Land data assimilation within SURFEX

Jean-François MAHFOUF [1]

Météo-France/CNRM/GMME/TURBAU

16 October 2009

---

[1] jean-francois.mahfouf@meteo.fr

Outline

- Scientific objectives
- Main features of a data assimilation system
- Land data assimilation within SURFEX
- Illustration with few examples
- Practical aspects
- Conclusions

Scientific objectives (1)

- Define the initial state of SURFEX prognostic variables (available in the `PREP.txt` file) by using observations informative about the soil/vegetation state.
- Applications :
  - Improved initial conditions for NWP atmospheric models using SURFEX (ARPEGE, ALADIN, AROME, ALARO, HARMONIE) [semi-coupled mode - forcing from NWP forecasts]
  - Improved description of the land carbon and water fluxes by constraining SURFEX with both realistic surface forcing (precipitation and radiation fluxes ; e.g. SAFRAN) and available observations [offline mode - forcing from observations or analyses]

Scientific objectives (2)

Recent satellite data informative about surface properties :

- Superficial soil moisture :
    - Passive microwave instruments (radiometers) : TMI/TRMM (X-band), AMSR-E/Aqua (C-band), WindSat/Coriolis (C-band), SMOS (L-band), SMAP (L-band)
    - Active microwave instruments (radars-scatterometers) : SCAT/ERS (C-band), ASCAT/MetOp (C-band)
- Leaf area index : SEVIRI/MSG (LandSAF), SPOT-VEGETATION, MODIS/Aqua
- Skin surface temperature : SEVIRI/MSG (LandSAF)
- Surface albedo : SEVIRI/MSG (LandSAF), MODIS/Aqua
- Snow cover extent : SEVIRI/MSG (LandSAF), MODIS/Aqua

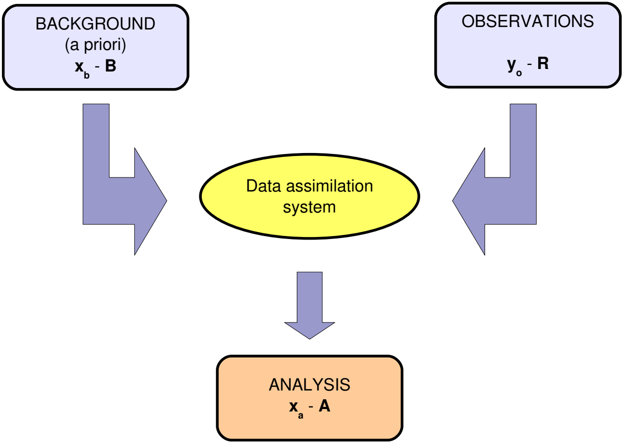
General features of data assimilation

- Optimal combination of an a-priori information (usually provided by a short-range forcecast) defined as $\mathbf{x}_b$ with an associated error σ_b (matrix of covariance errors $\mathbf{B}$ [2]) and available observations $\mathbf{y}_o$ with an associated error σ_o (matrix of covariance errors $\mathbf{R}$). The optimality criteria is usually defined as the minimum variance estimate.
- A number of hypotheses are also made in order to get the analysis for pratical applications : errors between the background and the observations are assumed to be uncorrelated, to have a zero mean and to be gaussian. The analysis is expressed as a linear combination of the background and of the observations. This state is called the "Best Linear Unbiased Estimate".

---

[2] defined as $\overline{(\mathbf{x}_b - \mathbf{x}_t)(\mathbf{x}_b - \mathbf{x}_t)^T}$

Schematic view of data assimilation (1)



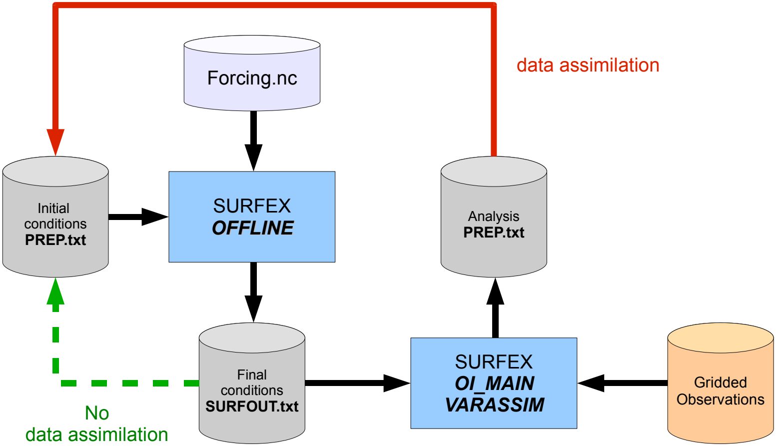
BACKGROUND
(a priori)
$\mathbf{x}_b$ - $\mathbf{B}$

OBSERVATIONS
$\mathbf{y}_o$ - $\mathbf{R}$

Data assimilation system

ANALYSIS
$\mathbf{x}_a$ - $\mathbf{A}$

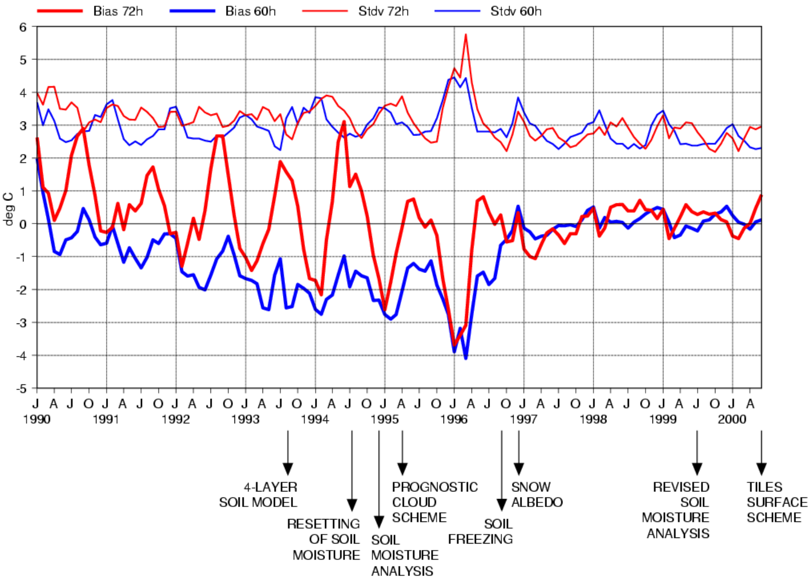
Schematic view of data assimilation (2)



Land data assimilation for NWP applications

- Provide an analysis of the prognostic variables described by land surface schemes (soil moisture contents and soil temperatures)
- Specific features of land data assimilation : long time scales w.r.t. atmospheric variables, lack of direct observations
- Possible solutions :
  - Past : soil climatology (no interannual variability), no dedicated analysis (soil state can drift)
  - Current : offline land surface schemes (NCEP and UKMO), assimilation of screen-level observations (T_{2m} and RH_{2m} from SYNOP) (MF, ECMWF, DWD, CMC, ALADIN, HIRLAM).
  - Future : improved assimilation techniques in order to allow the use of satellite observations (AMSR-E, ERS, ASCAT, SMOS), and of new soil/vegetation variables (e.g. vegetation biomass)

ECMWF forecast scores of T_{2m} over Europe



Going back to basics with scalars

x_b is an a-priori information with an error σ_b and y_0 an observation with an error σ_o, the analysis x_a is searched as a linear combination of those two pieces of information. In practice the analysis is performed in terms of *increments* (departure from the background), that is:

$$x_a - x_b = \alpha(y_0 - x_b)$$

The coefficient α (*optimal interpolation coefficient*) can be found by computing the variance of x_a and minimizing w.r.t. α:

$$\sigma_a^2 = (1 - \alpha)^2 \sigma_b^2 + \alpha^2 \sigma_o^2$$

Then if :

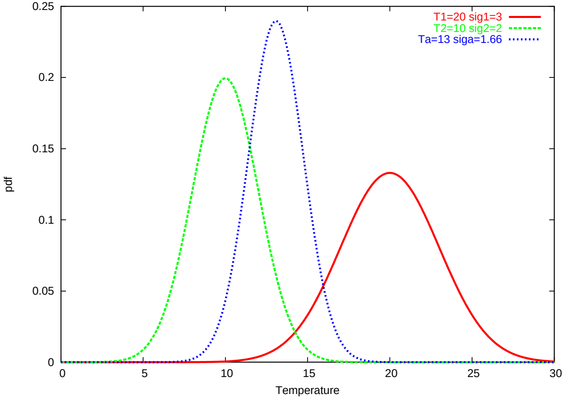
$$\frac{\partial \sigma_a^2}{\partial \alpha} = 0 \quad \Rightarrow \quad -(1 - \alpha)\sigma_b^2 + \alpha \sigma_o^2 = 0.$$

Finally

$$\alpha = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_o^2}$$

Simple analysis problem

Example : $x_b = T_1 = 20$ K with $\sigma_b = 3$ K, $y_o = T_2 = 10$ K with $\sigma_o = 2$ K then the analysis is $x_a = T_a = 13$ K with $\sigma_a = 1.66$ K



The variational equivalent

The analysis x_a is found as the minimum of a cost-function measuring the departure between the background x_b and the observation y_o (quadratic distance weighted by the corresponding errors):

$$J(x) = \frac{1}{2}\left[\frac{x - x_b}{\sigma_b}\right]^2 + \frac{1}{2}\left[\frac{y_o - x}{\sigma_o}\right]^2$$

- Show that the solution of $J(x) = 0$ is the same as before.
- Compute the analysis error σ_a defined as the inverse of the Hessian of J.

Being more general

- scalars describing model state x and observations y are replaced by vectors
- the (scalar) error variances are replaced by error covariance matrices
- the model state needs to be projected into the observation state $\mathcal{H}$ (not necessarily linear)
- the Jacobian matrix of $\mathcal{H}$ and its transpose ($\mathbf{H}$ and $\mathbf{H}^T$) allow linear transforms from one space to another

Current developments within SURFEX

- Data assimilation methods (derived from the Kalman filter theory):
  - Optimum Interpolation : "analytical" OI coefficients (similar to NWP operational soil analyses) - `OI_MAIN`
  - Extended Kalman Filter with and without cycling of the **B** matrix (EKF, SEKF) : "dynamical" OI coefficients (i.e. Kalman gain matrix) - Jacobians in finite differences - `VARASSIM`
  - Ensemble Kalman Filter : "dynamical" optimum interpolation coefficients - derived from the statistics of an ensemble of model integrations
- Analysis variables : prognostic variables of ISBA standard and ISBA-A-gs
- Observations : screen-level temperature and relative humidity (SYNOP data), superficial soil moisture (SCAT, ASCAT, AMSR-E), Leaf area index

Example: the EKF with ISBA-2L

- 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 $\mathbf{x}$ ($\mathbf{x_b}^t$) with a covariance matrix of background errors $\mathbf{B}$

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) :

$${\mathbf{x_a}}^t = {\mathbf{x_b}}^t + \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}(\mathbf{y}_o^t - \mathcal{H}({\mathbf{x_b}}^t))$$

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

$$\mathbf{H}_{ij} = \frac{\partial \mathbf{y}_i}{\partial \mathbf{x}_j}$$

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 $\mathbf{H}$:

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

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:

$$\mathbf{A} = (\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}$$

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 :

$$\mathbf{x_b}^{t+1} = \mathcal{M}(\mathbf{x_a}^t)$$

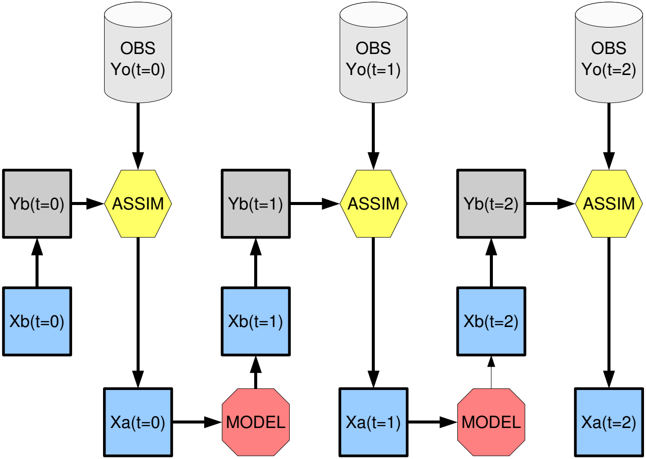
$$\mathbf{B}^{t+1} = \mathbf{M}\mathbf{A}^t\mathbf{M}^T + \mathbf{Q}$$

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

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

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

Data assimilation cycling



Optimum interpolation

$$
\begin{aligned}
w_g^a &= w_g^b + \alpha_1(T_{2m}^o - T_{2m}^b) + \alpha_2(RH_{2m}^o - RH_{2m}^b) \\
w_2^a &= w_2^b + \beta_1(T_{2m}^o - T_{2m}^b) + \beta_2(RH_{2m}^o - RH_{2m}^b) \\
T_g^a &= T_g^b + \mu_1(T_{2m}^o - T_{2m}^b) \\
T_2^a &= T_2^b + \mu_2(T_{2m}^o - T_{2m}^b)
\end{aligned}
$$

An analytical expression of the OI coefficients has been given by Giard and Bazile (2000) for the soil analysis at Météo-France (ALADIN), CMC and HIRLAM and by Douville et al.(2000) for the ECMWF soil analysis.

The Ensemble Kalman filter

The same equation as for the EKF is solved, but the matrices $\mathbf{BH}^T$ and $\mathbf{HBH}^T$ are obtained from an ensemble of predictions $(\mathbf{x_f})_i$ with i varying between 1 et N. One writes :

$$\mathbf{BH}^T \approx \overline{(\mathbf{x_b} - \overline{\mathbf{x_b}})(\mathcal{H}(\mathbf{x_b}) - \overline{\mathcal{H}(\mathbf{x_b})})^T}$$

$$\mathbf{HBH}^T \approx \overline{(\mathcal{H}(\mathbf{x_b}) - \overline{\mathcal{H}(\mathbf{x_b})})(\mathcal{H}(\mathbf{x_b}) - \overline{\mathcal{H}(\mathbf{x_b})})^T}$$

with the average operators :

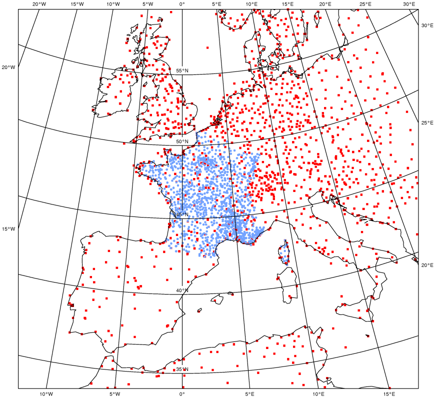
$$\overline{\mathbf{x}} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{x}_i \qquad \overline{\mathbf{x}\mathbf{y}^T} = \frac{1}{N-1}\sum_{i=1}^{N} \mathbf{x}_i \mathbf{y}_i^T$$

An advantage of this approach is that there is no need to get explicitely the various matrices entering in the computation of the Kalman gain.
A difficulty is the definition of the size of the ensemble (to get robust statistics) and of model errors in order to have an ensemble with enough spread.

Summary of current studies

- Comparative study over the ALADIN-France domain with the (S)EKF and the Météo-France OI using screen-level observations in July 2006 (Mahfouf et al., 2009; JGR)
- Assimilation of AMSR-E superficial soil moisture with the EKF and the (S)EKF over the ALADIN-France domain in July 2006 (Draper et al., 2009; JGR)
- Assimilation of ASCAT superficial soil moisture with the SEKF over the ALADIN-France domain coupled to the ALADIN 3D-Var (Mahfouf, 2009; QJRMS)
- Local scale assimilation of superficial soil moisture content and Leaf Area Index over the SMOSREX experimental site (southwestern France) (Sabater et al., 2008; AFM)

Surface network over ALADIN-France domain



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

Comparaison of two methods

Optimal interpolation

Analytical coefficients derived from Monte-Carlo experiments with a single column model (Bouttier *et al.*, 1993; Giard and Bazile, 2000)

EKF

Dynamical coefficients :

$$\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}$$

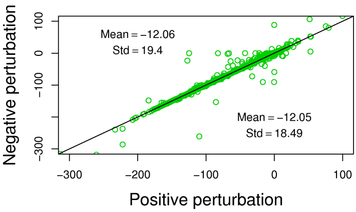
where the matrices **B** and **R** are prescribed and where the Jacobian of the observation operator **H** is obtained in finite differences :

$$\mathbf{H} \approx \frac{\mathbf{y}(w_2 + \Delta w) - \mathbf{y}(w_2)}{\Delta w}$$

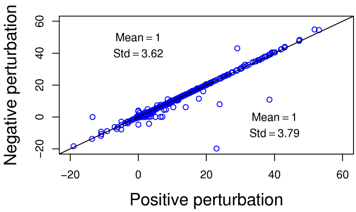
Linearity of the Jacobians

$[\partial T_{2m}/\partial w_2]$ $[\partial RH_{2m}/\partial w_2]$ $[\partial T_{2m}/\partial w_g]$ $[\partial RH_{2m}/\partial w_g]$

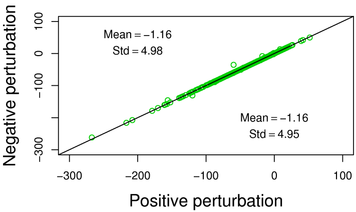
Jacobian dT_2m−dw_2 (K/m3/m3)



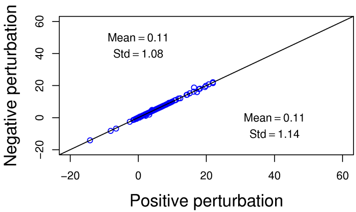
Jacobian dRH_2m−dw_2 (1/m3/m3)



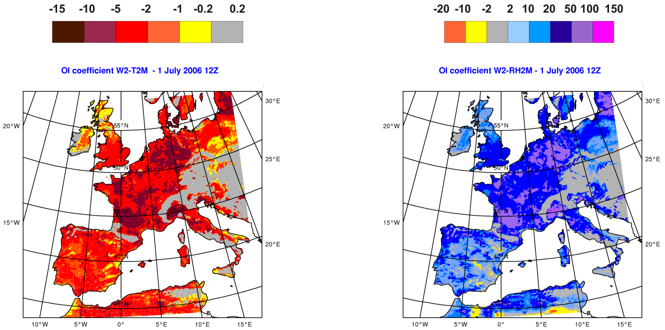
Jacobian dT_2m−dw_g (K/m3/m3)



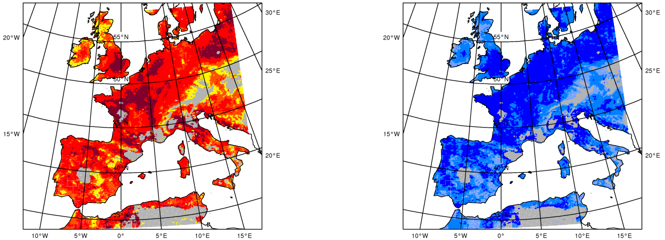
Jacobian dRH_2m−dw_g (1/m3/m3)



EKF gain matrix vs. OI coefficients


OI coefficient W2-T2M - 1 July 2006 12Z

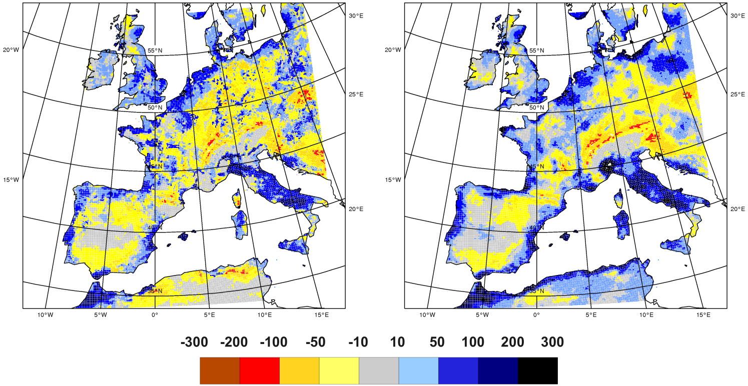
OI coefficient W2-RH2M - 1 July 2006 12Z


EKF coefficient W2-T2M - 1 July 2006 12Z

EKF coefficient W2-RH2M - 1 July 2006 12Z

Deep soil moisture increments in July 2006 (mm)

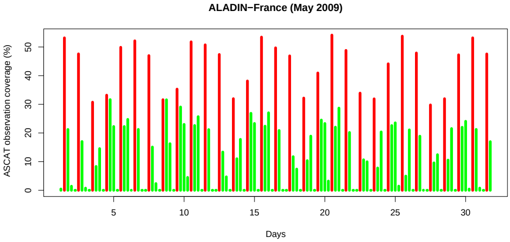
EKF increments - OI increments

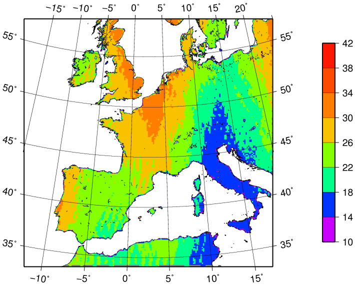


Assimilation of ASCAT soil wetness in ALADIN-France

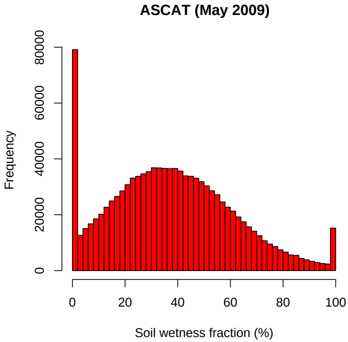
- Test : 3D-Var ALADIN with SURFEX (CY33T1)
- CANARI analysis for 2-m observations : soil temperature corrections
- Use of ASCAT products to correct soil moisture
- Period : 01/05/2009 - 28/05/2009
- Realistic background and observation errors - CDF matching bias correction scheme
- Data screening : topography index larger than 15 %, retrieval errors larger than 5 %, inland water bodies, urban fraction larger than 15 %
- Background check : data rejected when innovation too large

Availability of ASCAT products

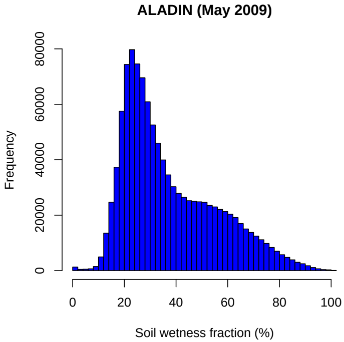




Comparaison w_g ASCAT vs. w_g ALADIN



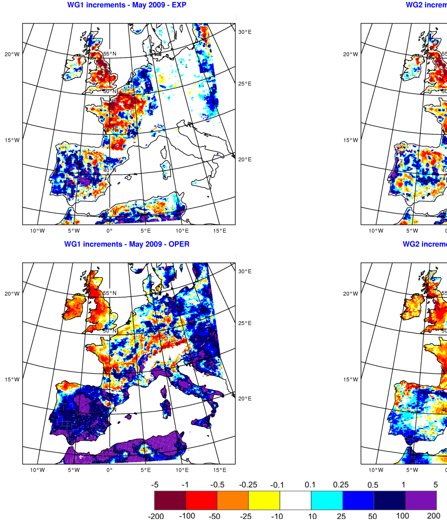
ASCAT (May 2009)



ALADIN (May 2009)

Root zone soil moisture increments in mm (May 2009)

OPER $(T_{2m} - RH_{2m})$ - EXP (w_g/w_{sat})



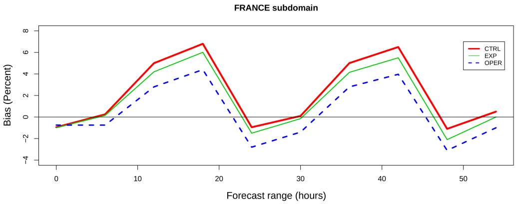
WG1 increments - May 2009 - EXP

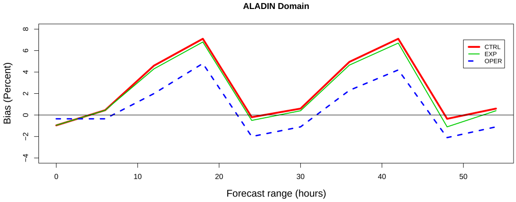
WG2 increments - May 2009 - EXP

WG1 increments - May 2009 - OPER

WG2 increments - May 2009 - OPER

Forecast scores (RH_{2m})





CTRL=no soil moisture analysis

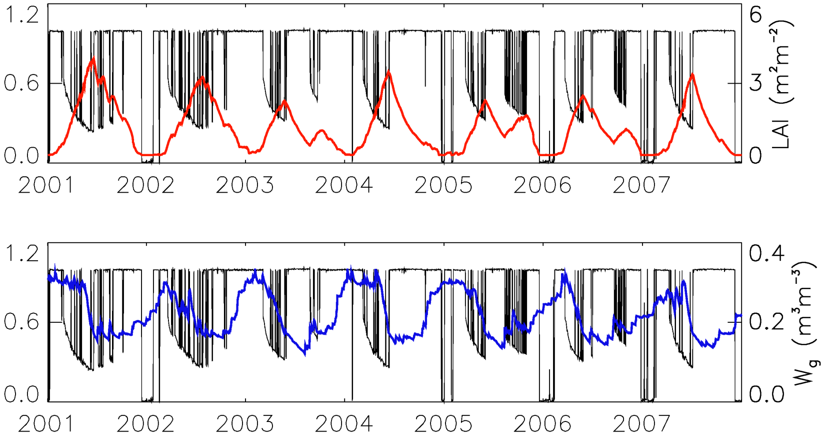
EXP=soil moisture analysis from ASCAT

OPER=soil moisture analysis from T_{2m} and RH_{2m}

Assimilation of w_g and LAI at local scale

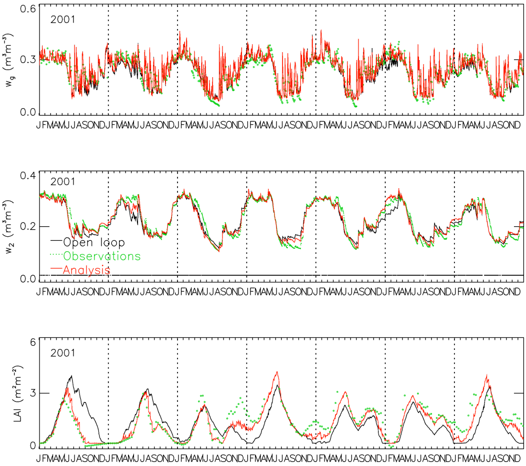
- Land surface scheme : ISBA A-gs (NIT version)
- Location : SMOSREX (experimental site near Toulouse)
- Period : 2001-2007
- Forcing : observations
- Observations : in-situ measurements of w_g (every 3 days) and LAI every 10 days
- Land data assimilation system : SEKF with w_2 and LAI as control variables
- Observation and background errors : $\sigma^o_{w_g}$=0.06 m^3/m^3, $\sigma^b_{w_2}$=0.02 m^3/m^3, $\sigma^o_{LAI}=0.2\times LAI_o$, $\sigma^b_{LAI}=0.2\times LAI_b$
- Experiments : FORCING_1 = observations / FORCING_2 = observations but rain set to zero

Jacobians $\partial LAI^t / \partial LAI^0$



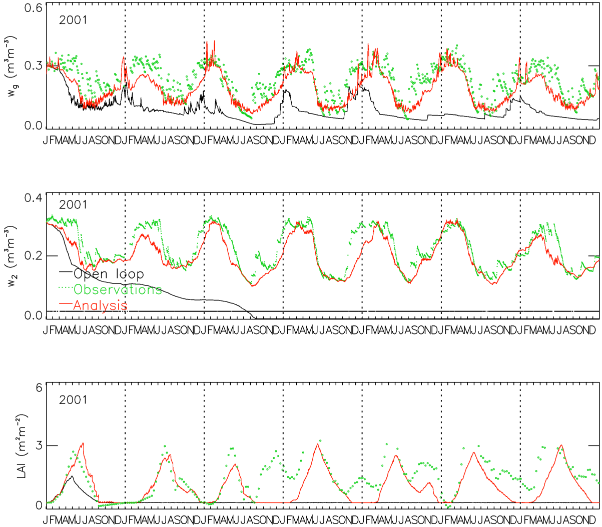
From Rüdiger et al. (2009; JGR)

FORCING_1



Albergel et al. (2009)

FORCING_2



Albergel et al. (2009)

Practical aspects (1)

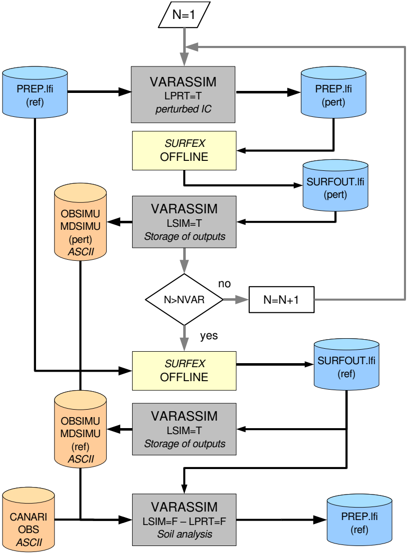
- Creation of executable files to run SURFEX (`OFFLINE`) and the analysis schemes : EKF (`VARASSIM`) and/or OI (`OI_MAIN`)
- The binaries `OFFLINE` and `OI_MAIN` are created automatically when installing SURFEX (from V4.0)
- The binary `VARASSIM` is generated from the FORTRAN files contained in the sub-directory `VARASSIM` of the directory `$SURFEX_EXPORT/src` (package available from J.-F. Mahfouf)
- The file `Makefile.SURFEX.mk` needs 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 fields from SURFEX outputs, writing of fields necessary for the analysis in temporary files, 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 the main script `run_ekf.sh`



Namelist `OPTIONS.nam` (1)

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

Namelist `OPTIONS.nam` (2)

Namelist block	Variable	Type	Description
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

Ongoing collaborations

- ALADIN/HIRLAM consortia : assimilation of T_{2m} and RH_{2m} with SURFEX OI for NWP models with SURFEX
- Météo-France and ZAMG : assimilation of ASCAT derived superficial soil moisture w_g with SURFEX EKF/(S)EKF in ALADIN
- Météo-France and University of Melbourne : assimilation of AMSR-E derived superficial soil moisture w_g with SURFEX EKF/(S)EKF in ALADIN
- Météo-France : assimilation of LAI and w_g with SURFEX/SEKF and ISBA-A-gs over France (FP7 GEOLAND2)
- MetNo/NILU : assimilation of T_{2m} and RH_{2m} with SURFEX EnKF (then satellite derived soil moisture from SMOS)

Areas of research

- Combined assimilation of conventional and satellite observations
- Inclusion of observed precipitation and radiation fluxes in the EKF
- Comparative studies of EKF vs. EnKF (downscaling issues)
- Improved specification of background and model errors (adaptive techniques)
- Coupling of SURFEX with a surface radiative transfer model for the assimilation of microwave brightness temperatures

Bibliography

Balsamo, G., F. Bouyssel, and. J. Noilhan, 2004: A simplified bi-dimensional variational analysis of soil moisture from screen-level observations in a mesoscale numerical weather prediction model. *Quart. J. Roy. Meteor. Soc.*, **130**, 895-915.

Balsamo, G., J.-F. Mahfouf, S. Bélair, and G. Deblonde, 2006: A global root-zone soil moisture analysis using simulated L-band brightness temperature in preparation for the HYDROS satellite mission. *J. Hydrometeor.*, **7**, 1126-1146.

Balsamo, G., J.-F. Mahfouf, S. Bélair, and G. Deblonde, 2007: A land data assimilation system for soil moisture and temperature : an information content study. *J. Hydrometeor.*, **8**, 1225-1242

Crow, W.T., and E.F. Wood, 2003: The assimilation of remotely sensed brightness temperature imagery into a land surface model using ensemble Kalman filtering: A case study based on ESTAR measurements during SGP97. *Adv. Water Resour.*, **26**, 137-149.

Douville, H., P. Viterbo, J.-F. Mahfouf, and A.C.M. Beljaars, 2000: Evaluation of optimal interpolation and nudging techniques for soil moisture analysis using FIFE data. *Mon. Wea. Rev.*, **128**, 1733-1756.

Draper, C.S., J.-F. Mahfouf, anf J.P. Walker, 2009: An EKF assimilation of AMSR-E soil moisture into the ISBA land surface scheme. *J. Geophys. Res.*, doi:10.1029/2008JD011650

Drusch, M., and P. Viterbo, 2007: Assimilation of screen-level variables in ECMWF's Integrated Forecast System: A study on the impact on the forecast quality and analyzed soil moisture. *Mon. Wea. Rev.*, **135**, 300-314.

Giard, D., and E. Bazile, 2000: Implementation of a new assimilation scheme for soil and surface variables in a global NWP model. *Mon. Wea. Rev.*, **128**, 997-1015.

Hess, R., 2001: Assimilation of screen-level observations by variational soil moisture analysis. *Meteor. Atmos. Phy.*, **77**, 145-154.

Mahfouf, J.-F., 1991: Analysis of soil moisture from near-surface parameters: A feasibility study. *J. Appl. Meteor.*, **30**, 506-526.

Mahfouf, J.-F., 2009: Assimilation of satellite derived soil moisture from ASCAT in an limited area NWP model. *Quart. J. Roy. Meteor. Soc.* (submitted)

Mahfouf, J.-F., K. Bergaoui, C. Draper, F. Bouyssel, F. Taillefer, and L. Taseva, 2009: A comparison of two off-line soil analysis schemes for assimilation of screen level observations. *J. Geophys. Res.*, **114**, D80105, doi:10.1029/2008JD011077

Muñoz-Sabater, J., L. Jarlan, J.-C. Calvet, F. Bouyssel, and P. De Rosnay, 2007: From near-surface to root-zone soil moisture using different assimilation techniques. *J. Hydrometeor.*, **8**, 194-206.

Muñoz-Sabater, J., C. Rüdiger, J.-C. Calvet, N. Fritz, L. Jarlan, and K.-H. Kerr, 2008: Joint assimilation of surface moisture and LAI observations into a land surface model. *Agric. For. Meteorol.*, **148**, 1362-1373

Reichle, R.H., J.P. Walker, R.D. Koster, and P.R. Houser, 2002: Extended versus ensemble Kalman filtering for land data assimilation. *J. Hydrometeor.*, **3**, 728-740.

Reichle, R. H., D. B., McLaughlin, and D. Entekhabi, 2002, Hydrologic data assimilation with the Ensemble Kalman Filter. *Mon. Wea. Rev.*, **130**, 103-114.

Rüdiger, C., C. Albergel, J.-F. Mahfouf, J.-C. Calvet, and J.P. Walker, 2009: Evaluation of Jacobians for Leaf Area Index data assimilation with an Extended Kalman Filter. *J. Geophys. Res.*, (in revision)

Seuffert, G., H. Wilker, P. Viterbo, M. Drusch, and J.-F. Mahfouf, 2004: The usage of screen-level parameters and microwave brightness temperature for soil moisture analysis. *J. Hydrometeor.*, **5**, 516-531.