Offline soil moisture analysis within the SURFEX-SODA framework

Viktor Tarjáni

Slovak Hydrometeorological Institute, Bratislava

September 18-20, 2019

Joint LACE Data Assimilation Working Days and ALADIN Data Assimilation basic kit Working Days

Outline

Upgrade highlights

- Prior to upgrade SURFEX from operational ALARO build based on `cy40t1_bf06` was used
- `srfx` suite scripts were written with possibility to easily switch back to previous version
- Downloaded the most actual commit (`1d8ef44`) of master branch from SURFEX git repository
- SURFEX V8.1 was built from downloaded source code
- ‘srfx’ suite scripts were adapted to work with SURFEX V8.1

Building SURFEX V8.1 on HPC2 → failed

- HPC2 spec:
  - cpu's: IBM Power7+
  - os: Red Hat (Fedora) Enterprise Linux
  - compiler: **GNU gcc 4.9.3**
- Getting following error during building:

```
spll_fadcpl_fort.f90:162:0: error:
unable to generate reloads for...
internal compiler error: in find_reloads, at reload.c
Please submit a full bug report,
make[1]: *** [spll_fadcpl_fort.o] Error 1
```

- Error is caused by bug in **old GNU gcc 4.9.3 compiler**
- Upgrade to newer compiler version on HPC2 was not allowed/possible

Building SURFEX V8.1 on HPC1 → success

- Solution: build on HPC1 with newer GNU gcc compiler
- HPC1 spec:
  - cpu's: IBM Power7 (predecessor of Power7+)
  - os: Gentoo Linux
  - compiler: **GNU gcc 7.3.0**
- Drawback: compiled binaries cannot be executed on HPC2 due to old *glibc* library installed on HPC2

Additional remarks

- SURFEX was built against the system *netcdf* and *gribapi* libraries ⇒ no need to compile those libraries distributed with SURFEX source code
- Alternative variants of SURFEX executables were built combining following build options:
  - `OPTLEVEL=DEBUG` and `O2`
  - `VER_MPI=NOMPI` and `MPICH2` and `MPIAUTO`
  - `VER_OMP=NOOMP`
- Following files were modified prior to build:
  - `Makefile.SURFEX.mk`
  - `Rules.LXgfortran.mk`

Problems during OFFLINE and SODA execution

- OFFLINE and SODA compiled with `VER_MPI=MPICH2` or `MPIAUTO` when executed with `poe SODA -procs > 1` and setting timeseries to FA or LFI:

```
*/*/* LFIOUV - Nom='SFXOUT.20190426_01h00.lfi'
***** LFIOUV - UNITE 34 NON FERMEE APRES
LA DERNIERE MODIFICATION*****
Program received signal SIGABRT: Process abort signal.
```

- Executing same without poe or setting time-series and restart file type to NC worked OK

Error reading CANARI file during SODA I

- File CANARI is FA file which stores gridded observation fields CLSTEMPERATURE and CLSHUMI.REL needed by EKF, EnKF and some additional fields necessary for OI method in SODA
- SODA (all variants built) exited with following error when gridded observations are to be red from CANARI file:

```
* LFIOUV - Nom='CANARI'
* FAITOU - CDNOMC='canari'
* FAITOU - UN ARTICLE DU CADRE A UNE LONGUEUR
INATTENDUE, UNITE 19 *
Program received signal SIGABRT...
```

- Same error has occurred even if original CANARI analysis output was replaced with pseudo-CANARI file (short range ALARO forecast – history or full-pos file)

Error reading CANARI file during SODA II

- Prior to SURFEX upgrade reading gridded observations from pseudo-CANARI file (prepared with full-pos) worked well
- Should the error be caused by incompatibility between cy40t1's XRD39 and XRD44 bulit during SURFEX v8.1? Olda suggested it should be rather problem with porting SURFEX. Should be examined in future.
- Currently temporary solution was implemented: CANARI FA file is converted first to ASCII file which SODA is able to read without any problems.

Bug in soda.f90 I

- SODA which was built with `OPTLEVEL=DEBUG` option finished with following error
```
At line 906 of file ../MASTER/spll_soda.f90
Fortran runtime error: Index '3' of dimension 2
of array 'zwork' outside of expected range (1:2)
```
- If it was built with `OPTLEVEL=O2` it finished without error message but debugging using gdb suggest that incorrect results is written
- Bug was encountered in **soda.f90** file (commit `1d8ef44` in master branch) in loop over lines 934 - 943:

Bug in soda.f90 II

```
! Set observations used for possibly other tiles than n
DO IOBS = 1,NOBSTYPE
  SELECT CASE (TRIM(COBS(IOBS)))
    CASE ("T2M")
      ZT2M(:)=ZWORK(:,JJ)
    CASE ("HU2M")
      ...
  END SELECT
ENDDO
```

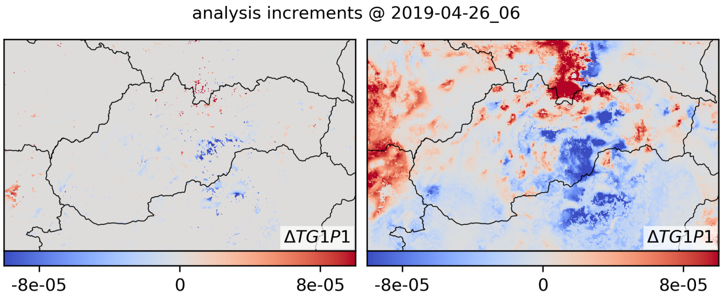
- This is the part of the code where gridded observations are read from ASCII file into internal arrays ZT2M, ZHU2M, ZSWE for all grid boxes (`NDIM_FULL`)
- There is other place (eg lines 928-932) where same ASCII observation file is read into XYO array, but only for grid boxes which contain NATURE tile.

Bug in soda.f90 III

- Thus, bug only affects reading obs. from ASCII file and not from CANARI FA file and in case of EKF or EnKF methods it only affects grid boxes without NATURE tile, i.e. those which are not assimilated in EKF/EnKF.
- In body of the cycle variable `JJ` is used instead of loop variable `IOBS`.
- Value of `JJ` is constant = 3 (as a result of using it as loop variable in some of the previous loops upon inspecting by `gdb` debugger) during cycle iteration which is outside of valid range.
- It not necessary leads to SODA crash but it surely leads to incorrectness of results in some parts of spatial domain
- It was fixed in user version (prefix VITA) by replacing `JJ` with `IOBS` in cycle body

Blank areas on analysis increment maps I

- Maps of analysis increment showed unexpectedly large portions of blank (zero-valued) areas over spatial domain
- 141 537 out of 150 801 grid boxes ($\approx$ 94%) had zero increments for each control variable



Blank areas on analysis increment maps II

- It was found that the problem is caused by setting CTOWN='NONE' instead of CTOWN='TEB' in &NAM_PGD_SCHEMES namelist which is used during PGD
- This is because setting CTOWN='NONE' effectively turns-off numerical scheme for each gridbox having FRAC_TOWN>0
- This was proved by showing that mask where analysis increments= 0 (in SODA output file) is identical to mask where FRAC_TOWN>0 (in PGD file).
- In fact this issue first occurs not in SODA but during OFFLINE execution (reference and all perturbed runs) when SURFEX counterpart of gridded observation (T2M, HU2M fields) are calculated. CTOWN='NONE' leads to filling all grid-boxes having FRAC_TOWN>0 with value XUNDEF = 1.E+20 and SODA subsequently excludes those grid-boxes from EKF analysis

Implemented configurations

3 basic configurations were implemented each aiming for specific purpose:

1. *oper*
   - Computes offline soil analysis which can be eventually used by operational ALARO
   - Spatial domain identical with ALARO (C+I zone only)
   - Uses only inputs from operational ALARO
2. *inca*
   - Prepares high-resolution soil analysis over Slovakia
   - Intended for specific applications, e.g. agriculture, etc.
   - Spatial domain identical with INCA-SK system
3. *1col*
   - Testing new ideas
   - Sensitivity studies
   - Comparison performance of EKF and other methods under various, but controlled, conditions
   - Allows maximum control over many parameters and inputs to analysis

Main features summary

<table>
<tr><th></th><th>oper</th><th>inca</th></tr>
<tr><td>horizontal resolution</td><td>4.5 km</td><td>1 km</td></tr>
<tr><td>grid size</td><td>614×565</td><td>501×301</td></tr>
<tr><td>gridded obs.</td><td>CANARI analysis</td><td>INCA-SK analysis</td></tr>
<tr><td>forcing height</td><td colspan="2">20m</td></tr>
<tr><td>forcing TA, QA, WIND, DIR</td><td colspan="2">ALARO-SK</td></tr>
<tr><td>forcing PS</td><td colspan="2">ALARO-SK</td></tr>
<tr><td>forcing CO2</td><td colspan="2">const. 0.000620</td></tr>
<tr><td>forcing surface radiation</td><td>ALARO-SK</td><td>ALARO-SK | GR_AV</td></tr>
<tr><td>forcing surface precipitation</td><td>ALARO-SK</td><td>INCA-SK</td></tr>
</table>

oper setup

- Spatial domain identical with operational ALARO
- Uses only inputs from operational model
- Forcing from 20m level ALAD assimilation fields
- Gridded observations: operational CANARI 2m analysis (`CLSTEMPERATURE`, `CLSHUMI.RELATIVE`)

inca setup

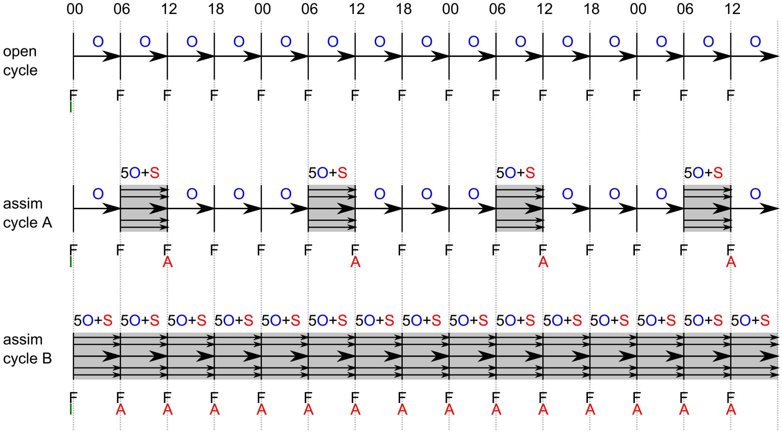
- Spatial domain identical with INCA-SK
- Uses most accurate combination of inputs
- Forcing from 20m level ALAD field (T, W), precipitation from INCA-SK analysis, global radiation: now from ALARO assimilation file but could be global radiation analysis (it however misses LW part and partitioning to DIR and DIF)
- Gridded observations: INCA-SK 2m analysis
- Important difference is that INCA-SK gridded observations are more observation oriented while CANARI gridded observations are strongly constrained by model dynamics and physics

Implemented assimilation cycling options I

1. Pseudo-assimilation (open) cycle:
   $HH_f = \{00, 06, 12, 18\}$, HH_a = None
2. Assimilation cycle A:
   $HH_f = \{00, 06, 12, 18\}$, $HH_a = \{12\}$
3. Assimilation cycle B:
   $HH_f = \{00, 06, 12, 18\}$, $HH_a = \{00, 06, 12, 18\}$

- Reforcing is performed at same hours HH_f for each option
- They differ in when analysis is done: In option 1 it is not done at all, in option 2 it is done once per day at 12 UTC and in option 3 analysis is done each 6 hour

Implemented assimilation cycling options II



I - initialization (cold start)

F - reforcing

A - analysis (reinitialization)

O - executing OFFLINE

S - executing SODA

Gridded observations I

- Gridded observations are in fact output of screen level (2m) analysis of CANARI or INCA-SK system. This fact gives them some specific properties when compared with conventional observations
- They form regular horizontal mesh, which ensures that during EKF analysis each SUREFX grid-box has corresponding observation (T2M, HU2M) at same horizontal location, thus innovation vector can be calculated consistently without the need of horizontal interpolation

$$\text{innov} = \mathbf{y}_b - \mathbf{y}_o$$

$\mathbf{y}_o$ gridded obs. vector (CANARI, INCA)
$\mathbf{y}_b$ SURFEX counterpart of gridded obs.

Gridded observations II

- They differ from conventional observations in another important aspect: they are observations as seen by observer in model world. For example T2M calculated by CANARI is temperature at 2m above model surface, which however differs from actual surface (orography) and this difference could be significant especially in complex mountain regions. Actual lapse rate, for example, is taken into account in CANARI vertical interpolation
- Gridded observations are not just simple horizontally interpolated observations to model grid but hey involve also some background information which, in case of CANARI and INCA-SK analysis, is short range (6h) forecast of ALARO model

Gridded observations III

- If there is some systematical deviation between $\mathbf{y}_o$ and $\mathbf{y}_b$ it will go through analysis and will produce deviation in analysis increment (e.g. biased analysis)
- It is important to know exact procedure how CANARI CLS analysis is calculated ($\mathbf{y}_o$) and how CLS values are calculated in SODA, SURFEX ($\mathbf{y}_b$) and made these calculations as consistent as possible
- Similar arguments hold when INCA-SK analysis is used as gridded observations
- In practice there are 2 main options in oper setup what to use as gridded observations
  1. Original CANARI analysis
  2. 6 hour forecast from ALARO previous assimilation or eventually production cycle

Gridded observations IV

Option 1) is however preferable because in option 2) large error in CLS fields can build up even during short time usually when soil variables are not properly represented in ALARO model

Forcing – general aspects I

- Forcing acts as upper boundary condition for SURFEX numerical schemes
- SURFEX computes not only surface state but also part of the atmosphere bellow forcing level (prognostically or diagnostically by interpolation between surface and forcing level)
- Indirect soil analysis method (originally proposed by Mahfouf) as is EKF, requires SURFEX CLS variables (T2M, HU2M) being sensitive to change in soil state. To ensure this forcing level has to be put higher then 2m above surface level for temperature and humidity i.e., TA and QA forecast need to be taken from higher vertical level then gridded observations (eg. 20m in our case)

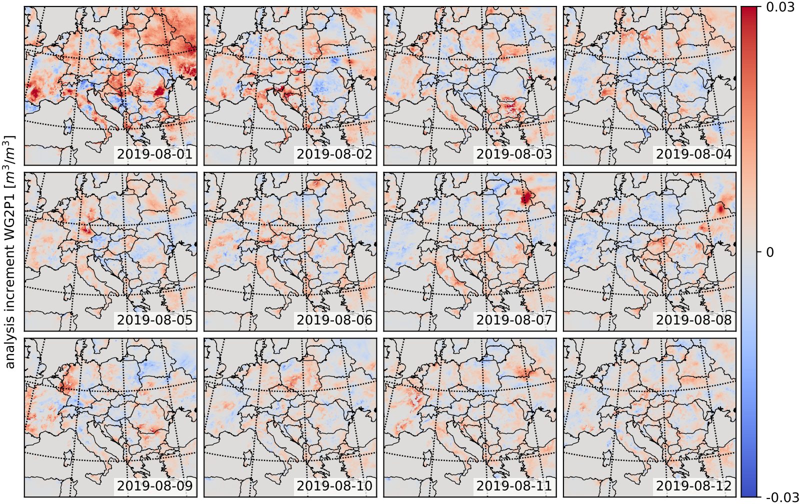
Forcing – general aspects II

- In order to let SURFEX calculate whole surface layer (the bottom 10 to 25 m of the troposphere) forcing level for temperature and wind both should be put at the top of surface layer
- In all configurations forcing level is set to 20m above surface (e.g. forcing fields are taken from that level)
- Forcing series cover 6h long temporal interval with 1h step during which Jacobi matrix **H** (eventually **M**) is calculated.
- In reanalysis mode forecasts from ALARO-SK assimilation cycle (based on long cut-off inputs) are used to prepare forcing
- Preparing forcing for both oper and inca configurations involves vertical full-pos to 20m H level, but inca setup requires additionally horizontal full-pos

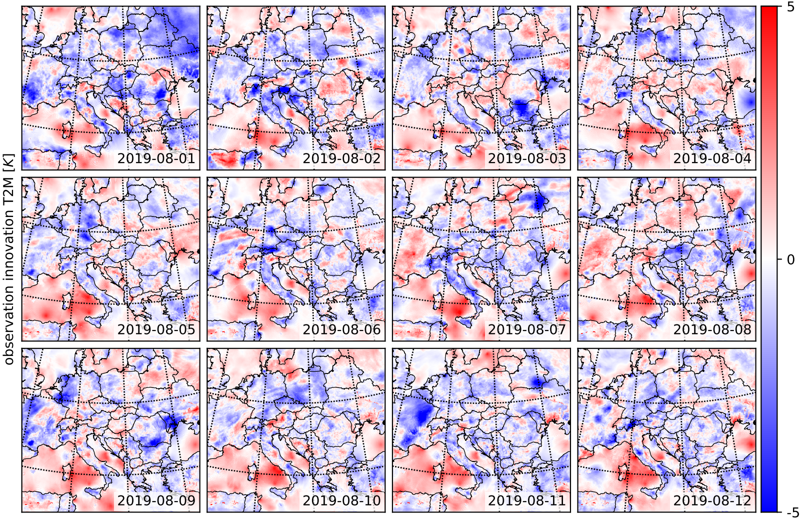
Demonstration run

- Demo EKF analysis cycling was done for first 12 days of August 2019
- It aims to demonstrate feasibility of cycled EKF analysis implementation
- Cycling mode A was tested for 'oper' configuration (i.e. analysis performed each day at 12 UTC).
- It was confirmed that computational time is acceptable for operational usage
- Size of output files (netcdf in our case) would need reduction in size by discarding writing of unnecessary fields and by turning on the compression
- Temporal evolution of analysis increments and observation innovation show physical relevance
- Thorough evaluation of performance would be done, by calculating verification scores of CLS variables, longer temporal interval, response of analysis to severe precipitation events, etc.

WG2 increments evolution



T2M observation innovation



RMS analysis increments

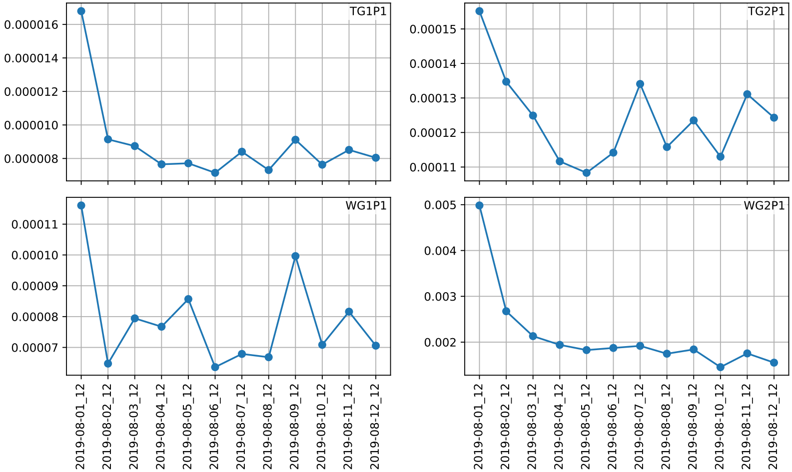


Figure: Temporal evolution of RMS analysis increments (averaged over whole spatial domain)

RMS observation innovations

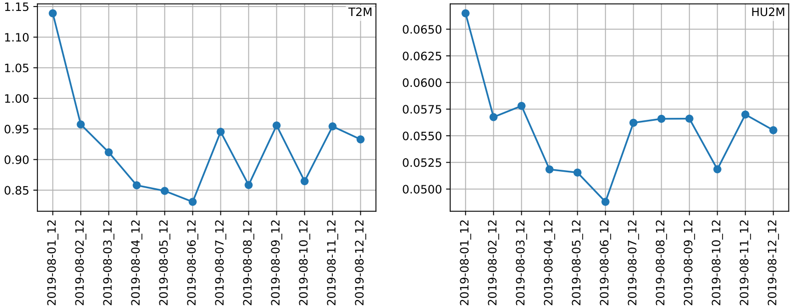


Figure: Temporal evolution of RMS observation innovations (averaged over whole spatial domain)

Thank You