A land surface analysis scheme based on an Extended Kalman Filter within SURFEX

Jean-François MAHFOUF

Météo-France/CNRM

March 16, 2009

Scientific objectives (1)

- Improve the analysis of soil prognostic variables (temperature and soil moisture content) in NWP models (ALADIN, AROME, ARPEGE)
- Current status in ALADIN consortium : either interpolation from ARPEGE soil analysis or dedicated soil analysis based on optimal interpolation (CANARI) using screen-level temperature and relative humidity (Mahfouf, 1991 ; Giard and Bazile, 2000)

Scientific objectives (2)

- Main weaknesses of the current soil analysis: difficult to assimilate new observation types and to initialize additional prognostic variables :
  - Satellite data informative about soil moisture state (AMSR-E, ERS, ASCAT, SMOS)
  - Precipitation analyses (raingauges, radars) and surface downward radiation fluxes derived from satellite measurements (LandSAF)
  - New versions of the land surface scheme ISBA (additional soil layers, inclusion of photosynthesis and plant growth)
- Proposal : Extended Kalman Filter (EKF) (*Simplified 2D-Var of Balsamo et al. (2004)*) in the offline version of SURFEX with a 6-hour assimilation window.

Definitions

- Control vector $\mathbf{x}$ (dimension N_x) : prognostic variables of the land surface scheme ISBA $\mathbf{x} = (w_g, w_2, T_s, T_2)$ $[N_x = 4]$
- Forward model $\mathcal{M}$: land surface scheme ISBA (`OFFLINE`) : $\mathbf{x}^t = \mathcal{M}(\mathbf{x}^0)$
- Observations $\mathbf{y}_o$ (dimension N_y) : T_{2m}, RH_{2m}, w_g with a covariance matrix of observation errors $\mathbf{R}$
- Observation operator $\mathcal{H}$: Model counterpart of observations : $\mathbf{y}^t = \mathcal{H}(\mathbf{x}^t)$ (e.g. vertical interpolation scheme in the SBL)
- Background state : short-range (6-h) forecast of $\mathbf{x}$ ($\mathbf{x_f}^t$) with a covariance matrix of background errors $\mathbf{B}$

Remark : In the SURFEX-EKF the observation operator $\mathcal{H}$ also includes the forward model propagation : $\mathbf{y}^t = \mathcal{H}(\mathbf{x}^0)$

EKF equations (1)

An analysis state $\mathbf{x_a}^t$ is given by an optimal combination (minimum variance) of the observations and the background (short-range forecast) :

$$\mathbf{x_a}^t = \mathbf{x_f}^t + \mathbf{BH}^T(\mathbf{HBH}^T + \mathbf{R})^{-1}(\mathbf{y}_o^t - \mathcal{H}(\mathbf{x_f}^t)) \quad (1)$$

whre $\mathbf{H}$ is the Jacobian matrix $(N_y \times N_x)$ of the observation operator $\mathcal{H}$:

$$\mathbf{H}_{ij} = \frac{\partial \mathbf{y}_i}{\partial \mathbf{x}_j} \quad (2)$$

Remark: in SURFEX, we use a finite difference approach where the input vector $\mathbf{x}$ is perturbed N_x times to get for each `OFFLINE` integration a column of the matrix $\mathbf{H}$:

$$\mathbf{H}_{ij} \simeq \frac{\mathbf{y}_i(\mathbf{x} + \delta x_j) - \mathbf{y}_i(\mathbf{x})}{\delta x_j} \quad (3)$$

where δx_j is a small increment value added to the j-th component of the $\mathbf{x}$ vector (defined in `OPTIONS.nam`).

EKF equations (2)

The analysis state in characterized by an analysis error covariance matrix:

$$\mathbf{A} = (\mathbf{I} - \mathbf{KH})\mathbf{B} \tag{4}$$

where $\mathbf{K}$ is the Kalman gain matrix

The analysis is cycled by propagating the time the two quantities $\mathbf{x_a}$ et $\mathbf{A}$ up to next time where observations are available :

$$\mathbf{x_f}^{t+1} = \mathcal{M}(\mathbf{x_a}^t) \tag{5}$$

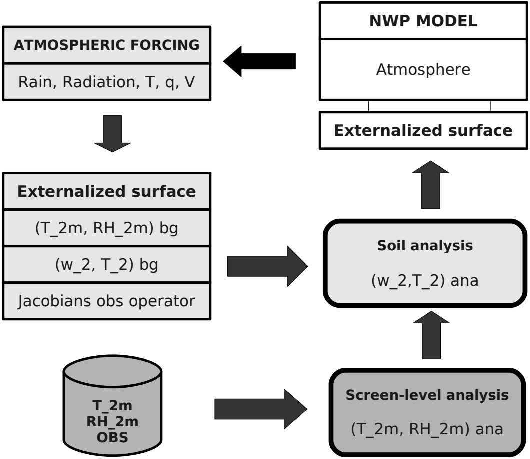
$$\mathbf{B}^{t+1} = \mathbf{M}\mathbf{A}^t\mathbf{M}^T + \mathbf{Q} \tag{6}$$

Jacobian matrix $\mathbf{M}$ of the forward model $\mathcal{M}$, between time t and time $t = 0$ (obtained like $\mathbf{H}$):

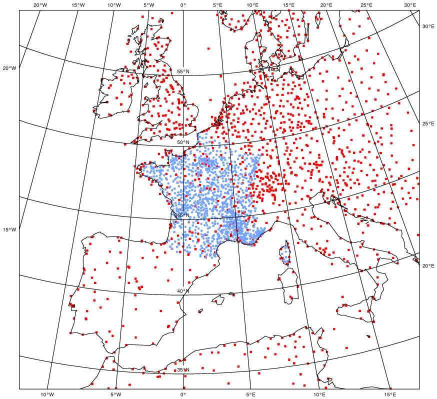
$$\mathbf{M}_{ij} = \frac{\partial x_i^t}{\partial x_j^0} \tag{7}$$

A new matrix $\mathbf{Q}$, representing the model error covariance matrix, needs to be defined.

Coupling between surface and atmospheric analyses

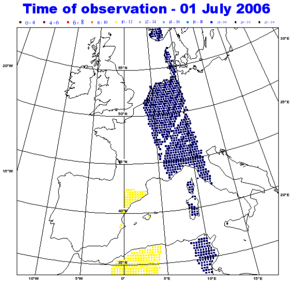


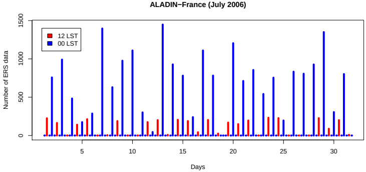
Surface network over ALADIN-France domain



blue = RADOME (1000) - red = SYNOP (1000)

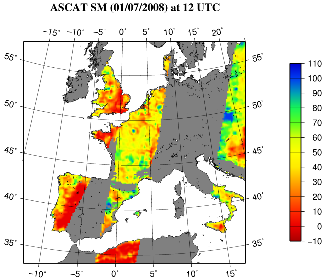
ERS superficial soil moisture (1993-present)

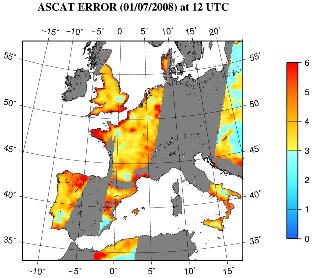




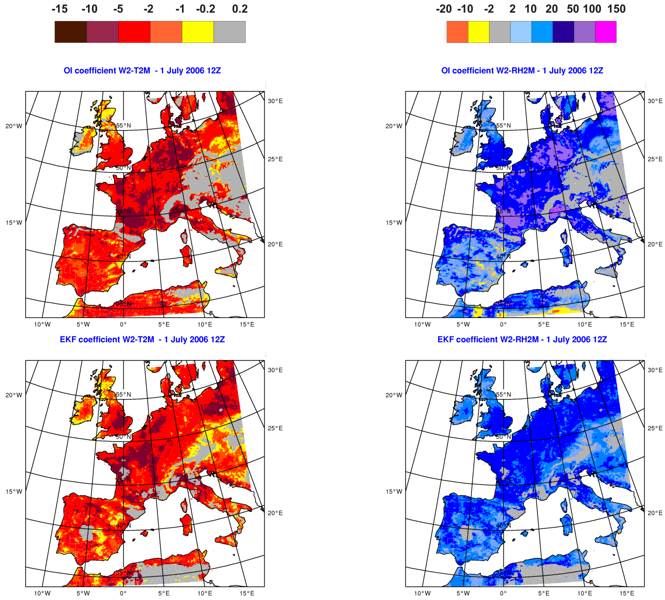
Requirements : projection on model grid - error specification - bias correction

ASCAT superficial soil moisture (2007-present)

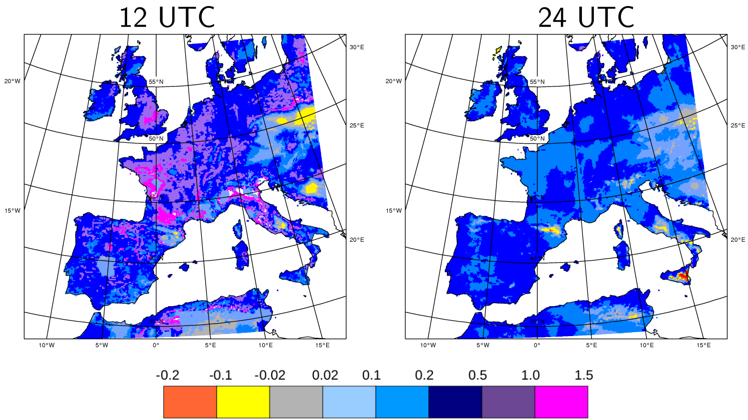




EKF gain matrix vs. OI coefficients

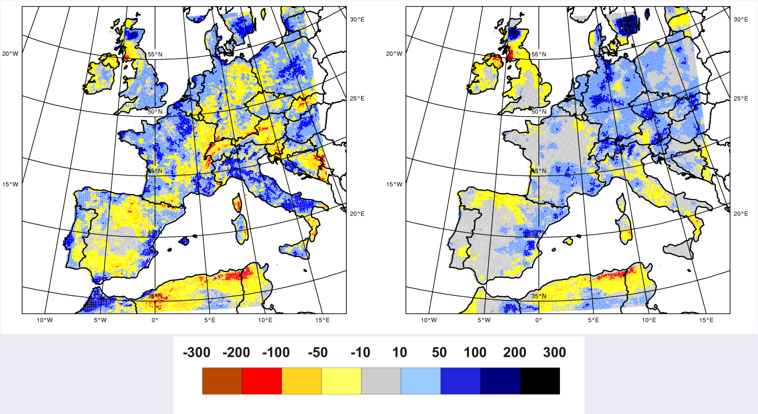


Jacobian $\partial w_g / \partial w_2$



Root-zone soil moisture increments (mm) July 2006

assimilation ($T_{2m}+RH_{2m}+w_g$) - assimilation (w_g)



Pratical aspects (1)

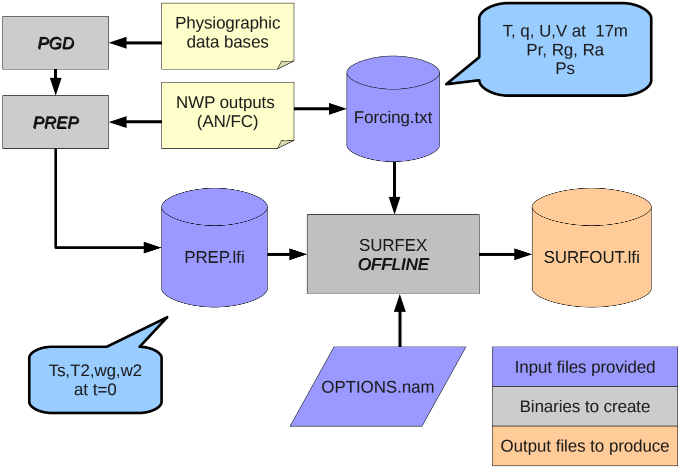
- Creation of executable files to run SURFEX (`OFFLINE`) and the EKF (`VARASSIM`)
- The binary `OFFLINE` is created automatically when installing SURFEX.
- The binary `VARASSIM` is generated from the FORTRAN files contained in the sub-directory `VARASSIM` of the directory `$SURFEX_EXPORT/src`.
- The subdirectory `MYSRC` contains the subroutines that you have modified with respect to the standard versions.
- The file `Makefile.SURFEX.mk` need to be changed in order to add the creation of `VARASSIM`

Practical aspects (2)

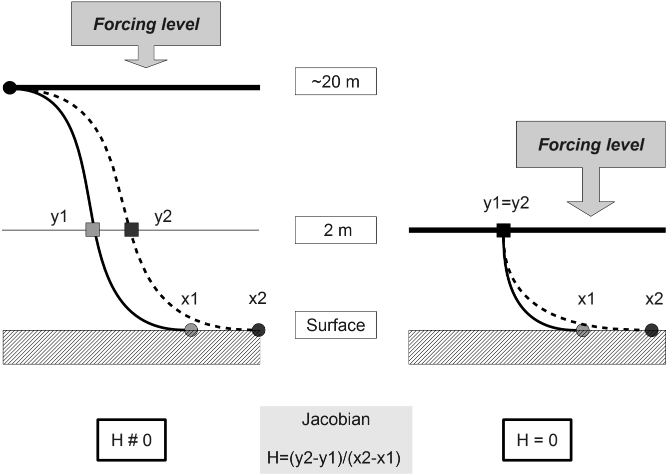
The directory `VARASSIM` contains the following files :

- `varassim.f90` : main program that performs the various steps of the assimilation : definition of initial perturbed states, reading of fields from SURFEX outputs, writing of fields necessary for the analysis, and finally the surface analysis.
- `choldc.f90` : Cholesky decomposition (part I)
- `cholsl.f90` : Cholesky decomposition (part II)
- `inverse_matrix.f90` : explicit computation of an inverse matrix after Cholesky decomposition.
- `trans_chaine.f90` : Transformation of an integer into a character .
- `get_file_name.f90` : gets the name of files for the current assimilation window.

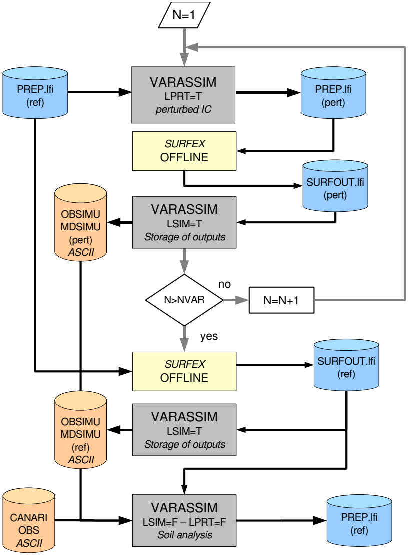
Flowshart of SURFEX offline



Jacobians in the SBL



Flowshart of the main script `run_ekf.sh`



Namelist OPTIONS.nam

Namelist block	Variable	Type	Description
NAM_IO_VARASSIM	LPRT*	**F**	to perform analysis
		T	to define δx_i and store $\mathbf{x} + \delta x_i$ at t=0
	LSIM*	**F**	to perform analysis
		T	to write the simulated observations $\mathcal{H}(\mathbf{x})$ and the evolved state vector $\mathbf{x}$
	LBEV*	**F**	to perform analysis
		T	to evolve of the **B** matrix
	LBFIXED	F	to evolve of the **B** matrix
		T	to keep the **B** matrix constant with time
NAM_OBS	NOBSTYPE	integer	Number of possible observation types *This value must be consistent with the obs file*
	YERROBS(1)	real	Observation error for T_{2m} in K
	YERROBS(2)	real	Observation error for RH_{2m} (no units)
	YERROBS(3)	real	Observation error for w_g (fraction of SWI)
	INCO(i)	integer	1 if observation type included
			0 if observation type excluded
NAM_VAR	IVAR*	**1**	Control variable of interest
	NVAR*	**1**	Number of control variables (dimension of control vector)
	XVAR_M(i)	character	Control variable identifier in PREP file
	PREFIX_M(i)	character	Control variable prefix in PREP.txt file
	XSIGMA_M(1)	real	(Initial) BG error for w_2 (fraction of SWI)
	XSIGMA_M(2)	real	(Initial) BG error for w_g (fraction of SWI)
	XSIGMA_M(3)	real	(Initial) BG error for T_s (K)
	XSIGMA_M(4)	real	(Initial) BG error for T_2 (K)
	TPRT_M(1)	real	Size of perturbation of w_2 for finite Jacobians The perturbation δx writes $x\times$ TPRT_M
	TPRT_M(2)	real	Size of perturbation of w_g for finite Jacobians
	TPRT_M(3)	real	Size of perturbation of T_s for finite Jacobians
	TPRT_M(4)	real	Size of perturbation of T_2 for finite Jacobians
	INCV(i)	integer	1 if element of control vector included
			0 if element of control vector excluded
	SCALE_Q	real	Definition of the matrix **Q** of model errors as fraction of the initial diagonal **B** matrix

Table 1: Description of each variable in the namelist OPTIONS.nam for the blocks relative to the Land Data Assimilation System. The elements with stars (*) should be kept at their value in bold - their actual values are defined by the script run_ekf.sh

Directory structure

A number of directories should be created and/provided :

- repforcing : Directory where the forcing data are stored (9pt sample provided in ASCII for July 2008)
- represults : Directory where the results will be stored
- reprun : Directory where the script run_ekf.sh is executed
- repobs : Directory where the observations are stored (9pt sample provided in ASCII for July 2008 with SYNOP and ASCAT data)
- repnamel : Directory where the namelist is located (namelist OPTIONS.nam provided)
- repanalyse : Directory where the initial conditions are stored (PREP_9pts.lfi provided for a set of 9 points within the ALADIN-France domain on 02 July 2008 at 00Z)
- repbin : Directory where the binary files to execute SURFEX and the EKF are located