Possible evolutions on SURFACE OFFLINE DATA ASSIMILATION (SODA) – HIRLAM perspective.

Notes from the discussion during the HIRLAM Management Meeting on the 29th of January 2014

Present : Ulf Andrae, Laura Rontu, Inger-Lise Frogner, Xiaohua Yang, Mariano Hortal, Jelena Bojarova (referent); Jeanette Onvlee (excused)

This note is to summarize our discussion on the topic raised by Jean-Francois Mahfouf regarding the future of the SODA development. First of all, we apologize that our input on these questions comes so late. We hope these notes are of use when the discussion about surface data assimilation continues in the coming ALADIN-HIRLAM all staff workshop and at other forums.

HIRLAM vision of the surface data assimilation

Current Operational status

Spatialization tool : CANARI analysis (T2m, H2m, snow depth) based on the optimum interpolation using homogeneous structure functions

Land Surface scheme : 1D-OI -main (update of soil temperature and soil humidity from the screen level observations from the output of CANARI). OI-main is called from the CANARI in order to benefit from the parallel environment and achieve necessary level of computational efficiency.

Current research status:

Spatialisation tool: MESCAN analysis (T2m, Rh2m, snow depth) based on the optimal interpolation using analytically derived non-homogeneous and unisoptopic (sea-land mask and height altitude) structure functions + prototype of 10m wind downscaling system and prototype of the precipitation field downscaling system; Lake and sea-ice spatialisation still to be designed

Land Surface scheme: 1D-EKF and various flavours of it. Presently available for soil temperature and soil moisture; EKF is in the development stage for lakes (off-line 1D-experiment; implementation in the HARMONIE environment still to be done); EKF for snow is delayed due to lack of dedicated manpower resources. Currently EKF is using screen-level observations as input, but this is an inherited outdated property from OI-main and is not a conceptual limitation of the EKF scheme.

Medium range vision of surface data assimilation

Spatialisation tool: The aim: spatialisation tool allowing orography-dependent structure functions and setup consistent with the EnVAR/4DEnVAR approach for UA DA. 2DEnVAR based on the variational solution of the problem using flow-dependent structure functions, derived from the historical ensemble of the forecast perturbation. A prototype of the 2DEnVAR exists, more tuning of the scheme is required in order to obtain the quality of MESCAN. Currently the scheme has been restricted to screen level analysis of T and RH, however this is not a conceptual limitation. The scheme can easily be extended to other variables using the HarmonEPS framework.

Land Surface scheme: The aim is1D-EnKF. The prototype of the EnKF for snow, temperature

and soil mosture exist in the off-line version (collaboration with NILU). The EnKF scheme requires validation. The EnKF scheme for lakes and sea-ice still to be developed.

Long range vision of surface DA:

Coupled SU + UA DA: J_tot = J_SU_(4D)EnVAR + J_UA_4DEnVAR hopefully using OOPS framework

EKF versus EnKF

We completely agree that the EKF is a very expensive and I/O-demanding scheme. This scheme might very easily become unsuitable for operational implementation unless the questions about the optimal design are addressed very thoroughly. At the same time we consider that it is a flexible scheme well suited as a reference because it is based on the analytical solution up to certain extent. EnKF scheme is I/O-demanding scheme also but it is cheaper computationally when the cost of the ensemble production is excluded, or more exactly the cost of ensemble production is shifted from the DA part to the the forecast part. HIRLAM experience shows that the generation of the ensemble perturbations is the most costly part of the ensemble based DA schemes. It is not so straightforward to relate the ensemble of "forcing" UA perturbations, which is always rank-deficient, to the 2DEnVAR + EnKF scheme. A substantial tuning might be required for the 2DEnVAR + EnKF scheme to achieve an optimal performance, and thus it is necessary to have a reference scheme (MESCAN + EKF). The EnKF scheme is based on the statistical linearisation technique and dubious results might appear in case of strongly non-linear systems. Again, the reference is needed as a sanity check. On the middle term perspective in HIRLAM community we see the EKF scheme as a reference scheme for surface DA.

Surface data assimilation

In HIRLAM community we consider the surface data assimilation to be an essential part of the efficient data assimilation system for mesoscale processes. The surface processes are those who have a long memory for small size LAM domains. The surface conditions trigger meso-scale phenomena. The proper surface conditions are essential for the efficient use of radiances. The realistic surface conditions are needed for the proper modelling of the lower part of the atmosphere in order to predict the meso-scale phenomena. The efficient spatialisation tool is a way towards nowcasting and a very short range NWP applications. Unfortunately the surface data assimilation is several years behind in the state-of-art development in comparison to the UA DA. We completely agree that the situation is not satisfactory here.

SURFEX model.

Surface Externalisée is a surface modelling platform developed by MF in cooperation with the scientific community. SURFEX is a very flexible software which includes physics models for various natural land surfaces, including snow, for urban area, lakes and ocean (and sea-ice, in the development). SURFEX can be used hydrological simulations and for simulation of chemistry and aerosol surface processes. SURFEX can be extended to allow the analysis of land surface variables. The spirit of the SURFEX environment is flexibility. SURFEX can be used both in a stand alone mode concentrated on modelling and analysis of particular processes (snow, lakes, …) and in a coupled mode with the atmospheric model for full-dimensional NWP applications. The computational efficiency of SURFEX environment is a weak point of the system as is the usual

situation for flexible scientific platforms. In the HIRLAM community we consider that the flexibility, concerning the tasks which can be addressed, is an important and valuable property of the SURFEX platform. At the same time we agree that a certain compromise might be needed in order to assure the acceptable level of the computational efficiency. This compromise should be agreed between different partners which use the SURFEX platform, which is also the aim of the SURFEX steering committee where they are represented.

SODA (Surface Offline Data Assimilation) environment.

The SODA interface was proposed during the surface workshop initiated by Jean-Francois Mahfouf, March 2012, in order to create a common environment that could accommodate different data assimilation schemes developed at that time for 1D SU DA, namely OI_main and the EKF scheme, in a consistent way. Trygve Aspelien (from HIRLAM) is an active person behind this development. In the mean time dedicated efforts were spent at MF to improve the computational efficiency of the SURFEX scheme and the surface data assimilation (Philippe Marguinaud). The solution proposed by Philippe was somewhat in conflict with SODA design implemented by Trygve in SODA. This lead to a divergence in the developments of land surface data assimilation between the MF and HIRLAM. During the HMG-CSSI meeting of April 2013, it has been decided that JF Mahfouf and C Fischer will organise a web conference between the various people involved in order to redefine a common strategy. The teleconf meeting took place in May 2013 and the following longer term actions have been agreed

1) HIRLAM/MF/ALADIN(?) will make an (technical) analysis of the current SODA framework, and agree on a recommendation on code organization to fulfill SURFEX/AROME++ needs (e.g. FA file format, optimization).
2) After this recommendation the code can be re-written to fulfill the recommendation by parties with time and interest (MF says need is not imminent, but HIRLAM and possibly others might be interested)
3) When re-written it should be a joint scientific co-operation on the "new" framework (SODA?) asa part of the common work plan.

During summer 2013 Trygve rewrote the SODA interface in a way that on the limits fulfil the NWP needs from the HARMONIE side. A new branch was made and MF side (Philippe) was provided with a test set to demonstrate the HARMONIE use of SODA. This “new” SODA version allows to choose OI_main or EKF for various tiles present in a grid box and it uses the SURFEX initialisation and I/O output consistent with those used in AROME. Technically both OI_main and the EKF schemes are possible, but certainly there are still drawbacks in the design and in the implementation. This is the usual status of any new development where a very small number of people is actively involved (we must admit, Trygve alone in the SODA design and implementation) and where ambitions are very high: various models – various data assimilation techniques - various file formats - various types of observations - online/offline - … The development of the “new” interface, under the working name “SODA” (can be changed if wished), is far away from being finished. The work done by Trygve can be seen only as a first step in fostering the development designed to fulfil the common needs of the research and operational groups in the NWP. Most of the essential questions such a strategy for coupling between the UA ensemble forcing and the 2DEnVAR+EnKF environment, computationally effiecient I/O design, remote sensing observations, coupled UA+SU DA, consistency with the OOPS/IFS code redesign, and so on, have not yet been addressed when designing the SODA interface. Currently, the SODA interface is tried in HARMONIE within the common CY38. Within this cycle, testing and development of SODA will be possible for a larger number of research groups..

In the HIRLAM community we consider that the optimal way to foster the developments, which are able to fulfil needs of most research and operational groups, is to work together addressing common problems by common efforts. One should avoid to develop jointly as parallel as possible systems in the common code environment.

Answers to the questions:

1. ***How land surface data assimilation should/could evolve in the medium term with SURFEX?***

In the medium term (a few years), "land surface data assimilation" in the HARMONIE framework relates in principle to the soil , snow and forest (multiple energy balance), sea and sea-ice, lake and urban area modellings. In the section "HIRLAM vision of land surface data assimilation" we discussed this question in more details.. In summary, in the HIRLAM consortia we aim at creation of a the unified with the UA DA framework of the ensemble based data assimilation (2DEnVAR + EnKF under forcing from HarmonEPS). The long term goal is the coupled SU + UA DA based on 4DEnVAR, hopefully using OOPS framework for the design. Extension of EKF from soil to lakes and snow is considered an essential short-term development task.

2. ***Can we still have a common framework for land data assimilation within SURFEX that could be suitable for most applications?***

We hope that "yes". SURFEX is a platform used by a wide scientific community. For HIRLAM the most relevant are the applications related to the NWP and climate modelling. Any scientific research carried in the HIRLAM community ultimately aims at the operational application. An important technical question here is how to develop a surface data assimilation framework which is not merely a collection of existing separate tools but a framework similar to the prognostic SURFEX itself. SURFEX contains the common interfaces and tools to handle input and output, allowing the parametrizations to focus to handling of the diverse physical processes. We should study to which extent, for example, it is possible to share common EKF tools between the evolving soil, lake and snow assimilation schemes, which all are tied to different prognostic schemes (models).

3. ***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 ?***

The flexibility is in the spirit of the SURFEX concept. SURFEX contains everything for research, but can be simplified and consequently optimised for operational NWP needs. This is not trivial, and has not yet been achieved in general for SURFEX in-line coupled to the (atmospheric) NWP model. Flexibility of the design comes usually at the price of computational efficiency and maintainability of the system. In the surface data assimilation to obtain a compromise between the flexibility and optimisation becomes even more challenging task than for pure prognostic system. In the HIRLAM community we would put the NWP requirements and possibility to run the

assimilation “offline” and in the “coupled” to an atmospheric model modes on the “high priority topics list”.

4. 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 ?

SODA is an interface, thus not causing, hopefully even not increasing additionally, the computational expenses and high I/O demand of the EKF scheme. In fact any perturbation based data assimilation scheme is I/O-demanding and computationally expensive if the cost to generate ensemble is taken into account. Current prototype of the SODA interface is the first step towards the common environment trying to address common needs in a consistent and a flexible way. This work is far away from being finished and the design is certainly imperfect in many senses. On the other hand, optimisation of the EKF processes is to be taken more seriously already now when developing the basic soil, snow and lake EKF, which is now ongoing the stand-alone SURFEX/SODA environment.

5. ***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).***

In the HIRLAM community we aim towards the 2DEnVAR + EnKF scheme and MESCAN (as it is now) + EKF is intended to be our starting point for further developments, which will be compared against this baseline. It is not straightforward to relate the 2DEnVAR + EnKF scheme to the forcing perturbations of the ensemble of forcing perturbations from HarmonEPS. A general problem is that the atmospheric perturbations are only weakly related to the prognostic variables inside soil, lake or snow. E.g. we may need a perturbation of a deep lake temperature instead of a screen-level temperature in the nearby forest or over the frozen lake surface.

6. 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 ?

Externalisation of DA environment making it independent of SURFEX is a natural extension/continuation of the development which is in line with the OOPS paradigm adopted for UA DA. In HIRLAM we know very little about PALM and how compatible this system is with the OOPS developments. On one side we are short of manpower and we should avoid duplicating work as much as possible. We should check out PALM and see what it could offer. It might be inefficient to develop SODA with a PALM like functionalities and keep the design open to all-kinds of DA algorithms. On another side, as experience with OOPS shows, a new design tool always come at price of work needed to introduce it, to learn it and to maintain it. OOPS is the ultimate goal. The HARMONIE experience with OOPS induced implications on the code refactoring indicates that the externalisation is a longer term solution. A middle term solution must exist as well. We stress again that our long term goal is the coupled SU+UA DA utilizing (most probably) the 4DEnVAR algorithm under OOPS framework.

7. 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.

Due to a number of reasons land surface data assimilation lies several years behind the state-of-art of the upper air data assimilation. The handling of observations is totally unsatisfactory. The potential of the remote sensing observations is explored very poorly in the HARMONIE community and the situations should be improved urgently. Some first suggestions have emerged in recent lake studies, see e.g. Kheyrollah Pour et. al., 2014 and Kurzeneva, 2014 (both under review for Tellus A lake issue, available upon request from the authors). The observations and spatialization are parts of the DA system. Now we have three DA systems: UA (VAR), skin (OI), and surface (OI,EKF). It would be nice to start the journey towards a unified system by bringing the skin and surface together with an EnsVar-system as a proof of concept.

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

In the HIRLAM community we agree that a close collaboration between the data assimilation scientists and the system people is essential.

Proposal from the HIRLAM side

1. ***Arrange a subgroup for surface data assimilation discussion, based on the questions raised by Jean-Francois Mahfouf, during the ALADIN-HIRLAM All Staff Workshop in Bucharest***
2. ***Arrange a one-two day Workshop on the status and further coordination of evolution of the land surface DA (coordinated by ALADIN/MF ?) in fall 2014 somewhere in central Europe.***