# Possible evolutions on SURFACE OFFLINE DATA ASSIMILATION (SODA)


## Jean-François Mahfouf


## 18 Décember 2013

In this document, I try to summarize the current status on developments related to surface data assimilation within SURFEX, and to raise a number of questions regarding future evolutions.

### *Context*

Almost two years ago a workshop was held at Météo-France on land surface data assimilation developments for various applications using SURFEX as a modelling platform. During the course of this workshop we agreed to develop a common interface for land surface developments that could accommodate various applications, in particular the possibility of having different assimilation schemes such as the Optimum Interpolation (OI : used to NWP applications) and also an Extended Kalman Filter (EKF : developed for NWP applications but also for land surface monitoring). It was recognized during the course of the workshop that other flavours of KF schemes such as STAEKF and EnKF (developed in the ALADIN and HIRLAM consortia) could also be introduced in that unified framework. Other foreseen extensions were the use of new surface models such as Flake or snow schemes.

### *Undertaken evolutions*

At the time of the workshop, only the OI scheme was available within SURFEX since it was already used for operational applications with the AROME model. The EKF (named VARASSIM in the SURFEX world) was not part of SURFEX at that time. It was decided to introduce the EKF in SURFEX version 7. The common interface named SODA and developed by Trygve Aspelien (MetNo) was introduced in SURFEX v7.2 and entered in the common atmospheric cycle CY38. SODA was proposing two flavours of the assimilation : the OI and the EKF. Moreover the assimilation was split in independent modules according to the various tiles present in a grid box : NATURE (ISBA) with a separate treatment for SNOW, TOWNS (TEB), LAKES and OCEANS.

### *Current difficulties*

The use of SODA in CY38 raised a first difficulty coming from the fact that in the mean time Philippe Marguinaud (GMAP) realised that the OI scheme was quite inefficient from a computing point of view (mono-processor task with a large amount of inputs/outputs). The OI scheme (named OI_MAIN) was then put back in the ARPEGE/IFS environment under the CANARI option (atmospheric OI scheme) in order to benefit from a parallelized computing environment and to avoid reading and writing fields on files (stored in buffers instead). A consequence was that OI_MAIN was called from two different subroutines, leading to maintenance issues. Another difficulty was the initialisation in SODA that was done separately for each tile, leading to a loss in computing efficiency. A recoding of SODA has

been done by Trygve during summer 2013 to circumvent these issues. This new SODA version should be made available in SURFEX v8.

Even though SODA goes in the direction of unifying the land surface assimilation environment by putting together the OI and EKF schemes, it is still incomplete and does not fulfil a number of needs from users interested in land data assimilation. This is currently the case for Alina Barbu (CNRM/GMME), who is developing an EKF version for land surface monitoring in GEOLAND2 and IMAGINES European projects. Indeed, the EKF version in SODA is only suited for the assimilation of screen level variables (temperature an humidity at 2m) and has to be extended to other quantities such as LAI or the superficial soil moisture in order to be suitable to these projects. In SODA, the observations are hard-coded as arguments of various routines and it is currently not flexible enough to accommodate easily new observation datasets. Another difficulty concerns the land surface scheme: the current version of EKF in SODA does not allow to consider the "patches" (various surface covers over the NATURE tile) that are compulsory for the use of ISBA versions with dynamical vegetation. Finally, SODA has only been tested with LFI file format, that is not used in most offline applications (ASCII and netcdf) that also tends to be replaced by FA format for NWP applications. Given these difficulties, the current offline version of EKF (VARASSIM) will co-exist for some time with the SODA version.

## *Conclusions and planned activities*

During the workshop in 2012, a need was recognized to unify within SURFEX various developments undertaken on land surface assimilation. So far, SODA has partly fulfilled this requirement by putting under the same computing environment the OI and EKF schemes that could run in offline and coupled modes, but limited to one version of ISBA and also to the use of screen-level observations. Currently SODA is used by HIRLAM for their NWP applications. A number of developments are undertaken in the ALADIN consortium (RMI, OMSZ) on the use of the EKF system but without using SODA so far. The technical developments required to use SODA for GMME applications are too important to become an interesting short term solution.

From the knowledge of future activities within the ALADIN and HIRLAM consortia regarding surface data assimilation, it would be useful to foster developments that could fulfil the needs of most research and operational groups.

## *Questions*

How land surface assimilation should or could evolve in the medium term within SURFEX ?
Can we still have a common framework for land data assimilation within SURFEX that could be suitable for most applications ?
Can we accommodate all the following constraints : various analysis schemes, various land surface models, diverse observational datasets, various file formats, allow to run the assimilation "offline" and "coupled" to an atmospheric model, NWP requirements (computational efficiency and scalability).
If not, what are the various aspects that could be developed in a common environment ?
For NWP applications the major constraint is computational efficiency. Since the EKF is quite time consuming and requiring significant amount of I/Os, is SODA compatible with NWP requirements ?

The use of surface schemes with many prognostic variables means that the EKF could be replaced without any significant cost increase by an EnKF (we can also recall that atmospheric systems will also move towards ensembles, providing perturbed input forcings to land surface schemes).

Would it be an interesting solution to have a full externalisation of the data assimilation system (like the PALM environment) that would be totally independent of SURFEX ?

How should the observations be handled in such data assimilation system ? The current approaches are very crude with either outputs from the CANARI OI or externalized spatialisation tool (nearest grid point). This is also the case for data selection, quality controls, and bias correction schemes.

It seems to me that part of these questions are technical should not be only addressed by scientists.