

Summary report on DAsKIT (ACCORD) video-conference, 26 January 2021

Participants: *Mohand Ouali Ait Meziane (Algeria), Alex Deckmyn, Idir Dehmous (Belgium); Andrey Bogatchev, Boryana Tsenova, Milan Tsankov, Konstantin Mladenov (Bulgaria); Fatima Hdidou, Zahra Sahlaoui (Morocco), Malgorzata Szczech-Gajewska, Marcin Kolonko (Poland); Maria Monteiro (Portugal); Alina Dumitru (Romania), Haythen Belghrissi, Wafa Khalfaoui (Tunisia); Yelis Cengiz (Turkey); Alena Trojakova (OPLACE, Czech Republic); Claude Fischer (ACCORD PM, France).*

The main topics of this video-conference were:

1. 'Tour de Table': which are your actual and short-term local needs concerning the DAsKIT setup scripts ? (ecFlow query results)
2. What and How
3. Suite design: 'assim' vs, 'prod' (Alena Trojakova)
4. Tentative roadmap
5. AOB

Motivation:

According to task 8.5 of the working package DA8 of the ACCORD RWP 2021, " the cycling in assimilation is generally arranged in a scripting system. For this the HARMONIE scripting or a part of it may be used, but also simpler cycling scripts used with LACE"; the deliverable for this task consists of scientific reports. The opportunity to organise this meeting was reinforced by the fact that for many countries the availability of a shared set of scripts, already for code version CY43T2_bf10, would boost their actual progress on implementing a local DA cycling.

Terminology:

DAsKIT basic scripts set - set of scripts which describe, in a clear way, the DA workflow for a particular DA algorithm; it should be suitable to be implemented locally by any country (and further developed locally to fulfill the local operational requirements), written under very low technical requirements. It can describe the analysis and assimilation algorithm for one network, for instance. It can be materialised as a 'Korn (or bash) shell set of scripts' or some more sophisticated tool like an 'ecFlow set of scripts'.

Tour de Table: actual and short-term local needs for DAsKIT setup scripts

ALGERIA

A testing cycling suite at CY43T2 should be set up in belenos (Météo-France computing platform) with bash shell scripts, close to what they have locally, since the local HPC infrastructure is not functional at the moment; near-future plans include the move to ecFlow.

BELGIUM

Operational & development AROME DA suites (surface and combined solutions, respectively) have been designed under ecFlow (version 4), although they still keep their old bash shell scripts (for '001' and 'e927' configurations); near-future plans include the migration of actual ecFlow suites to ECMWF platforms.

BULGARIA

A testing suite has been assembled for surface DA cycling; near-future plans include the move to ecFlow on the new HPC.

MOROCCO

Testing and pre-operational cycling suites have been designed for CY41T1 under shell scripts but an issue was encountered for OI_MAIN; near-future plans include the move to CY43T2, preferably with ecFlow.

POLAND

Operational & development scripts (including DA developments) is done with local bash shell scripts: local "nwpmaker" system handles features of ecFlow with warning system of problems, delays and failures; near-future plans do not include migration to ecFlow.

PORTUGAL

Testing suites at CY43T2, for dynamical adaptation, surface DA and combined DA cycling, have been designed and adapted from LACE (ARSO) "createsuites(ecFlow)" python tool at ECMWF, since the local machine has been declared as obsolete; near-future include further cleaning and building of a pre-operational cycling at ECMWF based on the on-going work.

ROMANIA

Planned to start to work on a DA cycling in the near future.

TUNISIA

Testing a 3-hour combined DA cycling using shell scripts + old SMS/ecFlow suite from ARSO.

TURKEY

Operational scripts under ecFlow environment but for experiments Bash scripts are used; plans include to move to ecFlow at ECMWF and the computation of a B-matrix by ensemble technique at ECMWF.

Main conclusions:

1. Overall, it would be useful if the workflow of the combined baseline DA solution, at each new code version, could be commonly shared as a basic scripts set, to be implemented according to the local needs and constraints, since it would save time and boost local DA teams progress.
2. At this stage, the adoption of a common sharing platform urges because most of the countries are now preparing the local implementation of their previously tested DA systems under a cycling (pre-operational) mode.
3. Most of the countries are already using or planning to implement in the short term, the scheduler ecFlow from ECMWF to run their local cycling suites; however, a few teams (Poland, for instance) rely on their fully validated local scheduler systems.
4. ecFlow scripts seem to fulfill the actual needs¹ to commonly share the combined DA solution workflow. In fact, as illustrated in a crude way in Figure 1, one can see it should be possible to recreate local Korn shell scripts from an original ecFlow.ecf scripts set.

¹ See slide 2 of presentation

https://docs.google.com/presentation/d/135gey2l8RWaKchbIH9RslpkGgh3MFyXg6yO9WK59elo/edit#slide=id.gb7d3745abe_0_0

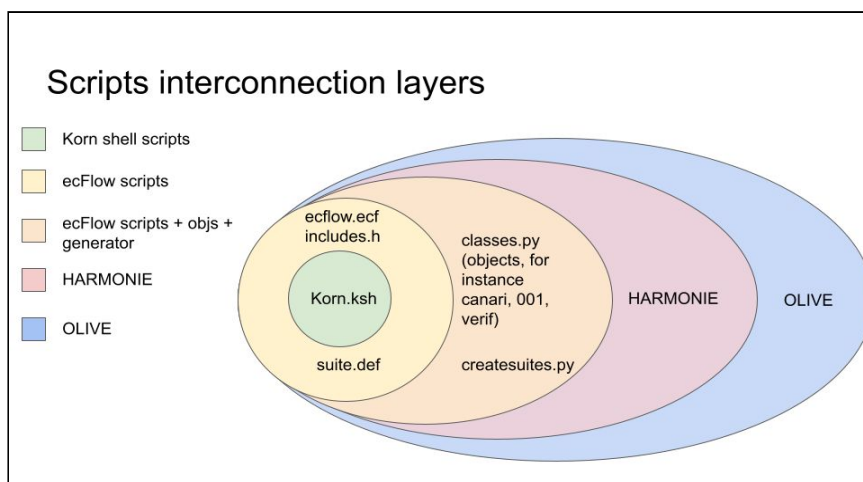


Figure 1 - Possible DA workflow platforms and its contents interconnection.

5. The latest observation together with the fact that some countries are already investing on ecFlow scripts for the same purpose, under an effort that can be shared with some LACE countries (Slovenia, for instance), easily leads us to believe an ecFlow script's set could be a useful platform for a common sharing of the combined DA workflow.
6. One should note however, that although ecFlow is available from ECMWF and Météo-France computing platforms, it may not be possible to use its visual interface remotely. This is the reason why ecFlow is less attractive if used remotely on Météo-France's platforms. And one should note, for instance, that some teams (like Algeria, which is not an ECMWF member state) will continue to rely on Météo-France's platforms while waiting for a solution for their local HPC infrastructures.
7. Moreover, it was mentioned that the local usefulness of ecFlow depends on the ability of local teams to install it and configure it on its client-server facility (between their front-end and HPC machines).
8. The local DA cycling suite design should take into account on: the availability and timeliness of global LBCs, number of available 'real-time' observations, and local users constraints for forecast time delivery, being the 'cut-off' time of a DA cycling suite understood as "how long it waits after nominal analysis time", e.g. when the 12UTC analysis starts at 14:30UTC, the cut-off is of 2:30 hours.
9. It was seen the global model produces an analysis by two ways: under an 'assimilation' cycling mode, where the cut-off is said 'long' to wait for a higher amount of observations and the model is just executed to produce a first guess to the next analysis; and a 'production' mode, which has a 'short' cut-off and takes advantage of the first guess available from the 'assimilation' cycling to initialise and produce a long model run.
10. It was also seen that, for AROME-France (since its 3-hour cycling design), the 'assimilation' mode is no longer separated from the 'production' mode, while these modes are still separated among the LACE DA systems. One could argue the added value to the local forecast of separating modes due to the availability of more observations however, since a single assimilation performance, on a 'production' mode suite, can still be a good solution in terms of forecast quality with the advantage that simplifies the local suite's design. In this way, countries are encouraged to check their local availability & timeliness of observations and the local constraints for the forecast delivery to make an estimate of the need for the separated 'assimilation' and 'production' modes.
11. Another aspect to be considered is the Incremental Analysis Update (IAU)². "The IAU process incorporates analysis increments into a model integration in a gradual manner. It

² [https://doi.org/10.1175/1520-0493\(1996\)124<1256:DAUIAU>2.0.CO;2](https://doi.org/10.1175/1520-0493(1996)124<1256:DAUIAU>2.0.CO;2)

does this by using analysis increments as constant forcings in a model's prognostic equations over a 3-h period (for instance) centered on an analysis time. Note that Météo-France is not using IAU in its original definition but in a special manner to enhance quality of the production forecast via using the next analysis which is already available, e.g. 00UTC production starts from 00UTC analysis and it is using also 01UTC one via IAU. However, it was discussed that for the time being this will not be considered to make the local implementation task easier. In this way, local cycling design should start with a 3-hour simple cycling.

12. As foreseen activities with the commonly shared set of ecFlow scripts, it was mentioned also the group should explore the possibility to commonly share the ecFlow python classes (suite objects) and ecFlow 'createsuites' tool (suites generator), above the simple ecFlow scripts layer sharing (see Figure 1) and, moreover, the possibility to extend this set of scripts with the B-matrix computation procedures (at ECMWF).
13. A propos of B-matrix computation and modeling, it was mentioned a dedicated video-meeting should be organised where Pierre Brousseau (Météo-France) would be invited.

Tentative roadmap for the 1Q2021:

- A. Continue the work to extend ARSO/ecFlow scripts @ECMWF in collaboration with ARSO (Portugal);
- B. Migrate RMI ecFlow scripts to ECMWF (Belgium);
- C. Discuss with other partners (bullet A, HARMONIE, Météo-France) to take advantage of the latest DA cycling features in order to converge to a final suite's set design that should be shared on DASKIT wiki page up to the end of 1Q2021;
- D. Keep an updated version of the ecFlow scripts under development (bullet A) in hendrix (Météo-France), so that the countries may, in the meantime, have some tool to look at, if needed;
- E. From point D, and from previously available DAsKIT Korn-shell scripts for CY40T1 in belenos (see 'testbed' in /home/gmap/mrpe/monteiro/public/DAsKIT2018/surfDAexer), give support to the implementation of Korn shell scripts in belenos so that they may be shared by other countries too (Algeria, Portugal) .

Recommendations:

- I. ecFlow is a client-server workflow package that enables users to run a large number of programs (with dependencies on each other and on time) in a controlled environment. It provides reasonable tolerance for hardware and software failures, combined with restart capabilities. It is used at ECMWF to run all their operational suites across a range of platforms. It is an OpenSource APACHE 2 License package. Countries are welcome to get further info details from: <https://confluence.ecmwf.int/display/ECFLOW>.
- II. According to suggestion from ECMWF³, a good starting point suggestion is to read the Newsletter article and the e-learning module. Paths for this information on ecFlow can be

³ with DAsKIT acknowledgements to Dominique Lucas.

found at:

Articles in the ECMWF newsletter. See e.g.:

<https://www.ecmwf.int/en/elibrary/14594-newsletter-no-129-autumn-2011>

The ecflow home page:

<https://confluence.ecmwf.int/pages/viewpage.action?pageId=13205615>

The ecflow e-learning module:

https://www.ecmwf.int/assets/elearning/ecflow/ecflow1/story_html5.html.

- III. Turkey will join Portugal in the efforts to extend the DAsKIT set of scripts for B-matrix computation by an ensemble technique.
- IV. Turkey will deliver a wiki page platform to allow the sharing of scripts documentation.