Surface assimilation: status and plans

R. Hamdi
(Royal Meteorological Institute of Belgium)

With help from: J-F Mahfouf, F. Bouyssel, L. Auger

and input from: Adam Dziedzic, Karim Bergaoui

2nd AROME-Training courses Lisbon (Portugal) 4-7 March 2008

Plan

- **Introduction**

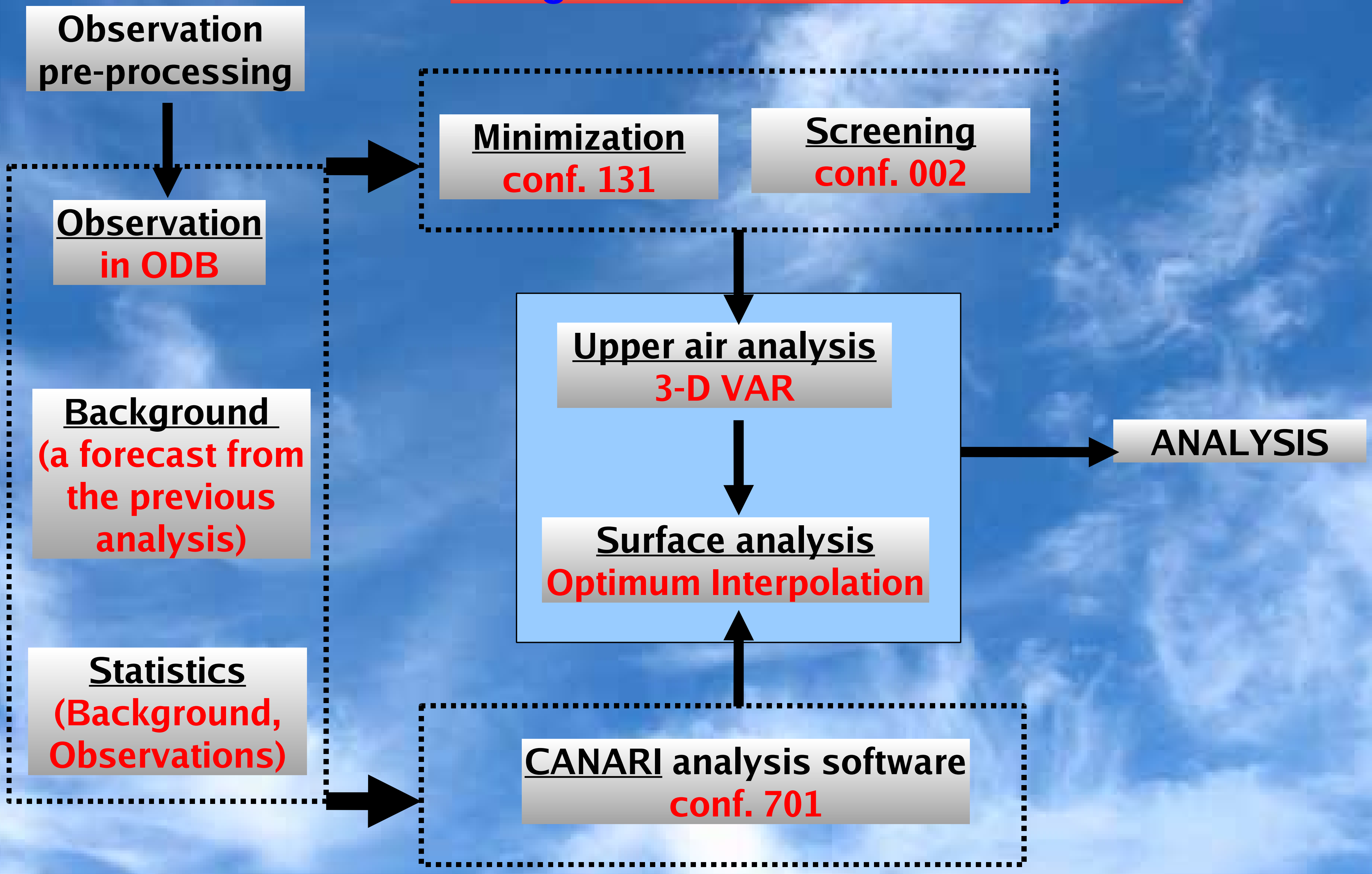
- **Surface analysis by Optimum Interpolation using CANARI software**

- **Moving away from Optimum Interpolation using SURFEX software**

- **Work to be done**

Introduction

The global data assimilation system

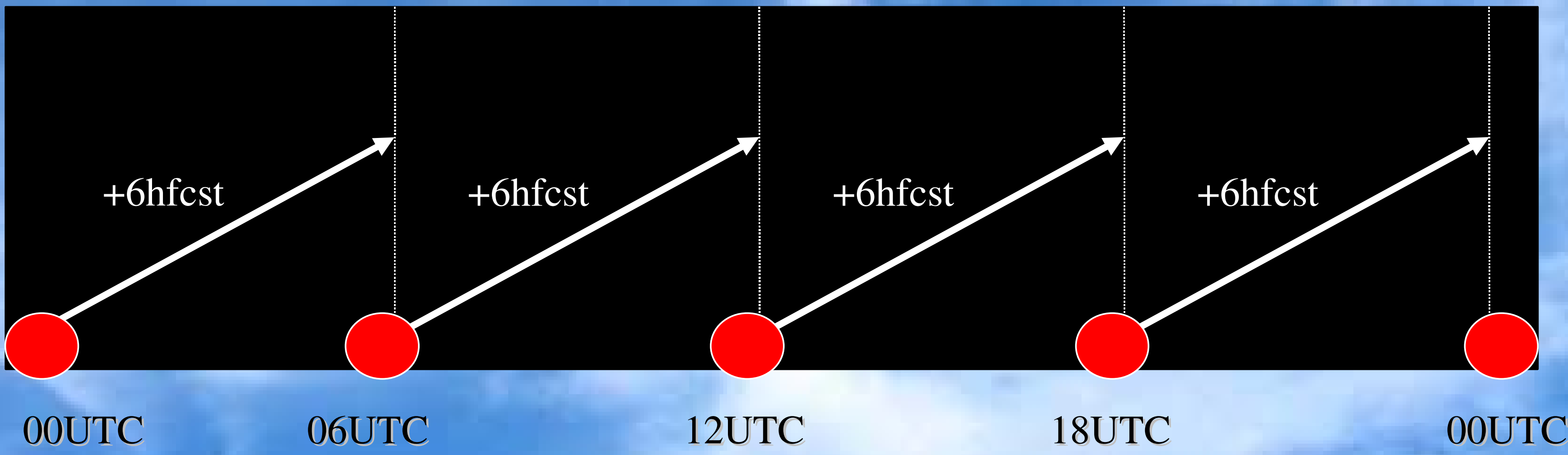


Surface analysis and upper-air analysis

- For the time being surface analysis are performed separately from upper air analysis. In theory a single analysis would be better but it is much more difficult to implement: 1) definition of **B** between upper air and surface variables, 2) time scale evolutions may be different, ...

- For the time being several surface analysis are used for different surface parameters (Soil temperature and Soil moisture, Snow, SST, Sea ice, ...

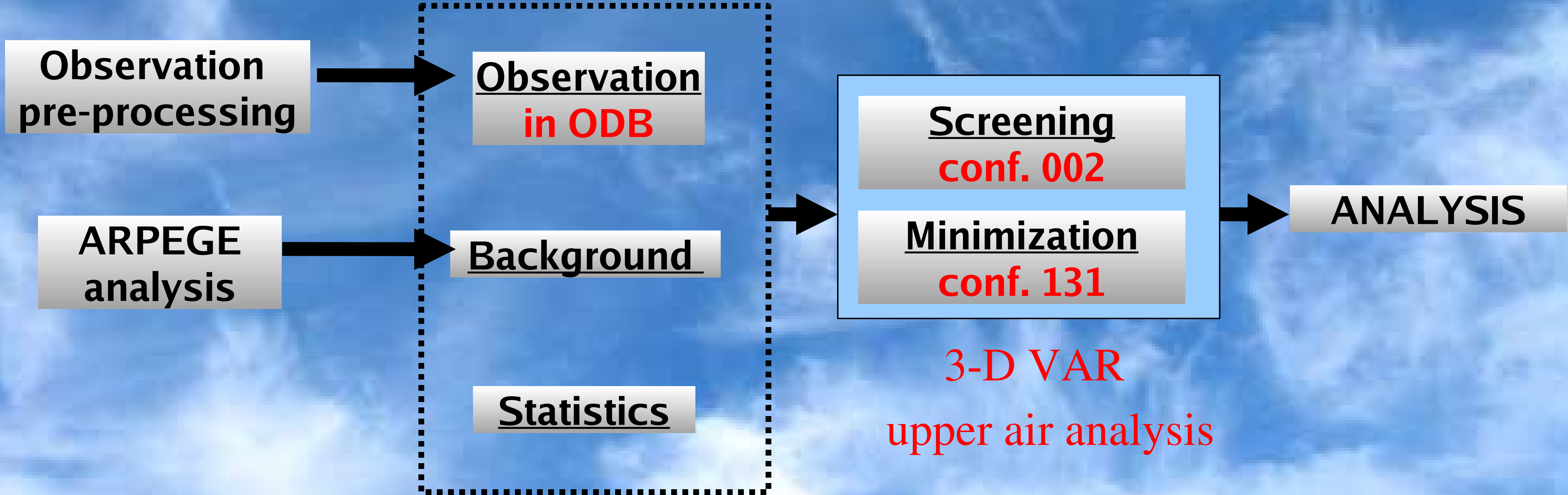
- Atmospheric analysis and several surface analysis are done separately and combined to provide the final analysis for the forecast.



(6h sequential analysis is just an example)

Surface analysis for ALADIN initialization: current status

- In most ALADIN countries the **CANARI** step is omitted in the operational practice, and the information from observations is used **only** via the global ARPEGE surface analysis projected to the ALADIN grid.

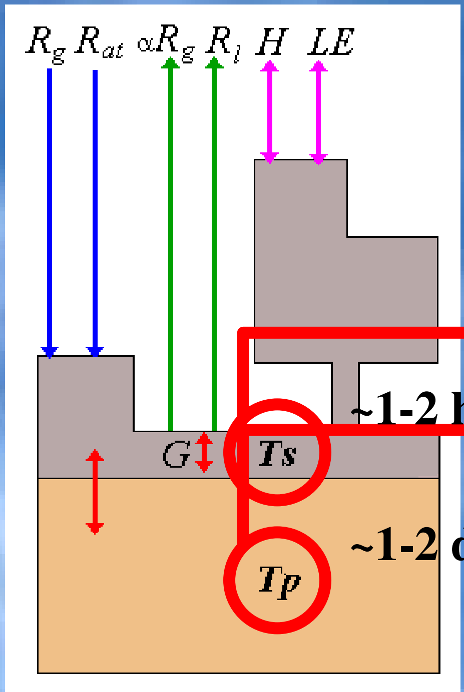


Surface analysis by Optimum Interpolation using CANARI software

Surface Parameterization scheme (ISBA)

- The **ISBA-2L** scheme evolves 4 prognostic variables. (Giard and Bazile 2000)

Energy



Analysis
- surface temperature
- mean soil temperature
- superficial soil water content
- total soil water content

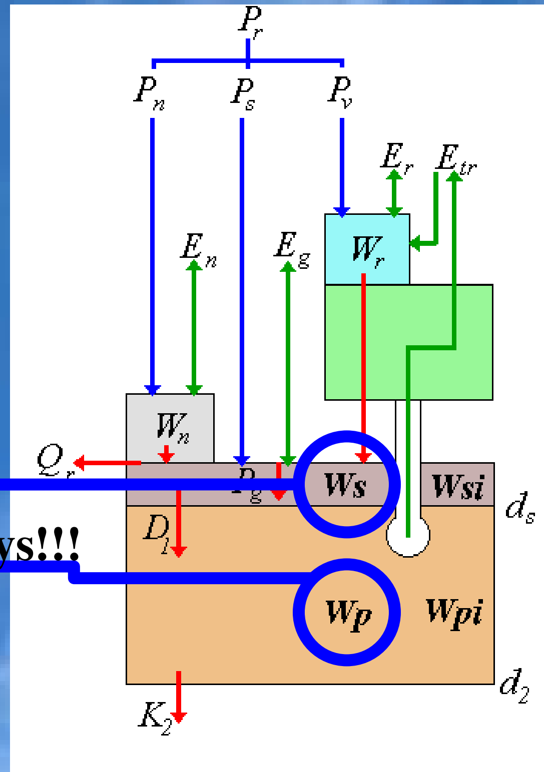
$$\frac{\partial T_s}{\partial t} = C_T(R_n - H - LE) - \frac{2\pi}{\tau}(T_s - T_2)$$

$$\frac{\partial T_2}{\partial t} = \frac{1}{\tau}(T_s - T_2)$$

$$\frac{\partial w_g}{\partial t} = \frac{C_1}{\rho_w d_1}(P_g - E_g) - \frac{C_2}{\tau}(w_g - w_{geq})$$

$$\frac{\partial w_2}{\partial t} = \frac{1}{\rho_w d_2}(P_g - E_g - E_{tr}) - \frac{C_3}{\tau}\max\left[0., (w_2 - w_{fc})\right]$$

Water



- Research versions: interactive vegetation module (Calvet et al. 1998), sub grid-scale runoff and sub-root layer (Boone et al. 1999), explicit 3-layers snow scheme (Boone and Etchevers 2001)...

Soil moisture content

- Volumetric water content (m^3 of water / m^3 soil): W_p.
- Saturation water content or porosity = maximum amount of water that a given soil can hold : W_{sat}=450 mm (for 1 m soil depth)
- Water content at field capacity = value above which evaporation takes place at potential rate-when water excess has drained away : W_{fc}=300 mm
- Water content at wilting point = value below which plants cannot extract soil water from their root system : W_{wilt}=200 mm
- Soil Wetness Index = normalized soil moisture content :

$$SWI = \frac{(W_p - W_{wilt})}{(W_{fc} - W_{wilt})}$$

Link between soil moisture and surface boundary layer

- The main interaction of soil moisture and SBL is due to evaporation and vegetation transpiration process.

- Under strong solar radiation at the surface, soil moisture becomes very important because it determines the repartition of incoming net radiation into sensible and latent heat flux.

Net radiation at the surface :

$$R_n = R_G(1-\alpha_t) + \epsilon_t(R_A - \sigma T_s^4) = H + LE + G$$

- For $W_p < W_{wilt}$: LE is almost negligible

- For $W_p > W_{fc}$: LE =potential evaporation

- Importance of soil moisture and temperature analysis because the impact of a prescribed initial error in the soil moisture field may degrade significantly the forecast during long period **up to several days.**

Available observations for surface analysis

- **Precipitations observations (rain gauges, radars) :**
  + direct link with the variations of soil water content

- Satellite observations:
  + global coverage
  + infrared: clear sky, low vegetation, geostationary satellites : high temporal and spatial resolutions (energy budget), strong sensitivity to low level wind, surface roughness
  + microwave: active and passive instruments measure directly the soil moisture in the first few centimeters (scatterometer (ERS,ASCAT), passive or active radiometers (SMOS, AMSR): resolution ~20/40km, frequency ~0.3/1 per day

- 2m observations **(temperature and humidity):**
  + good global coverage of existing network
  + close links with the fields in the ground in specific meteorological conditions

Optimum Interpolation: basic theory

- Based on Best Linear Unbiased Estimation (BLUE) :

$$X^A = X^G + \underbrace{BH^T\left(HBH^T + R\right)^{-1}}_{K}\left(Y - HX^G\right)$$

- with
  - **X^A:** analyzed state vector
  - **X^G:** background state vector
  - **Y:** observation vector
  - **H:** observation operator (model space to observation space)
  - **B:** background error covariance matrix
  - **R:** observation error covariance matrix
  - **K:** gain matrix

Optimum Interpolation: implementation in the CANARI software (I)

1) **Optimum Interpolation of T_{2m} and RH_{2m} using 2m observations interpolated at the model grid-point by a 2m analysis (2-D CANARI OI)**

$$\Delta T_{2m} = T_{2m}{}^{a} - T_{2m}{}^{b} \qquad \Delta RH_{2m} = RH_{2m}{}^{a} - RH_{2m}{}^{b}$$

2) **Correction of 4 surface parameters (T_s, T_p, W_s, W_p) using 2m increments between analysed and forecasted values.**

$$T_p^{a} - T_p^{b} = \Delta T_{2m} / 2\pi \qquad T_s^{a} - T_s^{b} = \Delta T_{2m}$$

$$W_s^{a} - W_s^{b} = \alpha_{WsT} \Delta T_{2m} + \alpha_{WsRH} \Delta RH_{2m}$$

$$W_p^{a} - W_p^{b} = \alpha_{WpT} \Delta T_{2m} + \alpha_{WpRH} \Delta RH_{2m}$$

OI coefficients

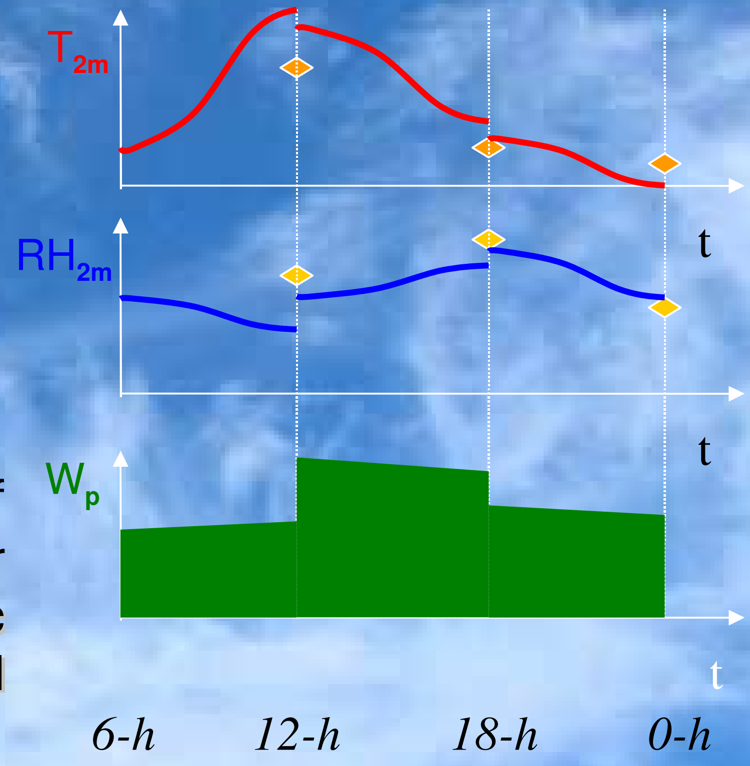
$$\alpha_{Ws/pT} = \frac{\sigma^b_{Ws/p}}{\Phi \sigma^b_{T2m}} \left\{ \left[1 + \left(\frac{\sigma^a_{RH2m}}{\sigma^b_{RH2m}} \right)^2 \right] \rho_{T2m,Ws/p} - \rho_{T2m,RH2m} \rho_{RH2m,Ws/p} \right\}$$

$$\alpha_{Ws/pRH} = \frac{\sigma^b_{Ws/p}}{\Phi \sigma^b_{RH2m}} \left\{ \left[1 + \left(\frac{\sigma^a_{T2m}}{\sigma^b_{T2m}} \right)^2 \right] \rho_{RM2m,Ws/p} - \rho_{T2m,RH2m} \rho_{T2m,Ws/p} \right\}$$

$$\Phi = \left[1 + \left(\frac{\sigma^a_{T2m}}{\sigma^b_{T2m}} \right)^2 \right] \left[1 + \left(\frac{\sigma^a_{RH2m}}{\sigma^b_{RH2m}} \right)^2 \right] - \rho^2_{T2m,RH2m}$$

- Very strong dependency of these background error statistics to **physiographic** properties and **meteorological conditions.**

Sequential analysis (every 6h)



Optimum Interpolation: implementation in the CANARI software (II)

- OI coefficient values were originally calculated in 1-D experiments with a MonteCarlo method, and for a single grid point location under anticyclonic summer conditions.

- For the operational implementation a set of regressions allow to define the OI coefficients to each grid point according to : vegetation fraction, LAI/R_{smin}, soil texture, and local solar time. (Bouttier et al. 1993a, b)

- In the operational global surface analysis, the meteorological conditions have been taken into account through a weighting function that ranges between 0 and 1.

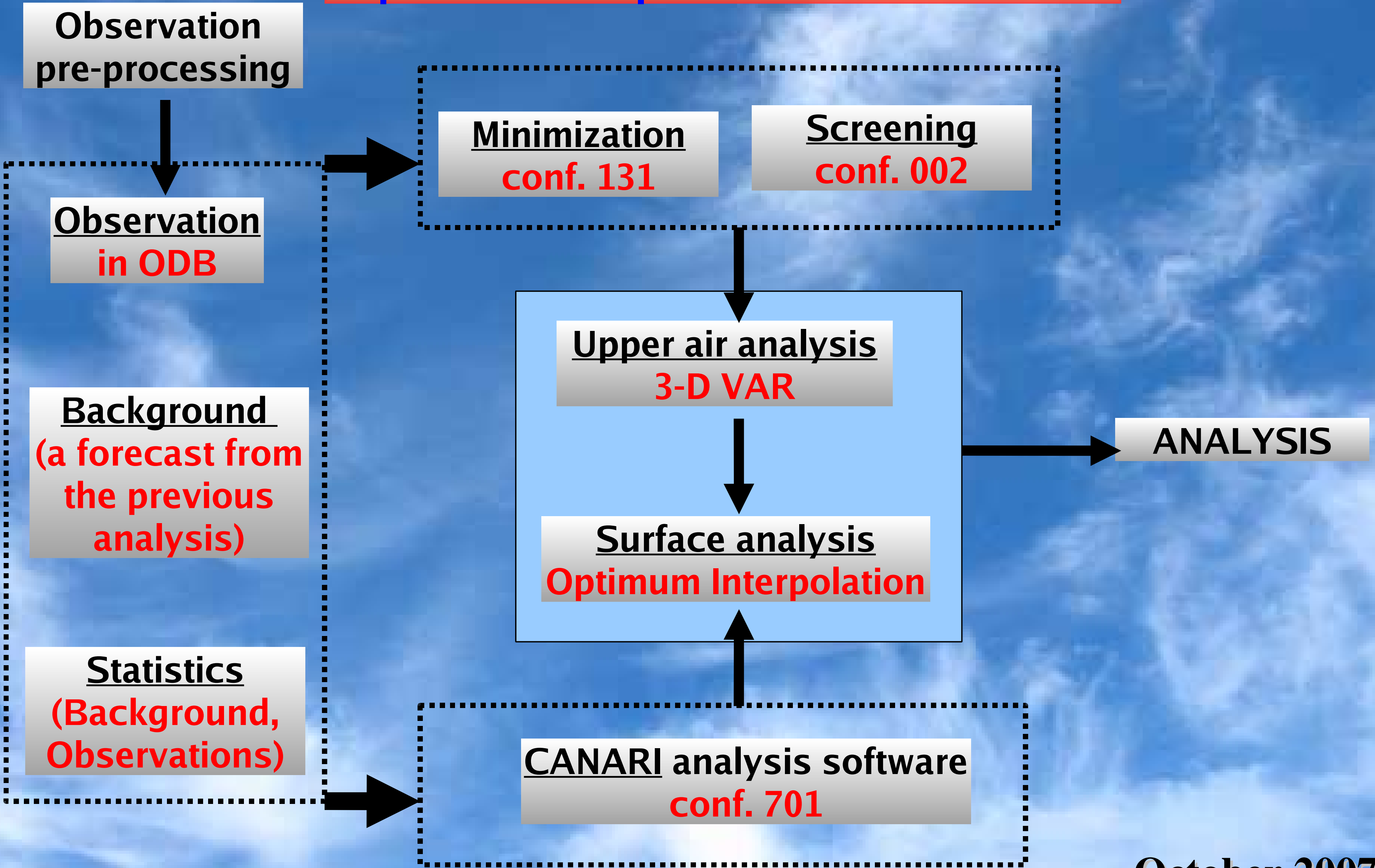
$$\alpha_{Wp/sT/RH} = f\,(t,\ veg,\ LAI/Rs_{min},\ texture,\ atmospheric\ conditions)$$

- This function evaluates several quantities and modulate the OI coefficients on grid points exceeding given thresholds:

- Giard and Bazil (2000)

Model Fields		Threshold
Min solar time duration	J_min	6 h
Max wind velocity	Vmax	10 ms^{-1}
Max precipitation	P_max	0.3 mm
Min surface evaporation	E_min	0.001 mm
Max soil ice	W_imax	5.0 mm
Presence of snow	Sn_max	0.001 kg/m2

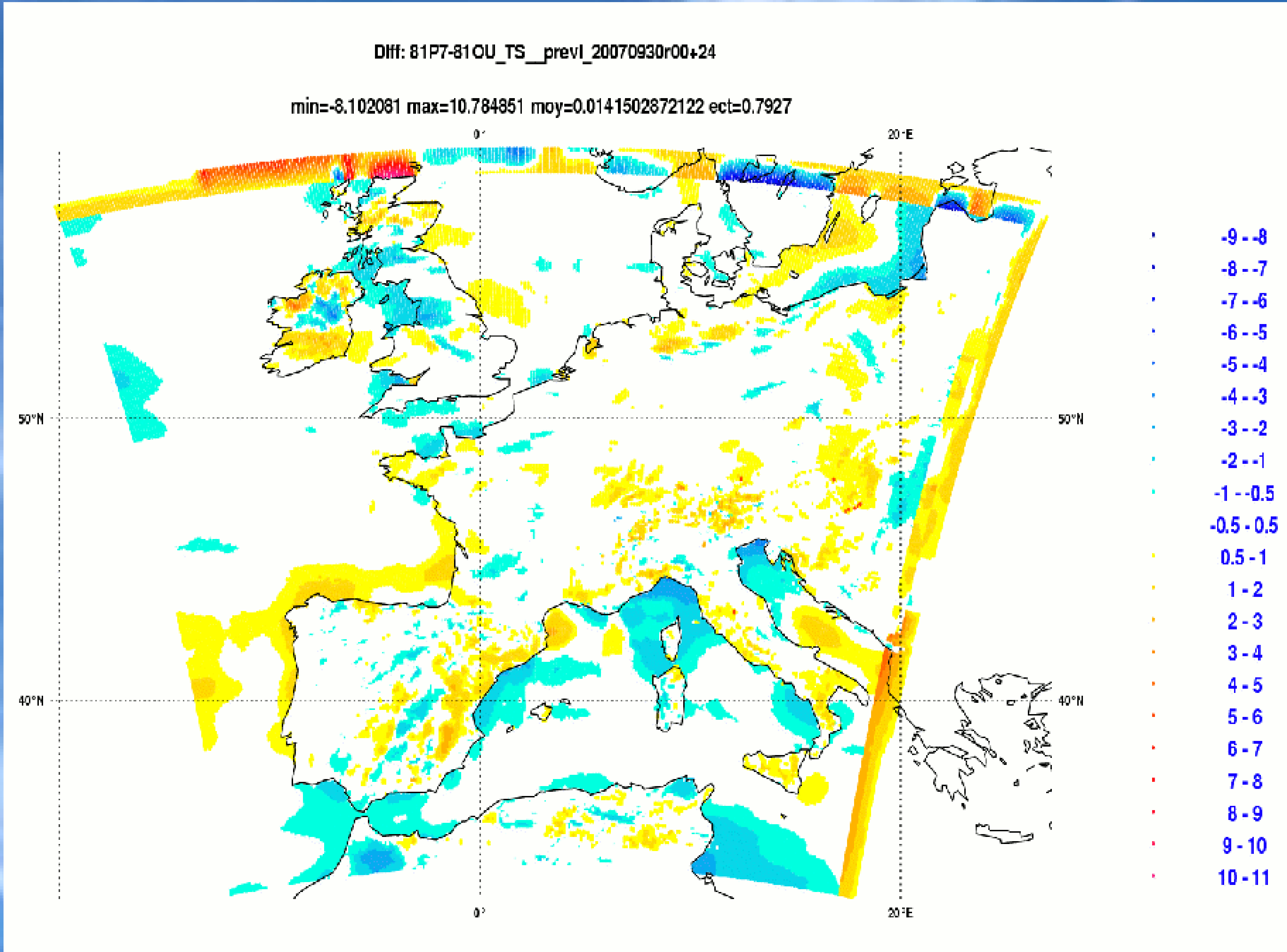
Optimum Interpolation: some results



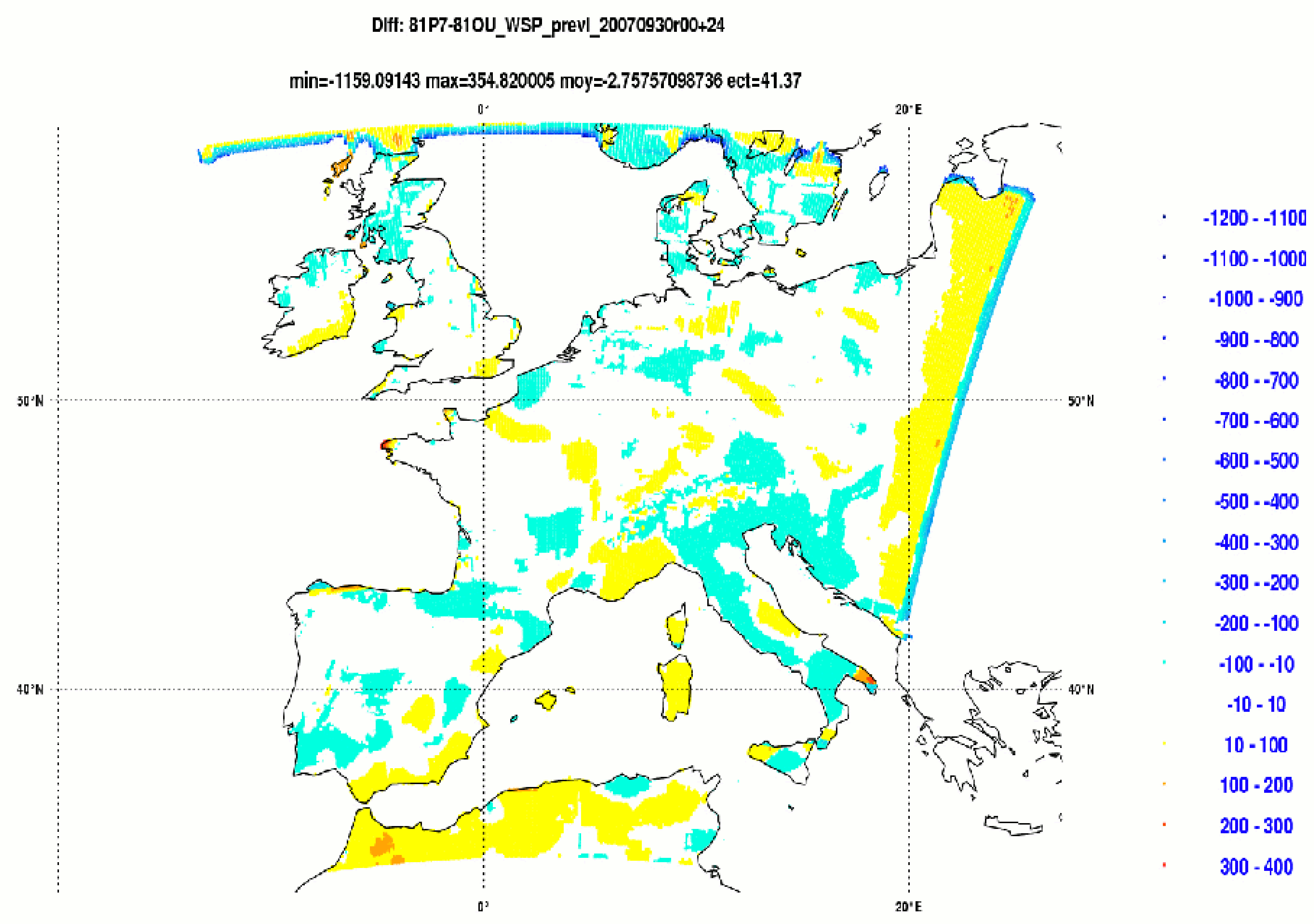
October 2007

Adam Dziedzic, Ludovic Auger, Françoise Taillefer

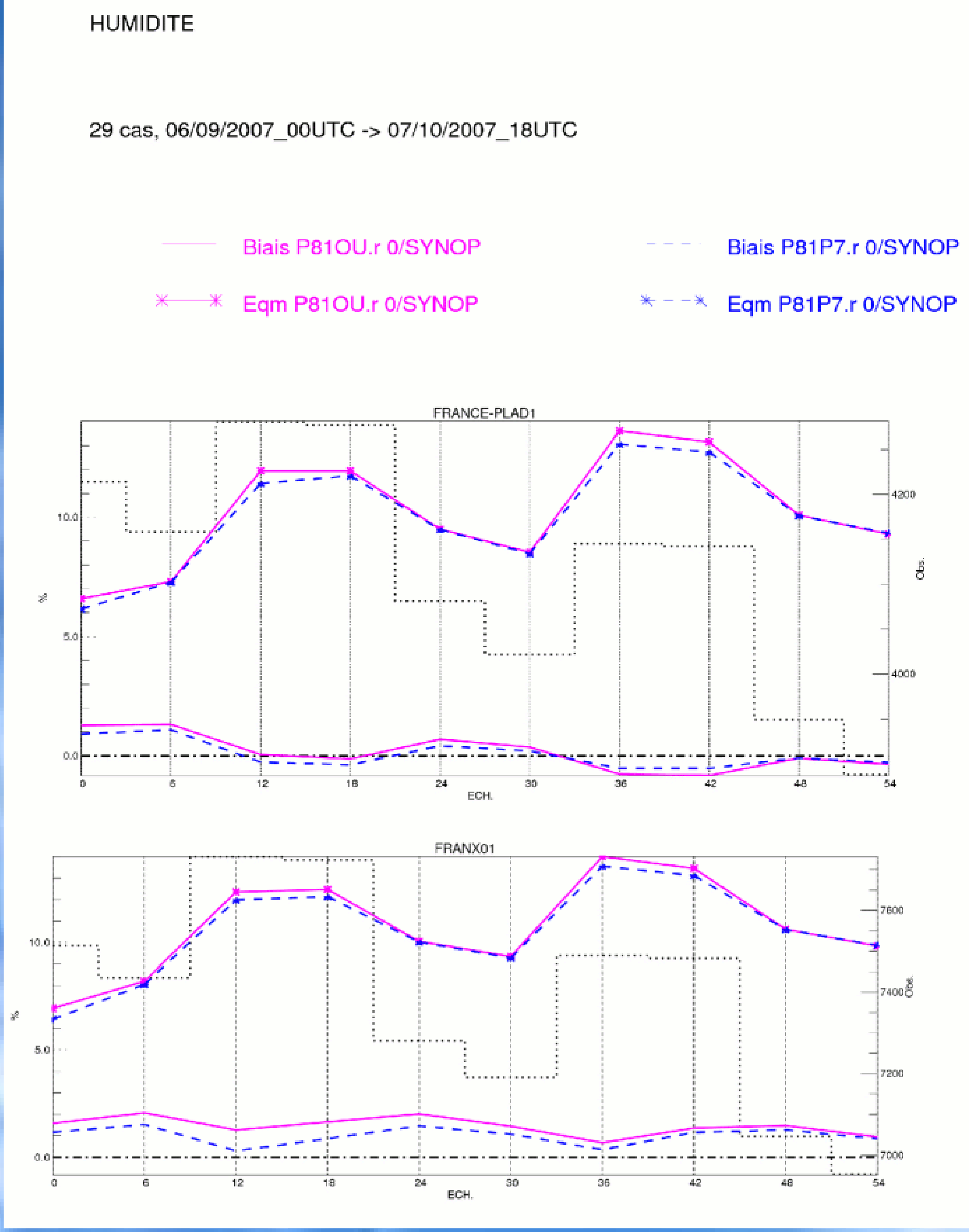
Optimum Interpolation: some results

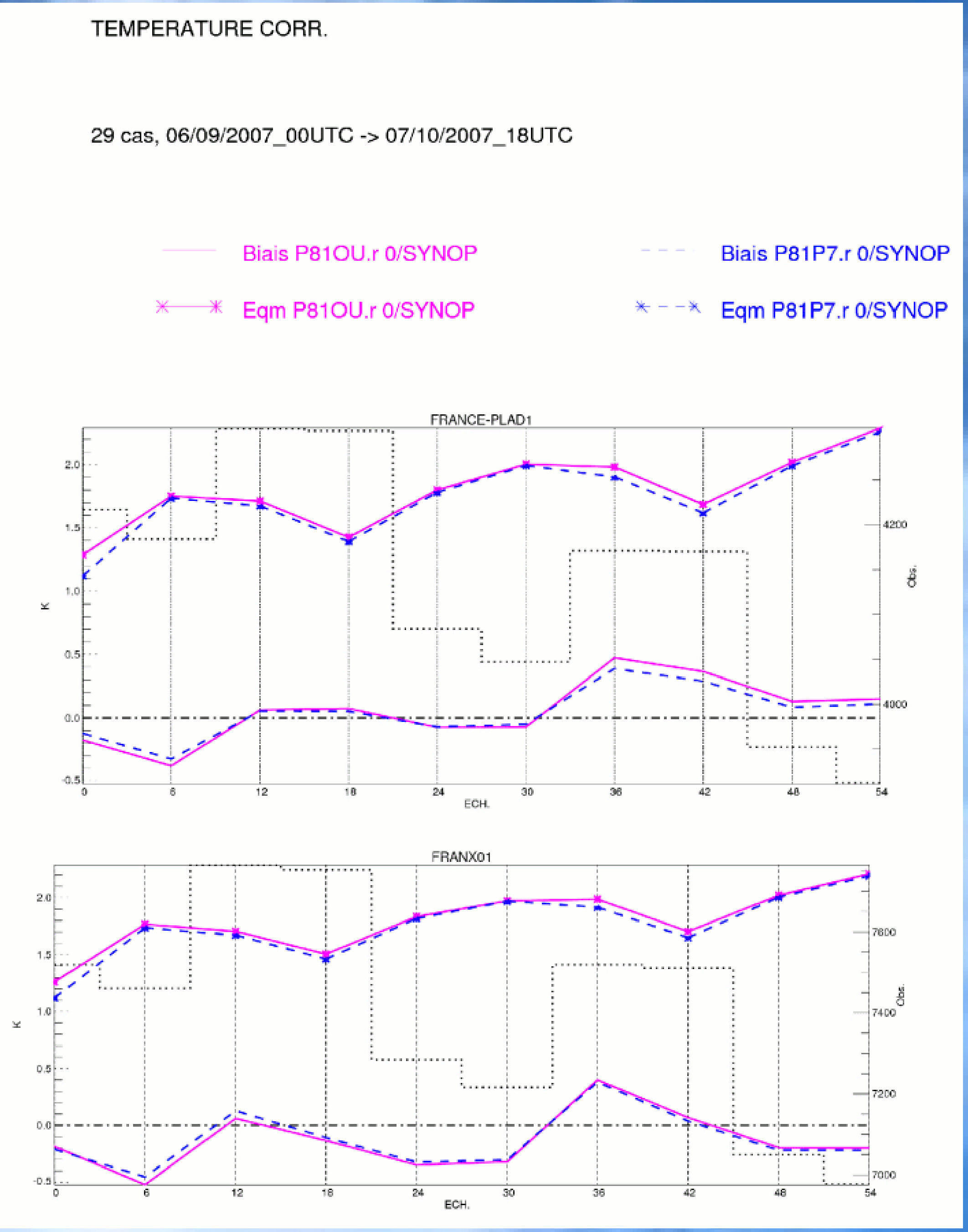


Optimum Interpolation: some results



Optimum Interpolation: some results





Optimum Interpolation: some weaknesses

- **Linear observation operator:** The link between current available observations (T2m, HU2m) and soil variables is complex (non linear relation with surface evaporation fluxes, dependence with the vertical interpolation scheme and with the prognostic equation for the land surface scheme).

- **The OI coefficients** are obtained by statistical equations and it is difficult to correct objectively (Factor 6 reduction of OI coefficient on W_p).

- **Arbitrary thresholds:** Need to impose arbitrary thresholds to avoid soil analysis in conditions where forecast errors at screen level are not related to soil moisture.

- **Difficult to consider new variable to analyse:** The current OI coefficients for T_s and T_p are taken from the previous global analysis (since Mahfouf (1991) did not perform Monte-Carlo experiments for these variables). The complexity of the ISBA scheme will increase for future NWP applications: ISBA-3L, ISBA-DF, ISBA-Ags, ISBA-SBL,...

- **Difficult to consider new observations:** Some observations that are more directly informative about soil/vegetation state than screen level parameters are available...

Moving away from Optimum Interpolation using SURFEX software

Proposal

- **Framework:** Offline version of SURFEX.
- **Forcing:** Short-range forecasts from the atmospheric model that will use the soil analyses as initial conditions.
- **Method:** Advanced assimilation techniques.
- Most of the deficiencies of the Optimum Interpolation previously mentioned are gone.
- Develop an analysis of soil prognostic variables suitable for the various NWP models (ARPEGE, ALADIN, AROME, ALARO), that can assimilate various observation types (conventional observations, satellite data, precipitation, surface radiative fluxes).
- This system will make the soil analysis independent of the definition of the land surface scheme and could evolve towards EnKF or Var approaches.

Vertical interpolation scheme in SURFEX

The vertical interpolation of variables at 2 meters can be done in **three ways** in SURFEX:

1. Diagnostic ways:

- Use of analytic formulation proposed by **Geleyn (1988)** that requires the knowledge (specification) of surface variables (temperature and humidity) that may be difficult to obtain when considering a tile approach for the surface (already in ISBA without tiles, the various evaporation flux components make the definition of surface specific humidity rather complex).

- Use of the Monin-Obukhov similarity theory to produce profiles that depend only on surface fluxes (various stability functions can be chosen, particularly in stable regime) and the upper air level in the surface boundary layer **Paulson (1970).**

2. Prognostic way:

- Couple SURFEX to a surface boundary layer scheme and put the first air level at 2 m (Canopy model) **Masson and Seity (2008)** and **Hamdi and Masson (2008)**.

The Extended Kalman Filter (I)

- We consider a control vector **X** (N_x) that represents the prognostic equations of ISBA ($\mathcal{M}$) which evolves with time as: $X_f^t = \mathcal{M}(X^0)$, with N_x=4 and $X=(W_p,W_s,T_p,T_s)$ the forecast X_f^t is characterized by a (forecast) background error covariance matrix **B**.

- An N_y dimensional observation vector Y_o is available at the regularly spaced discrete times $t_k=t_0+k.\tau$, k=1,2,... characterized by an error covariance matrix **R** and assimilation interval τ (06h for example).

- An observation operator $\mathcal{H}$ maps from model to observed variables $Y_f^t = \mathcal{H}(X_f^t)$. For example, it can be a vertical interpolation scheme for T_{2m} and HU_{2m} or a microwave radiative transfer model for brightness temperatures.

- Under the Tangent Linear (TL) hypothesis, the $\mathcal{H}$ operator can be expressed by its first-order Taylor expansion $\mathcal{H}(X+\delta X)=\mathcal{H}(X)+H.\delta X$ **H** is the Jacobian matrix of $\mathcal{H}$

- A new value X_a^t is obtained by an optimal combination of the observations and the background (BLUE):

$$X_a^t = X_f^t + BH^T(HBH^T+R)(Y_o^t - \mathcal{H}(X_f^t))$$

The Extended Kalman Filter (II)

- In this low dimensional problem, the Jacobian matrix **H** is obtained by finite differences:

$$H_{ij}[N_x raws, N_y columns] = \frac{\partial y_i}{\partial x_j} \simeq \frac{y_i(x+\delta x_j) - y_i(x)}{\delta x_j}$$

- The input vector x is perturbed N_x times to get for each integration a column of the matrix **H.**

- The analysis is cycled by propagating in time the two quantities $\mathbf{X_a}^t$ and $\mathbf{A}$ up to the next time where observations are available: $\mathbf{X_f}^{t+1} = \mathcal{M}(\mathbf{X_a}^t)$ and $\mathbf{B}^{t+1} = \mathbf{M}\mathbf{A}^t\mathbf{M}^T + \mathbf{Q}$, where $\mathbf{M}$ is the Jacobian matrix of the model, $\mathbf{A} = (\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}$ is the analysis error covariance matrix, and $\mathbf{Q}$ a new matrix representing the model error covariance matrix which needs to be defined.

- For the Simplified Extended Kalman Filter (SEKF), we take: $\mathbf{B}^{t+1} = \mathbf{B}^t = \mathbf{B}$ and $\mathbf{Q} = 0$.

- The SEKF has been coded within SURFEX by K. Bergaoui (Tunisia) in spring 2007 and is currently under scientific evaluation at Météo-France and IRM (Belgium).

The SEKF for W_P analysis within SURFEX

- Let us consider a SURFEX guess **S** and a perturbed state **S'**, where $\mathbf{W_P}$ has been modified by a small quantity $\boldsymbol{\delta}\mathbf{W_P}$.

- The SURFEX integrations provide the 2m forecast sensitivity evaluated at time $\mathbf{t_1}$ at which the observations of temperature, $\mathbf{T_{2m}}$ and relative humidity $\mathbf{RH_{2m}}$ are available: $\delta T_{2m}^{(1)} = T_{2m}^{S'(1)} - T_{2m}^{S(1)}$ and $\delta RH_{2m}^{(1)} = RH_{2m}^{S'(1)} - RH_{2m}^{S(1)}$.

- For a 06h assimilation interval, the matrices **B**, **R** and $\mathbf{H^T}$ used for the computation of **K** are:

$$B = (\sigma^2_{W_p})$$

$$R = \begin{pmatrix} \sigma^2_{T_{2m}} & 0 \\ 0 & \sigma^2_{RH_{2m}} \end{pmatrix}$$

$$H^T = \begin{pmatrix} \dfrac{\delta T^1_{2m}}{\delta W_p^{(0)}} \\ \dfrac{\delta RH^1_{2m}}{\delta W_p^{(0)}} \end{pmatrix}$$

- If we expand the analysis correction according to the BLUE equation:

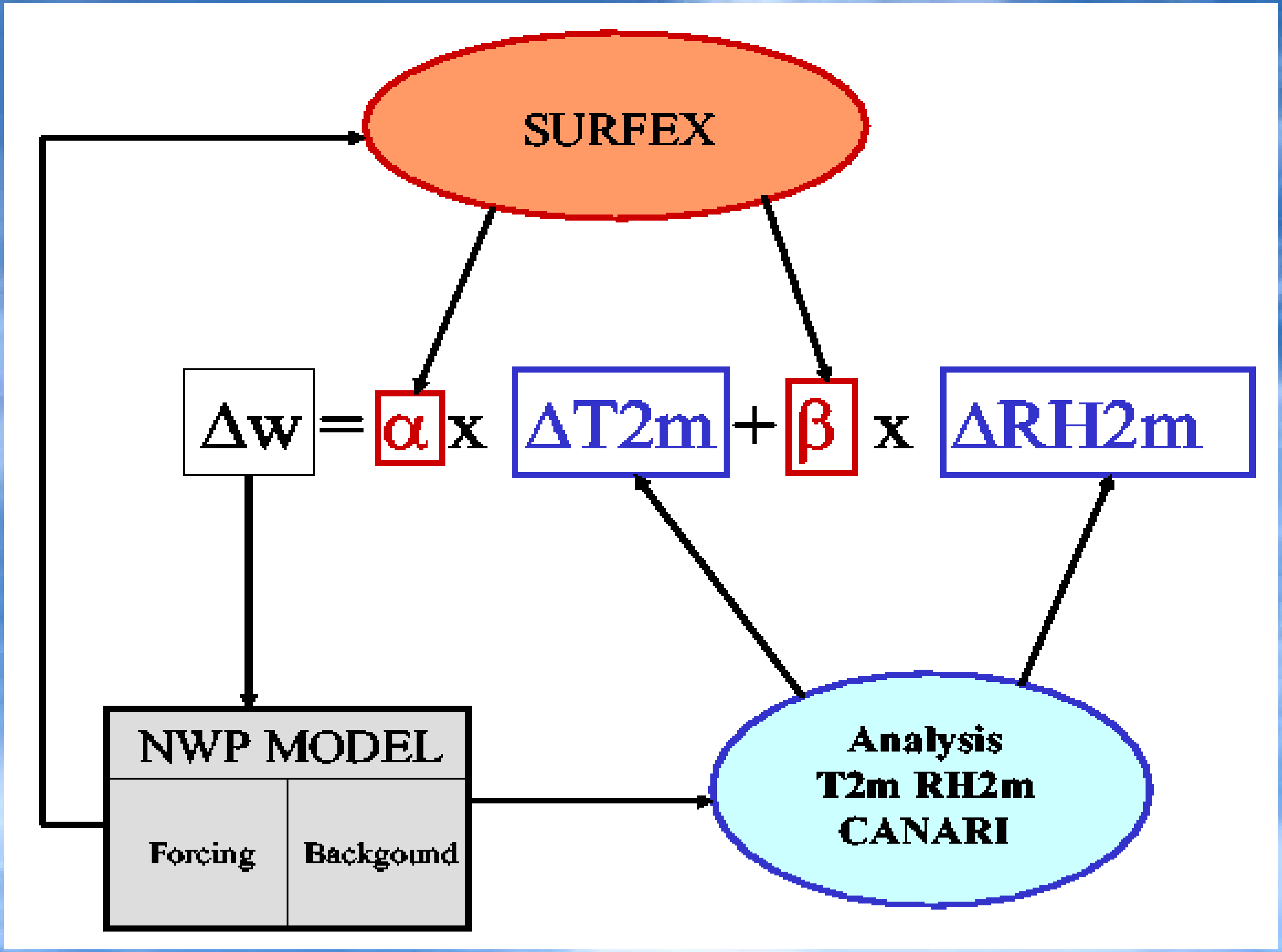
$$W_p^a = W_p^f + (k_1 \quad k_2) \begin{pmatrix} \Delta T^1_{2m} \\ \Delta RH^1_{2m} \end{pmatrix}$$

- $\mathbf{k_1}$ and $\mathbf{k_2}$ are the elements of the gain matrix. $\Delta T_{2m}{}^1$ and $\Delta RH_{2m}{}^1$ are innovation vectors at the time $\mathbf{t_1 = t_0 + 06h}$.

OI Coef.

$$W_p^a - W_p^b = \alpha_{WpT} \Delta T_{2m} + \alpha_{WpRH} \Delta RH_{2m}$$

Coupling between atmospheric model and offline SURFEX



Feasibility study within SURFEX (I)

Description of the "twin experiments" framework:

- Define a reference simulation from which "**perfect observations**" are generated.
- Perturb the W_P initial condition in order to run an "**open loop**" simulation.
- Start from the perturbed W_P initial conditions, and assimilate the "**perfect observations**".

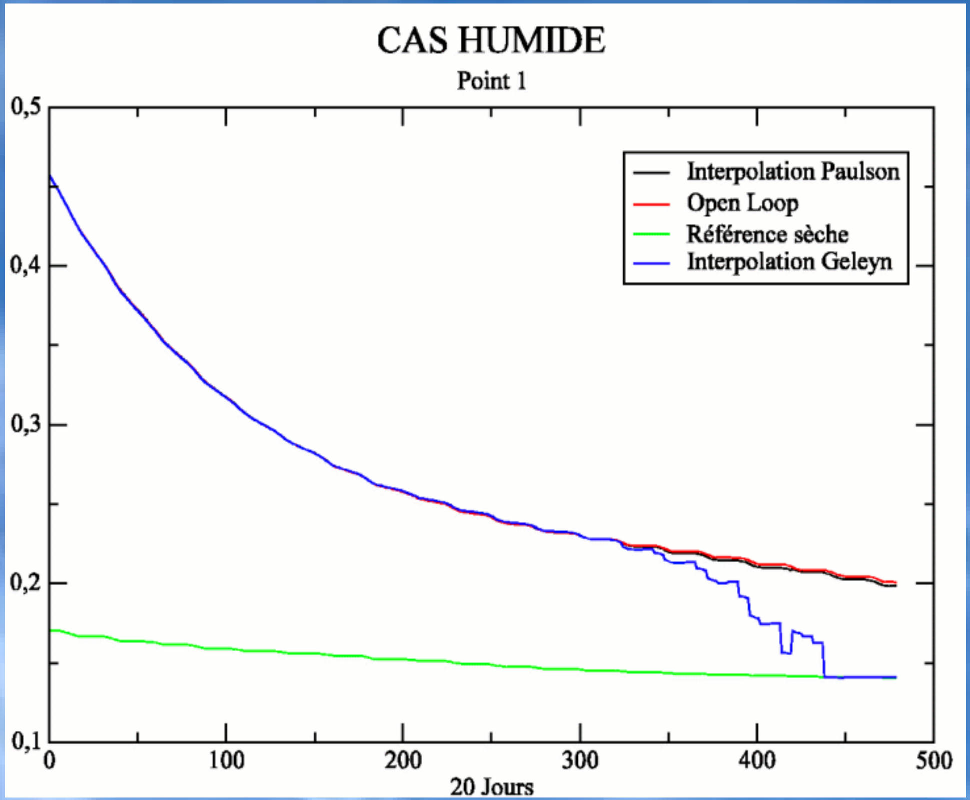
If the assimilation scheme works properly, the system should get away from the "**open loop**" simulation and get closer to the "**reference simulation**".

Experimental set-up: (SURFEX as close as possible to ALADIN configuration)

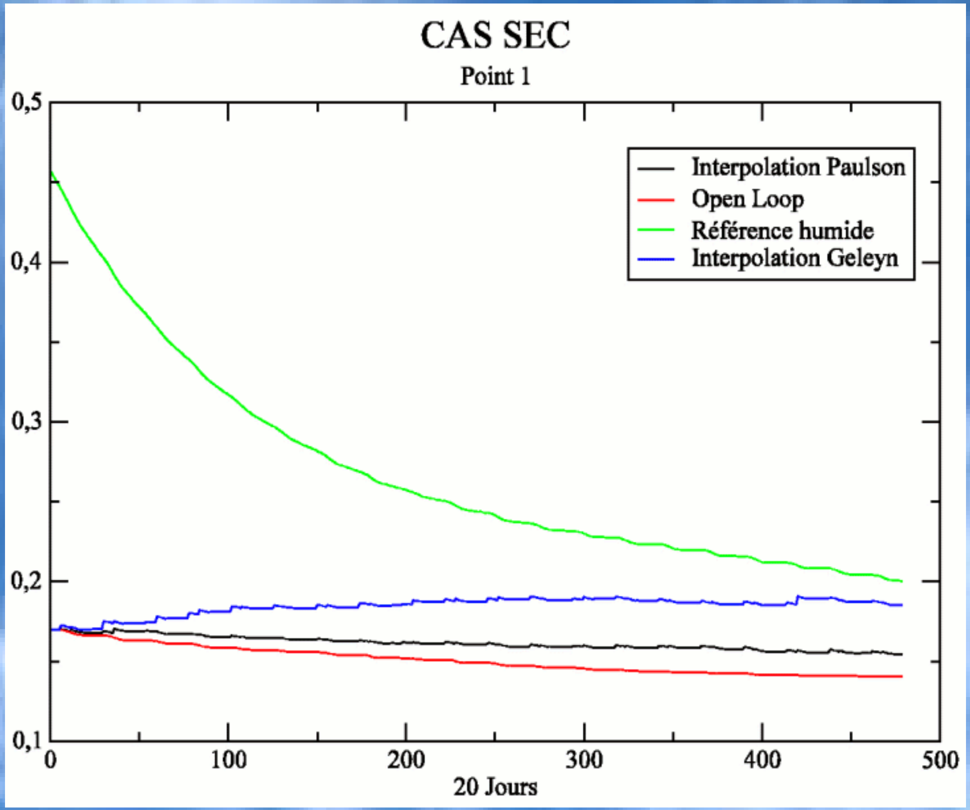
- Reference simulations: Soils at (i) wilting point, (ii) saturation.
- Perturbed simulations: Initial conditions at (i) saturation, (ii) wilting point.
- Atmospheric forcing: hourly short-range forecasts (0-6h) over the ALADIN-France domain on July 2006.
- $\sigma(T_{2m})=1$ K, $\sigma(RH_{2m})=10$ %, $\sigma(SWI)=0.1$, $\delta_{SWI}=10^{-3}$

Feasibility study within SURFEX (II)

Soil at wilting point



Soil at saturation



June 2007

- K. Bergaoui, J-F Mahfouf, R. Hamdi

Work to be done

Work to be done

- A **SEKF** is now available within SURFEX for the analysis of soil moisture but needs analysis increments from a 2-D spatial interpolation tool like CANARI OI.
- The first results over ALADIN-France (done by J-F Mahfouf) are encouraging.
- Reduce remaining inconsistencies between SURFEX and ALADIN-ISBA.
- Couple the soil analysis with the atmospheric analysis to allow feedback.
- Improve the efficiency of offline SURFEX version.
- The vertical interpolation scheme in the surface boundary layer.
- The use of SURFEX should help moving from a **SEKF** to an Ensemble Kalman Filter, where the matrices $\mathbf{BH}^T$ and $\mathbf{HBH}^T$ are obtained from an ensemble of predictions ($\mathbf{X}_f$) with i varying between 1 and N and thus there is no need to get explicitly the various matrices entering in the computation of the Kalman gain $\mathbf{K}$.
- Inclusion of a microwave emission model within SURFEX.
- Inclusion of improved radiative and precipitation forcing.