DA (coupling with ARPEGE).

reported issues:
Canari crashes when the radar data are added into the ODB (ADDVIEWDB("ecset" : db="ECMA") –- 'You cannot put data back into a different database from the one you got');
All DBZH data are rejected in AROME screening

BULGARIA

data acquisition:
Conventional: SYNOP, TEMP from GTS (BUFR), local SYNOP (converted to BUFR).

data pre-processing:
new tools to split compressed BUFR file into single BUFR file have been developed in-doors using eccodes; duplications are then removed using the same tools.

monitoring tools:
OBSMON, MANDALAY ported locally.

verification tools:
HARP ported and some tests just started; local surface verification tool.

surface DA:

BATOR (CY43T2_bf10): SYNOP (locally); AMDAR (tests in beaufix);
OI_MAIN (CY40T1_bf07) for AROME-BG was migrated from beaufix (Météo-France HPC platform) and cycled for 2 weeks with BUFR data and ODB validation; a newcomer has arrived and efforts are being put onto its training;
Plans:migration of surfDA for AROME-BG to the new machine using local SYNOP.
<u>operational systems:</u>
CY43T2 by dynamical adaptation is running in operational suite since November 2020:
ALADIN (105L/5km/72-48h/00,06,12,18UTC runs) and
AROME(60L/2.5km/36h/00,06,12,18UTC runs).
Plans: To restart the work on DA.

MOROCCO
<u>data acquisition:</u>
Conventional: GTS BUFR SYNOP (local and automatic), TEMP and AMDAR from GTS (BUFR);
Non-conventional: local GPS, ATOVS (BUFR).
<u>data pre-processing:</u>
GTS and local BUFR file for conventional observations and GPS data.
<u>monitoring tools:</u>
OBSMON installation: on-going in the local machine; MANDALAY on the local machine.
<u>verification tools:</u>
HARP not yet.
<u>combined surface+upper-air DA:</u>
OI_MAIN plus 3D-Var are now being cycled in the new machine (CY43T2, EDA B-matrix, SYNOP observations (GTS+automatic), 3-hour cycling, January-February testing period).
Plans: testing of new observations: TEMP, AMDAR, GPS, satellite; validation of the combined DA (CY43T2) and pre-operational implementation; implementation of OBSMON on the new machine.

POLAND
<u>data acquisition:</u>
OPLACE data is used; local SYNOP is available.
<u>data pre-processing:</u>
conversion of local SYNOP to BUFR.
<u>monitoring tools:</u>
OBSMON installed and tested with DAsKIT WD data.
<u>verification tools:</u>
HARP-v2 runs for DA cycle but not in use.
<u>surface DA:</u>
6-hour testing cycling of a surface DA based on CANARI (not SURFEX) for ALARO (CY43T2_bf10) with OPLACE data, new LBC 4 km; 66-hour forecasts, introduction of new precipitation types; AROME CY43T2 under test.
Plans: to implement CANARI cycling for ALARO on the new super-computer first and then with AROME.
<u>operational systems:</u>
ALARO CY43T2_bf10 (newLBC, 70L, 2.5km; 4x a day, up to 72h);
AROME CY43T2_bf10 (70L, 2.5km; 4x a day; up to 30h).

PORTUGAL
<u>data acquisition:</u>
Conventional: SYNOP, TEMP, AMDAR from GTS (BUFR);

Non-conventional: OIFS radar data.

data pre-processing:
local handling of duplications and amends (FORTRAN): SYNOP and TEMP (missing validation); following local implementation of SAPP (SYNOP ECMWF BUFR data).

monitoring tools:
home-made (metview plotting for SYNOP); local OBSMON_V3.3.2 (shiny part) and MANDALAY (CY40T1_bf07) implementation is on-going.

verification tools:
local (home-made IPRODS-IVERIF) surface verification tool; implementation of pre-MONITOR@ECMWF and Webgraf@ECMWF with the support of Slovenia.
Plans: HARP not implemented.

surface DA:
3-hour cycling of a standalone surface DA scheme (CANARI_OI_MAIN, CY40T1_bf07, AROME, 60L, 2.5km, locally; CANARI_OI_MAIN, CY43T2_bf10@ECMWF); on-going validation of 48-hour forecasts of AROME-PT2 (CY43T2_bf10, 60L) initialised by surface DA using as reference the same AROME-PT2 model configuration, initialised by dynamical adaptation for the two periods: WINTER: 10dez2018-10fev2019 (cold and rainy period); SUMMER: 01ago2018-09set2018 (extreme temperatures).

combined surface+ upper-air DA:
BATOR testing (CY43T2_bf10, ported to ECMWF): SYNOP, TEMP, AMDAR;
combined DA (OI_MAIN+3D-Var, SYNOP, B-matrix computed by AEARP downscaling and tested over Météo-France platforms for AROME-PT2 during a 20-day period on a combDA experiment (CY42T2, AROME DA VarBC) , with conventional and Iberian radars), has been cycled @ECMWF for a Summer testing period with an adaptation of Slovenia (ARSO) scripts (ported to ecgate); HOOF tool was added for pre-processing; pre-'monitor' verification tool has been plugged-in; other local diagnostic tools (metview based) have been migrated to ECMWF platforms too.

operational systems:
dynamical adaptation of AROME-PT2 (CY40T1_bf07, 60L, 2.5km, locally; CY43T2_bf10 has been successfully implemented into pre-operational model twice a day);
Plans: on-going validation of combDA @ECMWF with conventional as well as OIFS HDF5 radar data from Iberia.

reported issues:
screening and minimization is passing all Portuguese radar data but Jo/n values are not compliant with those obtained in Météo_france platforms.

ROMANIA

data acquisition & pre-processing:
OPLACE data is in use; obsoul_merge.pl deals with duplications.

monitoring tools:
none;

verification tools:
MET from NCARS; LAEF-verification tool adapted for deterministic purposes.

combined surface + upper-air DA:
BATOR (CY40) was tested for conventional data;
CANARI+3D-Var (CY40).

TUNISIA

data acquisition & pre-processing:
OPLACE;
Conventional: local SYNOP; TEMP and WIND PROFILER;

POP-RMI is used to pre-process conventional data.
<u>monitoring tools:</u>
OBSMON and MANDALAY implemented on the local machine.
<u>verification tools:</u>
HARP installed; its test should be started during the first half of 2021.
<u>surface DA:</u>
BATOR (CY43T2_bf10, new HPC): SYNOP;
surface DA (OI_MAIN) has been implemented in beaufix but not yet on the local machine;
Plan: move to CY43T2_bf10 in the new HPC.
<u>upper-air DA:</u>
6-hour DA cycling (ALADIN, AROME).
<u>combined surface+ upper-air DA:</u>
B-matrix has been computed by the ensemble method and tested in beaufix for AROME;
On-going implementation of a combined surface + 3D-Var DA, with a Jk component on the new HPC platform: V-matrix computation; namelist tuning; code modifications (adapting Jk ALADIN existing routines for AROME).
Plans: cycling of AROME 3D-Var on the new HPC; Jk validation blending to overcome the "sparseness" of observations on Tunisian domain; Enhance Local Data Base Observation for DA; use of more observations: AMDAR, loca GNSS.

TURKEY
    <u>data acquisition:</u>
Conventional observations: SYNOP, AIREP and TEMP from GTS (BUFR), local SYNOP (conversion to BUFR), SHIP;
Non-conventional observations: AMSUA (Metop & NOAA), MHS (METOP & NOAA) and AMV (METEOSAT).
<u>data pre-processing:</u>
SAPP BUFR data (SYNOP) is being used in BATOR (see 2020 LACE DAWD & DAsKIT WD for details).
<u>monitoring tools:</u>
OBSMON has been installed on desktop and tested with provided observations (shiny part is OK; SqLite part not installed); OBSMON is not installed on supercomputer yet: a python script has been created to visualize MANDALAY output.
<u>verification tools:</u>
Currently HARMONIE verification is used.
Plan: HARP implementation.
<u>surface DA:</u>
BATOR (CY43T2_bf10): tested with local SYNOP observations;
AROME-TR (cy43t2_bf10, 72 levels, 1.7 km 3 hourly surf assim cycling with SAPP Synop data, coupling with ECMWF-IFS) is running in the pre-operational mode; CANARI namelist settings: LMESCAN=.TRUE., REF_A_H2=100000.,REF_A_T2=100000., REF_S_H2=0.1, REF_S_T2=1.6,REF_S_SST=0.8, REF_A_SST=200000., REF_A_SN=50000., REF_S_SN=5., LAECHK=.FALSE.,
<u>combined surface + upper-air DA:</u>
B-matrix has been calculated from AEARP at CY43T2 by the ensemble method for ALARO-Tr;
Setting-up of a combined surface (OI_MAIN) + 3D-Var DA to AROME-Turkey is now on-going: to compute a new B-matrix for the AROME combined solution with couplings from ECMWF is starting using ZAMG (Austria) set of scripts;
DAsKIT wiki was created during the DAsKIT programme and will be maintained under the scope of the ACCORD WP DA8.

<u>operational systems:</u>
CY43T2_bf10 is operational since September 2019 for dynamical adaptation of AROME (72L, 1.7km and 48-h lead time, ECMWF-IFS coupling);
CY40T1_bf07 is operational for ALARO (60L, 4.5km and 72-h lead time);
6-hour combined surface + upper-air DA cycling for ALARO CY40T1 was stopped due to lack of computing power (at 00, 06, 12, 18UTC network times), at 4.5km, 60 levels and with LBC from ARPEGE
<u>reported issues:</u>
when coupling AROME-Tr with an analysis obtained by surface DA, the model dries too much.

Acknowledgments: to Alena Trojakova (CHMI, LACE), Benedikt Strajnar (ARSO, Slovenia) and Roger Randriamampianina (ACCORD DA Area Leader) for the enriched discussion and hints.

Main conclusions:

1. By the end of 2020, along the creation of the new ACCORD consortium, the DAsKIT programme was transformed into the ACCORD RWP Working Package DA8.

2. The goal of this working package is to build ACCORD DA capacities through the coordination of a common effort and support to the local implementation of DA procedures.

3. 10 countries are now part of the group: Algeria, Belgium, Bulgaria, Estonia, Morocco, Poland, Portugal, Romania, Tunisia, Turkey.

4. As expressed by the new ACCORD DA Area Leader, Roger Randriamampianina, from the ACCORD point of view, "all the participants on the ACCORD DA activities are at the same level although some difficulties may appear due to different local constraints and facilities available. Therefore, one should be able to work together; however the working mechanism that will allow it is still being discussed by the Management Group (MG), but already some ideas are in the air. For instance, at least two ACCORD topics have been identified which can immediately involve DAsKIT countries and which can, at the same time, support the increase of local capacities. Those are: 1) contribute directly to OBSMON development; 2) contribute in development of the verification tool which allows to use all available observations in DA, including non-conventional ones like, for example, radiances. So that it will be possible to perform the verification against all types of observations (Roger and Eoin Whelan have the very preliminary version of this tool). Roger invited the group to contribute to the preparation for the EUMETSAT next generation observations. The latter depends on the "readiness"of the available local DA system. For those who are members of EUMETSAT, financial support might be available for a short skills development awards visit (from few weeks to few months) between ACCORD centres.

5. In the first quarter of 2021, all countries from the former programme are working with the CY43T2; ideally, newcomer Romania, should catch up the cycle version;

6. from those countries (already at CY43T2), at least three are cycling a combined solution of CANARI_OI_MAIN + 3D-Var DA: Belgium, Morocco and Portugal (the last one @ECMWF). Besides,

7. concerning the implementation of the DAsKIT set for surface it has entered into operations in Belgium and pre-operational in Turkey, after some tuning of the OI/CANARI structure functions (MESCAN against default). However, according to the results shown during this meeting, special attention should be paid to Turkey settings, since scores have shown the initialization by surface DA dries too much the model along the day, along the period under analysis (March 2021).

8. At this moment, most of the countries are still focusing their efforts on the local implementation of a 3D-Var algorithm or a combined algorithm of surface+3D-Var, namely: Algeria, Belgium, Morocco, Portugal, Tunisia and Turkey. This implementation is done by porting a previous set (Algeria, Turkey), by building the scheme from scratch (Belgium), or by migrating it from a reference environment (Morocco, Portugal, Tunisia). Bulgaria and Poland are still blocked due to lack of computer power.

9. Moreover, countries start to invest on the re-computation of their existing B-matrices (Belgium, to move from an NMC B-matrix to an EDA B-matrix, for instance); Turkey by adapting a set of scripts from Austria to this computation.

10. The second main concern of DAsKIT countries is the handling of locally available data. Main observation types are: SYNOP, TEMP, AMDAR and WindProfilers (Tunisia) under BUFR format from GTS or from local networks. However, some countries are now concerned with the handling of non-conventional data (for instance, ASCAT, GNSS (Algeria, Morocco); MODE-S (Belgium); radar volumetric data (Belgium, Portugal)).

11. So far, countries have created pre-processing tools based on eccodes (ECMWF) to handle BUFR data, as it is the example of Belgium and Bulgaria. Pre-processing consists of removing duplicates from corrections and amends and filtering a certain type of BUFR template over a particular geographical area. Pre-processing is applied to SYNOP, TEMP and AMDAR, but removing duplicates on TEMP is a process missing to the most part of the DAsKIT countries.

12. in particular, Tunisia (as a non-ECMWF member country) has successfully implemented and tested the Belgium pre-processing PYTHON tool (POP-RMI), being user friendly and simple to install; while

13. Turkey is successfully using SYNOP BUFR data coming from SAPP on the local (CY43T2-export, modified by HIRLAM modset from Eoin Whelan) BATOR version.

14. MANDALAY (CY40T1 or CY43T2) has been implemented and tested with demo data in almost all countries (still on-going for Portugal) and no issues have been reported so far. Turkey has their own set of scripts to do some data monitoring with MANDALAY;

15. OBSMON has been implemented and tested in almost all the countries with demo data (still on-going for Algeria and Portugal), and Belgium has started to use it on a regular basis;

16. HARP has still not been implemented in most of the countries and more countries start to invest on MONITOR (HARMONIE system).

17. DAsKIT group has now an upgraded version of the original set of scripts (testbed over the Iberian domain, @CY43T2_bf11) in Météo-France platforms for both: surface DA (by the CANARI_OI_MAIN method) and combined DA (CANARI_OI_MAIN + 3D-Var).

18. A recent inquire and discussion on the shared scripts' form has revealed the majority of the countries are now willing to move to a solution based on the ecFfow scheduler (see detailed report on
http://www.umr-cnrm.fr/aladin/IMG/pdf/daskit_setup_scripts_setting.pdf).

19. In this way, Alex Deckmyn and Maria Monteiro started an analysis on the possible sharing of the DAsKIT scripts set under ecFlow environment, by comparing two different scripting systems: 'NodeRunner' (available from RMI) and 'createsuites' (available from ARSO and adapted to AROME DA @ECMWF platforms). From this on-going work it was seen that by reorganising and cleaning some parts of NodeRunner and, by adopting universally created python objects with specific DA tasks (like CANARI mapping) - a concept which is behind ARSO scripts - it could be possible to share at least part of some ecFlow tool not only among DAsKIT, but also among some countries inside LACE. This aspect would minimise the maintenance effort and would optimise the sharing, without losing other properties of the tool related to its portability and easy handling and flexibility either in operations or research modes. It was stressed the tool is not meant to share the namelists which should be locally organized according to local expertise and needs, starting from some reference environment like Météo-France in opposition to what is done in HARMONIE, where the namelists are automatically generated; this tool is just meant to share the DA workflow under the ecFlow environment. A first version/prototype of such a tool may be ready for sharing in the next month.

20. Finally, a presentation of the DAsKIT wiki web site prepared and maintained by TSMS http://212.175.180.89/daskit/index.php/ , was performed by Yelis Cengiz: this tool was created under the scope of DAsKIT with the purpose to keep track of its achievements under an easy-friendly and organized way. Is therefore complementary to a forum mechanism, like the one LACE is maintaining. The tool will be shared only among ACCORD participants. As kick-off editors, DAsKIT participants from the former programme, were invited to edit different sections, taking into account the tasks they had along the programme. This wiki should be used, for instance, to provide clear guidelines on how to get the different types of data, since now many countries try to get new sources of data. Participants were invited to be really active in this way of sharing the information, since the more active, the better they can help the newcomers and help each other.

Perspectives and recommendations:

1. DAsKIT participants are invited to register at DAsKIT wiki page by sending an email to ycengiz@mgm.gov.tr where should write their "name, surname and institution" as the mail content and by writing "DAsKIT wiki login" in the mail subject;

2. Alex Deckmyn is invited to finish and deliver a new version of NodeRunner as soon as possible, so that DAsKIT countries may try an ecFlow set of scripts to enshort their local efforts onto a way to keep up-to-date the local DA workflow.

3. A task force onto the extension of NodeRunner to B-matrix computation is now a possibility. In fact some working momentum already exists: Yelis is now trying to adapt a set of scripts at ECMWF starting from ZAMG scripts; Ouali is willing to migrate those scripts afterwards to Metéo-France platforms; Idir is willing to recompute Belgium B-matrices under EDA method. During the meeting discussion it was suggested also that

Pierre Brousseau could be contacted in order to provide the status on the externalization of their vortex python scripts for B-matrix computation.

4. Concerning the issue found on Turkey surface DA, it was suggested the use of ECMWF blacklist (Roger offered to provide the actual ECMWF blacklist) and the assessment of quality of the observations in use. A propos, it was said Patrick Samuelsson, the actual surface Area Leader should eventually participate in the DAsKIT meetings.

5. To get acquainted with the tool latest developments, was also a recommendation from the DA Area Leader and the following message by Paulo Medeiros (OBSMON developper, SHMI) was shared:

"Dear Colleagues,

The Obsmon plotting tool is now open-source and [available on GitHub](). Please visit <https://github.com/Hirlam/obsmon> for more information and to download the code.

If you use Obsmon and wish to suggest new features, report a bug or ask for support, please browse the "issues" tab in the GitHub interface and, if your request doesn't match any of the existing ones, feel free to open a new one.

To receive notifications when the code repository is updated, please click in "Watch" in the GitHub interface.

If you wish to collaborate in the development and maintenance of the code, you are of course more than welcome.

N.B.: The previously used repository <git@hirlam.org:Obsmon> will no longer receive updates.

Best wishes,
Paulo."

6. Finally, Ouali shared the following information about GPSSOLMETHOD: the "key" controls the way the selection of gps observations values is done. In case of mean it calculates the mean of the values of the station that exist in the timeslot. You have the possibility to set the key to "CENT", meaning centrally that it will only take one value (nearest in the timeslot). The default value is null i.e not selection of gps observations. In fact, the "key" controls the value which is used by screening (if "CENT", then it picks up the closest observation on the time window; is "mean" option is set, then an averaged value of that station observation is used, inside the time window. The default value is "null" which means no selection.