A-LAEF ecFlow TC2 Suite

(technical documentation)

Martin Belluš (SHMÚ)
martin.bellus@shmu.sk

October 2019, Bratislava
(updated in February 2021)

Introduction

A meso-scale ensemble system A-LAEF (ALARO - Limited Area Ensemble Forecasting) based on ALADIN canonical model configuration ALARO (Termonia et al., 2018) has been developed in the frame of RC LACE consortium (Regional Cooperation for Limited Area modelling in Central Europe). It is focused on short range probabilistic forecasts and profits from the advanced multi-scale ALARO physics (which is adapted to perform on horizontal mesh-sizes of 2-10 km). Its main purpose is to provide a probabilistic forecast for the national weather services of RC LACE partners (Wang et al., 2017) who could not achieve that with their own HPC resources. It also serves as a reliable source of probabilistic information applied further into the downstream models of hydrology, energy industry and even in the nowcasting.

New A-LAEF system is a sequel to the ALADIN-LAEF (Aire Limitée Adaptation dynamique Développement InterNational - Limited Area Ensemble Forecasting), which was operational on ECMWF HPCF as TC2 application since 2011 (Wang et al., 2011).

A-LAEF system

In the A-LAEF system we use several strategies to simulate the uncertainty of the initial conditions and of the numerical model, while the perturbations at the boundaries are prescribed by the downscaled information from driving global EPS.

a) IC perturbations

The surface and soil prognostic fields uncertainty in the initial conditions of A-LAEF system is simulated by the ensemble of surface data assimilations - ESDA (Belluš et al., 2016). Each ensemble member has its own data assimilation cycle with randomly perturbed screen-level measurements.

The uncertainty of the upper-air part of the initial conditions used in the A-LAEF system is currently simulated by the upper-air spectral blending (Derková and Belluš, 2007; Wang et al., 2014). It combines the large-scale perturbations provided by the driving global ensemble (ECMWF EPS) with the small-scale perturbations coming from A-LAEF first guess within the pseudo-assimilation cycle. It is going to be replaced by so-called ENS BlendVar (ensemble of blending + 3DVar) at Phase II.

b) Model perturbations

The model uncertainty is simulated by the combination of ALARO multi-physics (four different setups of micro-physics, deep and shallow convection, radiation and turbulence schemes) and the stochastic perturbation of physics tendencies (Wang et al., 2019).

c) LBC perturbations

The A-LAEF system is coupled to the global ECMWF EPS. The perturbed lateral boundary conditions are retrieved from the first 16 ENS members with a coupling frequency of 6h to account for the uncertainties at the domain boundaries.

A-LAEF system specifications can be seen in Table 1, while the integration and postprocessing domains are shown in Figure 1.

Tab.1: *A-LAEF system specifications.*

Code version	cy40t1
Horizontal resolution	4.8 km
Vertical levels	60
Number of grid points	750x1250
Grid	linear
Time step	180s
Forecast length	72 h (00/12 UTC)
Members	16+1
IC perturbation	ESDA [surface], blending (Phase I) / ENS BlendVar (Phase II) [upper-air]
Model perturbation	ALARO-1 multi-physics + surface SPPT
LBC perturbation	ECMWF ENS

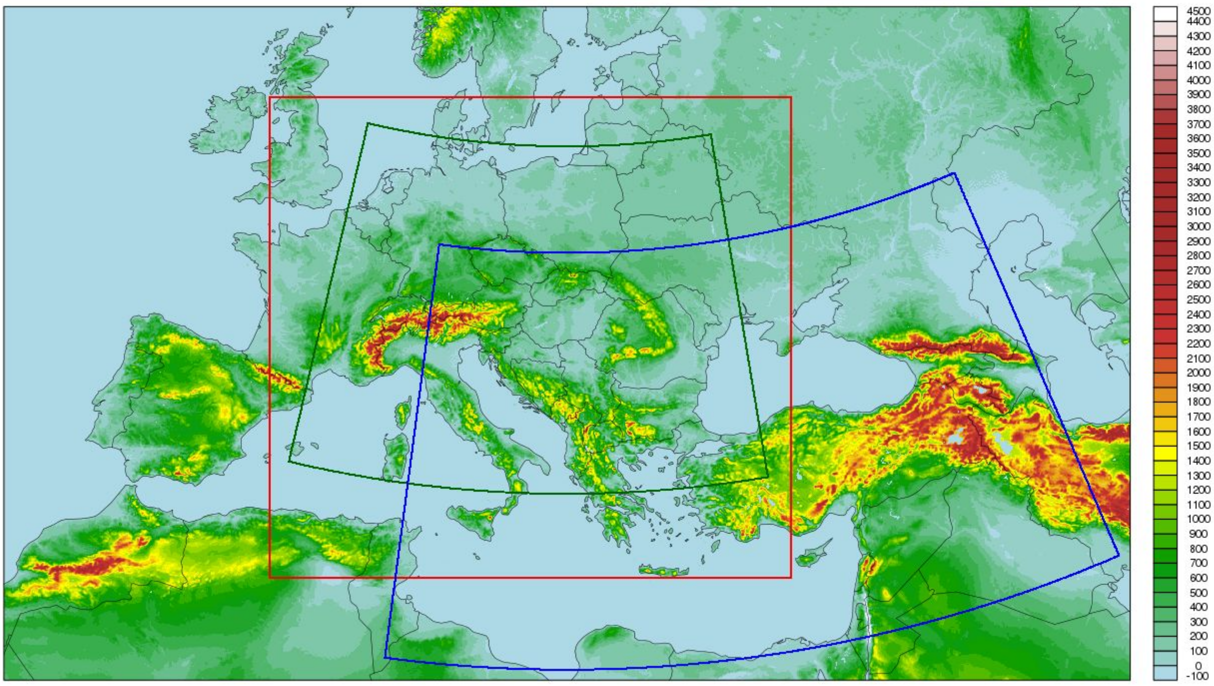

Fig.1: *A-LAEF integration domain with model orography and boundaries of the operational postprocessing domains - common LACE (red), CZ (green), TR (blue).*

Technical specifications

Suite definition file is generated by Python script (create_laef.py). The ECF_MICRO character used for the code preprocessing was redefined to “^” (caret). The reason is to have the least interaction with the original Perl code. The default setting for ECF_MICRO is "%" (percent), which is used for the hash associative arrays in Perl and thus is widely spread in the code. Even now, all the original caret signs (e.g. in some regular expressions) must have been exchanged by "^^" (caret caret) to avoid their substitution by ecFlow preprocessing.

The "laef" suite has three essential families: "admin" to make an easy switch between the computational clusters (SCHOSTs) and file systems (STHOSTs); "RUN_00" and "RUN_12" for 00 and 12 UTC cycles respectively. Each of the last two contains five other families for processing input data ("ecpds"); fetching observations ("getobs"); running ensemble ("main"); dissemination of products ("dissemination") and final cleaning/mirroring of the data ("closure"). The family "main" has subfamilies for each member ("MEM_<nn>") involving preparation of the coupling files (task "getlbc"), surface data assimilation (task "canari"), upper-air spectral blending by DFI (task "blend"), model integration (task "laeff") and GRIBs production (task "make_gribs").

All the ecFlow scripts (Perl application tasks, Python script to generate the "laef" suite definition file and include files) are stored on ecgate server. The "laef" suite is generated at virtual machine ecgb-vecf and also loaded into the ecFlow server running on that VM to avoid the load on the interactive node on ecgate.

ecgate: /home/ms/laef/zla/ecf

drwxr-x---. 3 zla laef 4096 Oct 22 11:09 **app**	ecFlow tasks
drwxr-x---. 3 zla laef 4096 Oct 21 16:16 **def**	suite definition
drwxr-x---. 2 zla laef 4096 Oct 21 13:10 **include**	head, tail, pbs
drwxr-x---. 4 zla laef 4096 Oct 9 07:12 **laef**	suite*

(*) auto-generated by Perl script app/link_apps_to_tasks.pl

On the other hand, the compiled binaries, namelists and input files must be located on the HPCF cluster (cca/ccb), because this is where the jobs are executed under the queuing system (PBS).

cca/ccb: /$STHOST/tcwork/zla/lb/app_LAEF5F

drwxr-x--- 3 zla laef 4096 Apr 24 11:08 **blend**	blending namelists
drwxr-x--- 3 zla laef 4096 Apr 24 11:11 **canari**	surface namelists
drwxr-x--- 3 zla laef 4096 Nov 9 14:18 **fpos**	fullpos namelists
drwxr-x--- 3 zla laef 4096 Apr 24 11:18 **getlbc**	namelist for c903
drwxr-x--- 3 zla laef 4096 Apr 24 11:18 **laeff**	e001 namelists
drwxr-x--- 2 zla laef 4096 Oct 22 11:42 **setup**	main laef config*

(*) Conf_app.pm and Support.pm Perl modules

cca/ccb: /$STHOST/tcwork/zla/lb

drwxr-xr-x 2 zla laef 4096 Oct 23 12:57 **BIN**	all executables*
drwxr-x--- 5 zla laef 4096 Apr 24 13:13 **CLIM**	clim files
drwxr-x--- 4 zla laef 4096 Oct 16 15:09 **LBC**	input from ECMWF
drwxr-xr-x 3 zla laef 4096 Jun 28 12:19 **LAEF5F**	A-LAEF output
drwxr-x--- 2 zla laef 4096 Oct 9 06:55 **bin**	Perl tools

(*) sources and compilation: /perm/ms/at/kmxy/packs/40t1_LAEF5_FULL

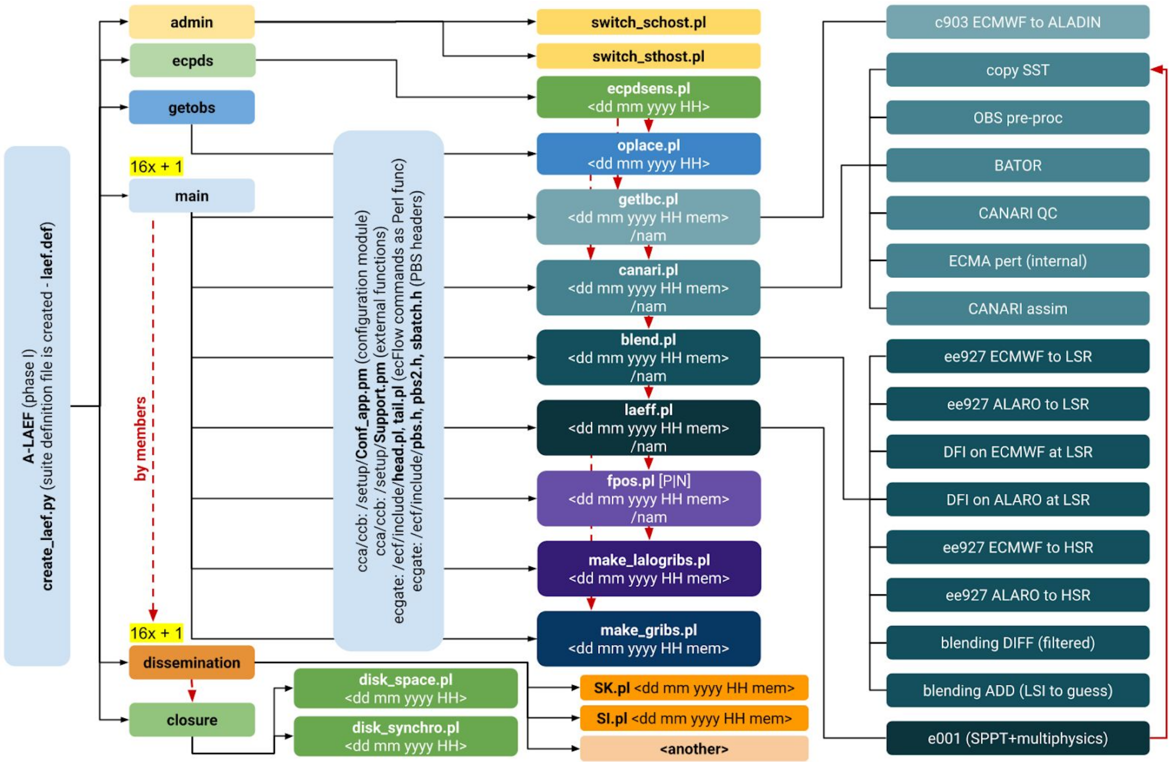

Fig.2: *New A-LAEF suite built under the ecFlow environment. Suite definition file is generated by Python code, while all tasks, include-files and configuration modules are written in Perl. Task dependencies are denoted by the red dashed arrows.*

Tab.2: *Overview of the "laef" suite requirements.*

1. Run time of suite	Start time	• 00:30 UTC (RUN 00) • 12:30 UTC (RUN 12)
	End time	• Preferably before 4:00 and 16:00 UTC.
	Availability requirements	• As soon as possible, but there is no deadline.
2. Input data	Dependence on ECMWF forecast products	• ENS 00/12 UTC 6-hourly output to STEP=84 for the first 16 members and control run (for 12h-lagged coupling).
	Dependence on other data	• OPLACE obsouls*
	How is this transferred to ECMWF?	• ectrans Gateway: ecaccess.ecmwf.int Host: ftp.met.hu
	Backup procedure?	• Not applicable.

3. HPC compute resources	Number of nodes/CPUs (maximum)	• 17x(8 nodes / 288 CPUs)
	Times at which peak number of CPUs are needed	• For ~2 hours (01:00~03:00 UTC; 13:00~15:00 UTC).
4. HPC storage resources	Input data volume	• 350 GB per cycle / 700 GB per day
	Output data volume	• 1 TB per cycle/ 2 TB per day
	Permanent storage requirements (for scripts, libraries and binaries)	• 10 GB on $TCWORK/lb
	Number of days of data to be stored	• 3 days of input data (2.1 TB) • 3 days of output data (6 TB)
5. Data handling storage requirements (archiving)	Data to be stored on MARS	• None
	Data to be stored in ECFS	• To be decided.
6. Volume of data to be transferred from ECMWF to local site(s)		• ~80 GB per cycle.
7. Software requirements	ECMWF software	• ecFlow, ECaccess, ecCodes.
	Other software	• Specific software and libraries for running ALARO CMC model chain.

(*) observations are retrieved from OPLACE archive (operational database of RC LACE)

Tab.3: *Contacts.*

Emergency contact	Martin Belluš	martin.bellus@gmail.com	+421 918 412 792	24x7

Tab.4: *Description of the "laef" suite.*

Family	Task	Sys	Description
admin	switch_schost	ecgb	Switching between clusters (cca/ccb).
	switch_sthost	ecgb	Switching between FSs (sc1/sc2).
RUN_00|12			00 and 12 UTC runs
ecpds	ecpdsens	cca/ ccb	Processing local ECPDS stream (ENS/DET input global fields).
getobs	oplace	cca/ ccb	Fetching OBS files from OPLACE. Trigger("../ecpds == complete")
main/MEM_00|01|..16	getlbc	cca/ ccb	Creating LBCs for A-LAEF domain from global ENS gribs by conf 903. Trigger("../../ecpds == complete")
	canari	cca/ ccb	Surface assimilation using perturbed OBS by conf 701. Trigger("../../getobs == complete and getlbc == complete")
	blend	cca/ ccb	Upper-air spectral blending by DFI. Trigger("canari == complete")
	laeff	cca/ ccb	Model integration by conf e001. Trigger("blend == complete")
	fpos_P|N	cca/ ccb	Offline LATLON fullpos (multidomain). Trigger("laeff == complete")
	make_gribs	cca/ ccb	Converting FA files to GRIBs (Lambert). Trigger("laeff == complete")
	make_lalogribs	cca/ ccb	Converting FA files to GRIBs (LATLON). Trigger("fpos_P == complete and fpos_N == complete")
dissemination/MEM_00|01|..16	SK, SI, RO	cca/ ccb	Sending Lambert GRIBs via ECPDS to RC LACE members. Trigger("../../main/MEM_xx/make_gribs == complete")
	CZ, TR	cca/ ccb	Sending LATLON GRIBs via ECPDS to Czech Rep. and Turkey. Trigger("../../main/MEM_xx/make_lalogribs == complete")

closure	disk_space	cca/ ccb	Old data cleaning (also mirrored files). Trigger("../dissemination == complete")
	disk_synchro	cca/ ccb	Incremental synchronization (mirroring). Trigger("../main == complete")

Tab.5: *Location of the "laef" suite scripts and data.*

Data	Host	Location
Suite scripts	ecgate	/home/ms/laef/zla/ecf/laef
A-LAEF executables	cca/ccb	/$STHOST/tcwork/zla/lb/BIN
A-LAEF namelists	cca/ccb	/$STHOST/tcwork/zla/lb/app_LAEF5F
A-LAEF configuration	cca/ccb	/$STHOST/tcwork/zla/lb/app_LAEF5F/setup
Model climatology files	cca/ccb	/$STHOST/tcwork/zla/lb/CLIM
Input ENS data	cca/ccb	/$STHOST/tcwork/zla/lb/LBC/ECPDS/YYYYMMDDHH
A-LAEF output data	cca/ccb	/$STHOST/tcwork/zla/lb/LAEF5F/TCC
Job output files	cca/ccb	/$STHOST/tcwork/zla/sb/ECF/laef

The time duration of individual A-LAEF tasks within the whole suite (+72h forecast) is shown in table 6. Processes like ensemble of surface data assimilation (ESDA) and upper-air spectral blending are not dependent on the forecast length and their duration is marginal. OBS retrieval from OPLACE archive is negligible and the preparation of coupling files - although for full forecast length (+72h) - is still fast when using 903 configuration in NFPSERVER mode (with the loop on files inside). The only time demanding task is the model integration itself. For the configuration involving 288 CPUs (8 nodes) it takes about 2 hours and 20 minutes to finish the 3-day forecast.

Tab.6: *Time duration of individual tasks for +72h forecast length and their usual finishing times for 00 and 12 UTC run.*

Task	CPUs	+72h forecast	00 UTC	12 UTC
ecpdsens	1	00:22:00	00:52-00:54	12:52-12:56
oplace	1	00:00:30	00:53-00:55	12:54-12:56
getlbc	36	00:06:00	00:57-00:59	12:57-12:59
canari	288	00:02:45	01:00-01:02	13:00-13:05

blend	288	00:08:00	01:08-01:11	13:08-13:14
laeff	288	02:20:00	03:30-03:40	15:30-15:40
fpos_P|N	2x36	00:18:00	03:50-04:00	15:50-16:00
make_gribs	18	00:05:00	03:35-03:45	15:35-15:45
make_lalogribs	18	00:05:00	03:55-04:05	15:55-16:05

Tab.7: *Technical characteristics of the main tasks running on HPCF (not including preprocessing and GRIBs production).*

Task	CPUs	Wall (s)	Output	SBUs	Total SBUs
	values per member				16+1 mem (+72h, 2x/day)
canari	288	~180	518 MB	~232	~7888
blend	288	~480	518 MB	~620	~21080
laeff	288	~8420	44 GB	~11000	~374000
fpos_P|N	2x36	2x ~1080	3 GB	~350	~11900
Total SBUs consumption per year (an approximation only)					~151 mio

References

Termonia, P., Fischer, C., Bazile, E., Bouyssel, F., Brožková, R., Bénard, P., Bochenek, B., Degrauwe, D., Derkova, M., El Khatib, R., Hamdi, R., Mašek, J., Pottier, P., Pristov, N., Seity, Y., Smolíková, P., Spaniel, O., Tudor, M., Wang, Y., Wittmann, C., Joly, A., 2018. The ALADIN System and its Canonical Model Configurations AROME CY41T1 and ALARO CY40T1. Geoscientific Model Development. 1–45. doi: 10.5194/gmd-2017-103.

Wang, Y., M. Belluš, C. Wittmann, M. Steinheimer, F. Weidle, A. Kann, S. Ivatek-Šahdan, W. Tian, X. Ma, S. Tascu, and E. Bazile, 2011: The Central European limited-area ensemble forecasting system: ALADIN-LAEF. Quart. J. Roy. Meteor. Soc., 137, 483–502.

Belluš, M., Y. Wang, F. Meier, 2016: Perturbing surface initial conditions in a regional ensemble prediction system. Mon. Wea. Rev. 144:3377-3390.

Wang, Y., M. Belluš, J. Geleyn, X. Ma, W. Tian, and F. Weidle, 2014: A new method for generating initial perturbations in regional ensemble prediction system: blending. Mon. Wea. Rev. 142: 2043-2059.

Wang, Y., M. Belluš, A. Ehrlich, M. Mile, N. Pristov, P. Smolikova, O. Spaniel, A. Trojakova, R. Brozkova, J. Cedilnik, D. Klaric, T. Kovacic, J. Masek, F. Meier, B. Szintai, S. Tascu, J. Vivoda, C. Wastl, Ch. Wittmann, 2017: 27 years of Regional Co-operation for Limited Area Modelling in Central Europe (RC LACE). Bulletin of the Am. Met. Soc., Vol. 99 Issue 7, 1415-1432.

Wang, Y., M. Belluš, F. Weidle, et al., 2019: "Impact of land surface stochastic physics in ALADIN-LAEF", Quarterly Journal of the Royal Meteorological Society, 1–19, doi: 10.1002/qj.3623.

Derková, M., M. Belluš, 2007: Various applications of the blending by digital filter technique in the ALADIN numerical weather prediction system. Meteorologicky casopis, 10, 27–36.