SURFEX
SURFEX_EKF for ALARO
SURFEX_EKF + 3DVAR
STAEKF for soil analysis

Recent activities on surface assimilation using SURFEX in the RMI

Hamdi R. with input from Annelies Duerinckx

LACE Data Assimilation Working Days, 18-20 June 2012

1. Development of a SURFEX EKF for ALARO and comparison of the offline and coupled version with the OI analysis.

2. Preliminary results on combining SURFEX EKF with 3dVar atmospheric assimilation.

3. A feasibility study of using a Short Time Augmented Extended Kalman Filter (STAEKF) for Soil Analysis.

The new land surface scheme of ALARO: SURFEX

- **Both ALADIN and ARPEGE relied on the ISBA scheme for the parameterization of the surface processes.** (Noilhan and Planton 1989 ; Mahfouf et al. 1995 ; Noilhan and Mahfouf 1996)

- **In 2000 Valéry Masson developed a scheme for simulation the interactions with urban areas and this scheme becomes part of the meso-NH surface model.** (TEB, Masson 2000)

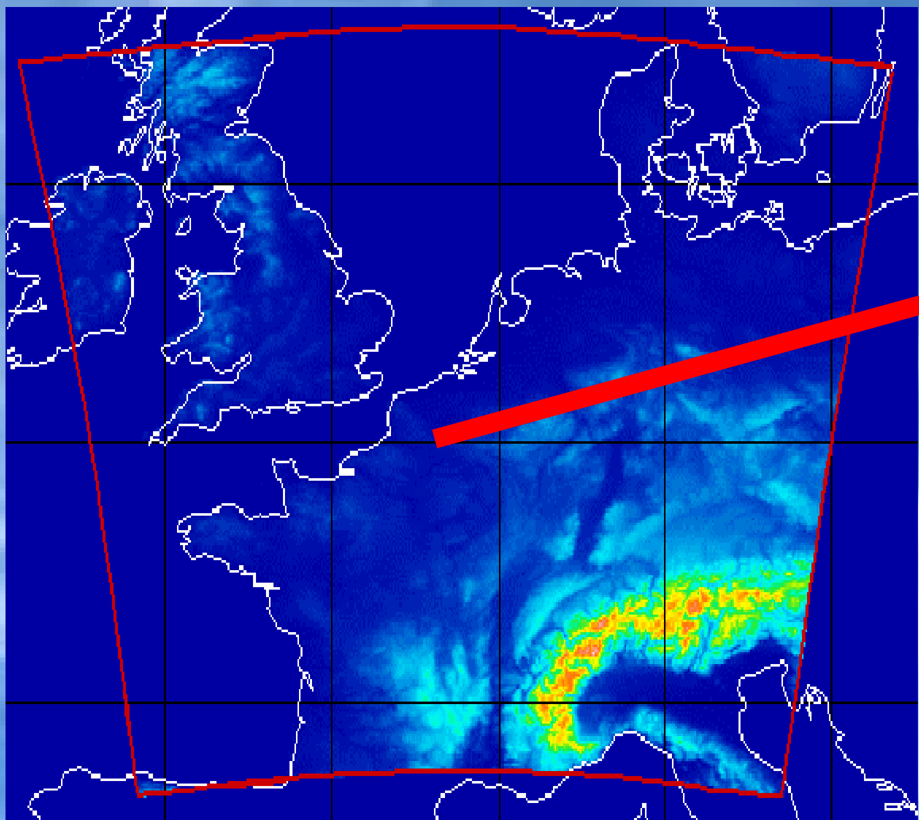
- **During the last decade, the surface scheme (ISBA+TEB) has been externalized from the atmospheric part of the meso-NH model following the approach of** (Best et al. 2004)

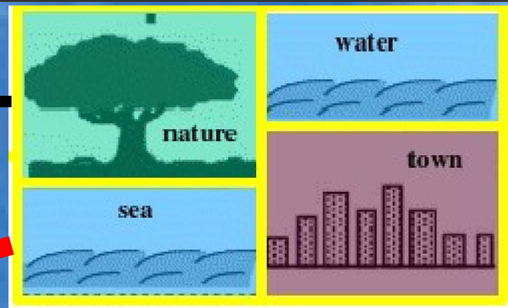
- **This led to the creation of the SURFEX scheme (SURface Externalisée) where the characteristics of the surface are specified by the ECOCLIMAP database** (Masson et al. 2003)

- **Additionally, more schemes has been added to SURFEX for: Sea and oceans (prescribed SST, ECUME, 1-D ocean model), lakes (prescribed ST, FLAKE), Surface boundary layer scheme CANOPY** (Masson and Seity 2009; Hamdi and Masson 2008)**...etc.**

Tiling

One important feature of the externalized surface:
each grid cell is divided into 4 elementary units called tiles according to the fraction of covers in the grid cell





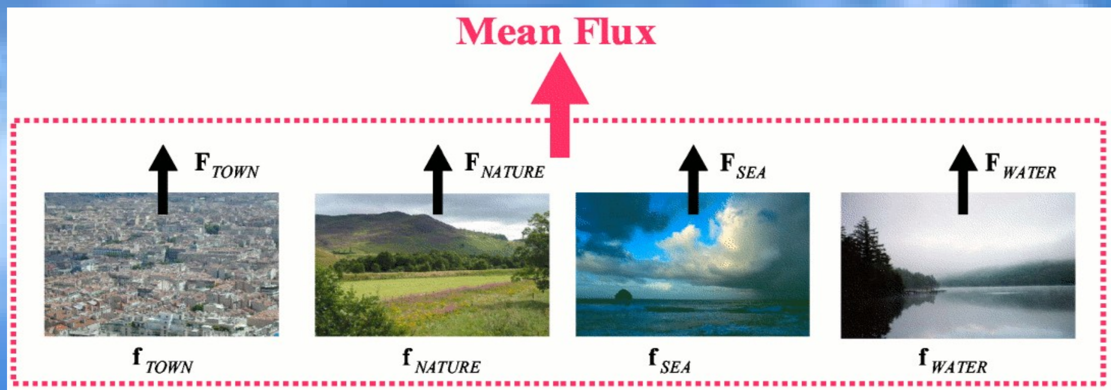
1: bare ground
2: rocks
3: permanent snow
4: deciduous forest
5: conifer forest
6: evergreen broadleaf trees
7: C3 crops
8: C4 crops
9: irrigated crops
10: grassland
11: tropical grassland
12: garden and parks

Surfex output as surface boundary conditions for atmospheric radiation and turbulent scheme.

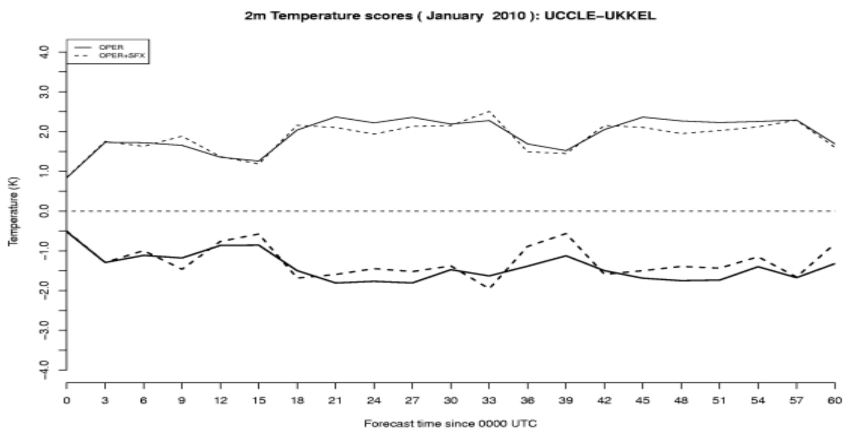
- albedo
- emissivity
- radiative temperature
- momentum flux
- sensible heat flux
- latent heat flux
- CO_2 flux
- chemical flux

ALARO model

- Atmospheric forcing
- Sun position
- Downward radiative flux



ALARO+FMR+SURFEX Vs ALARO+ACRANEB run over Belgium



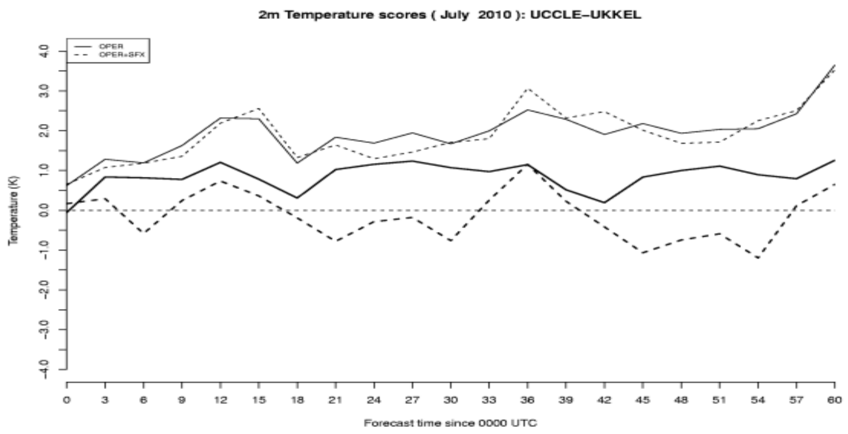


Table 2. the average daytime/nightime scores for the flat/high topography and coastal synoptic stations, sign (+) means improvement, sign (0) means neutral effect, and sign (-) means degradation of the scores.

	$Winter_{Night}$	$Winter_{Day}$	$Summer_{Night}$	$Summer_{Day}$
Temperature at 2m				
Flat	+	0	0	0
High	0	-	0	0
Coast	0	0	+	0
Wind speed at 10m				
Flat	0	0	0	0
High	0	0	0	0
Coast	0	0	0	0
Wind direction at 10m				
Flat	0	0	0	0
High	0	0	0	0
Coast	0	0	0	0
Relative humidity at 2m				
Flat	+	+	0	0

- Optimum Interpolation (OI)
  - Coefficients are derived using simple assumptions
- Extended Kalman Filter (EKF)
  - dynamical coefficients
  - Formulation: $\mathbf{x}_t^a = \mathbf{x}_t^b + \mathbf{B}\mathbf{H}^T\left(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R}\right)^{-1}\left[\mathbf{y}_t^o - \mathcal{H}(\mathbf{x}_0^b)\right]$

$\mathcal{H}$: the observation operator

includes a model propagation

H: the Jacobian matrix of the observation operator:

$$\mathbf{H} = \frac{\partial \mathbf{y}_t}{\partial \mathbf{x}_{t_0}}$$

H is calculated by a finite difference approach:

$$H_{ij} = \frac{y_i(\mathbf{x} + \delta x_j) - y_i(\mathbf{x})}{\delta x_j}$$

- Finite difference approach:

$$H_{ij} = \frac{y_i(\mathbf{x} + \delta x_j) - y_i(\mathbf{x})}{\delta x_j}$$

- Calculation of $y_i(\mathbf{x} + \delta x_j)$ and $y_i(\mathbf{x})$:
  - Perturb a component x_j of the control vector **x**
  - Perform run with the perturbed surface field to calculate corresponding y_i

    Offline: surface scheme decoupled from atmospheric model

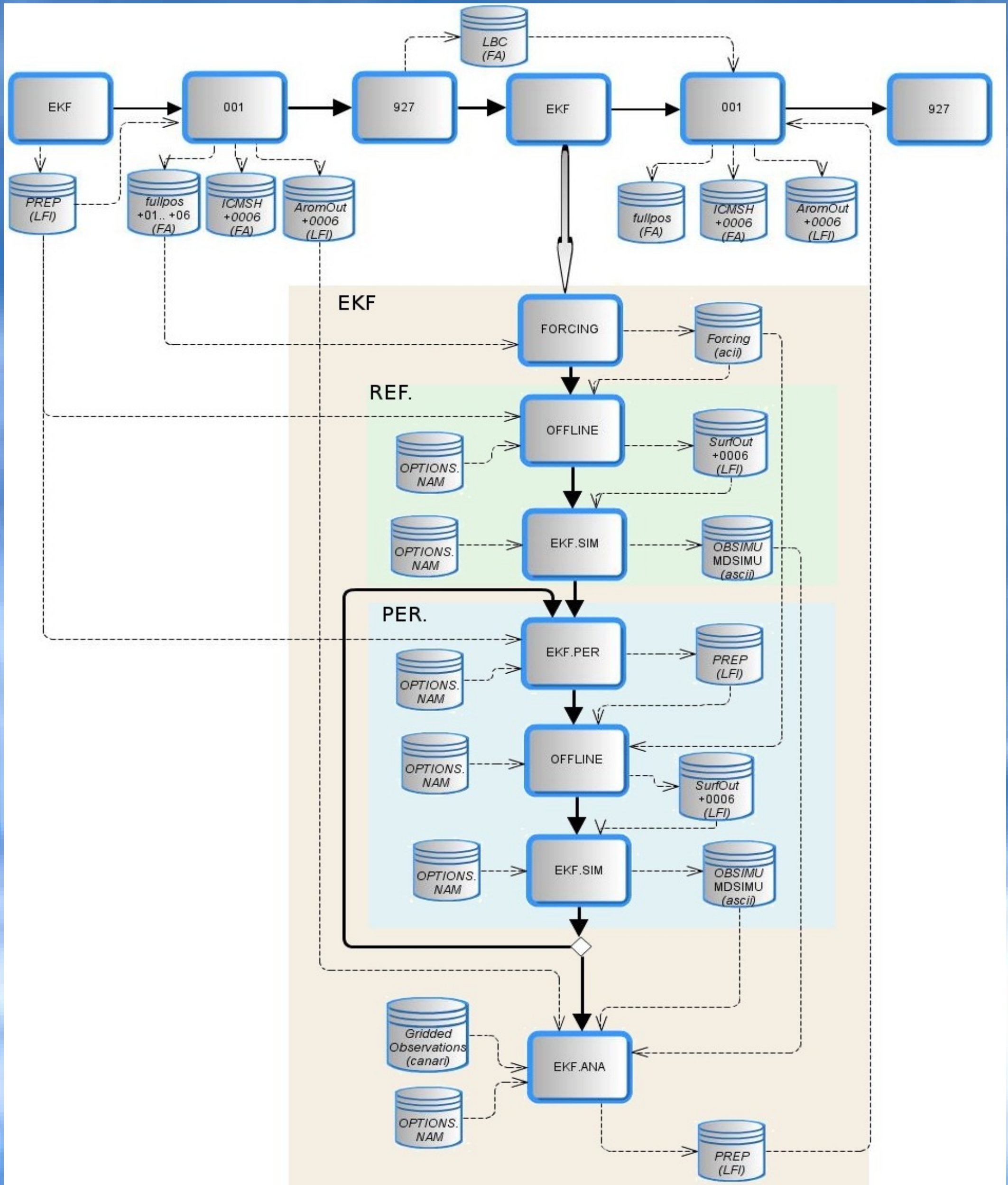
    the forcing is taken from the lowest model level

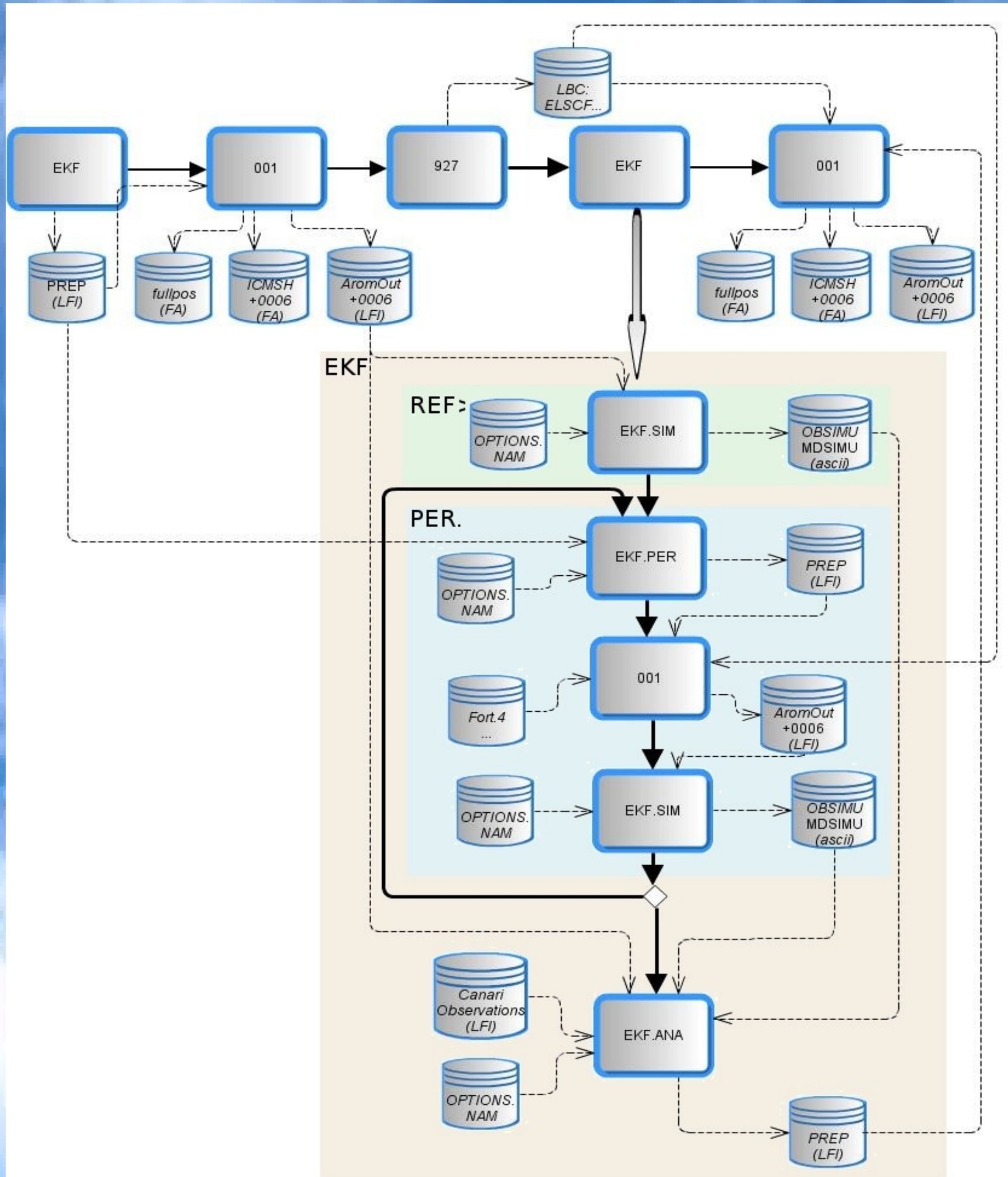
    Coupled: surface scheme coupled to atmospheric model

    interaction is possible between atmosphere and surface

- In Balsamo et al. (2007,JH), the offline approach is compared to the coupled approach for a single day during the summer with the GEM model. The gain components are found to be smaller for the offline approach but with similar patterns as the coupled approach.
- In Balsamo et al (2004,QJRMS) it is mentioned that for coupled experiments the choice of the perturbation size is important. Small perturbations can lead to a noisy H matrix and inaccurate corrections.
- This presentation:
  - Implementation of offline and coupled SURFEX EKF for ALADIN
  - Comparison between offline and coupled EKF: jacobians, gain, increments
  - Comparison of forecast scores of EKF, OI and runs without assimilation

- In Balsamo et al. (2007,JH), the offline approach is compared to the coupled approach for a single day during the summer with the GEM model. The gain components are found to be smaller for the offline approach but with similar patterns as the coupled approach.
- In Balsamo et al (2004,QJRMS) it is mentioned that for coupled experiments the choice of the perturbation size is important. Small perturbations can lead to a noisy H matrix and inaccurate corrections.
- This presentation:
  - Implementation of offline and coupled SURFEX EKF for ALADIN
  - Comparison between offline and coupled EKF: jacobians, gain, increments
  - Comparison of forecast scores of EKF, OI and runs without assimilation


EKF
001
927
EKF
001
927
LBC (FA)
PREP (LFI)
fullpos +01.. +06 (FA)
ICMSH +0006 (FA)
AromOut +0006 (LFI)
fullpos (FA)
ICMSH +0006 (FA)
AromOut +0006 (LFI)
EKF
FORCING
Forcing (acii)
REF.
OFFLINE
OPTIONS. NAM
SurfOut +0006 (LFI)
EKF.SIM
OPTIONS. NAM
OBSIMU MDSIMU (ascii)
PER.
EKF.PER
OPTIONS. NAM
PREP (LFI)
OFFLINE
OPTIONS. NAM
SurfOut +0006 (LFI)
EKF.SIM
OPTIONS. NAM
OBSIMU MDSIMU (ascii)
Gridded Observations (canari)
OPTIONS. NAM
EKF.ANA
PREP (LFI)


LBC: ELSCF...
EKF
001
927
EKF
001
PREP (LFI)
fullpos (FA)
ICMSH +0006 (FA)
AromOut +0006 (LFI)
fullpos (FA)
ICMSH +0006 (FA)
AromOut +0006 (LFI)
EKF
REF.
OPTIONS. NAM
EKF.SIM
OBSIMU MDSIMU (ascii)
PER.
EKF.PER
OPTIONS. NAM
PREP (LFI)
001
Fort.4 ...
AromOut +0006 (LFI)
EKF.SIM
OPTIONS. NAM
OBSIMU MDSIMU (ascii)
Canari Observations (LFI)
OPTIONS. NAM
EKF.ANA
PREP (LFI)

- Prognostic variables:
  - Superficial water content (wg1)
  - Root zone water content (wg2)
  - Surface temperature (Tg1)
  - Deep soil temperature (Tg2)
- Observations:
  - Screen level temperature (T2m)
  - Screen level relative humidity (RH2m)
- Additional information:
  - LBC data from Aladin France
  - Assimilation interval τ = 6 hours, with assimilation at 00,06,12,18 UTC
  - Forecasts with surfex + alaro and inline fullpos (interval 1 hour)
  - Background error covariance matrix **B** is kept constant (WG set to 0.1,TG set to 2 K)
  - Error covariance matrix **R:** T2m set to 1K, RH2m set to 10%

Experiments:

- Free run: (no assimilation)

  Surface field from 6h forecast of previous run

- Open loop: (no assimilation)

  Surface is interpolated from Arpege analysis

- Surface Assimilation runs:

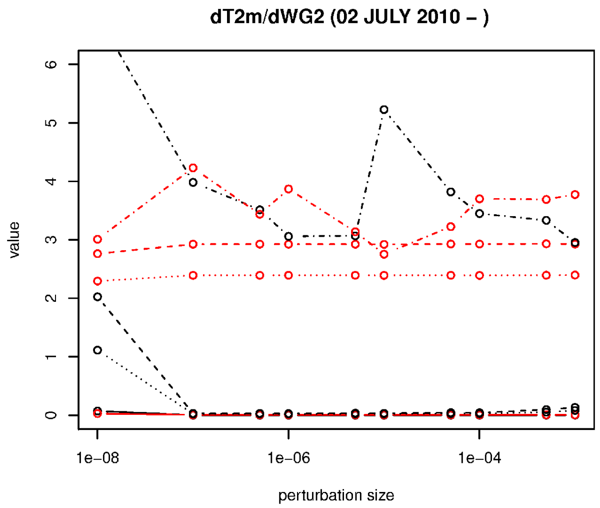
  Optimum Interpolation

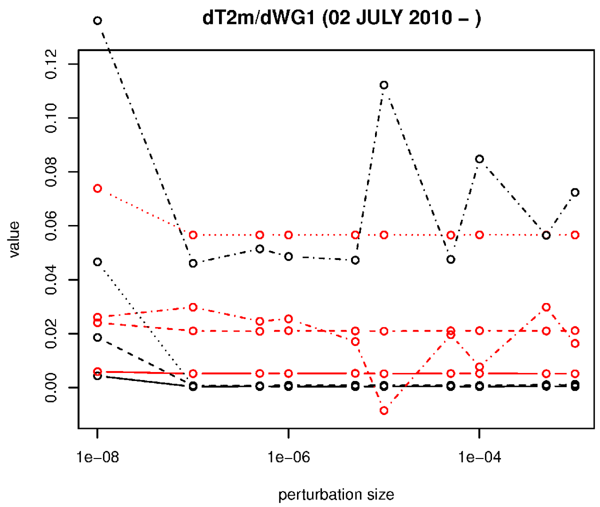
  EKF with offline jacobian calculation

  EKF with coupled jacobian calculation

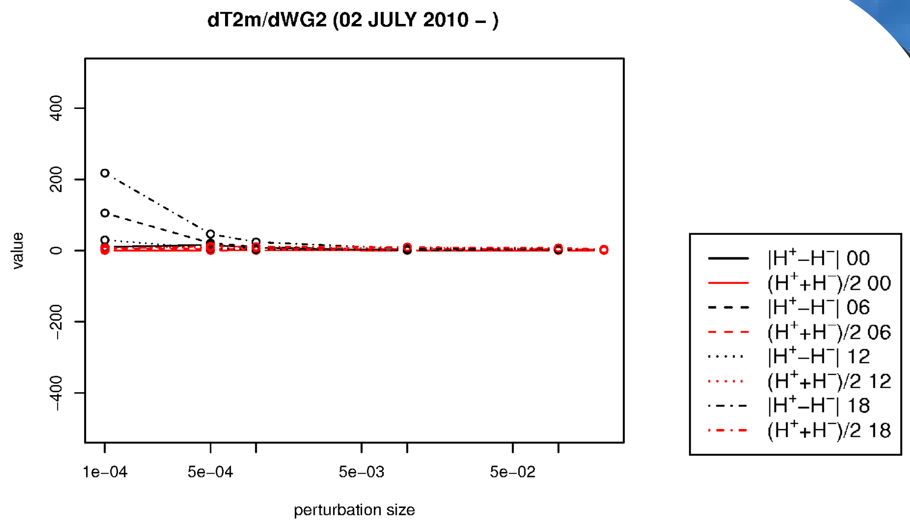
  → Surface guess for assimilation is taken from 6h forecast of previous run

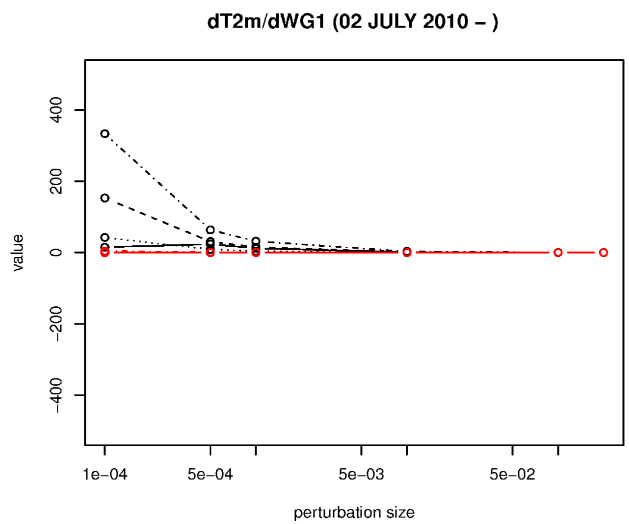
EKF offline





EKF Coupled

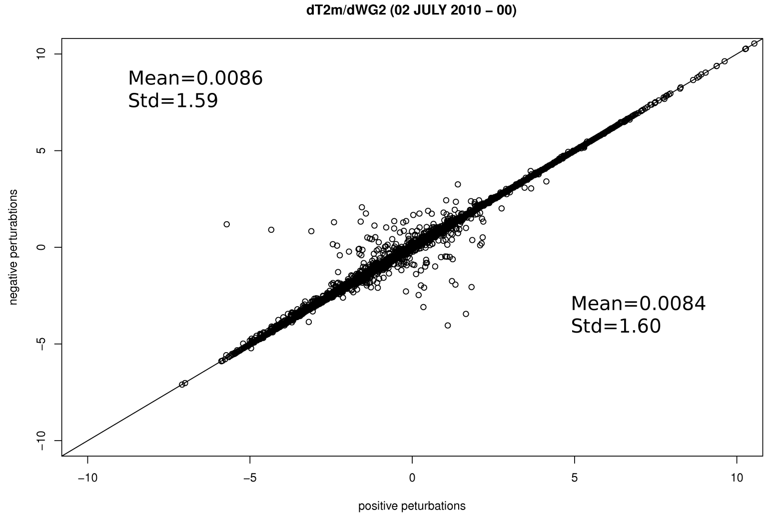


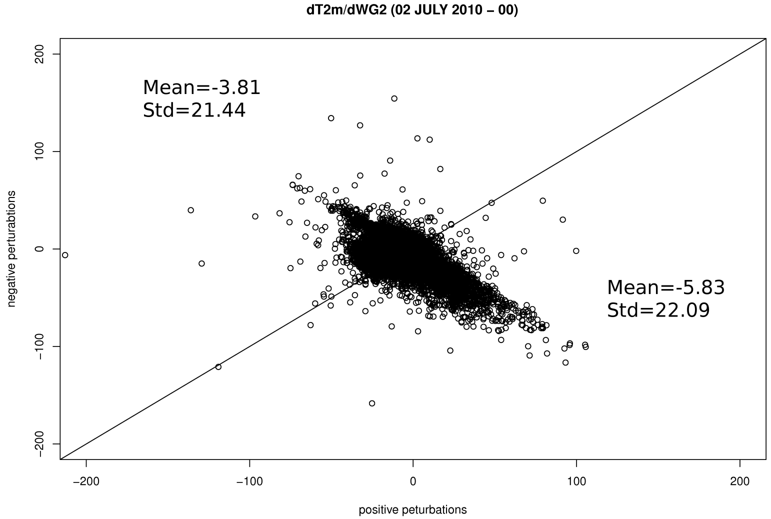


EKF offline

EKF Coupled

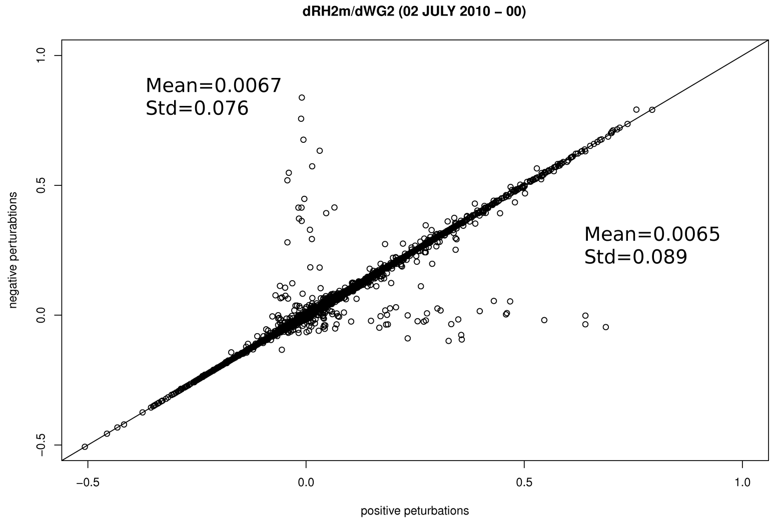
Pert. size: 10^{-4}

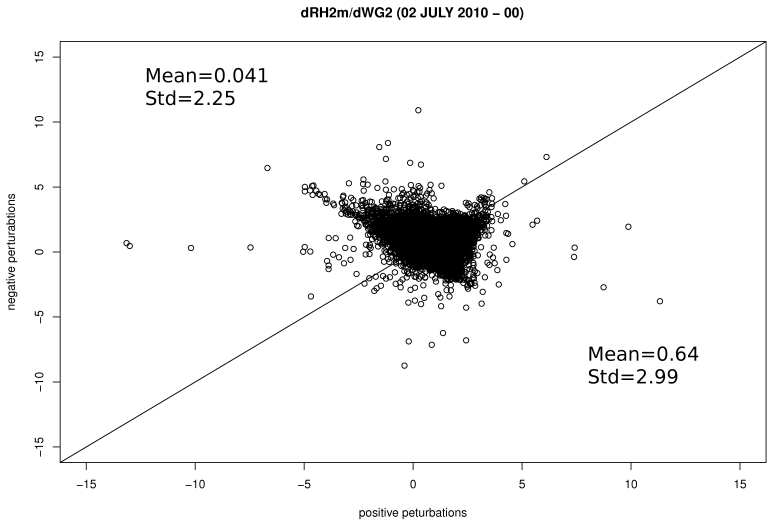




Pert. size: 10^{-1}

Pert. size: 10^{-4}



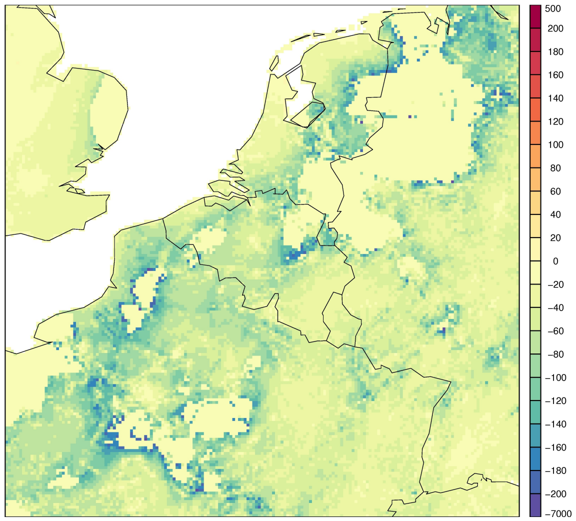


Pert. size: 10^{-1}

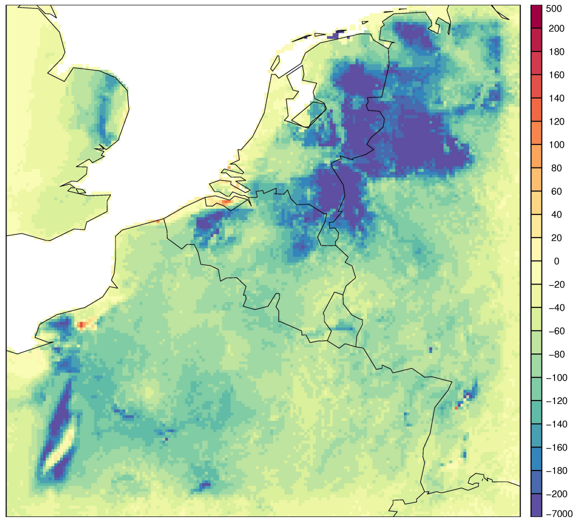
	EKF offline	EKF coupled

T2m-WG2

T2m–WG2 jacobians for 02 July 2010 1200

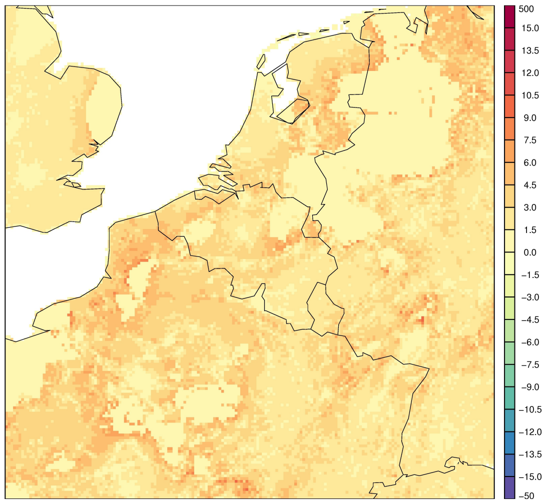


T2m–WG2 jacobians for 02 July 2010 1200

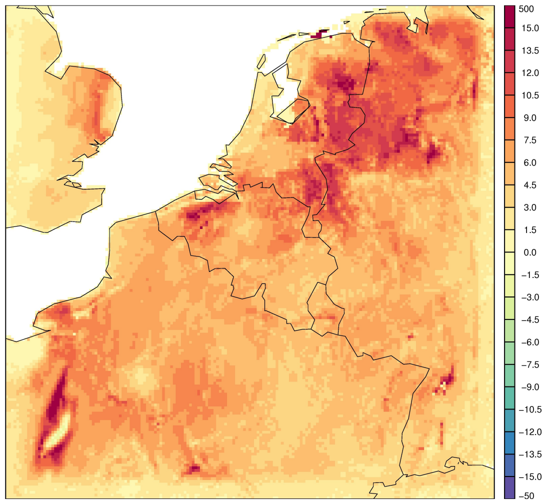


RH2m-WG2

RH2m–WG2 jacobians for 02 July 2010 1200



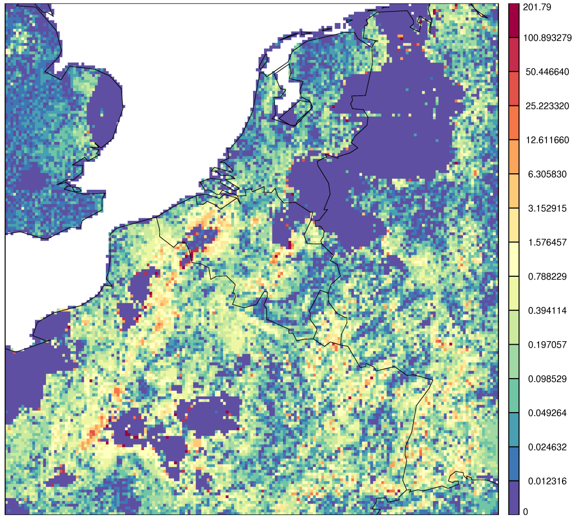
RH2m–WG2 jacobians for 02 July 2010 1200



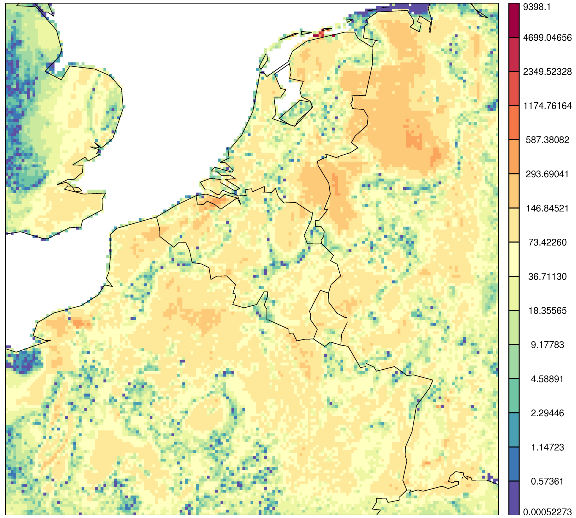
EKF offline | EKF coupled

T2m-WG2

|H+ – H–| for dT2m/dWG2 02 July 2010 1200

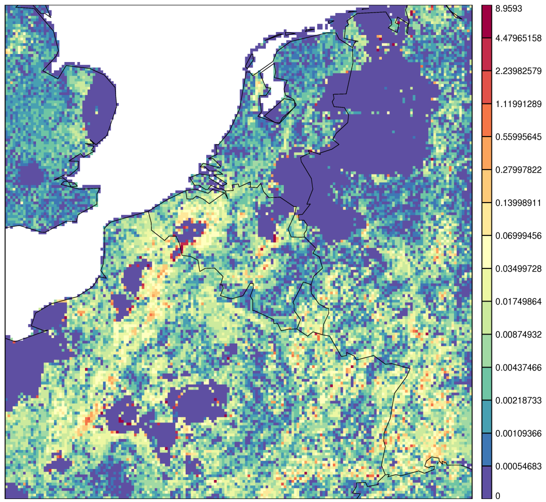


|H+ – H–| for dT2m/dWG2 02 July 2010 1200

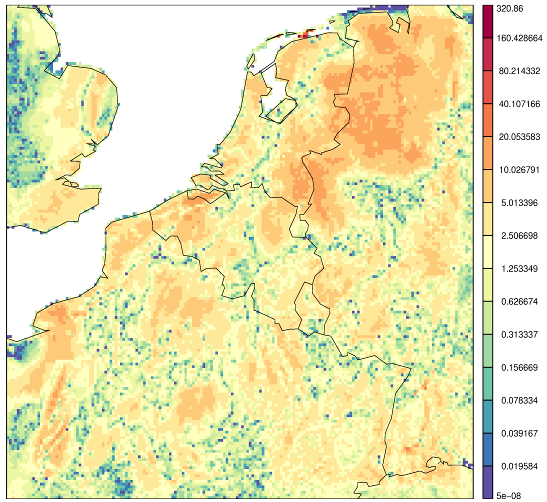


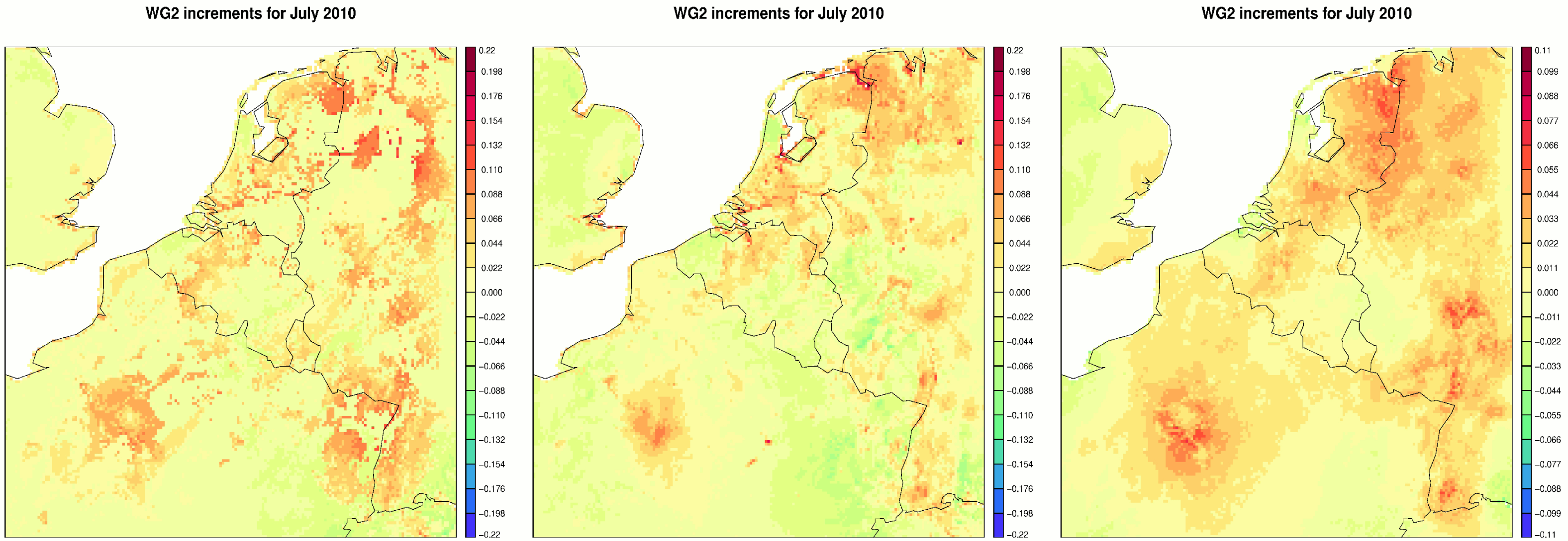
RH2m-WG2

|H+ – H–| for dRH2m/dWG2 02 July 2010 1200



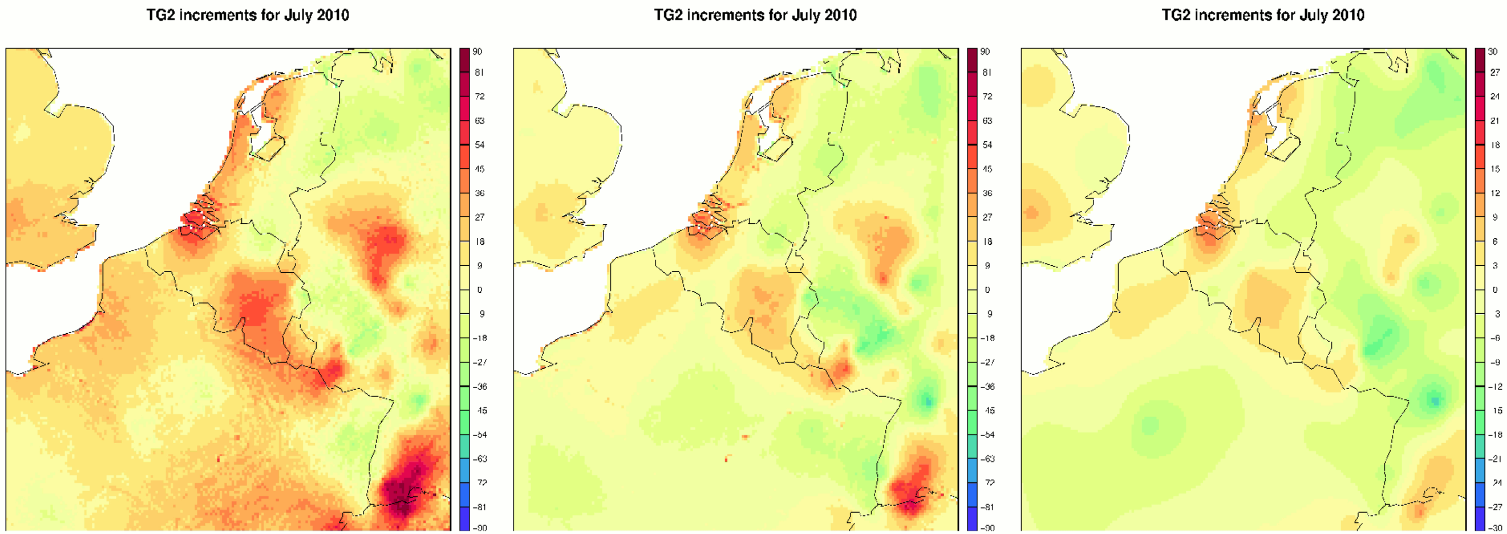
|H+ – H–| for dRH2m/dWG2 02 July 2010 1200





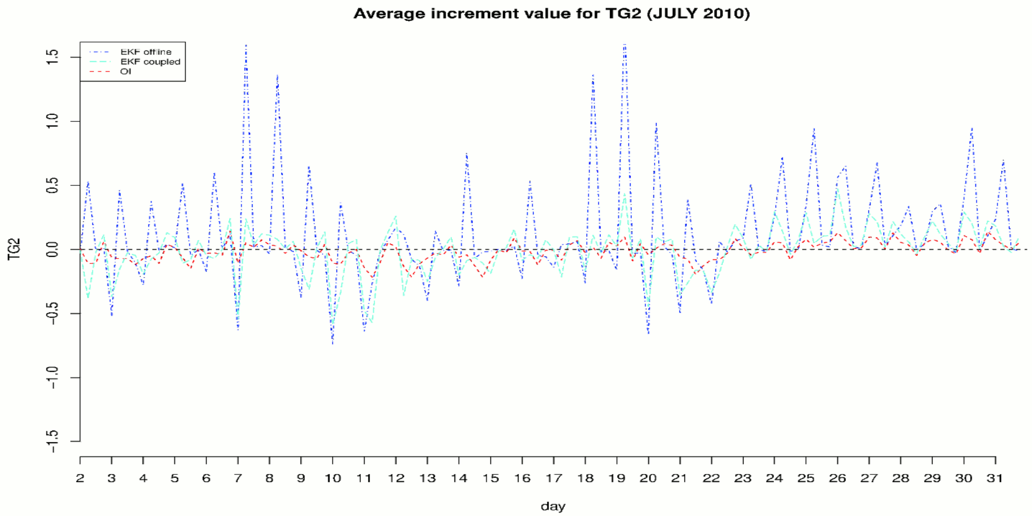
(24.1) offline EKF (24.2) coupled EKF (24.3) OI

Figure 24: Cummulative increments for July 2010, WG2. The incrementscale for OI is two times smaller than for both EKF.

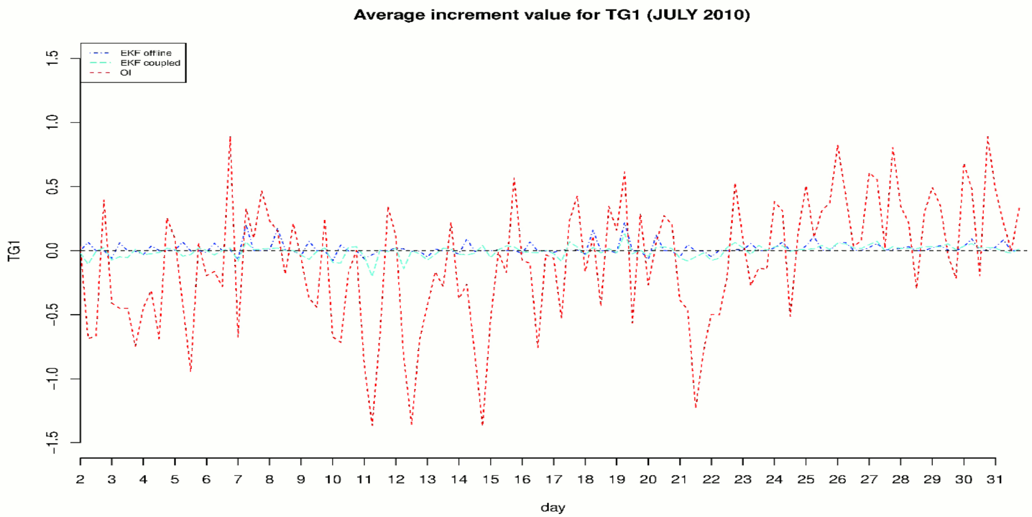


(26.1) offline EKF (26.2) coupled EKF (26.3) OI

Figure 26: Cummulative increments for July 2010, TG2. The incrementscale for OI is three times smaller than for both EKF.

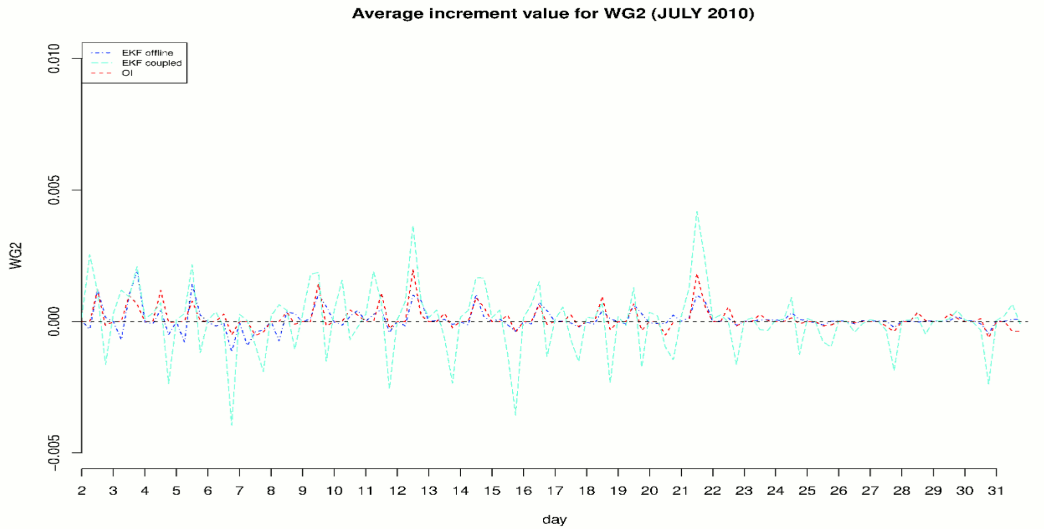


(28.1) Evolution of the TG2 increments averaged over the domain for July 2010

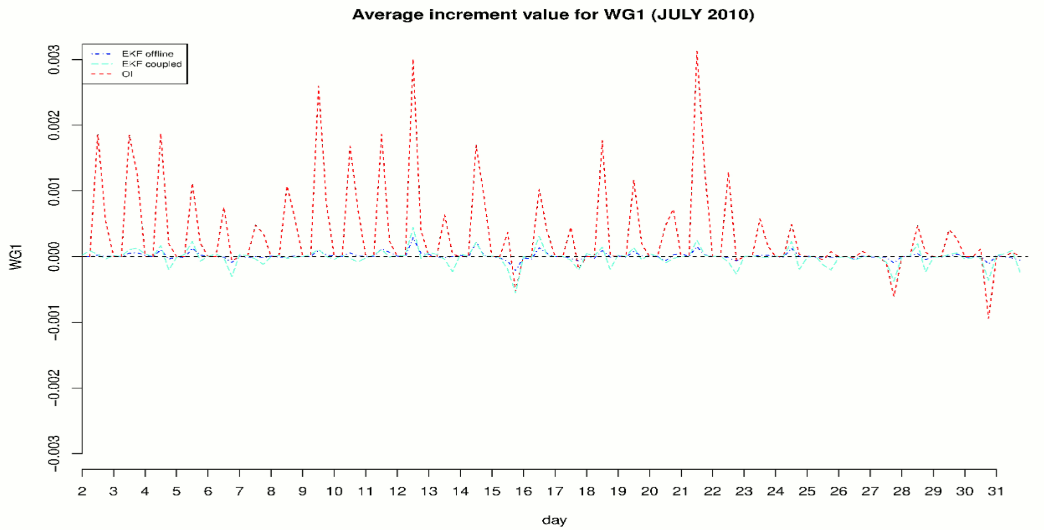


(28.2) Evolution of the TG1 increments averaged over the domain for July 2010

Figure 28



(28.3) Evolution of the WG2 increments averaged over the domain for July 2010



(28.4) Evolution of the WG1 increments averaged over the domain for July 2010

Figure 28: Evolution of the increments of the soil prognostic variables

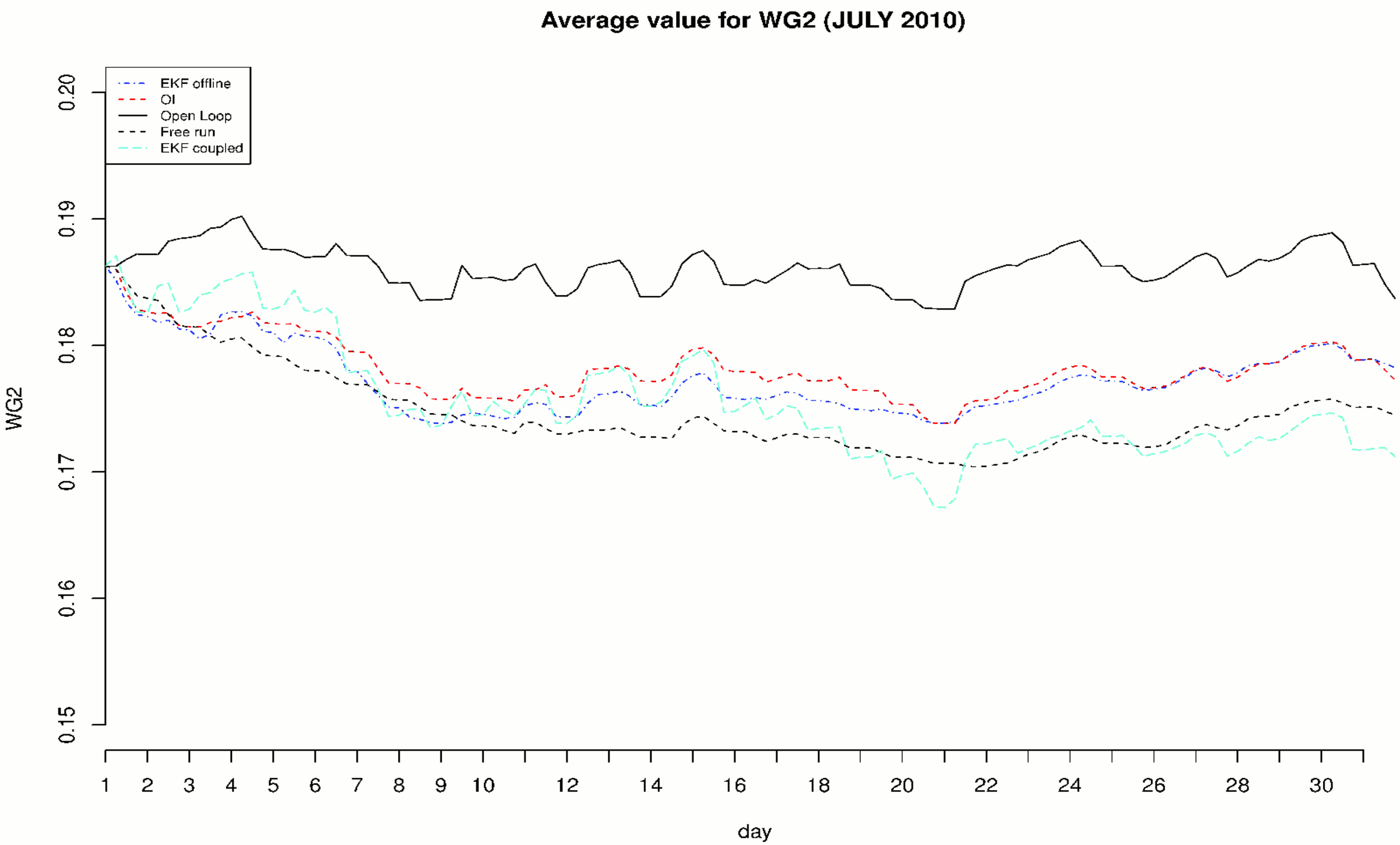
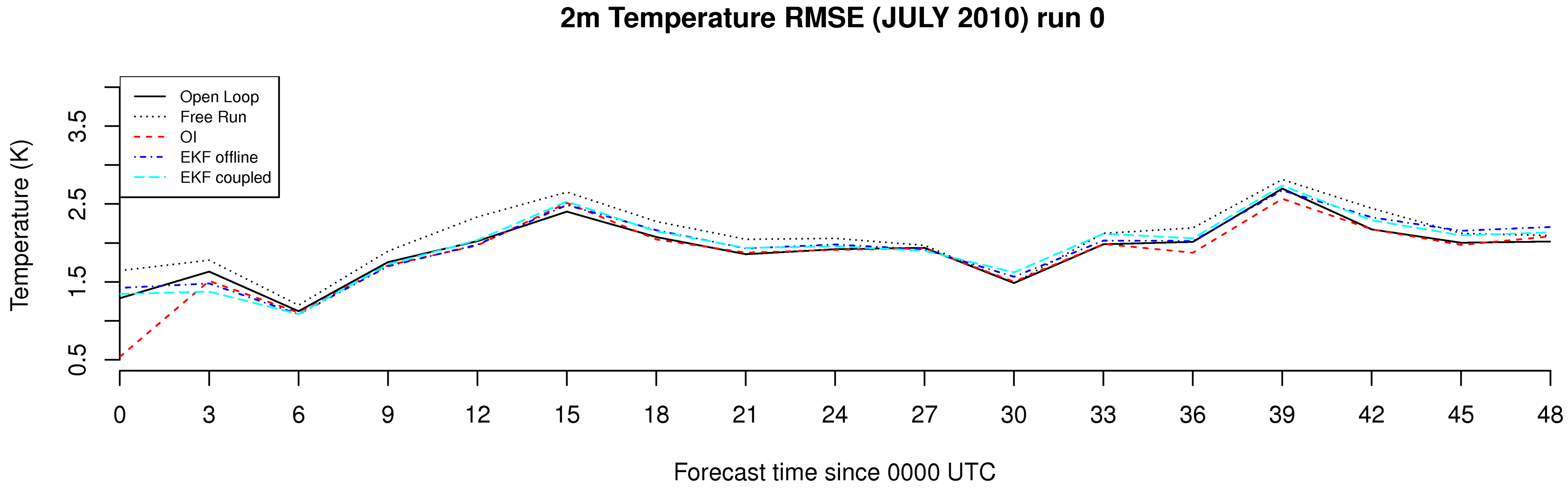
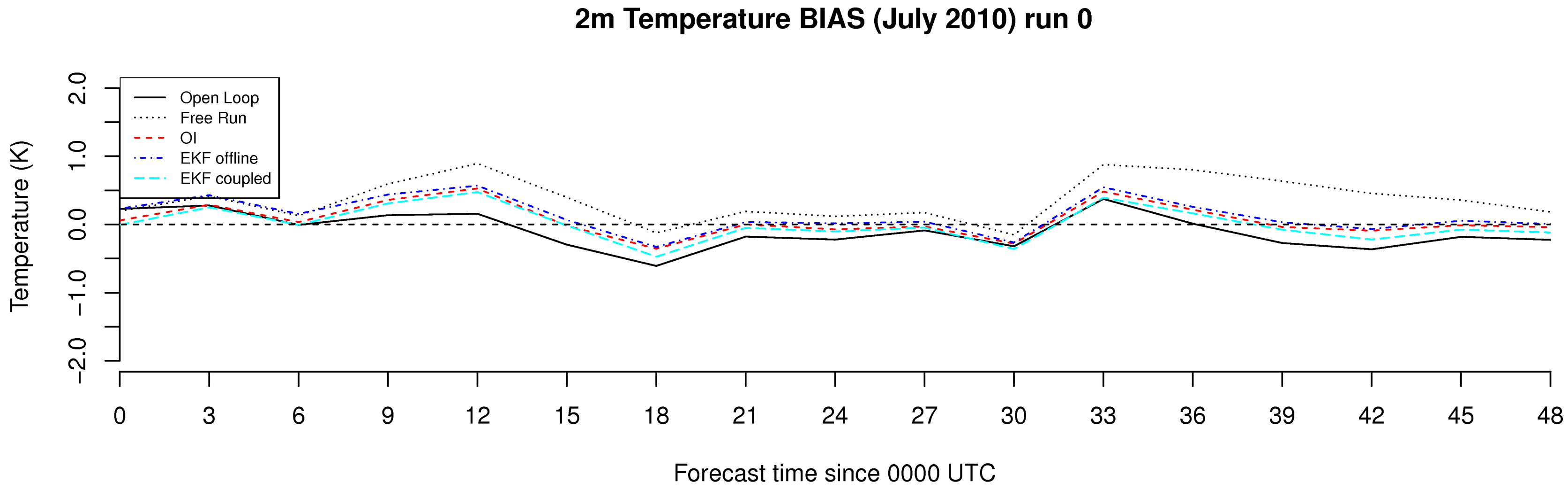
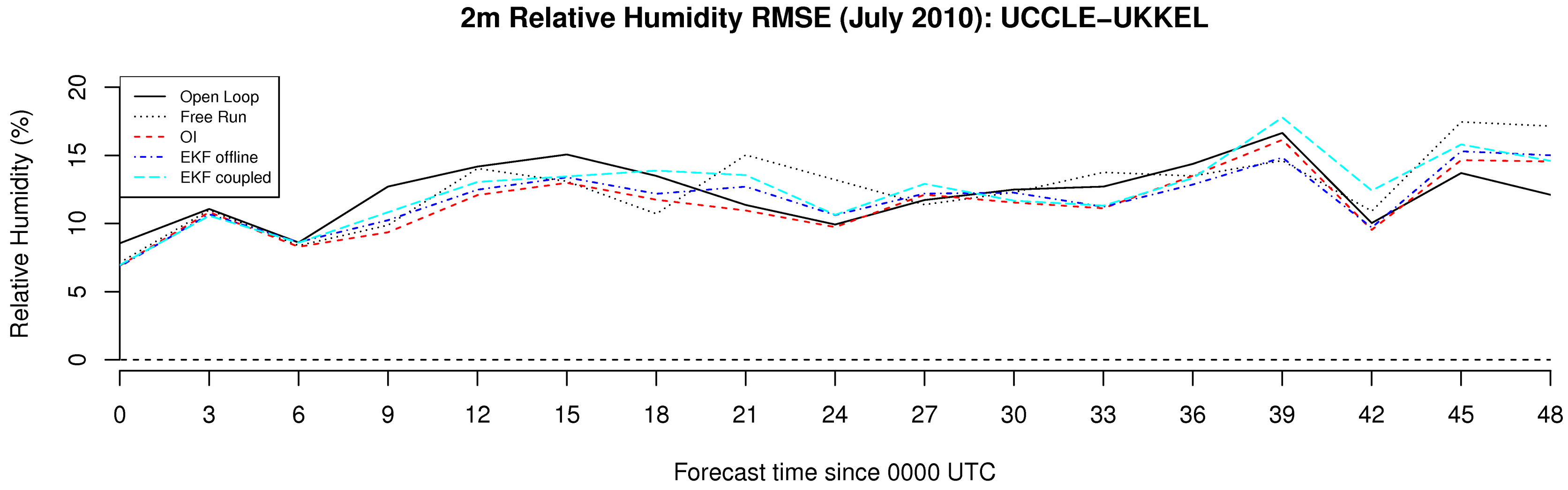
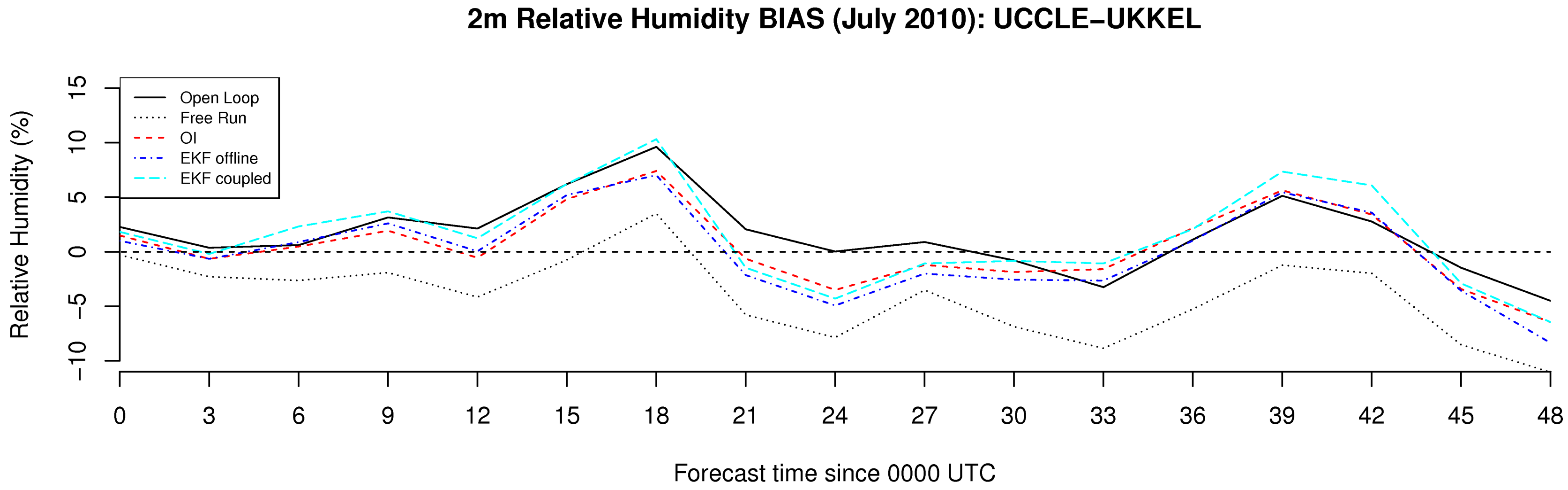


Figure 32: Evolution of the WG2 values averaged over the domain for July 2010


2m Temperature RMSE (JULY 2010) run 0
Open Loop
Free Run
OI
EKF offline
EKF coupled
Temperature (K)
0.5
1.5
2.5
3.5
0
3
6
9
12
15
18
21
24
27
30
33
36
39
42
45
48
Forecast time since 0000 UTC


2m Temperature BIAS (July 2010) run 0
Open Loop
Free Run
OI
EKF offline
EKF coupled
Temperature (K)
−2.0
−1.0
0.0
1.0
2.0
0
3
6
9
12
15
18
21
24
27
30
33
36
39
42
45
48
Forecast time since 0000 UTC


2m Relative Humidity RMSE (July 2010): UCCLE–UKKEL
Relative Humidity (%)
Forecast time since 0000 UTC
Open Loop
Free Run
OI
EKF offline
EKF coupled
0
5
10
15
20
0 3 6 9 12 15 18 21 24 27 30 33 36 39 42 45 48


2m Relative Humidity BIAS (July 2010): UCCLE–UKKEL
Relative Humidity (%)
Forecast time since 0000 UTC
Open Loop
Free Run
OI
EKF offline
EKF coupled
−10
−5
0
5
10
15
0 3 6 9 12 15 18 21 24 27 30 33 36 39 42 45 48

1. Development of a SURFEX EKF for ALARO and comparison of the offline and coupled version with the OI analysis.

2. Preliminary results on combining SURFEX EKF with 3dVar atmospheric assimilation.

3. A feasibility study of using a Short Time Augmented Extended Kalman Filter (STAEKF) for Soil Analysis.

Experiments:

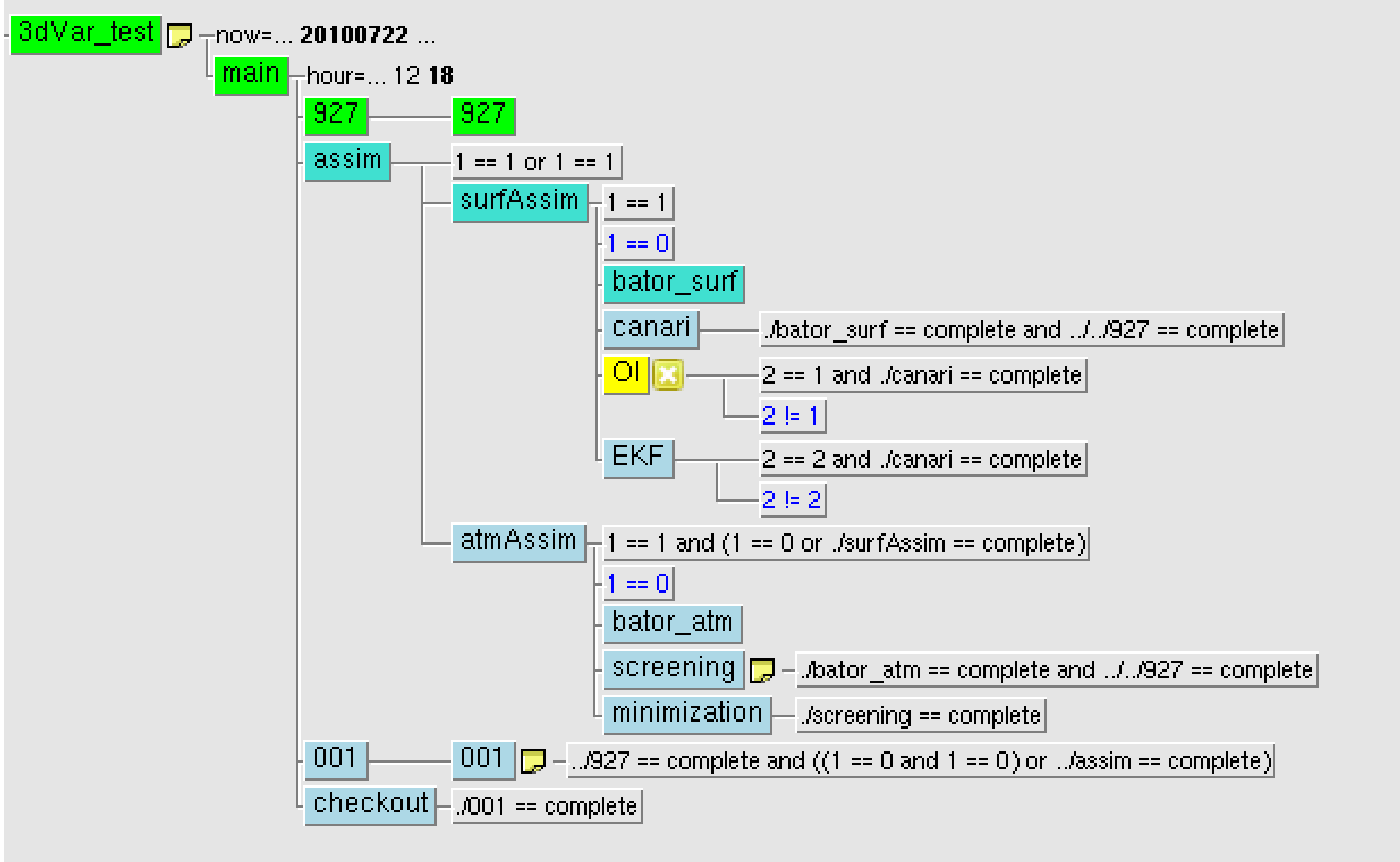
- **Free run**: (no assimilation)

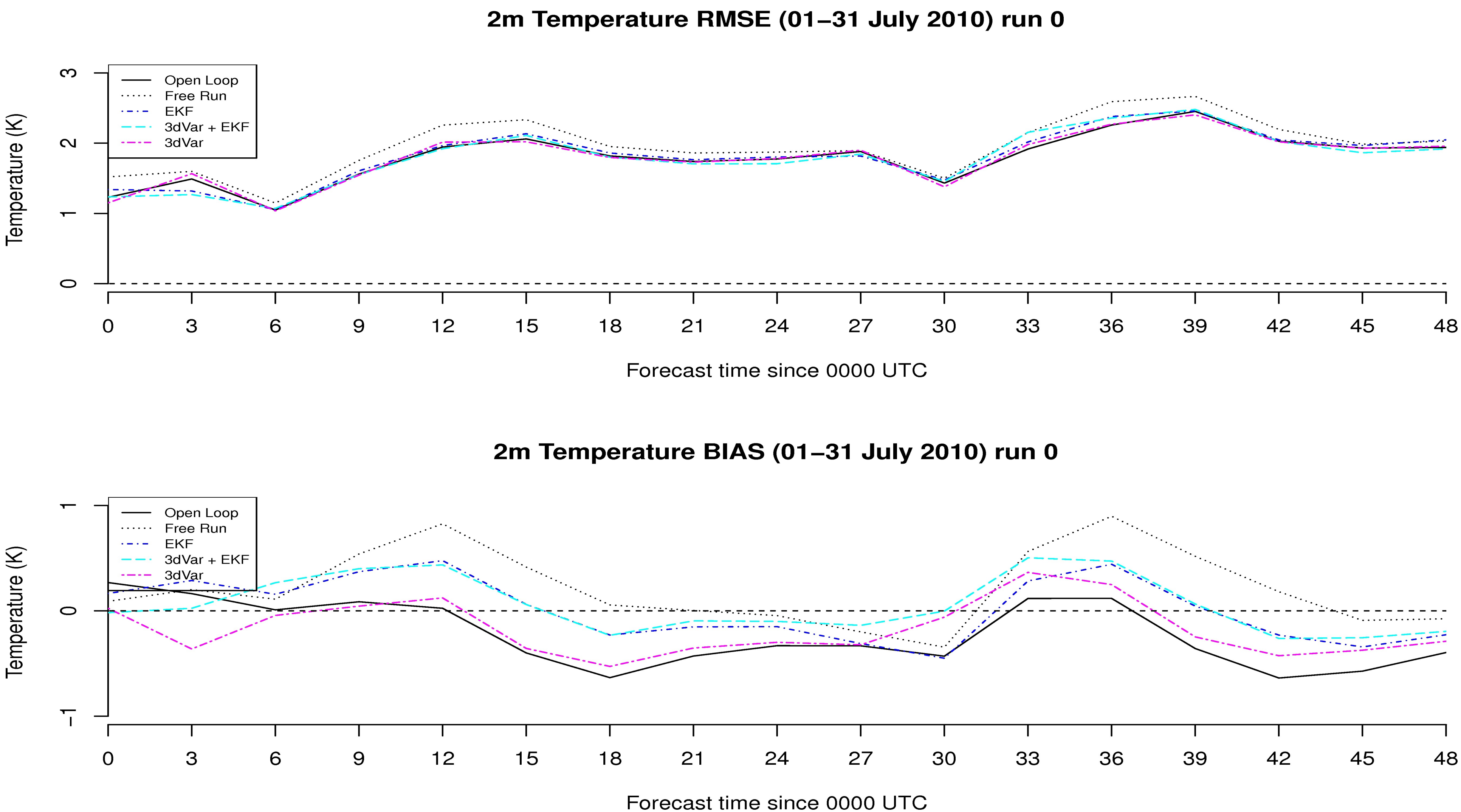
  Atmospheric fields from ARPEGE analysis. Surface field from 6h forecast of previous run

- **Open loop**: (no assimilation)

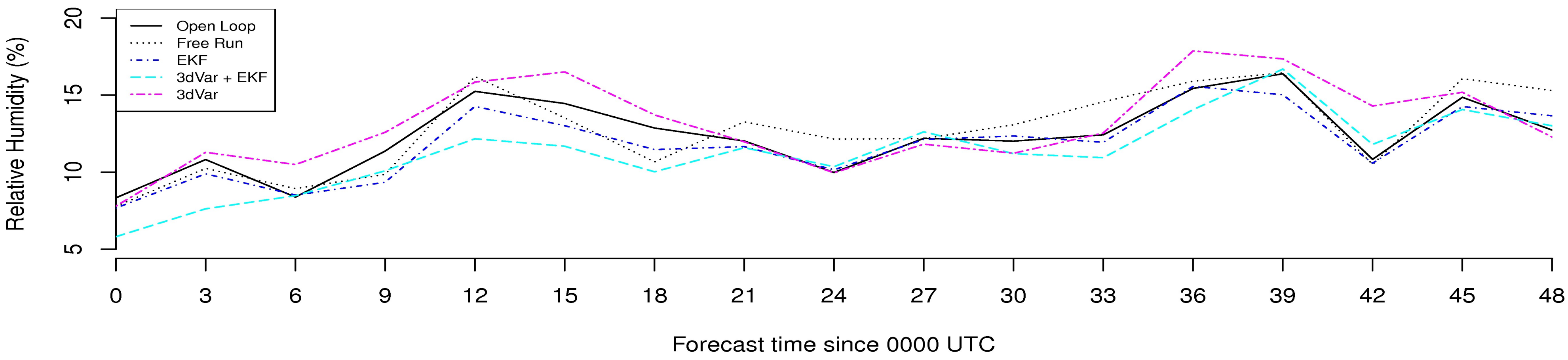
  Atmospheric fields and surface fields are interpolated from Arpege analysis

- **EKF** with surface assimilation.
- **3DVAR** With atmospheric assimilation using only conventional observations (no satellite, no radar)
- **3DVAR+EKF** surface assimilation is done before atmospheric assimilation.

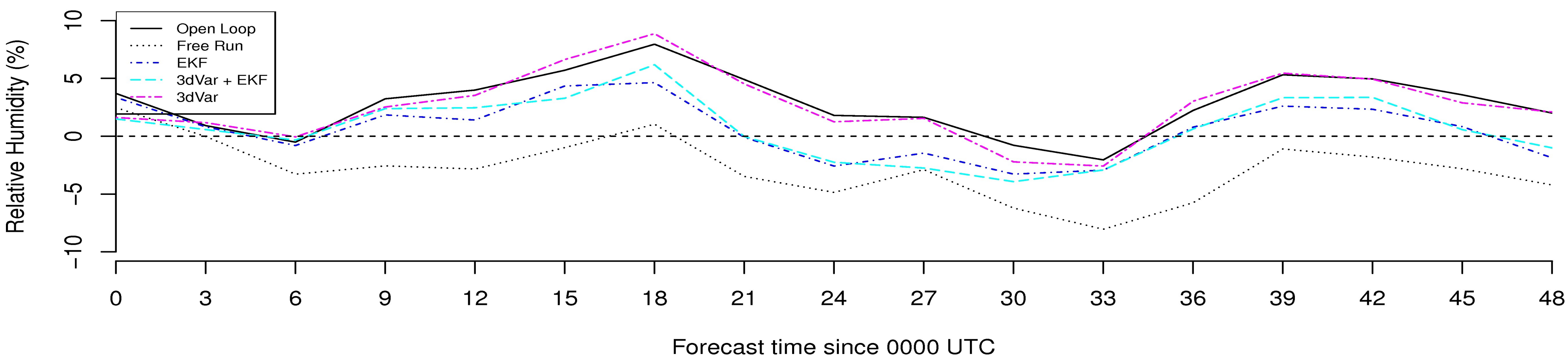

3dVar_test
now=... 20100722 ...
main
hour=... 12 18
927
927
assim
1 == 1 or 1 == 1
surfAssim
1 == 1
1 == 0
bator_surf
canari
./bator_surf == complete and ../../927 == complete
OI
2 == 1 and ./canari == complete
2 != 1
EKF
2 == 2 and ./canari == complete
2 != 2
atmAssim
1 == 1 and (1 == 0 or ../surfAssim == complete)
1 == 0
bator_atm
screening
./bator_atm == complete and ../../927 == complete
minimization
./screening == complete
001
001
../927 == complete and ((1 == 0 and 1 == 0) or ../assim == complete)
checkout
./001 == complete


2m Temperature RMSE (01–31 July 2010) run 0
Temperature (K)
Open Loop
Free Run
EKF
3dVar + EKF
3dVar
0
1
2
3
0 3 6 9 12 15 18 21 24 27 30 33 36 39 42 45 48
Forecast time since 0000 UTC
2m Temperature BIAS (01–31 July 2010) run 0
Temperature (K)
Open Loop
Free Run
EKF
3dVar + EKF
3dVar
−1
0
1
0 3 6 9 12 15 18 21 24 27 30 33 36 39 42 45 48
Forecast time since 0000 UTC

2m Relative Humidity RMSE (01–31 July 2010) run 0



2m Relative Humidity BIAS (01–31 July 2010) run 0



1. Development of a SURFEX EKF for ALARO and comparison of the offline and coupled version with the OI analysis.

2. Preliminary results on combining SURFEX EKF with 3dVar atmospheric assimilation.

3. A feasibility study of using a Short Time Augmented Extended Kalman Filter (STAEKF) for Soil Analysis.

- **Recently, Carrassi and Vannitsem (2011, QJRMS) introduced an alternative formulation of the EKF where the uncertain model parameters are estimated along with the system state variables.**

- **The algorithm, Short Time Augmented Extended Kalman Filter (STAEKF), uses a deterministic formulation for the model error dynamics (Nicolis, 2003, JAS).**

- **The same formulation has been used for the treatement of the error arising from the unresolved scales (Carrassi and Vannitsem, 2011, IJBC) and in the context of variational assimilation (Carrassi and Vannitsem, 2010, MWR).**

- **We undertake here a set of numerical twin experiments designed to test the STAEKF in estimating three land surface parameters: LAI, the albedo, and the minimum stomatal resistance RSmin.**

- **Assimilation of 2m temperature and relative humidity using an offline version of ISBA.**

- **The two-layers version of the land surface model ISBA.**

The model equations read:

$$\frac{\partial T_s}{\partial t} = C_T(R_n - H - LE) - \frac{2\pi}{\tau}(T_s - T_2)$$
$$\frac{\partial T_2}{\partial t} = \frac{1}{\tau}(T_s - T_2)$$
$$\frac{\partial w_g}{\partial t} = \frac{C_1}{\rho_w d_1}(P_g - E_g) - \frac{C_2}{\tau}(w_g - w_{geq})$$
$$\frac{\partial w_2}{\partial t} = \frac{1}{\rho_w d_2}(P_g - E_g - E_{tr}) - \frac{C_3}{d_2\tau} max[0., (w_2 - w_{fc})]$$

- **The model is available within a surface externalized platform (SLDAS, Mahfouf 2007).**

- **The state vector,** $X=(T_s, T_2, W_g, W_2)$ **and the equation can be formally written as a dynamical system,** $\frac{dx}{dt}=g(X,\lambda)$

- **The vector λ is taken to represent the set of model parameters.**

The forecast model, is augmented with P model parameters:

$$\mathbf{z}^f = \begin{bmatrix} \mathbf{x}^f \\ \lambda^f \end{bmatrix} = \mathcal{F}\mathbf{z}^a = \begin{bmatrix} \mathcal{M}\mathbf{x}^a \\ \mathcal{F}^\lambda \lambda^a \end{bmatrix}, \tag{1}$$

$\mathbf{z} = (\mathbf{x}, \lambda)$ is the augmented state vector. The augmented dynamical system $\mathcal{F}$ includes the dynamical model for the system's state, $\mathcal{M}$, and a dynamical model for the parameters $\mathcal{F}^\lambda$. In the absence of additional information, a persistence model for $\mathcal{F}^\lambda$ is often assumed so that $\mathcal{F}^\lambda = \mathrm{I}$ and $\lambda^f_{t_{k+1}} = \lambda^a_{t_k}$; the same choice has been adopted here.

The forecast/analysis error covariance matrix, $\mathbf{P}_z^{f,a}$, for the augmented system reads:

$$\mathbf{P}_z^{f,a} = \begin{pmatrix} \mathbf{P}_x^{f,a} & \mathbf{P}_{x\lambda}^{f,a} \\ \mathbf{P}_{x\lambda}^{f,a\,T} & \mathbf{P}_\lambda^{f,a} \end{pmatrix} \tag{2}$$

where the $I \times I$ matrix $\mathbf{P}_x^{f,a}$ is the error covariance of the state estimate $\mathbf{x}^{f,a}$, $\mathbf{P}_\lambda^{f,a}$ is the $P \times P$ parametric error covariance and $\mathbf{P}_{x\lambda}^{f,a}$ the $I \times P$ error correlation matrix between the state vector, $\mathbf{x}$, and the vector of parameters λ. These correlations are essential for the estimation of the parameters. In general one does not have access to a direct measurement of the parameters, and information are only obtained through observations of the system's state.

The forecast error propagation in the STAEKF is given by $\mathbf{P}_z^f = \mathbf{C}\mathbf{P}_z^a\mathbf{C}^T$, with $\mathbf{C}$ being the STAEKF forward operator defined as:

The short-time truncation of the dynamics

$$\mathbf{C} = \begin{pmatrix} \mathbf{M} & \frac{\partial g}{\partial \lambda}\Big|_{\lambda^a}\mathcal{T} \\ 0 & \mathbf{I}_P \end{pmatrix} \quad (3)$$

where $\mathbf{I}_P$ is the $P \times P$ identity matrix. Equation (3) embeds the key feature of the STAEKF; the presence of the term $\frac{\partial g}{\partial \lambda}\big|_{\lambda^a}\mathcal{T}$ allows for accounting for the contribution of the parametric error to the forecast error as well as to the error correlation between model state and parameters.

An augmented observation operator is introduced, $\mathcal{H}_z = [\mathcal{H} \quad 0]$ with $\mathcal{H}$ as for the standard EKF. Its linearization, $\mathbf{H}_z$ is now a $M \times (I + P)$ matrix in which the last P columns contain zeros. The augmented state and covariance update complete the algorithm and are equivalent to those of the EKF except that they refer now to the augmented system, and the gain matrix has dimension $(I + P) \times M$ (see Carrassi and Vannitsem, 2011a, for details).

Experimental Set-up

- **Observation system simulation experiments (OSSE).**

- **The forcing consist of 1-hourly air temperature, specific humidity, atmospheric pressure, incoming global radiation, incoming long-wave radiation, precipitation rate and wind speed relative to the ten summers in the decade 1990-1999 extract from ECMWF Re-analysis ERA40.**

- **ISBA is run in one offline single column mode for a 90 day period.**

- **The simulated observations are T2m and RH2m at 00, 06, 12 and 18 UTC.**

- **The initial Pf (B) and Pm (Q) required by the EKF, are taken from Mahfouf (2007).**
  diag(R)=(1,10-2) diag(Pf)=(1,1,10-2,10-2) diag(Pm)=(25.10-2,25.10-2,4.10-4,4.10-4)

- **Parametric errors are introduced by perturbing either alternatively or simultaneously, the Leaf Area Index, LAI , the albedo, and the minimum stomatal resistance.**

Experimental Set-up

- For each summer in the period 1990 - 1999, a reference trajectory is generated by integrating the model with LAI = 1 m2 /m2, albedo = 0.2, and Rsmin=94 s/m.

- Around each of these trajectories, Gaussian samples of 100 initial conditions and uncertain parameters are used to initialize the assimilation cycles.

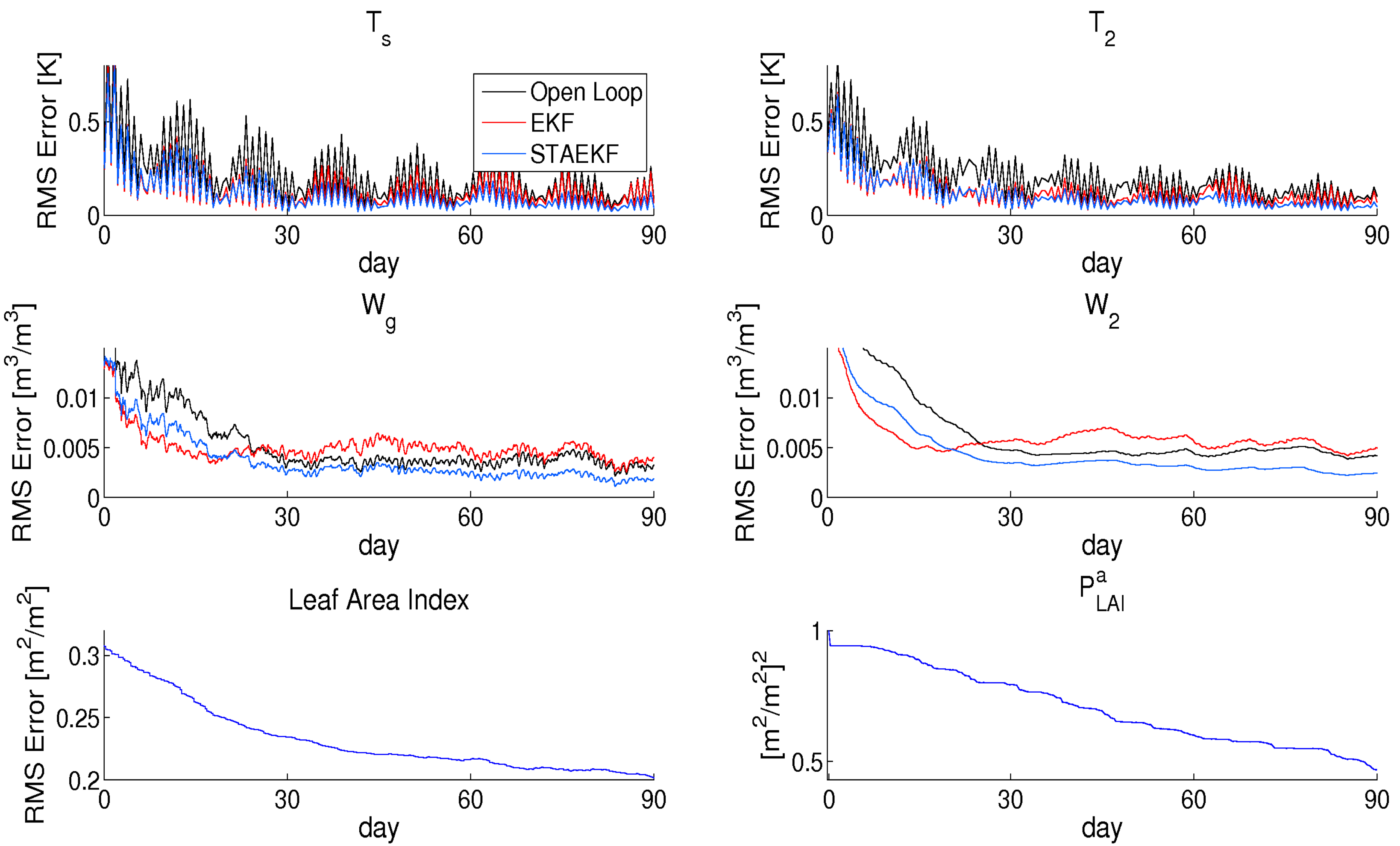
- The initial conditions are sampled from a distribution with standard deviation:

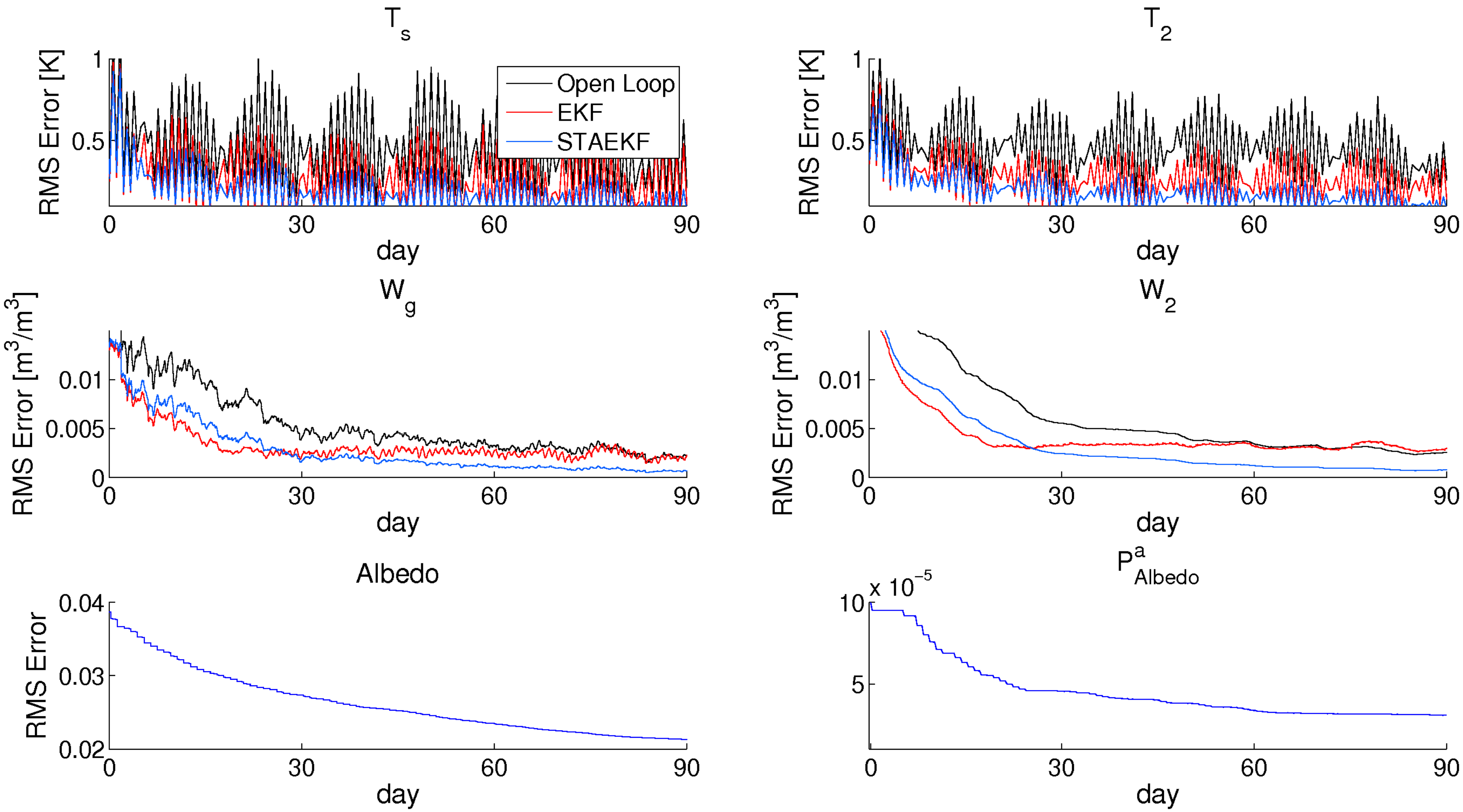
$$(\sigma T_s , \sigma T_2 , \sigma w_g , \sigma w_2) = (5, 5, 1, 1)$$

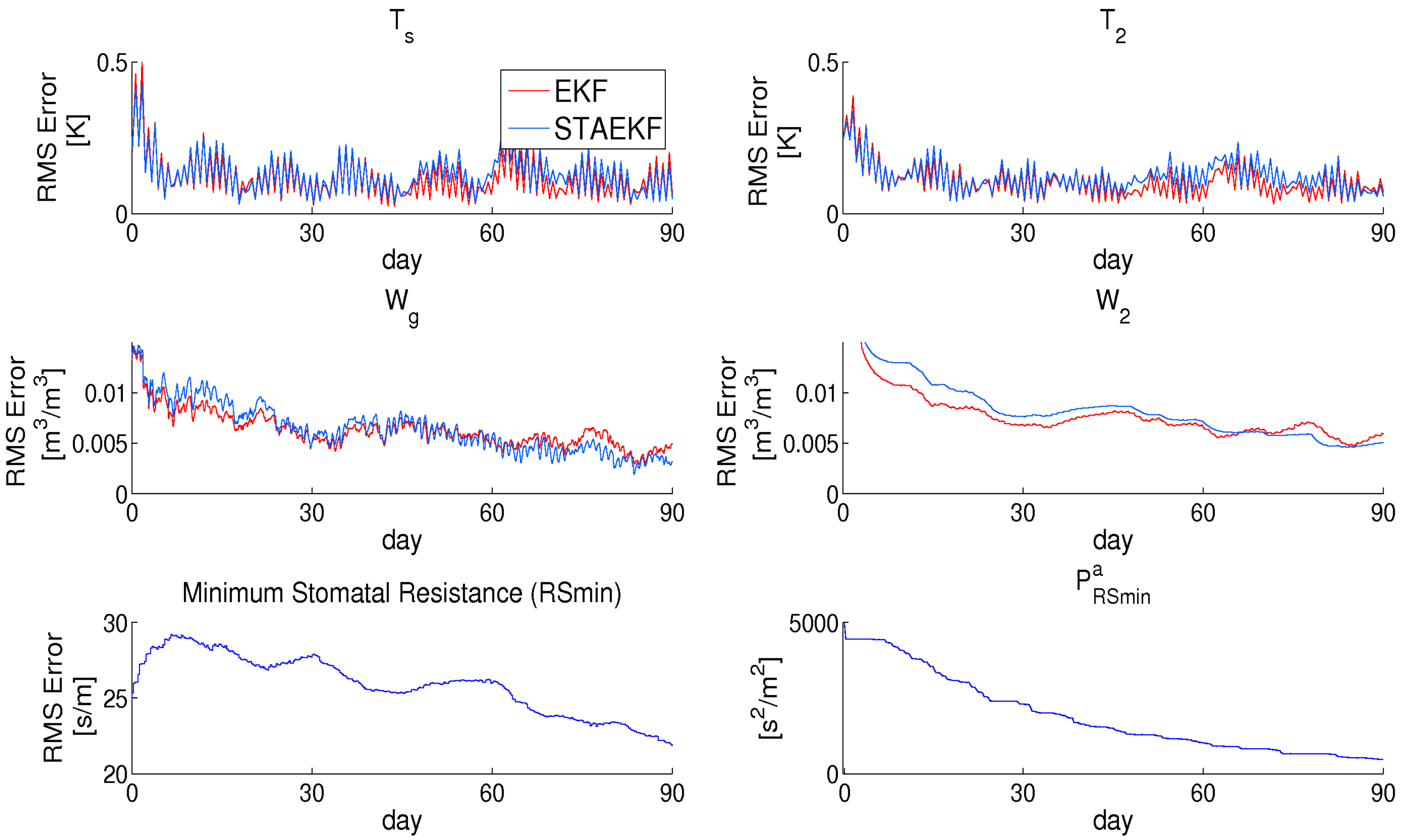
- LAI, albedo, and RSmin are sampled with standard deviations:

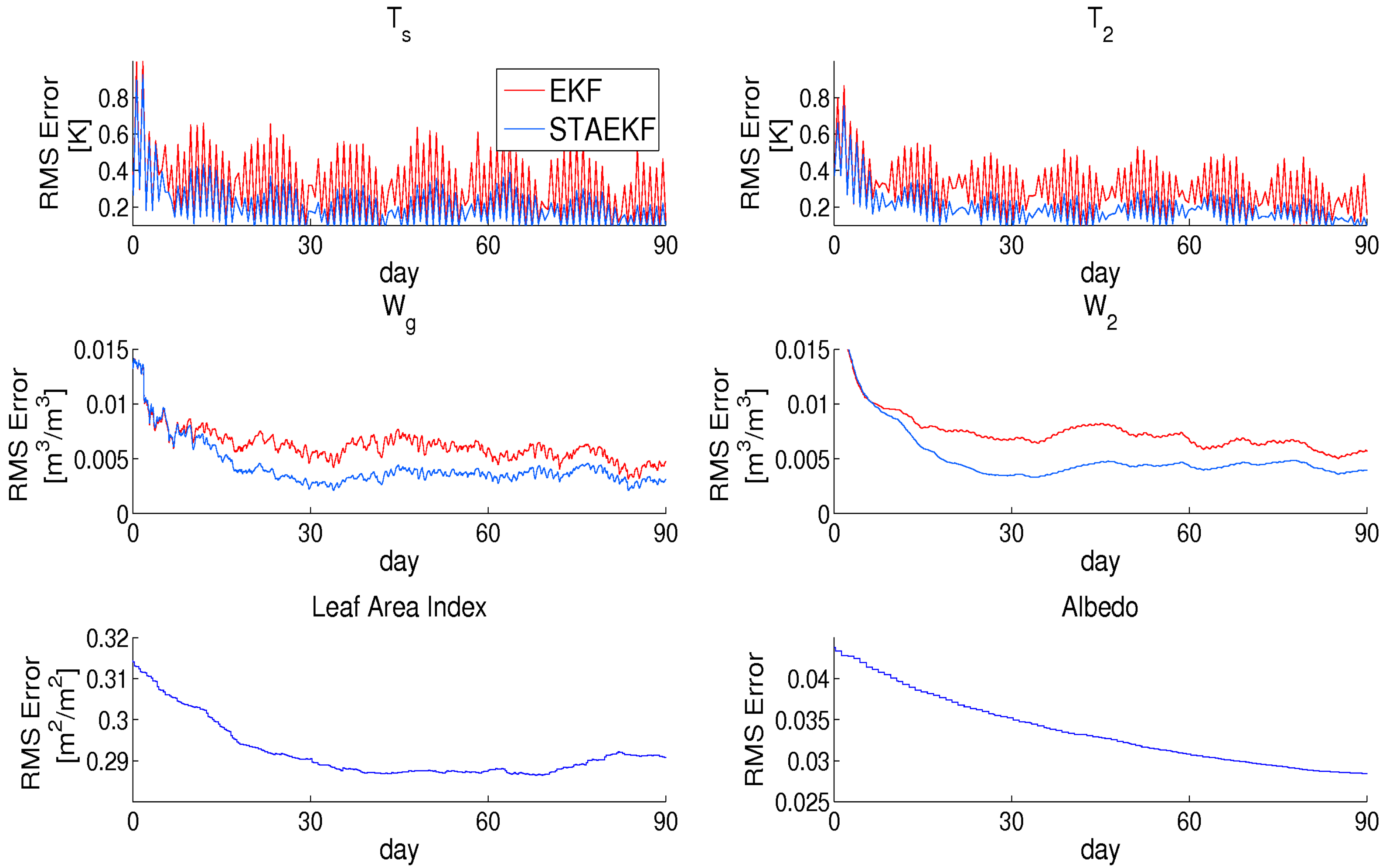
$$\sigma_{LAI} = 0.5 \quad \sigma_{albedo} = 0.1 \quad \sigma_{RSmin} = 50$$

- PLAI=1 Palbedo=10-4 PRSmin=5000 in the STAEKF, while Pax is taken as in the EKF; Px,λ is initially set to zero.


T
s
Open Loop
EKF
STAEKF
RMS Error [K]
0.5
0
0
30
60
90
day
T
2
RMS Error [K]
0.5
0
0
30
60
90
day
W
g
RMS Error [m3/m3]
0.01
0.005
0
0
30
60
90
day
W
2
RMS Error [m3/m3]
0.01
0.005
0
0
30
60
90
day
Leaf Area Index
RMS Error [m2/m2]
0.3
0.25
0.2
0
30
60
90
day
P
a
LAI
[m2/m2]2
1
0.5
0
30
60
90
day


T_s
RMS Error [K]
1
0.5
0
30
60
90
day
Open Loop
EKF
STAEKF
T_2
RMS Error [K]
W_g
RMS Error [m^3/m^3]
0.01
0.005
W_2
Albedo
RMS Error
0.04
0.03
0.02
P^a_Albedo
x 10^-5
10
5


T s
RMS Error [K]
EKF
STAEKF
day
T 2
RMS Error [K]
W g
RMS Error [m3/m3]
W 2
Minimum Stomatal Resistance (RSmin)
RMS Error [s/m]
P a RSmin
[s2/m2]


T_s
T_2
EKF
STAEKF
RMS Error [K]
RMS Error [K]
0.8
0.6
0.4
0.2
0
30
60
90
day
W_g
W_2
RMS Error [m^3/m^3]
0.015
0.01
0.005
Leaf Area Index
Albedo
RMS Error [m^2/m^2]
0.32
0.31
0.3
0.29
RMS Error
0.04
0.035
0.03
0.025

- **Continuing the evaluation of the STAEKF, a scientific paper is submitted: Short Time Augmented Extended Kalman Filter for Soil Analysis: A feasibility study.**
  Carrassi A., Hamdi R., Termonia P., and Vannitsem S., 2012 ASL.

- **The STAEKF is able to reduce the parameter estimation errors.**

- **The accuracy of these estimates is inherently related to the type of parameter to be estimated.**

- **The model sensitivity to the specific parameter and the accuracy of the short-time approximation in the STAEKF.**

- **The rate of error convergence in the STAEKF is related to the initial parametric error variance.**

- **Implementing the STAEKF in SURFEX and study its behaviour within ALARO.**