Comparing the SEKF with the DEnKF on a land surface model

David Fairbairn, Alina Barbu, Emiliano Gelati, Jean-Francois Mahfouf and Jean-Christophe Caret

CNRM - Meteo France
Partly funded by European Union CORE-CLIMAX project

3rd December 2014

Aims and justification for study

- Simplified extended Kalman filter (SEKF, Mahfouf *et al.* (2009)) currently used within SURFEX framework to assimilate soil moisture observations in a land surface model;
- SEKF already improves soil moisture estimates compared with open loop (no assimilation);
- Ensemble Kalman filter (EnKF) may be a better DA method for the future;
- **Main aim of study is to compare the performance of the SEKF with an EnKF**.
- The two methods differ in the representation of the background-error covariance;
- We follow in the footsteps of similar work by Reichle and Koster (2003); Sabater *et al.* (2006); Mahfouf (2007).

Step by step development

Increase the domain size and model complexity as more knowledge is gained:

1. **Single site (SABRES in-situ obs)**;
2. **12 in-situ observation sites (SMOSMANIA network)**;
3. France domain;
4. New multi-layer model.

Land surface model: ISBA-Ags

- The surface layer ($WG1$) (top 1cm) is forced by precipitation and evaporation, and restored towards an equilibrium value (w_{geq});
- w_{geq} reached when gravity matches capillary forces;
- Forcing term causes rapid intermittent changes in $WG1$ (e.g. rainfall event), while restoration takes about a day.
- A deep layer $WG2$ (1-3m deep), called the 'root zone', exists below $WG1$;
- Root zone is also affected by vegetation transpiration and drainage;
- Changes in $WG2$ much slower than $WG1$, due to its larger depth;
- Includes vegetation dynamics (not important for this talk).

WG1 model and obs at Sabres site

Bias corrected (CDF matched) obs used for assimilation:

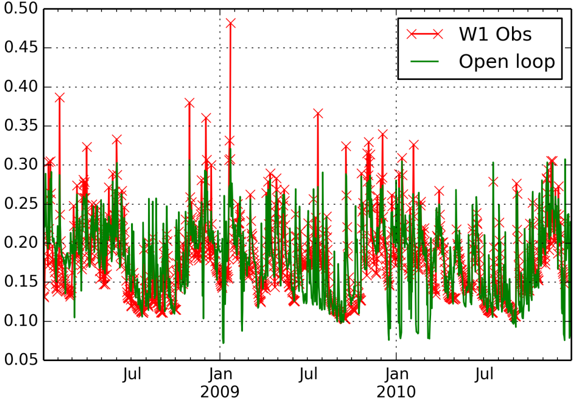

Figure 1: WG1 model and observations (m^3/m^3).

WG2 model and obs at Sabres site

Bias corrected (CDF matched) obs used for verification only:

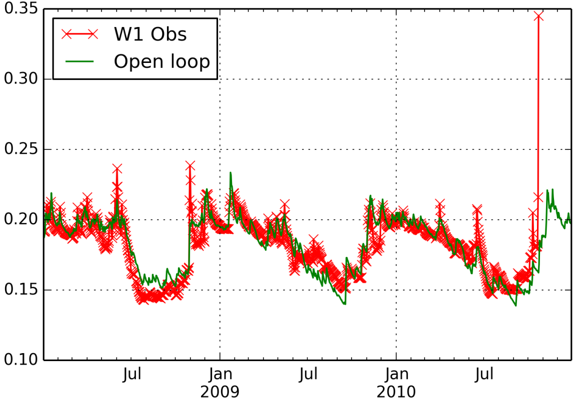

Figure 2: WG2 model and observations (m^3/m^3).

Land data assimilation system - LDAS

- Atmospheric forcing provided by mesoscale analysis at 8km resolution (SAFRAN);
- **Optimal interpolation** with screen-level temp/humidity obs provides soil moisture analyses for France NWP model (AROME);
- **Simplified extended Kalman filter (SEKF)** assimilates daily WG1 satellite obs (ASCAT) and LAI obs (SPOT-VEG) over France;
- **SEKF** used for monitoring carbon and water fluxes;
- Our study compares the **SEKF** with the **Deterministic Ensemble Kalman filter (DEnKF, Sakov and Oke (2008))**.

METEO FRANCE
Toujours un temps d'avance

Simplified Extended Kalman filter (SEKF)

- Background ($\mathbf{x}^b(t_i)$) is a nonlinear propagation of previous analysis:

$$\mathbf{x}^b(t_i) = M(\mathbf{x}^a(t_{i-1})) \tag{1}$$

- Analysis ($\mathbf{x}^a(t_i)$) is calculated using one gridpoint observation;
- Assimilated observation ($\mathbf{y}^o$) is weighted using Kalman gain ($\mathbf{K}$):

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{K}(\mathbf{y}^o - H(\mathbf{x}^b)), \tag{2}$$

where

$$\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}, \tag{3}$$

where $\mathbf{B}$ is a climatological and diagonal background-error covariance.

SEKF observation operator

- Model predicted observation $\mathbf{y} = H(\mathbf{x})$;
- Jacobian of observation operator for obs k and model point l calculated by finite differences:

$$\begin{aligned} \mathbf{H}^{kl} &= \frac{y^k}{\delta x^l} = \left[\frac{y^k(\mathbf{x} + \delta x^l) - y^k(\mathbf{x})}{\delta x^l} \right] \\ &= \left[\frac{H^k(M(\mathbf{x}^b(t_{i-1}) + \delta x^l)) - H^k(M(\mathbf{x}(t_{i-1})))}{\delta x^l(t_{i-1})} \right]; \end{aligned} \tag{4}$$

- In operational setup each grid-cell is split into 12 land-cover types (not relevant for this talk).

DEnKF

- Background ensemble calculated from previous analysis ensemble:

$$x_j^b(t_i) = M(\mathbf{x}_j^a(t_{i-1})), \text{ for } j = 1, .., m. \tag{5}$$

Background anomaly matrix $\mathbf{X}^b$ (of dimension $n \times m$) comes from m column vectors $\delta \mathbf{x}_j^b$:

$$\mathbf{X}^b = \frac{1}{\sqrt{m-1}} \left[\mathbf{x}_1^b - \overline{\mathbf{x}}^b \quad \ldots \quad \mathbf{x}_m^b - \overline{\mathbf{x}}^b \right], \tag{6}$$

where m is the number of ensemble members.

- Ensemble background-error covariance:

$$\mathbf{P}^b = \mathbf{X}^b (\mathbf{X}^b)^T. \tag{7}$$

- DEnKF halves Kalman gain in analysis perturbation update:

$$\mathbf{X}^a = \mathbf{X}^b - \frac{1}{2} \mathbf{K} \mathbf{H} \mathbf{X}^b. \tag{8}$$

- Deterministic analysis comes from ensemble mean:

$$\overline{\mathbf{x}}^a = \overline{\mathbf{x}}^b + \mathbf{K}(\mathbf{y}^o - \overline{H(\mathbf{x}^b)}). \tag{9}$$

METEO FRANCE
Toujours un temps d'avance

Experimental setup

- Firstly test methods at single site (Sabres);
- Then test methods over 12 sites;
- Daily assimilated WG1 (surface soil moisture) observations;
- WG2 (root-zone soil moisture) observations used for verification purposes;
- 12 ensemble members for DEnKF;
- Period 2007-2010 (first year spin-up);
- Observation error std $= 0.023m^3/m^3$ (half satellite observation error) estimated using Desrozier diagnostics (Desroziers *et al.*, 2005);
- Imperfect model - model error approximated in background-error covariance.

Background/Model error calibration

- SEKF **B** variances tuned to produce smallest analysis errors;
- SEKF **B** includes contribution from model error;
- DEnKF $\mathbf{P}^b$ collapses without allowing for model error;
- DEnKF model error tuned using additive noise and perturbed precipitation forcing;
- Additive noise sampled from Gaussian distribution (Mitchell *et al.*, 2002);
- Precipitation perturbations sampled from lognormal distribution (Mahfouf, 2007);

DEnKF model error calibration

- Time correlated noise ϕ introduced using 1^{st} order auto-regressive model (Mahfouf, 2007):

$$\begin{aligned} \phi(t_{i+1}) &= \nu\phi(t_i) + \psi\sqrt{1-\nu^2}, \\ \nu &= 1/(1+\Delta t/\tau), \end{aligned} \tag{10}$$

where ψ is Gaussian white noise and τ is the temporal correlation.

- Additive noise prescribed to analysis ensemble members:

$$\mathbf{x}_j^a \leftarrow \mathbf{x}_j^a + \phi_j \tag{11}$$

- Precipitation (Pr) perturbed for each ensemble member j:

$$\begin{aligned} Pr_{log} &= log(P_r + 1) \\ Pr'_j &= exp(Pr_{log} + \phi_j) - 1, \end{aligned} \tag{12}$$

where Pr'_j is the perturbed precipitation for ensemble member j.

Important algorithmic differences

Property	SEKF	DEnKF
Flow-dependent background-error covariance	No	Yes
Requires Jacobians of obs operator	Yes	No
Stochastic model error representation	No	Yes
Suffers from sampling error	No	Yes

Table 1: Important algorithmic differences. Green implies advantage and red implies disadvantage for a Gaussian system.

Results - Sabres site

SEKF and DEnKF give similar average performance for single site in terms of RMSE and correlation coefficient (CC):

Method	RMSE (m^3/m^3)	CC
Open loop	8.4×10^{-3}	0.90
SEKF	7.0×10^{-3}	0.91
DEnKF	6.9×10^{-3}	0.91

Table 2: WG2 RMSE and CC for Sabres site.

- Additive inflation significantly improves DEnKF performance;
- Our perturbed precipitation does not improve DEnKF performance (not shown);

DEnKF ensemble size - Sabres site

Ensemble size	DEnKF WG2 RMSE (m^3/m^3)
3	8.0×10^{-3}
6	7.3×10^{-3}
12	6.9×10^{-3}
20	7.0×10^{-3}
50	6.9×10^{-3}

Table 3: DEnKF WG2 RMSE for various ensemble sizes.

WG2 DEnKF analysis - Sabres site

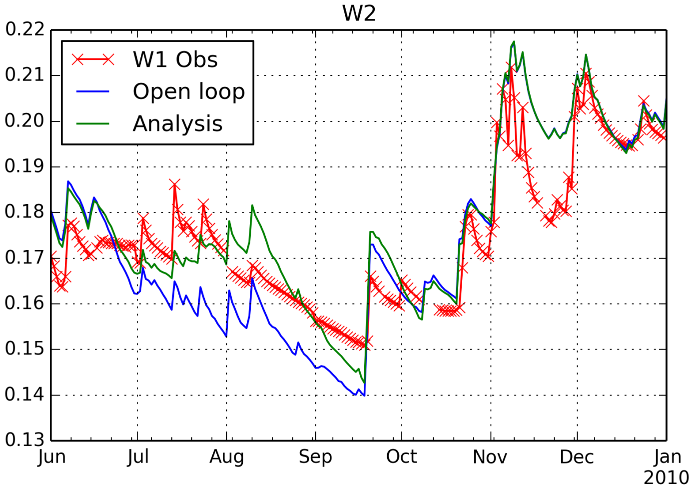

Figure 3: WG2 open loop, analysis and observations (m^3/m^3).

Performance averages - 12 sites

- Same background/model error calibration used for all sites;
- Analysis performs better than open loop for 10 of 12 sites;

Method	Average RMSE over 12 sites (m^3/m^3)	Average CC over 12 sites
Open loop	2.5×10^{-2}	0.81
SEKF	2.0×10^{-2}	0.89
DEnKF	2.0×10^{-2}	0.88

Table 4: WG2 RMSE and CC averaged over 12 sites.

WG2 monthly average RMSE - 12 sites

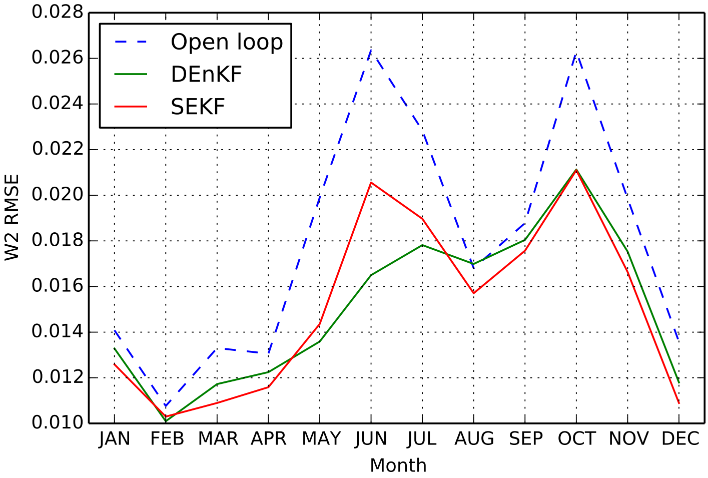

Figure 4: WG2 RMSE for DA methods (m^3/m^3).

METEO FRANCE
Toujours un temps d'avance

Possible explanations for the results

- If the system were Gaussian (or quasi-Gaussian), DEnKF with the correct model-error specification should perform better than SEKF;
- Is the problem with our DEnKF related to our model error representation, or is it a non-Gaussian issue with the system?
- Probably a combination of both;
- Potential non-Gaussian issues:
  1. **Forcing (mainly precipitation)**: Widely known to exhibit highly non-Gaussian behaviour;
  2. **Bounded system** also non-Gaussian (drainage of water above field capacity, little water loss below wilting point);
  3. **Highly nonlinear WG1** variable;
- Potential model-error specification issues:
  1. **Over-simplified stochastic representation** of precipitation uncertainty (lognormal distribution);
  2. No uncertainty representation of **other forcing parameters**: radiative forcing, wind, etc...
  3. No uncertainty representation of **model parameters**.

Conclusions

- DEnKF and SEKF give similar performance for single ground-based observation;
- Model and observation agreement unusually good for single site $\rightarrow$ analysis corrections small;
- Results show similar performance of the two methods averaged over 12 sites;
- Large changes in soil moisture content between the seasons could expain why WG2 analysis errors are twice as large in spring and autumn than in winter;
- DEnKF unable to improve on SEKF due to deficient model-error representation and/or non-Gaussian issues with model.

Future work

- Investigate issues caused by the limitations in the linear approximations made by the DA methods - See Alina's poster;
- Comparison between the two methods for different temporal observation frequencies (every 6 hours - every 3 days);
- Improve understanding and representation of model error;
- Investigate horizontal correlations with a 2D grid;
- Increase the vertical resolution of the model using a diffusive hydrological scheme.

METEO FRANCE
Toujours un temps d'avance

References

G. Desroziers, L. Berre, B. Chapnik, and P. Poli. Diagnosis of observation, background and analysis-error statistics in observation space. *Q. J. R. Meteorol. Soc.*, 131:3385–3396, 2005.

J.F. Mahfouf. L'analyse dans le sol a meteo-france. partie 1: Evaluation et perspectives a l'echelle locale. Meteo France technical report, 2007. Last accessed September 2014.

J.F. Mahfouf, K. Bergaoui, C. Draper, C. Bouyssel, F. Taillefer, and L. Taseva. A comparison of two off-line soil analysis schemes for assimilation of screen-level observations. *J. Geophys. Res*, 114:D08105, 2009.

H.L. Mitchell, P.L. Houtekamer, and G. Pellerin. Ensemble size, balance and model-error representation in an ensemble Kalman filter. *Mon. Weather Rev.*, 130:2791–2808, 2002.

R.H. Reichle and R.D. Koster. Assessing the Impact of Horizontal Error Correlations in Background Fields on Soil Moisture Estimation. *Journal of Hydrometeorology*, 4:1229–1242, 2003.

J.M. Sabater, L. Jarlan, J.C. Calvet, and F. Boyssel. From near-surface to root-zone soil moisture using different assimilation techniques. *Journal of Hydrometeorology*, 8:194–206, 2006.

P. Sakov and P. R. Oke. A deterministic formulation of the ensemble Kalman filter: an alternative to ensemble square root filters. *Tellus A*, 60:3035–3049, 2008.