Data Assimilation in SURFEX

Bertrand Bonan (CNRM, Météo-France)

Contact: bertrand.bonan@meteo.fr

SURFEX Training Course, 12th – 15th March 2024

Outline

- Bases of data assimilation in SURFEX
- Optimal Interpolation for continental surfaces in NWP
- Simplified Extended Kalman Filter for land surface monitoring
- Particle filter for snow model Crocus
- Further topics

Outline

- Bases of data assimilation in SURFEX
- Optimal Interpolation for continental surfaces in NWP
- Simplified Extended Kalman Filter for land surface monitoring
- Particle filter for snow model Crocus
- Further topics

Data assimilation

- *"The aim of data assimilation is to use measured observations in combination with a dynamical system model in order to derive accurate estimates of the* [past,] *current and future states of the system, together with estimates of the uncertainty in the estimate states"* (Nichols, 2010)

- **Why?**

  - Numerical models are not perfect. Several possible sources of errors and uncertainties for land surface models:

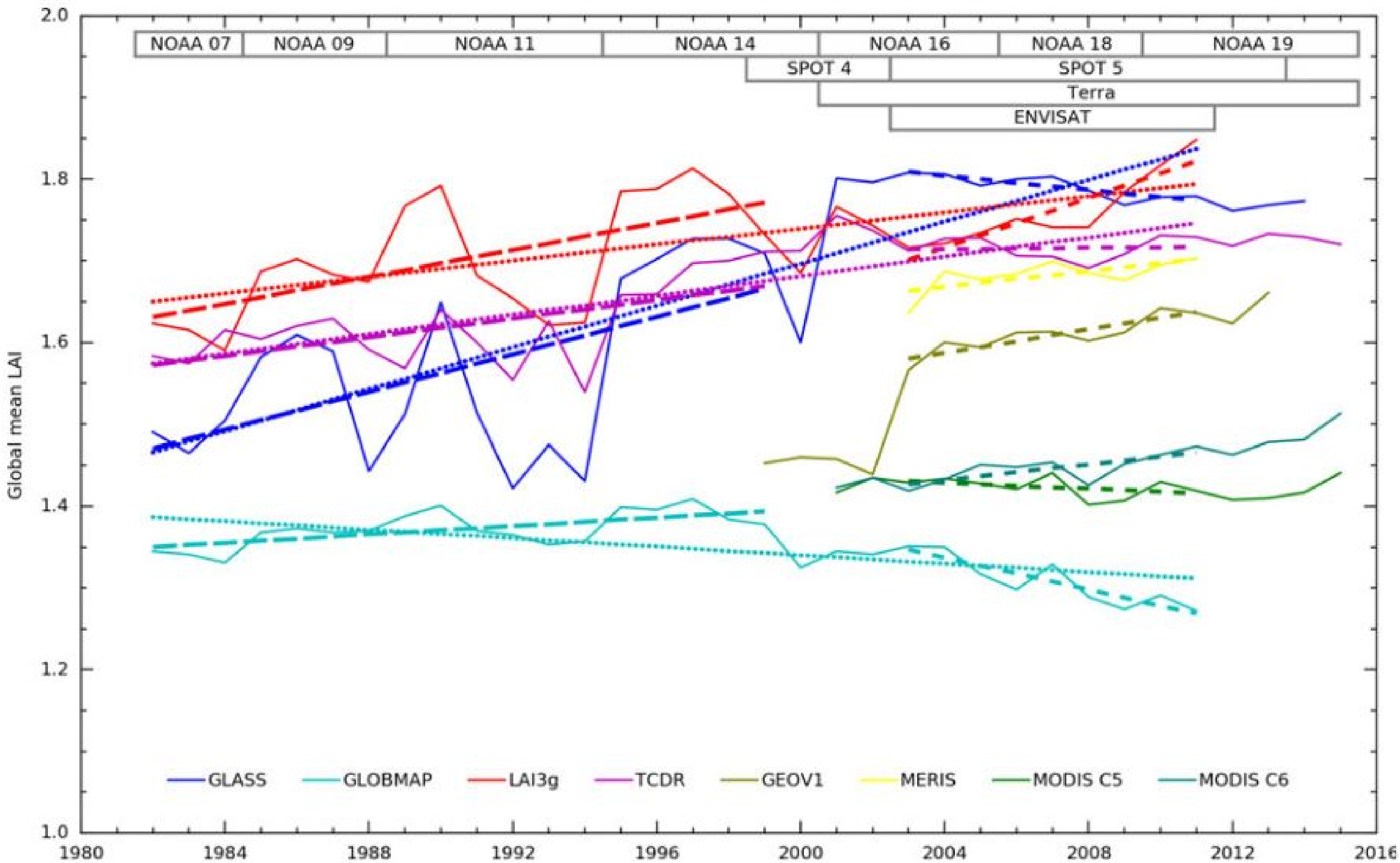
    - Initial conditions
    - Parameters (e.g. soil texture, land cover, …)
    - Atmospheric variables
    - Missing processes or inadequate parametrizations in the model
    - Numerical approximations / Errors in codes

  - We cannot fully observe the system. For example, global satellite observations of surface soil moisture (ASCAT, SMAP, SMOS, …) or terrestrial water storage (GRACE) are available. But we have no observations of the water distribution within the soil except locally (in situ measurements).

Data assimilation

- Observations are not perfect either

  Below is an example of global trends from several satellite-derived Leaf Area Index (LAI) (Jiang *et al.*, 2016). Which one is the best? Well it depends …



Data assimilation in SURFEX

- Historically, data assimilation routines were included in SURFEX to initialize accurately soil variables in the context of Numerical Weather Prediction (NWP) systems at Météo-France (Giard and Bazile, 2000)

- Nowadays in the current version of SURFEX *i.e.* **SURFEX v8.1**:
  - For the tiles **SEA**, **WATER** and **TOWN**,(sea, water or road) surface temperature can be updated using external files *(out of scope of this presentation)*
  - For the tile **NATURE**, several data assimilation routines are available:
    - ***Optimal Interpolation (OI)*** for surface initialization in NWP systems
    - ***Simplified Extended Kalman Filter (SEKF)*** either for surface initialization in NWP systems or for Land Data Assimilation Systems (LDASs) (Barbu *et al.*, 2014; Albergel *et al.*, 2017) in order to monitor land surface variables (focus on the water cycle in the soil and to the vegetation cycle)
    - Ensemble Kalman Filter for LDAS context *(out of scope of this presentation)*
    - **Particle Filter** for the CROCUS snow model

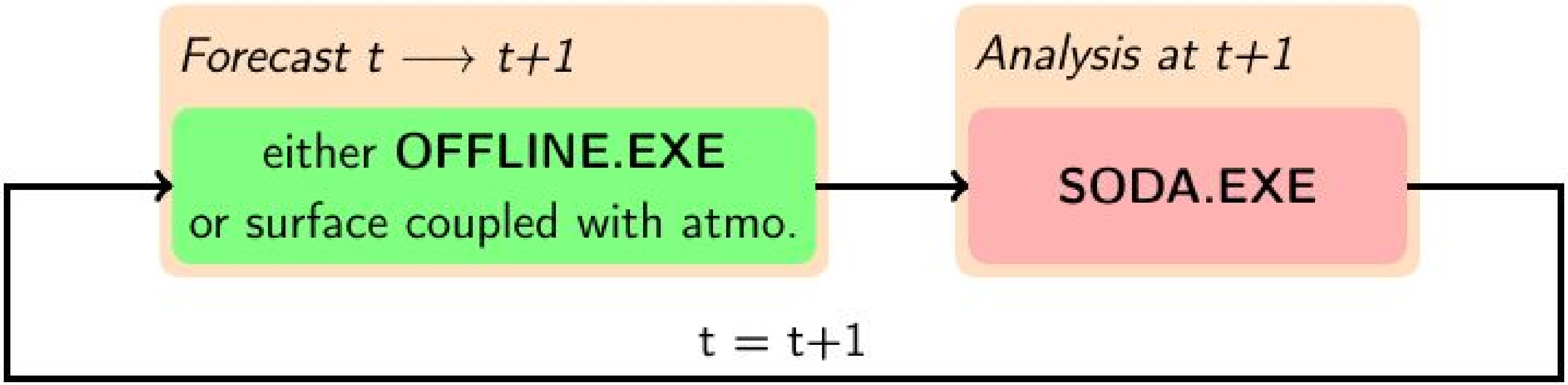
Who uses data assimilation with SURFEX?

- Two communities who interact with each other:
  - Several European national meteorological services (ACCORD community) for their NWP systems
    - Geosphere Austria
    - Finnish Meteorological Institute
    - MET Norway
    - OMSZ, Hungary
    - Royal Meteorological Institute of Belgium
    - SMHI, Sweden
    - (Météo-France)
    - …
  - Research teams and Environmental institutes to develop land surface monitoring systems
    - CNRM, Météo-France
    - Norwegian Institute for Air Research,
    - Delft University of Technology (TU Delft), the Netherlands
    - …

ACCORD
A Consortium for COnvection-scale modelling Research and Development

Basis of data assimilation within SURFEX

- The programme to perform surface data assimilation within SURFEX is **SODA.EXE**
- **SODA** stands for **S**URFEX **O**FFLINE **D**ATA **A**SSIMILATION
- The code associated with SODA including the various data assimilation algorithms is stored in the **ASSIM/** directory in the **src/** directory
- Data assimilation in SURFEX is performed on a grid point level *i.e.*:
  - assimilated observations have to be provided on the same model grid
  - for a given grid point, model variables are updated using only observations available for that grid point
- Data assimilation in SURFEX operates sequentially *i.e.*



Namelists in OPTIONS.nam

- **Requirement:** CNATURE = 'ISBA' in ***NAM_PGD_SCHEMES***

- Two main namelists related to data assimilation in SURFEX:

  - ***NAM_ASSIM***: General assimilation namelist used with SODA
  - ***NAM_OBS***: Specific namelist for the observations

- About ***NAM_ASSIM***: a full description can be found here https://www.umr-cnrm.fr/surfex/spip.php?article347

LASSIM	Logical	**T** or **F**	LASSIM need to be set to **T** for SODA
CASSIM_ISBA	Character(5)	**'OI'**, **'EKF'** (or **'ENKF'**), **'PF '**	DA approaches for tile NATURE
CFILE_FORMAT_FG	Character(5)	**'ASCII'** or **'FA'**	Format of the first guess file (OI)
CFILE_FORMAT_CLIM	Character(5)	**'ASCII'** or **'FA'**	Format of the climate file (OI)
LAROME	Logical	**T** or **F**	Case coupling surface with AROME atmospheric model for first guess file (OI)

Namelists in OPTIONS.nam

- About ***NAM_OBS***: a full description can be found here http://www.umr-cnrm.fr/surfex/spip.php?article344

CFILE_FORMAT_OBS	Character(5)	**'ASCII'** or **'FA'**	Format of the observation file
NOBSTYPE	Integer	≤ 5	Number of different observed variables
COBS_M(i) i = 1, …, NOBSTYPE	Character(10)	**'T2M'**, **'HU2M'**, **'WG2'**, **'LAI'** or **'SWE'**	Type of observed variables
XERROBS_M(i) i = 1, …, NOBSTYPE	Real		Observation error for **COBS_M(i)**
NNCO(i) i = 1, …, NOBSTYPE	Integer	**0** or **1**	If **1**, **COBS_M(i)** is assimilated If **0**, **COBS_M(i)** is not assimilated

- SODA can assimilate (depending on which algorithm is used) the following types of observations (case for NWP or land surface monitoring):
  - **'T2M'**: screen level air temperature at 2m
  - **'HU2M'**: screen level air relative humidity at 2m
  - **'WG2'**: surface soil moisture
  - **'LAI'**: leaf area index
  - **'SWE'**: snow water equivalent

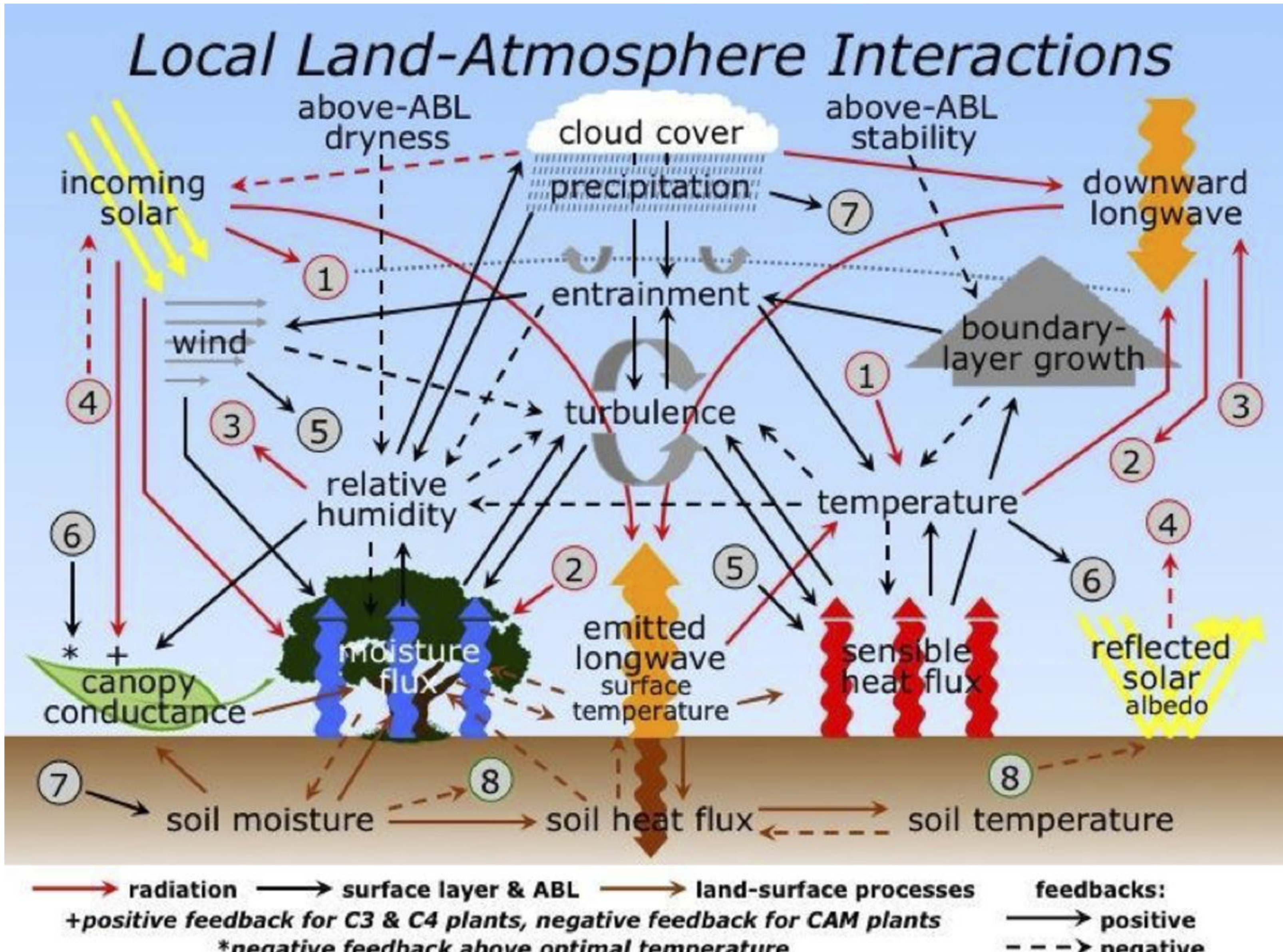
Outline

- Bases of data assimilation in SURFEX
- Optimal Interpolation for continental surfaces in NWP
- Simplified Extended Kalman Filter for land surface monitoring
- Particle filter for snow model Crocus
- Further topics

Land-Atmosphere interactions in NWP

Soil moisture, temperature and vegetation play a decisive role in land-atmosphere interactions

Figure from Mike Ek and Kevin Trenberth (UCAR/NCAR)



Assimilation for surface conditions in NWP

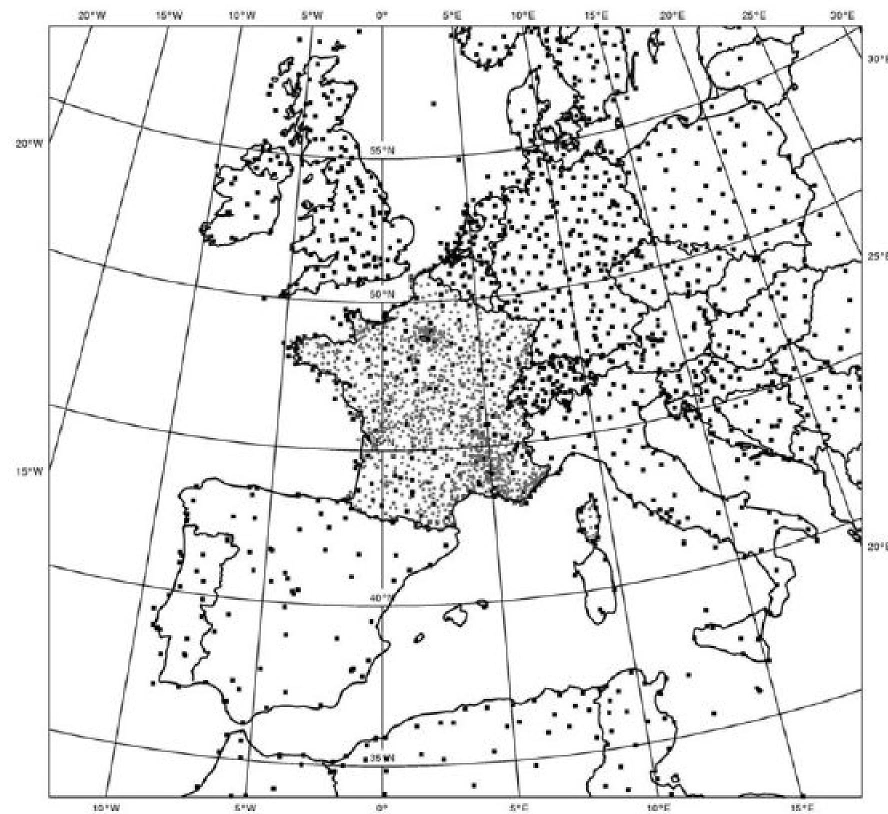
- Soil variables and their initialization are known to have a significant influence on numerical weather forecasts either at short or medium ranges (see *e.g.* Beljaars *et al.*, 1996). So it is important to initialize them correctly at a relatively low cost.

- **What kind of observations to assimilate?**
  - 2-m air temperature (T) and relative humidity (RH) available from weather stations (SYNOP, Surface Synoptic Observations)
  - Coiffier *et al.* (1987) shows the positive impact of using 2-m air temperature to initialise soil temperatures for weather forecasts
  - Mahfouf (1991) shows the positive impact of assimilating jointly 2-m air temperature and relative humidity to improve soil moistures for atmospheric predictions

- **A two step process**
  - Transform data from local site to horizontal grid => **screen level analysis**
  - Assimilate **screen level analysis** to correct land variables

An example of optimal interpolation for NWP

- To illustrate how works OI, we consider the experiment in Mahfouf *et al.* (2009):
  - **ISBA** is run in the following configuration: 2 layers of soil (CISBA = '2-L', CPHOTO = 'NON',NPATCH = 1) over the ALADIN-France domain (covers most Western Europe with a 9.5 km grid, see figure)
  - **ISBA** is forced with atmospheric ALADIN-France model
  - Period of analysis: July 2006
  - Observations are available on sites indicated on the figure
- **First step:** screen level analysis **(not included in SODA)**
- **Second step**: soil analysis with **SODA**



Step 1: Screen level analysis

- **2D Optimal Interpolation. This first step is performed <u>before using</u> SODA!**
- Screen level analysis increments are obtained for each variable as follows:

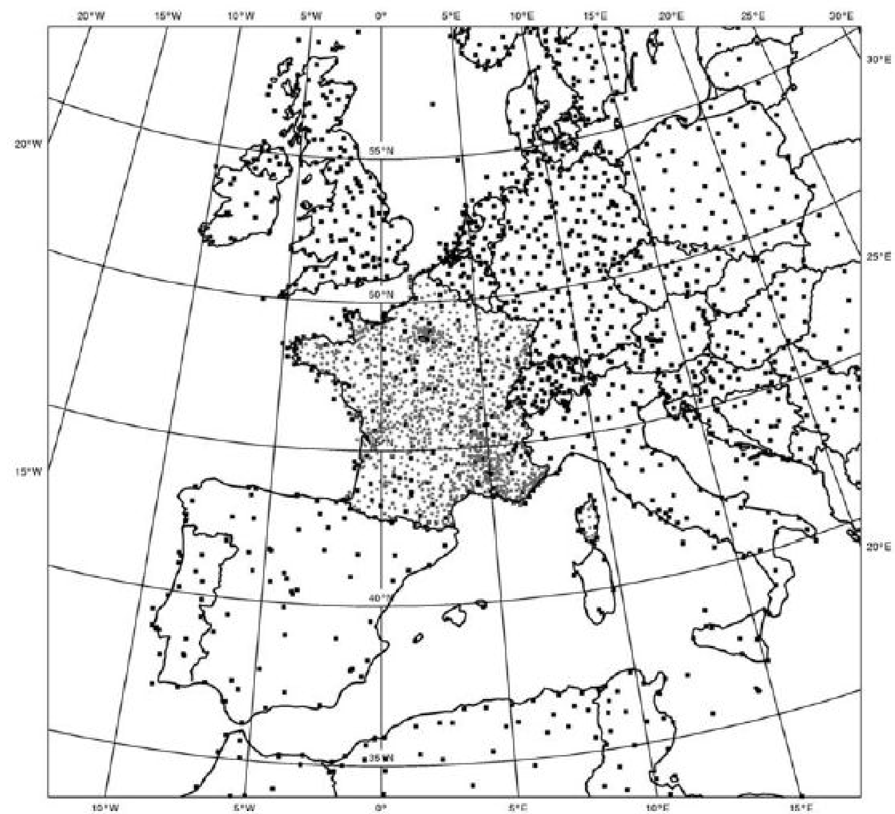
$$\Delta \mathbf{x}^a = \mathbf{W}\left(\mathbf{y}^o - \mathbf{H}\mathbf{x}^f\right)$$

with
  - $\mathbf{y}^o$ 2-m observations (either T or RH)
  - $\mathbf{x}^f$ 2-m forecasts from NWP system
  - $\mathbf{H}$ interpolation operator from model grid to observation locations
  - $\mathbf{W}$ weight matrix

and **for a given model grid point i**:

$$(\mathbf{B} + \mathbf{O})\,[\mathbf{W}]_{i:} = \mathbf{b}_i$$

with
  - $[\mathbf{W}]_{i:}$ i^th line of matrix $\mathbf{W}$
  - $\mathbf{O}$ observation covariance matrix (assumed diagonal)
  - $\mathbf{b}_i$, $\mathbf{B}$ background covariance vector and matrix depending on distance between grid point location i vs obs and distance between locations of obs



Step 2: Update surface variables with OI

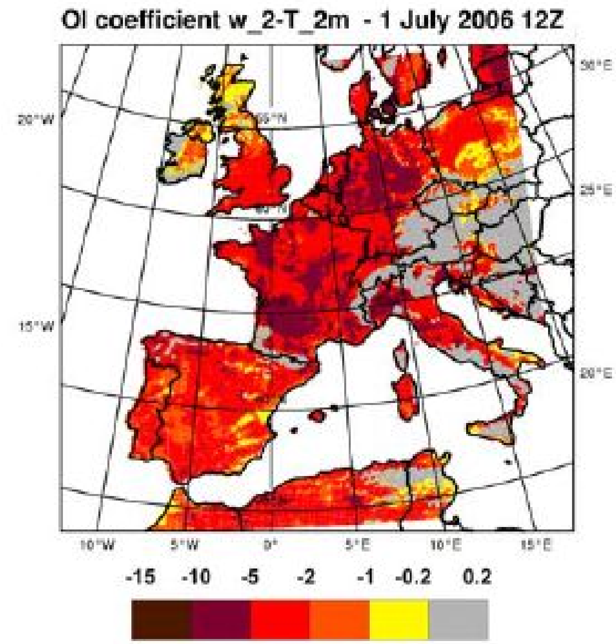
- **<u>Model variables to be estimated/updated</u>**:
  - Soil temperature T_{g1} and T_{g2} in the 2 layers
  - Soil moisture w_{g1} and w_{g2} in the 2 layers
- Screen level analysis increment ΔT_{2m} and ΔRH_{2m} to be assimilated (obtained from step 1)
- **<u>Analysis step follows</u>** <u>Giard and Bazile (2000)</u>:

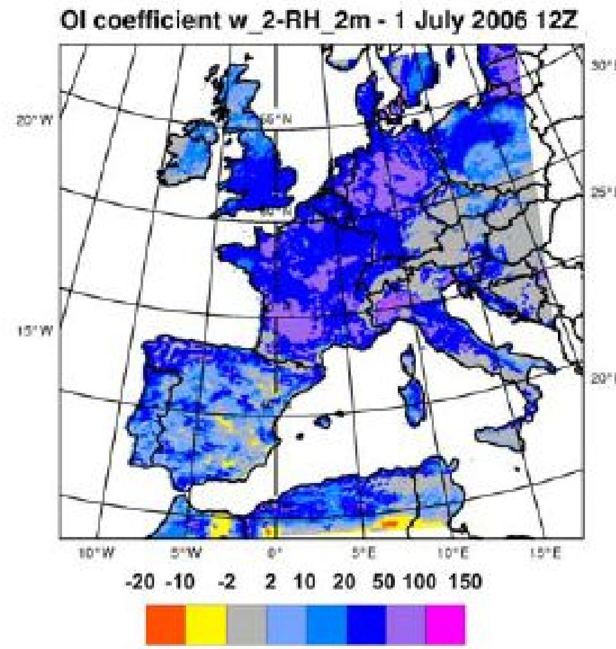
$$T_{g1}^{a} = T_{g1}^{f} + \Delta T_{2m} \qquad T_{g2}^{a} = T_{g2}^{f} + \frac{1}{2\pi}\Delta T_{2m}$$

$$w_{g1}^{a} = w_{g1}^{f} + \alpha_1 \Delta T_{2m} + \alpha_2 \Delta RH_{2m}$$

$$w_{g2}^{a} = w_{g2}^{f} + \beta_1 \Delta T_{2m} + \beta_2 \Delta RH_{2m}$$

- Maps of β_1 (top) and β_2 (bottom) for 1th July at 12hUTC





Optimal Interpolation in SODA

- OI is coded in `assim_nature_isba_oi.F90` and related files in **src/ASSIM**

- **Requirements:**

CNATURE = 'ISBA'	in ***NAM_PGD_SCHEMES***
CISBA = '2-L', '3-L'	in ***NAM_ISBA***
CPHOTO = 'NON'	
NPATCH = 1	
LASSIM_ISBA = T	in ***NAM_ASSIM***
CASSIM_ISBA = 'OI'	
COBS_M(i) = 'T2M', 'HU2M' or 'SWE'	in ***NAM_OBS***

- If in ***NAM_OBS***, CFILE_FORMAT_OBS = 'FA', the other variables in ***NAM_OBS*** are bypassed (hardcoded in **`soda.F90`**)

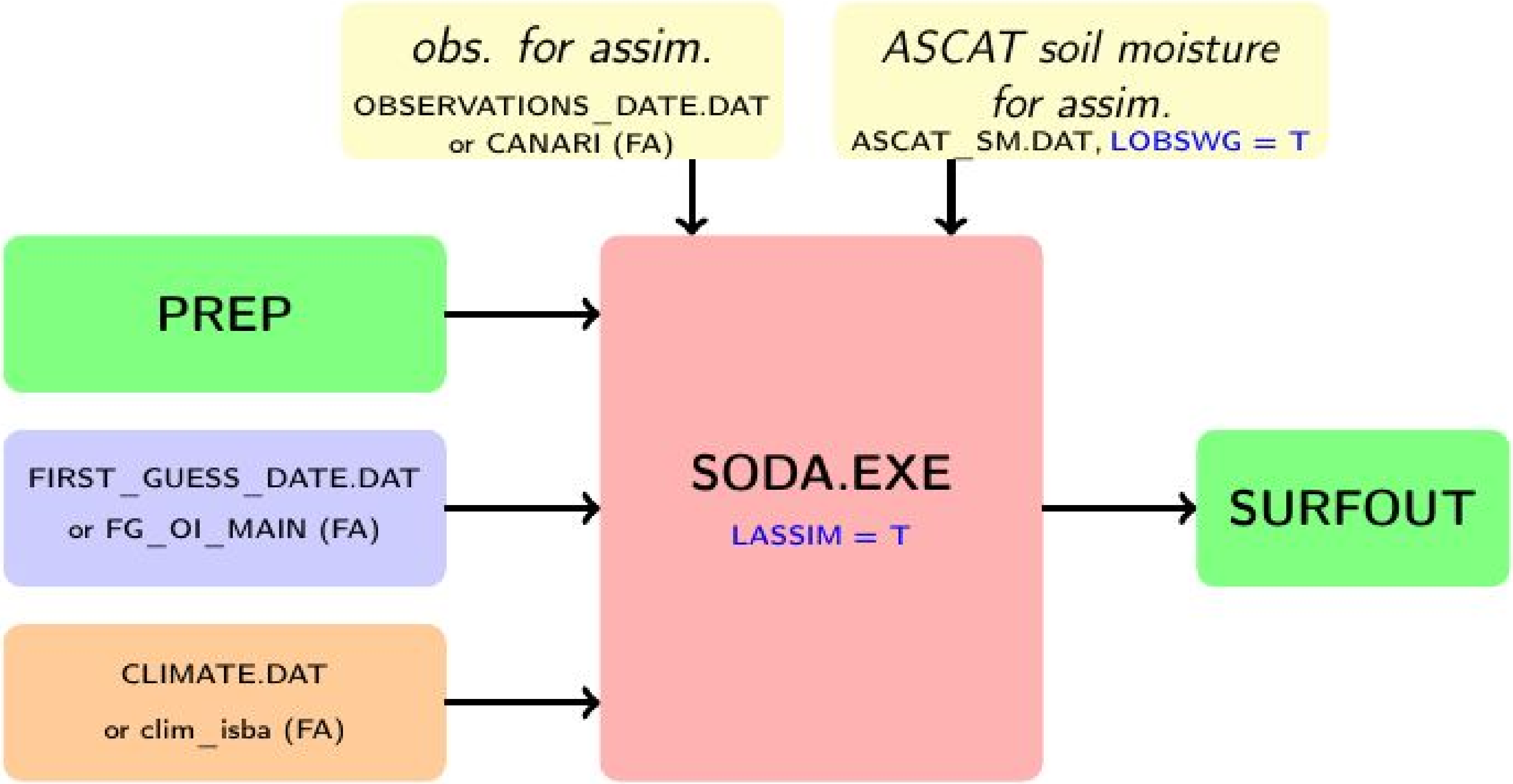
- Assimilated observations with OI are screen level air temperature, relative humidity at 2m, snow water equivalent and surface soil moisture (from ASCAT). The latest has to be specified in ***NAM_NACVEG*** by setting LOBSWG = T.

- ***NAM_NACVEG***: namelist setting parameters for OI. Description available here: https://www.umr-cnrm.fr/surfex/spip.php?article349

Optimal Interpolation scheme in SODA

For a given DATE: OI analysis with SODA involves several files

DATE is under the format ***yymmdd*H*hh*** (ex: 240315H09)



Input files for Optimal Interpolation

- Current state of the surface (forecast, before assimilation) in **PREP** file
- Other input files have their names ***hardcoded*** in **`soda.F90`**
- **FIRST_GUESS_DATE.DAT** (ASCII) or **FG_OI_MAIN** (FA) contains:
  - Amount of convective liquid precipitation
  - Amount of stratiform liquid precipitation
  - Amount of convective solid precipitation
  - Amount of stratiform solid precipitation
  - Cloud cover
  - Land-sea mask
  - Evaporation
- **CLIMATE.DAT** (ASCII) or **clim_isba** (FA) contains:
  - Climatology of surface temperature
  - Climatology of snow water equivalent

Input/Output files for Optimal Interpolation

- **OBSERVATIONS_DATE.DAT** (ASCII) or **CANARI** (FA):
  - Observations of 2m air temperature *for assimilation*
  - Observations of 2m air relative humidity *for assimilation*
  - Observations of snow water equivalent *for assimilation*
  - [case FA] Surface temperature
  - [case FA] Zonal wind
  - [case FA] Meridian wind
- **ASCAT_SM.DAT** (ASCII):
  - ASCAT-derived observations of surface soil moisture assimilated if LOBSWG = T
- Current state of the surface (after assimilation, analysis) in **SURFOUT** file

Outline

- Bases of data assimilation in SURFEX
- Optimal Interpolation for continental surfaces in NWP
- Simplified Extended Kalman Filter for land surface monitoring
- Particle filter for snow model Crocus
- Further topics

SEKF in SODA: compatibility with ISBA

- SEKF is coded in `assim_nature_isba_ekf.F90` and related files in **src/ASSIM**

- **Requirements:** CNATURE = 'ISBA' in ***NAM_PGD_SCHEMES***

  LASSIM_ISBA = T in ***NAM_ASSIM***
  CASSIM_ISBA = 'EKF'

  COBS_M(i) = 'T2M', 'HU2M', 'WG2' or 'LAI' in ***NAM_OBS***

- **Requirements if LAI is assimilated**: need to use ISBA version with prognostic LAI

  CPHOTO = 'NIT' in ***NAM_PGD_SCHEMES***
  NPATCH = 12

- Two namelists invoked by SEKF in SODA (in addition to ***NAM_ASSIM*** and ***NAM_OBS***):

  - ***NAM_IO_VARASSIM***: General SEKF options
  - ***NAM_VAR***: Namelist for the control variables for SEKF

Namelists in OPTIONS.nam

- About ***NAM_VAR***, a full description can be found here
  https://www.umr-cnrm.fr/surfex/spip.php?article343

NVAR	Integer	≤ 9	Number of control variables
NIVAR	Integer	$0 \leq$ NIVAR $\leq$ NVAR	Number of the perturbed variable
CVAR_M(i) i = 1, …, NVAR	Character(5)	**‘TG1’, ‘TG2’, ‘LAI’, ‘WG1’, ‘WG2’, ‘WG3’, ‘WG4’, ‘WG5’, ‘WG6’, ‘WG7’, ‘WG8’**	Type of control variables
XSIGMA_M(i) i = 1, …, NVAR	Real		Background error for **CVAR_M(i)** If background covariance matrix is fixed
XPRT_M(i) i = 1, …, NVAR	Real		Perturbation amplitude for **CVAR_M(i)**
NNCV(i) i = 1, …, NVAR	Integer	**0** or **1**	If **1**, **CVAR_M(i)** is updated by SEKF If **0**, **CVAR_M(i)** is not updated by SEKF

- **TGi** stands for soil temperature in layer i, **WGi** for soil moisture in layer i
- Obviously CVAR_M depends on CISBA. If CISBA = ‘2-L’, there are only 2 layers. If CISBA = ‘DIF’, we can control up to the first metre of soil (**WG1** to **WG8**)

Namelists in OPTIONS.nam

- About ***NAM_VAR***, if LAI is in CVAR_M, the namelist also includes

CBIO	Character(12)	**'BIOMA1', 'BIOMA2'**, …	Name of biomass variable related to LAI
XALPHA(i) i= 1, …, 12	Real	(0., 0., 0., 0.08203445, 0.07496252, 0.06846970, 0.06771856, 0.09744689, 0.09744689, 0.07164350, 0.17686594, 0.07164350)	Multiplicative coefficient transforming LAI after assimilation into CBIO for each patch (NPATCH = 12)

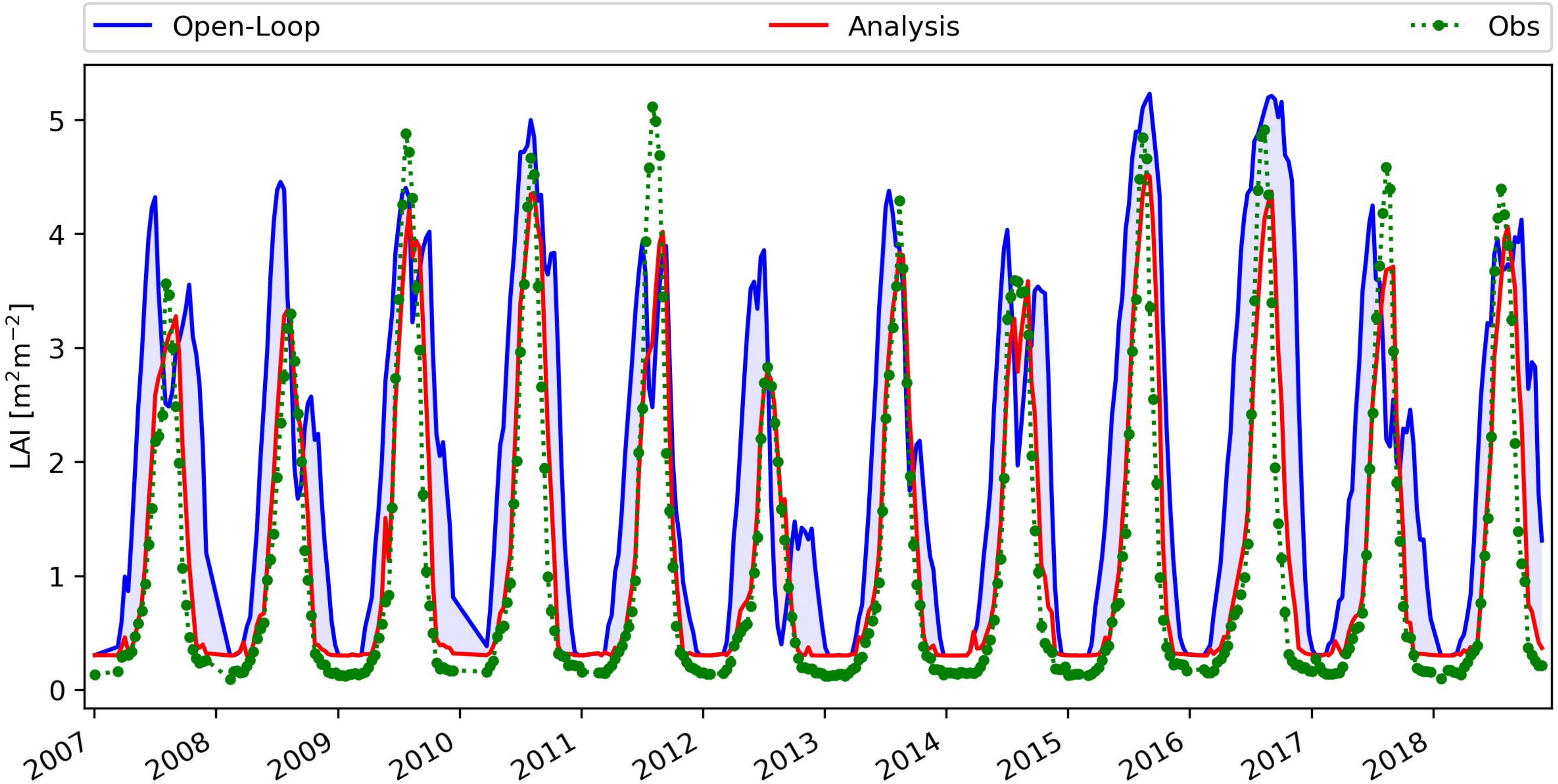
LAI is not a prognostic variable of ISBA but CBIO is. To make the assimilation of LAI in SEKF working, LAI needs also to be converted in CBIO for each patch after each assimilation step.

- About ***NAM_IO_VARASSIM***: a full description can be found here https://www.umr-cnrm.fr/surfex/spip.php?article345

LPRT	Logical	**T** or **F**	Need to be set to **T** for perturbed runs
LBEV	Logical	**T** or **F**	If **T**, full EKF. If **F**, SEKF
LBFIXED	Logical	**T** or **F**	If **T**, SEKF. If **F**, full EKF

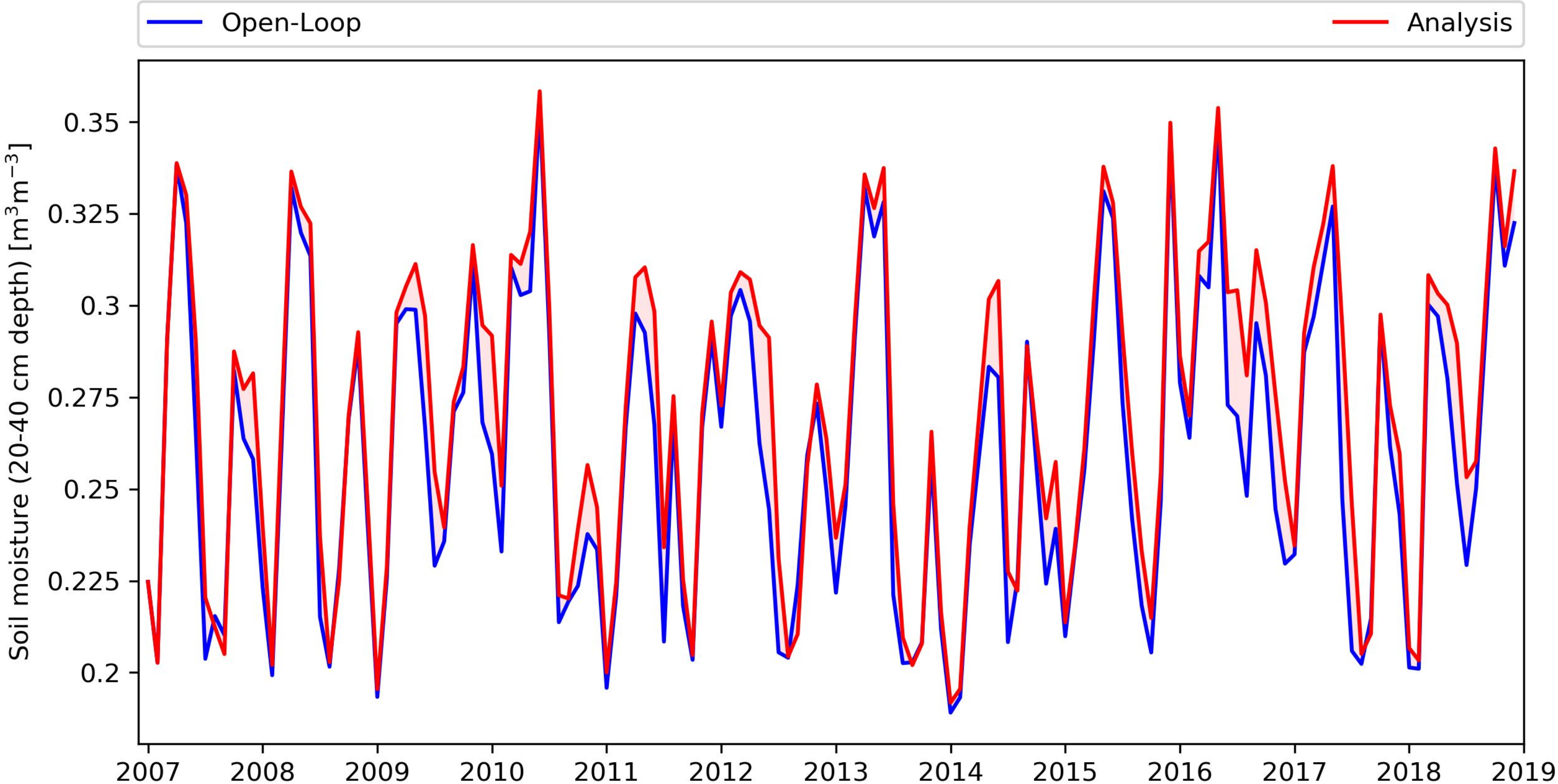
Land Data Assimilation System

- LDAS aim to improve our knowledge of surface variables (water, vegetation cycles) and their evolution by integrating information from satellite observations into a land surface model.
- Example of assimilation of LAI on a site near Lincoln, Nebraska, USA.



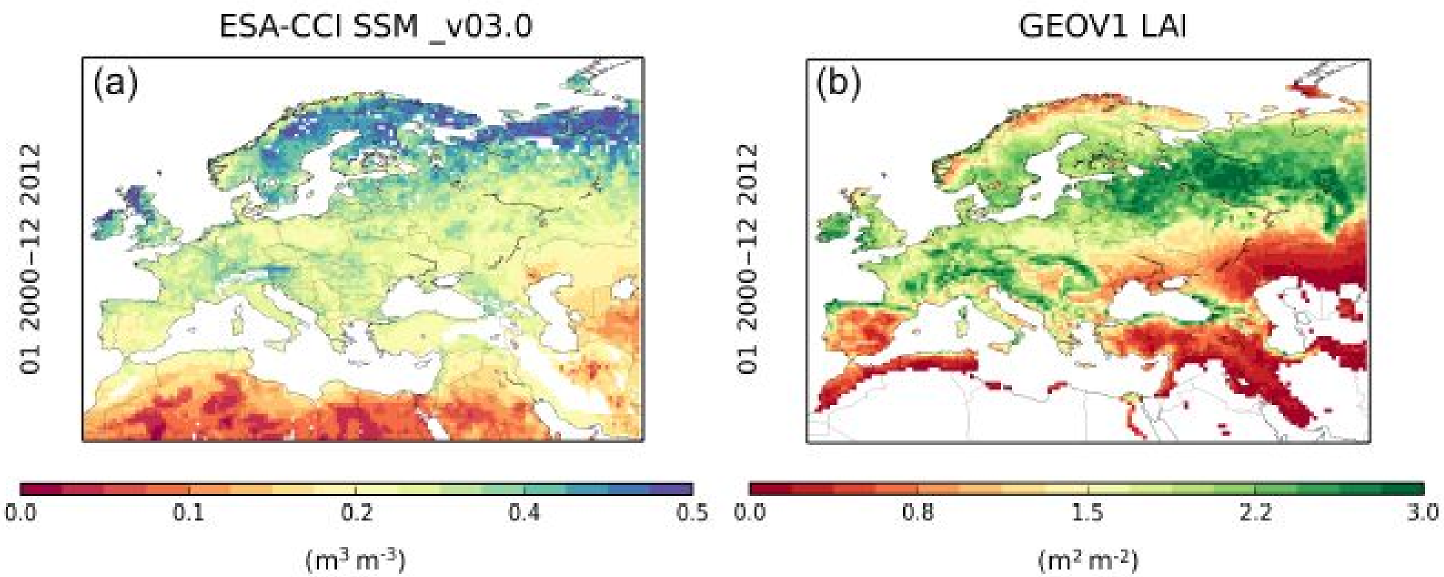
Land Data Assimilation System

- Assimilating LAI has an impact on soil moisture notably root-zone soil moisture
- Soil moisture (averaged monthly) in layer 20–40 cm depth for site Lincoln, Nebraska, USA.



SEKF for Land Data Assimilation System: an example

- To illustrate how works SEKF, we consider the experiment in Albergel *et al.* (2017):
  - Daily assimilation of ESA-CCI Surface Soil Moisture (a) and CGLS LAI V1 (b) satellite products into the ISBA land surface model (CISBA = ‘DIF’, CPHOTO = ‘NIT’, NPATCH = 12)
  - ISBA is forced with the WFDEI atmospheric dataset (offline mode)
  - Period of analysis: 2000 – 2012
  - Domain of analysis: Euro-Mediterranean region at 0.5° spatial resolution



First step: Retreating observations for assimilation

- Observations need to be averaged/rescaled on the model grid (here 0.5° spatial resolution)

- **Special case**: **assimilation of surface soil moisture (SSM)**.

  - Observed SSM need to be rescaled to the ISBA model climatology to avoid introducing any bias in the system (Reichle and Koster, 2004; Drusch *et al.*, 2005).

  - This is performed by applying a linear rescaling to match the observation mean and variance to the mean and variance of w_{g2}, the ISBA modelled soil moisture in the 2nd layer of soil i.e. 1-4 cm depth (Scipal *et al.*, 2008)

  - The linear rescaling is performed on a seasonal basis with a 3-month moving window (Draper *et al.*, 2009; Barbu *et al.*, 2014)

- Again, this needed pre-treatment is performed outside SODA and SURFEX.

- We can now produce the ASCII files **OBSERVATIONS_DATE.DAT** containing all the regridded and pre-treated observations to be assimilated

Second step: Assimilation at model grid cell level

Need to set correctly beforehand the following namelists:

- ***NAM_ASSIM***: General assimilation namelist used with SODA (SEKF selected)
- ***NAM_IO_VARASSIM***: General SEKF options
- ***NAM_OBS***: Specific namelist for the observations
- ***NAM_VAR***: Namelist for the control variables
  (i.e. model variables directly updated by assimilation)

Setting NAM_OBS for this experiment

- Below (left) is the setting of ***NAM_OBS*** for the experiment

```
&NAM_OBS        NOBSTYPE          = 2
                LOBSNAT           = F,
                CFILE_FORMAT_OBS  = "ASCII",
                COBS_M(1)         = 'WG2',
                COBS_M(2)         = 'LAI',
                XERROBS_M(1)      = 0.4,
                XERROBS_M(2)      = 0.2,
                NNCO(1)           = 1,
                NNCO(2)           = 1,
/
```

- An example of **OBSERVATIONS_DATE.DAT** is on the right:
  - The first column contains the retreated observations of surface soil moisture (`COBS_M(1) = 'WG2'`)
  - The second column contains the retreated obs. of leaf area index (`COBS_M(2) = 'LAI'`)
  - `999` indicates that there is no observation for that grid point

```
0.3447255194 1.986791015
0.3408581018 2.246661425
0.3250174522 2.180453539
0.3086568415 2.574210882
0.2868968844 2.447184086
0.3052057326 2.264844894
0.3285748363 2.256729603
0.3430277705 2.450858116
0.3357311189 2.461838245
999 2.29612112
999 2.467530966
999 2.417602301
0.3279027045 2.390994549
0.3240542114 2.821894169
0.3480122387 2.680686951
0.3638544083 2.563509941
0.362436831 2.777883291
0.334685117 2.556204796
999 2.438989639
999 2.219954729
999 2.330261707
999 2.372661352
999 2.501777887
999 2.815329313
0.3829928041 2.406977892
0.3627473116 2.439852953
0.355199039 2.778611898
0.3593625426 2.71013999
```

Control variables and NAM_VAR

- Simplified Extended Kalman Filter introduced by Mahfouf *et al.* (2007). Here adapted to NPATCH = 12.

- Below is the setting of ***NAM_VAR*** for control variables. Basically, the SEKF will update LAI and soil moisture from layer 2 (1 – 4 cm depth) to layer 8 (80 – 100 cm depth).

- Soil moisture in layer 1 (0 – 1 cm depth) is not controlled as it is mostly driven by precipitations

- Control vector $\mathbf{x}_{[p]}$ is defined for each patch p of the tile **NATURE**

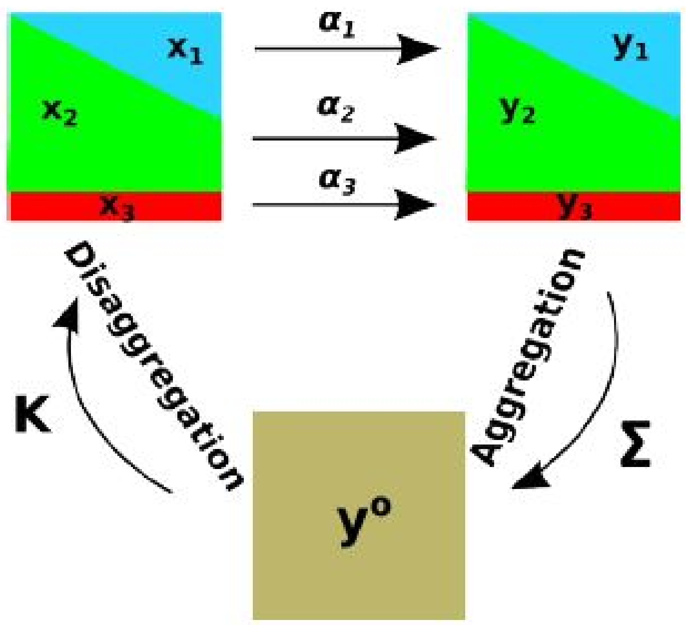
$$\mathbf{x}_{[p]} = \begin{pmatrix} LAI_{[p]} \\ w_{g2,[p]} \\ w_{g3,[p]} \\ \vdots \\ w_{g8,[p]} \end{pmatrix}$$

```
&NAM_VAR          NIVAR       = XXX-NIVAR-XXX,
                  NVAR        = 8,
                  CVAR_M(1)   = 'LAI',
                  CVAR_M(2)   = 'WG1',
                  CVAR_M(3)   = 'WG2',
                  CVAR_M(4)   = 'WG3',
                  CVAR_M(5)   = 'WG4',
                  CVAR_M(6)   = 'WG5',
                  CVAR_M(7)   = 'WG6',
                  CVAR_M(8)   = 'WG7',
                  CVAR_M(9)   = 'WG8',
                  NNCV(1)     = 1,
                  NNCV(2)     = 0,
                  NNCV(3)     = 1,
                  NNCV(4)     = 1,
                  NNCV(5)     = 1,
                  NNCV(6)     = 1,
                  NNCV(7)     = 1,
                  NNCV(8)     = 1,
                  NNCV(9)     = 1,
```

Control variables and model equivalent

- Control variables $\mathbf{x}_{[p]}$ are defined by patch **BUT observations are available at grid point level** (we assume here grid points = 100 % tile **NATURE**)
- Need to aggregate variables from patch level to grid point level to calculate model equivalent $\mathbf{y}$ of observations $\mathbf{y}^o$

$$\mathbf{y} = \sum_{k=1}^{12} \alpha_{[k]} \, \mathbf{H} \, \mathbf{x}_{[k]}$$



with $\alpha_{[k]}$ patch fraction of patch

and $\mathbf{H}$ selection operator

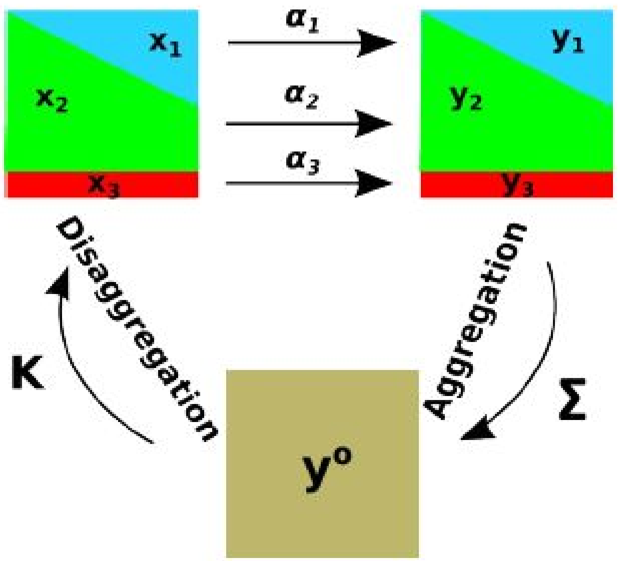
$$\mathbf{H} = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \end{pmatrix}$$

Simplified Extended Kalman Filter (SEKF)

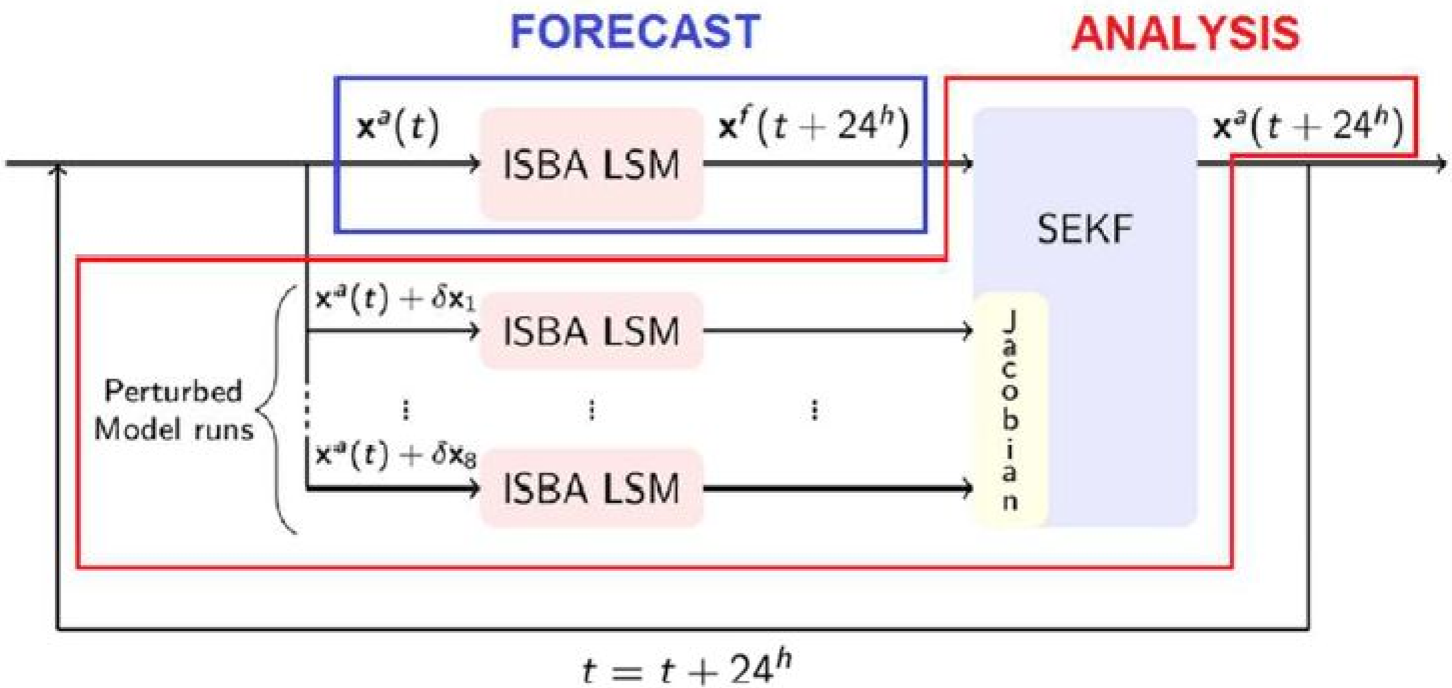
- Update step of SEKF is for patch p:

$$\mathbf{x}^{a}_{[p]} = \mathbf{x}^{f}_{[p]} + \mathbf{K}_{[p]} \left(\mathbf{y}^{o} - \mathbf{y}^{f}\right)$$

$$\mathbf{K}_{[p]} = \alpha_{[p]} \mathbf{B} \left(\mathbf{H}\mathbf{M}_{[p]}\right)^{T} \left(\sum_{k=1}^{12} \alpha^{2}_{[k]} \left(\mathbf{H}\mathbf{M}_{[k]}\right) \mathbf{B} \left(\mathbf{H}\mathbf{M}_{[k]}\right)^{T} + \mathbf{R} \right)^{-1}$$



- Jacobian matrices $\mathbf{M}_{[p]}$ are calculated using perturbed model runs

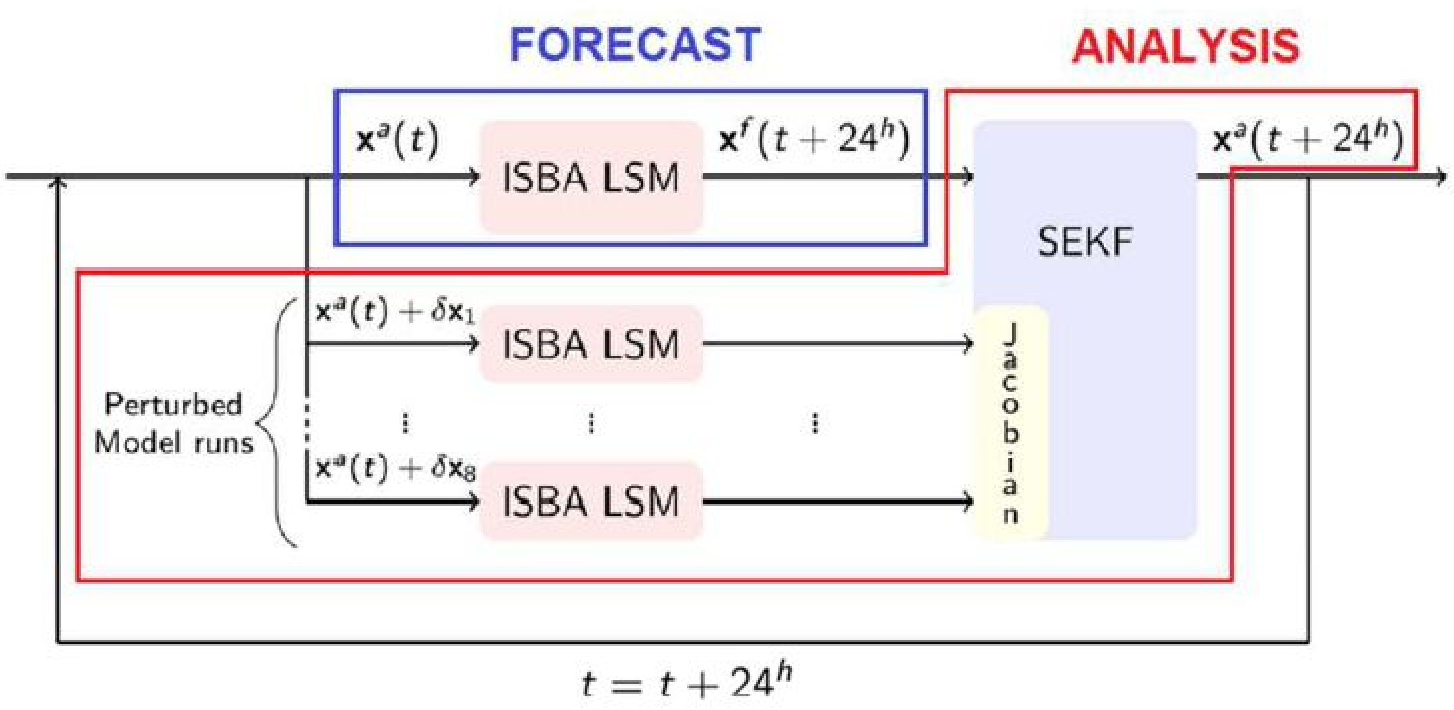


Perturbed model runs with SURFEX

- Jacobian matrices $\mathbf{M}_{[p]}$ are calculated using perturbed model runs

$$\left[\mathbf{M}_{[p]}\right]_j = \frac{\partial \mathbf{x}^f\,(t+24h)}{\partial x_j}$$
$$\approx \frac{\mathcal{M}_{[p]}\left(\mathbf{x}^a(t)+\delta x_j\right) - \mathcal{M}_{[p]}\left(\mathbf{x}^a(t)\right)}{\delta x_j}$$

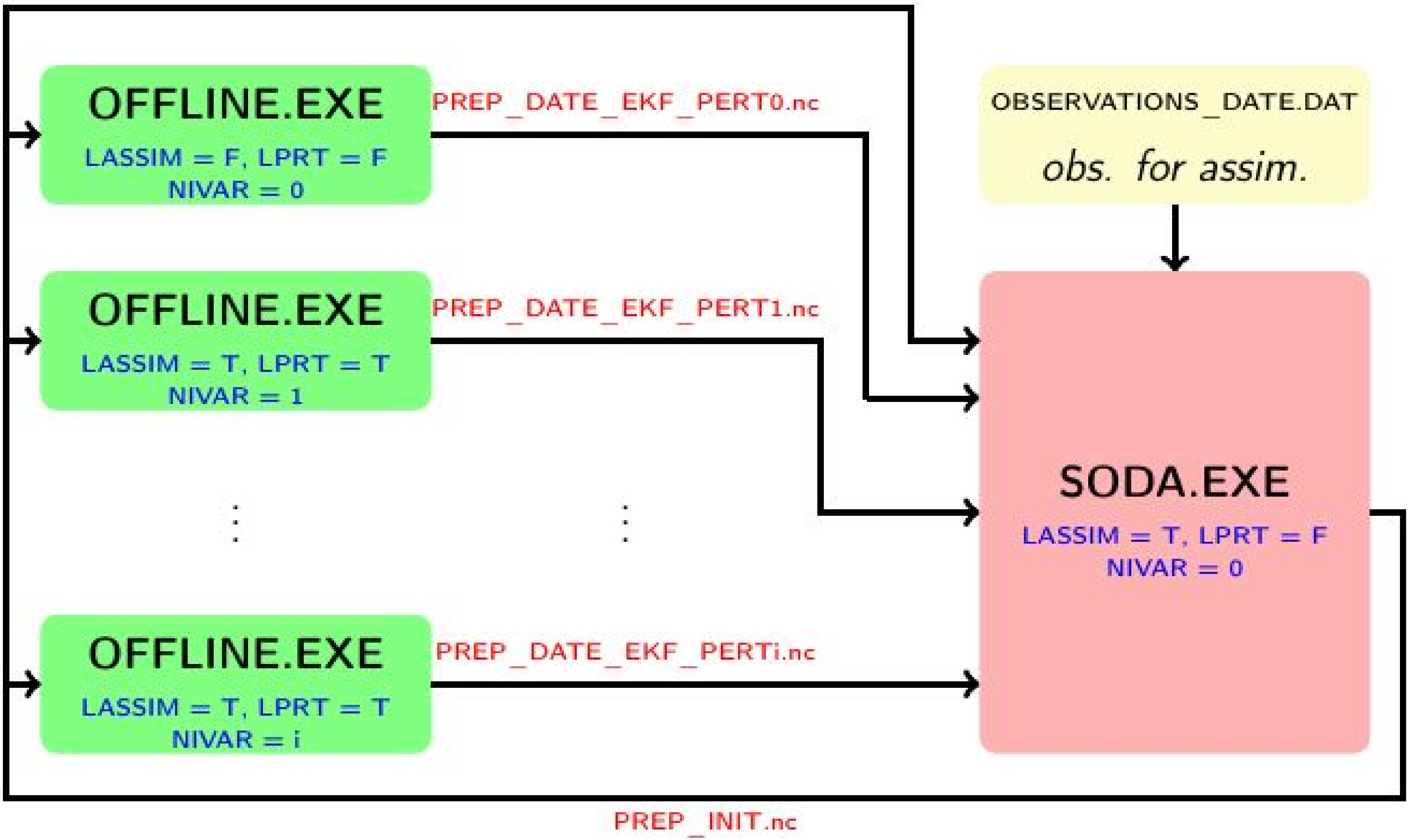
- Number of model runs for SEKF = 1 + Number of control variables per patch
  Here, it means 9 model runs!



SEKF scheme with SODA

For a given DATE: forecast from DATE-1 to DATE then analysis at DATE

DATE is under the format ***yymmdd*H*hh*** (ex: 240315H09)



Script (pseudo-code) for SEKF with SURFEX

For a given DATE: forecast from DATE-1 to DATE then analysis at DATE

```
set in    NAM_IO_OFFLINE: CPREPFILE="PREP_INIT"
          NAM_ASSIM: LASSIM=F
          NAM_IO_VARASSIM: LPRT=F
OFFLINE.EXE
mv SURFOUT.nc PREP_DATE_EKF_PERT0.nc

set in    NAM_ASSIM: LASSIM=T
          NAM_IO_VARASSIM: LPRT=T
for i = 1, …, NVAR
     set in    NAM_VAR: NIVAR=i # perturbed run with initially perturbed CVAR_M(i)
     OFFLINE.EXE
     mv SURFOUT.nc PREP_DATE_EKF_PERTi.nc

set in    NAM_IO_OFFLINE: CPREPFILE="PREP_DATE_EKF_PERT0"
          NAM_IO_VARASSIM: LPRT=F
          NAM_VAR: NIVAR=0
SODA.EXE
mv SURFOUT.nc PREP_INIT.nc
```

Script (pseudo-code) for SEKF with SURFEX

For a given DATE: forecast from DATE-1 to DATE then analysis at DATE

```
set in    NAM_IO_OFFLINE: CPREPFILE="PREP_INIT"
          NAM_ASSIM: LASSIM=F
          NAM_IO_VARASSIM: LPRT=F
OFFLINE.EXE
mv SURFOUT.nc PREP_DATE_EKF_PERT0.nc
```

CONTROL RUN aka FORECAST

```
set in    NAM_ASSIM: LASSIM=T
          NAM_IO_VARASSIM: LPRT=T
for i = 1, …, NVAR
     set in    NAM_VAR: NIVAR=i # perturbed run with initially perturbed CVAR_M(i)
     OFFLINE.EXE
     mv SURFOUT.nc PREP_DATE_EKF_PERTi.nc

set in    NAM_IO_OFFLINE: CPREPFILE="PREP_DATE_EKF_PERT0"
          NAM_IO_VARASSIM: LPRT=F
          NAM_VAR: NIVAR=0
SODA.EXE
mv SURFOUT.nc PREP_INIT.nc
```

Script (pseudo-code) for SEKF with SURFEX

For a given DATE: forecast from DATE-1 to DATE then analysis at DATE

```
set in    NAM_IO_OFFLINE: CPREPFILE="PREP_INIT"
          NAM_ASSIM: LASSIM=F
          NAM_IO_VARASSIM: LPRT=F
OFFLINE.EXE
mv SURFOUT.nc PREP_DATE_EKF_PERT0.nc
```

```
set in    NAM_ASSIM: LASSIM=T
          NAM_IO_VARASSIM: LPRT=T
for i = 1, …, NVAR
     set in    NAM_VAR: NIVAR=i
     OFFLINE.EXE
     mv SURFOUT.nc PREP_DATE_EKF_PERTi.nc
```

PERTURBED RUNS (can be run in parallel with CONTROL RUN)

```
set in    NAM_IO_OFFLINE: CPREPFILE="PREP_DATE_EKF_PERT0"
          NAM_IO_VARASSIM: LPRT=F
          NAM_VAR: NIVAR=0
SODA.EXE
mv SURFOUT.nc PREP_INIT.nc
```

Script (pseudo-code) for SEKF with SURFEX

For a given DATE: forecast from DATE-1 to DATE then analysis at DATE

```
set in    NAM_IO_OFFLINE: CPREPFILE="PREP_INIT"
          NAM_ASSIM: LASSIM=F
          NAM_IO_VARASSIM: LPRT=F
OFFLINE.EXE
mv SURFOUT.nc PREP_DATE_EKF_PERT0.nc

set in    NAM_ASSIM: LASSIM=T
          NAM_IO_VARASSIM: LPRT=T
for i = 1, …, NVAR
    set in    NAM_VAR: NIVAR=i
    OFFLINE.EXE
    mv SURFOUT.nc PREP_DATE_EKF_PERTi.nc
```

DATA ASSIMILATION

```
set in    NAM_IO_OFFLINE: CPREPFILE="PREP_DATE_EKF_PERT0"
          NAM_IO_VARASSIM: LPRT=F
          NAM_VAR: NIVAR=0
SODA.EXE
mv SURFOUT.nc PREP_INIT.nc
```

Perturbed model runs and NAM_VAR

- Jacobian matrices $\mathbf{M}_{[p]}$ are calculated using perturbed model runs

$$\left[\mathbf{M}_{[p]}\right]_j = \frac{\partial \mathbf{x}^f\left(t+24h\right)}{\partial x_j}$$
$$\approx \frac{\mathcal{M}_{[p]}\left(\mathbf{x}^a(t)+\delta x_j\right) - \mathcal{M}_{[p]}\left(\mathbf{x}^a(t)\right)}{\delta x_j}$$

- Size of perturbations δx_j is determined by XTPRT_M in ***NAM_VAR***

$$\delta LAI_{[p]} = \texttt{XTPRT\_M(1)}\, LAI^a_{[p]}$$

$$\delta w_{gi,[p]} = \texttt{XTPRT\_M(i+1)}\left(w_{fc} - w_{wilt}\right) \quad i = 2, \ldots, 8$$

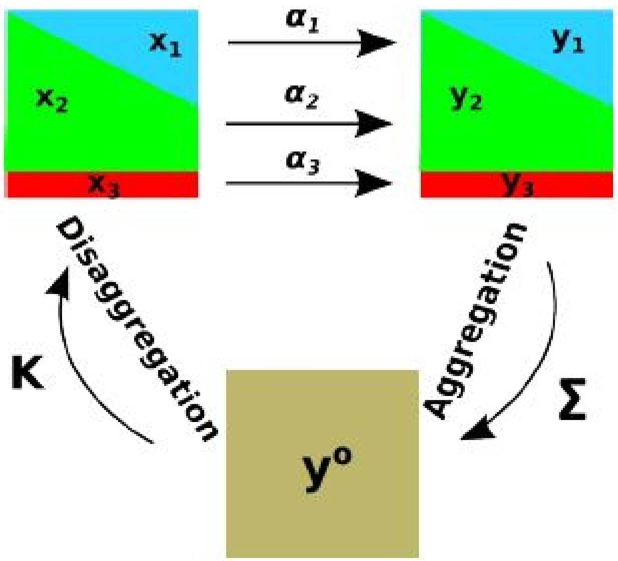
```
XTPRT_M(1)   = 0.001,
XTPRT_M(2)   = 0.001,
XTPRT_M(3)   = 0.001,
XTPRT_M(4)   = 0.001,
XTPRT_M(5)   = 0.001,
XTPRT_M(6)   = 0.001,
XTPRT_M(7)   = 0.001,
XTPRT_M(8)   = 0.001,
XTPRT_M(9)   = 0.001,
```

Simplified Extended Kalman Filter (SEKF)

- Update step of SEKF is for patch p:

$$\mathbf{x}^a_{[p]} = \mathbf{x}^f_{[p]} + \mathbf{K}_{[p]}\left(\mathbf{y}^o - \mathbf{y}^f\right)$$

$$\mathbf{K}_{[p]} = \alpha_{[p]}\mathbf{B}\left(\mathbf{HM}_{[p]}\right)^T\left(\sum_{k=1}^{12}\alpha^2_{[k]}\left(\mathbf{HM}_{[k]}\right)\mathbf{B}\left(\mathbf{HM}_{[k]}\right)^T + \mathbf{R}\right)^{-1}$$



- Observation error covariance matrix $\mathbf{R}$ is set in namelist ***NAM_OBS***

- Background error covariance matrix $\mathbf{B}$ is set in namelist ***NAM_VAR***

Observation error covariance matrix and NAM_OBS

- Below is the setting of ***NAM_OBS*** for the experiment

```
&NAM_OBS                NOBSTYPE         = 2
                        LOBSNAT          = F,
                        CFILE_FORMAT_OBS = "ASCII",
                        COBS_M(1)        = 'WG2',
                        COBS_M(2)        = 'LAI',
                        XERROBS_M(1)     = 0.4,
                        XERROBS_M(2)     = 0.2,
                        NNCO(1)          = 1,
                        NNCO(2)          = 1,
/
```

- Setting of the observation error covariance matrix $\mathbf{R}$ for assimilation:
  - Standard deviation for observations of surface soil moisture

    $$\sigma^o_{wg2} = \texttt{XERROBS\_M(1)}\,(w_{fc} - w_{wilt})$$

    with w_{fc} soil field capacity
    and w_{wilt} soil wilting point
  - Standard deviation for observations of LAI

    $$\sigma^o_{LAI} = \texttt{XERROBS\_M(2)}\,LAI^o$$

    with LAI^o observed LAI
  - No correlation between the two types of observations

Background error covariance matrix and NAM_VAR

- The background error covariance matrix $\mathbf{B}$ is determined by XSIGMA_M in ***NAM_VAR*** (see below for the current experiment)

$$\sigma^b_{LAI_{[p]}} = \begin{cases} \texttt{XSIGMA\_M(1)}\, LAI^f_{[p]} & LAI^f_{[p]} > 2.0\,\mathrm{m}^2.\mathrm{m}^{-2} \\ 0.4 & LAI^f_{[p]} \leq 2.0\,\mathrm{m}^2.\mathrm{m}^{-2} \end{cases}$$

$$\sigma^b_{w_{gi,[p]}} = \texttt{XSIGMA\_M(i+1)}\,(w_{fc} - w_{wilt}) \quad i = 2, \ldots, 8$$

- No correlation between variables is assumed in $\mathbf{B}$. The sensitivity of control variables to observations is entirely driven by the Jacobian matrices.

```
XSIGMA_M(1)  = 0.2,
XSIGMA_M(2)  = 0.2,
XSIGMA_M(3)  = 0.4,
XSIGMA_M(4)  = 0.2,
XSIGMA_M(5)  = 0.2,
XSIGMA_M(6)  = 0.2,
XSIGMA_M(7)  = 0.2,
XSIGMA_M(8)  = 0.2,
XSIGMA_M(9)  = 0.2,
```

Linking LAI analysis and biomass variable

- To be considered, the updated value of LAI for each patch $LAI^a_{[p]}$ has to be propagated to the prognostic leaf biomass variable in the **ISBA** land surface model

- This propagation is determined in ***NAM_VAR*** by CBIO and XALPH

$$\texttt{CBIO}_{[p]} = \texttt{XALPH(p)}\, LAI^a_{[p]} \quad p = 1, \ldots, 12$$

- **Warning: XALPH must be prescribed in accordance to vegetation parameters for Specific Leaf Area (SLA) used in laigain.F90**

```
CBIO          = 'BIOMA1',
 XALPH(1)     = 0.,
 XALPH(2)     = 0.,
 XALPH(3)     = 0.,
 XALPH(4)     = 0.08203445,
 XALPH(5)     = 0.07496252,
 XALPH(6)     = 0.06846970,
 XALPH(7)     = 0.06771856,
 XALPH(8)     = 0.09744689,
 XALPH(9)     = 0.09744689,
 XALPH(10)    = 0.07164350,
 XALPH(11)    = 0.17686594,
 XALPH(12)    = 0.07164350
```

Linking LAI analysis and biomass variable

- Leaf Area Index computed from leaf biomass reservoir (B_l) using Specific Leaf Area (SLA) in **laigain.F90**:

$$LAI = SLA \times B_l \quad \text{with} \quad SLA = e\,N_l + f$$

with: e (in $m^2\,kg^{-1}\,\%^{-1}$) and f (in $m^2\,kg^{-1}$) plasticity parameters, and N_l (in %) the nitrogen concentration of the leaf biomass B_l.

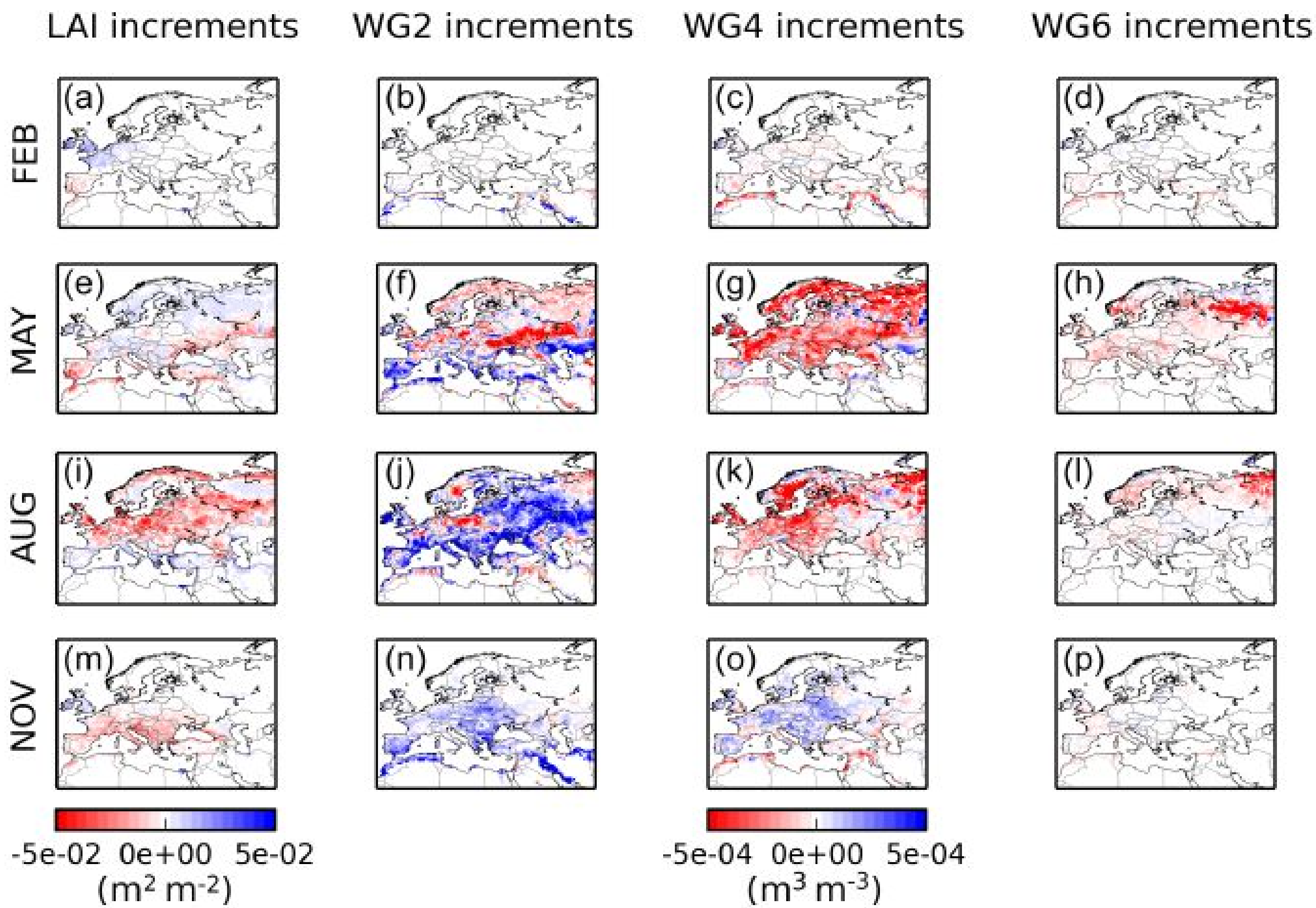
- e, f and N_l are defined by for the 19 vegetation type then averaged by patch depending on what is prescribed in `NPATCH`:
  - default values defined in **ini_data_param.F90** (`CE_NITRO`, `CF_NITRO` and `CNA_NITRO`) but can be prescribed in namelist ***NAM_DATA_ISBA***
  - N_l can vary with CO_2 concentration in **vegetation_evol.F90** if `LNITRO_DILU = T` in ***NAM_ISBA_AGSn*** (advanced version)

Impact of SEKF on control variables

- To measure the impact of the assimilation on control variables, it can be interesting to output analysis increments (analysis – forecast). This can be done by adding in CSELECT defined in ***NAM_WRITE_DIAG_SURFn*** ‘ANA_INCRx’ (see list below left)

Figure 5. Rows from top to bottom represent averaged analysis increments for all months of February, May, August and November over 2000–2012. From left to right, four control variables are illustrated: leaf area index and soil moisture in the second (w_2, 1–4 cm), fourth (w_4, 10–20 cm) and sixth (w_6, 40–60 cm) layer of soil, respectively.

```
'ANA_INCR1' ,
'ANA_INCR2' ,
'ANA_INCR3' ,
'ANA_INCR4' ,
'ANA_INCR5' ,
'ANA_INCR6' ,
'ANA_INCR7' ,
'ANA_INCR8' ,
```



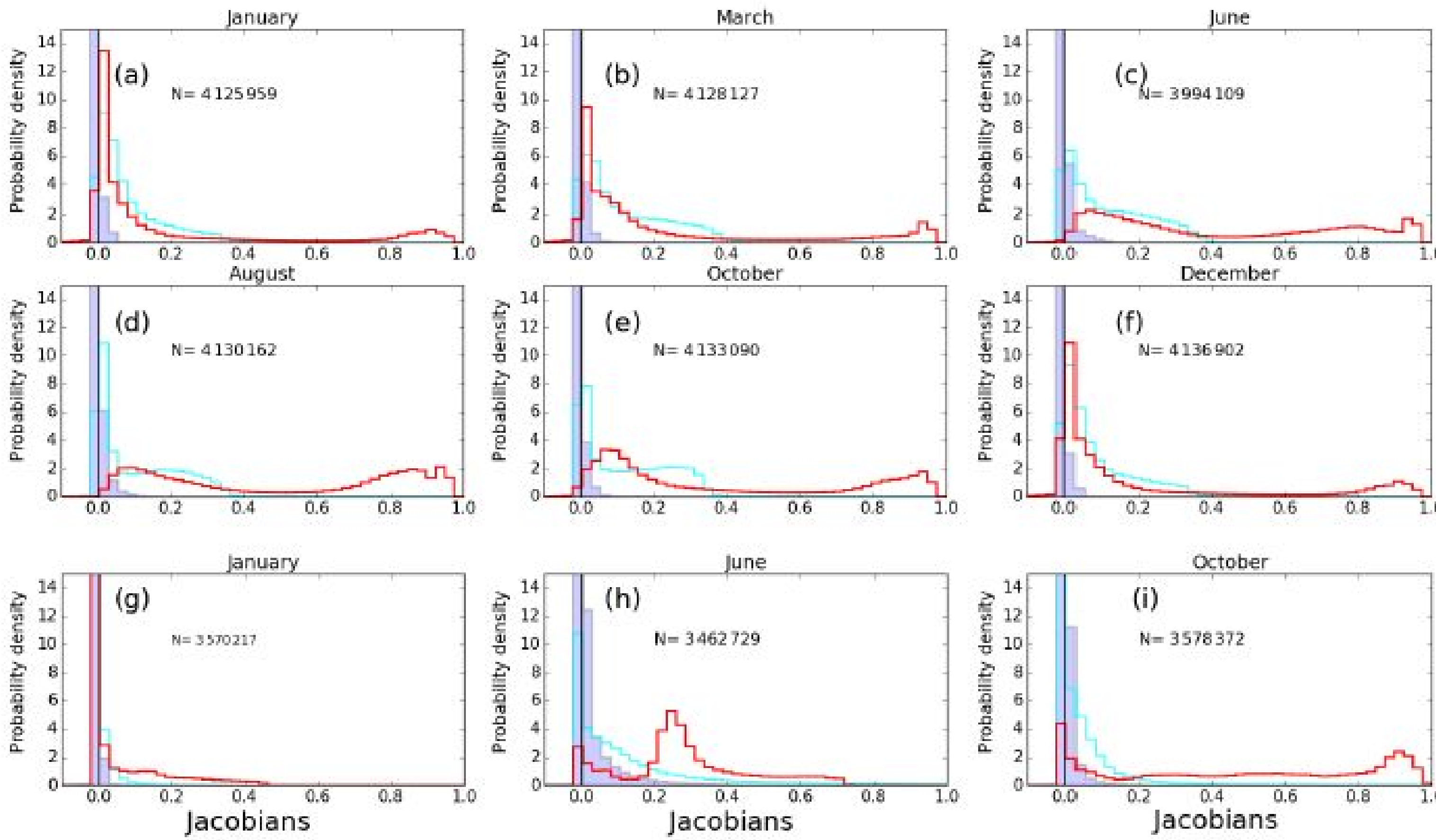
METEO FRANCE

Example of Jacobians

- It can be interesting to output Jacobians to understand the influence of observations on control variables. This can be done by adding in CSELECT defined in ***NAM_WRITE_DIAG_SURFn*** 'HOy_x_1' (see list below left)

Figure 4. Jacobian value distribution: (a–f), $\frac{\partial \mathrm{SSM}^t}{\partial w_2{}^0}$ (red line), $\frac{\partial \mathrm{SSM}^t}{\partial w_4{}^0}$ (cyan line) and $\frac{\partial \mathrm{SSM}^t}{\partial w_8{}^0}$ (blue line) for all months of January, March, June, August, October and December over 2000–2012, (g–i), $\frac{\partial \mathrm{LAI}^t}{\partial \mathrm{LAI}^0}$ (red line), $\frac{\partial \mathrm{LAI}^t}{\partial w_4{}^0}$ (cyan line) and $\frac{\partial \mathrm{LAI}^t}{\partial w_8{}^0}$ (blue line) for all months of January, June and October over 2000–2012. A black solid line represents a value of 0.

```
'HO1_1_1'     ,
'HO1_2_1'     ,
'HO1_3_1'     ,
'HO1_4_1'     ,
'HO1_5_1'     ,
'HO1_6_1'     ,
'HO1_7_1'     ,
'HO1_8_1'     ,
'HO2_1_1'     ,
'HO2_2_1'     ,
'HO2_3_1'     ,
'HO2_4_1'     ,
'HO2_5_1'     ,
'HO2_6_1'     ,
'HO2_7_1'     ,
'HO2_8_1'     ,
```



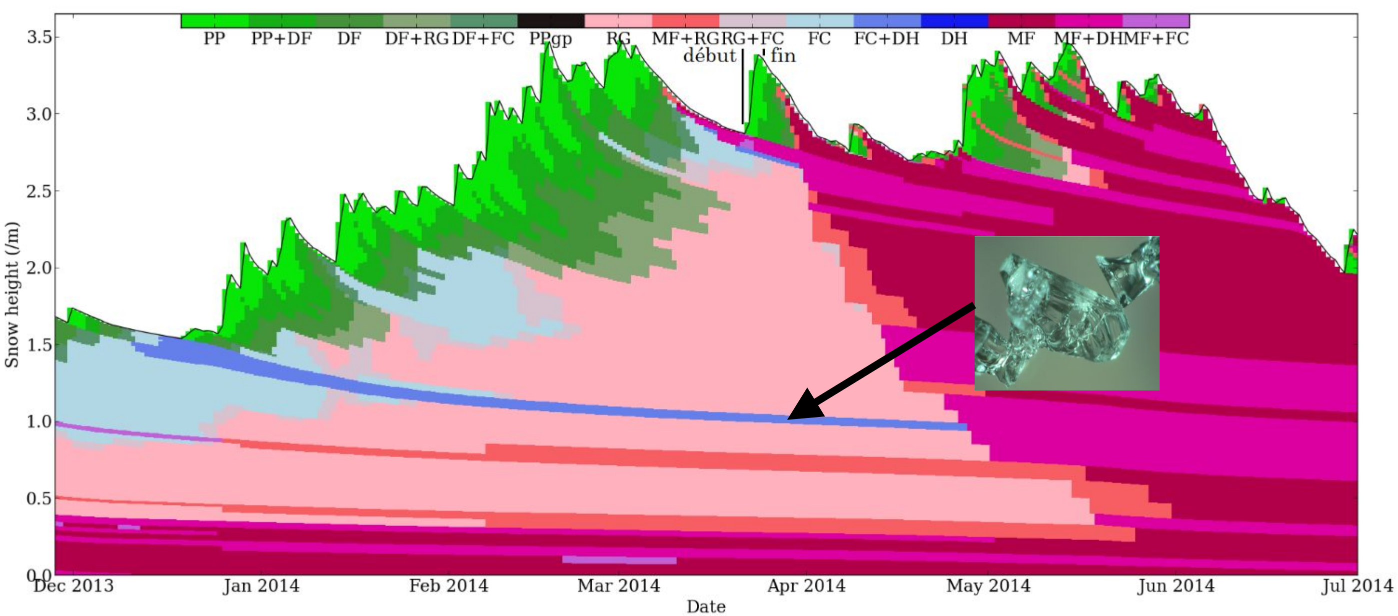
Outline

- Bases of data assimilation in SURFEX
- Optimal Interpolation for continental surfaces in NWP
- Simplified Extended Kalman Filter for land surface monitoring
- Particle filter for snow model Crocus
- Further topics

Crocus Snow model

The following slides have been provided by Matthieu Lafaysse (CNRM/CEN, Grenoble)

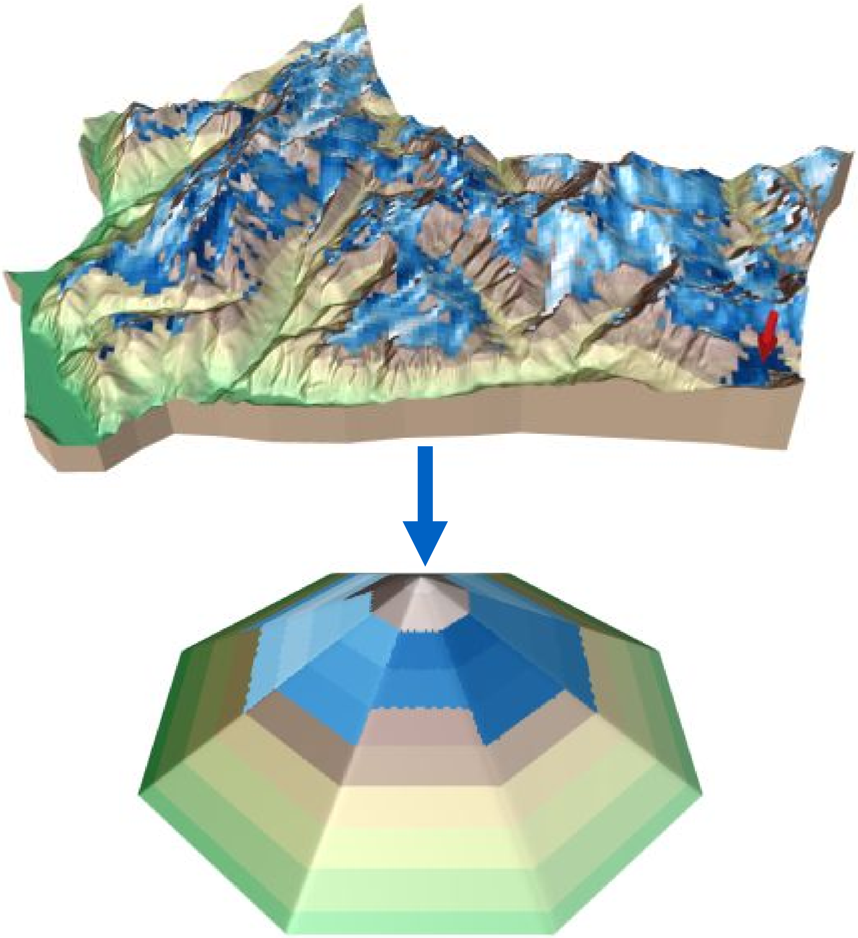
- Crocus (Brun *et al.*, 1989, 1992; Vionnet *et al.*, 2012) is a detailed snowpack model with:
  - More **detailed vertical layering** than usual snow schemes in NWP or climate applications
  - Explicit snow **microstructure** properties (**metamorphism**)
  - Improved heat diffusion / energy budget / phase change, compaction, liquid water percolation

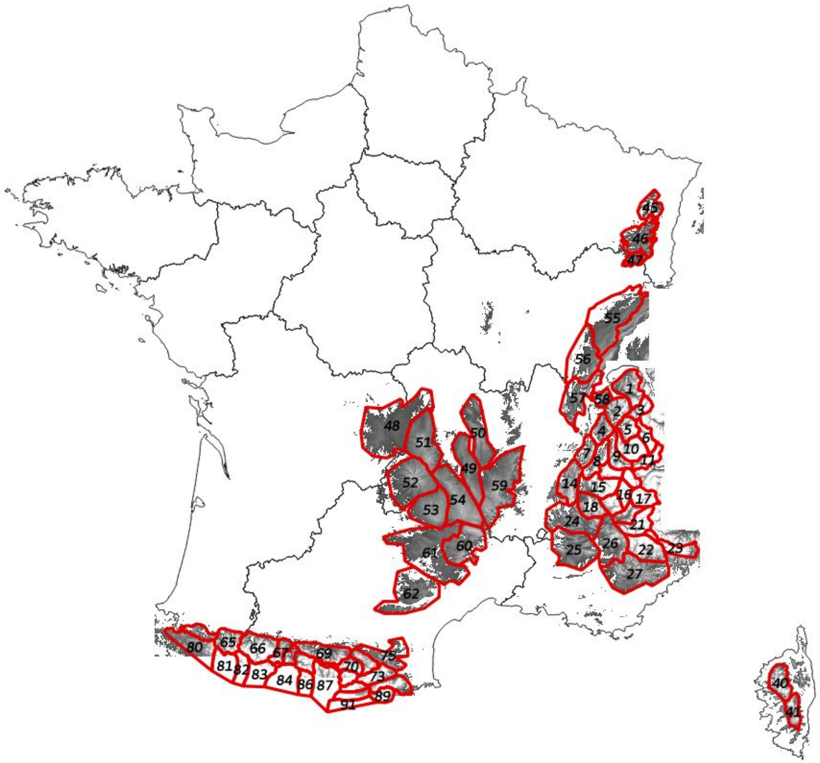


Each colour displays a different type of snow microstructure

Modelling approach

- **By topographic classes**, i.e. we assume homogenous areas where meteorological and snow conditions <u>only depend on</u> elevation, aspect and slope **for each massif** (Durand *et al.*, 1999; Vernay et *al.*, 2022)

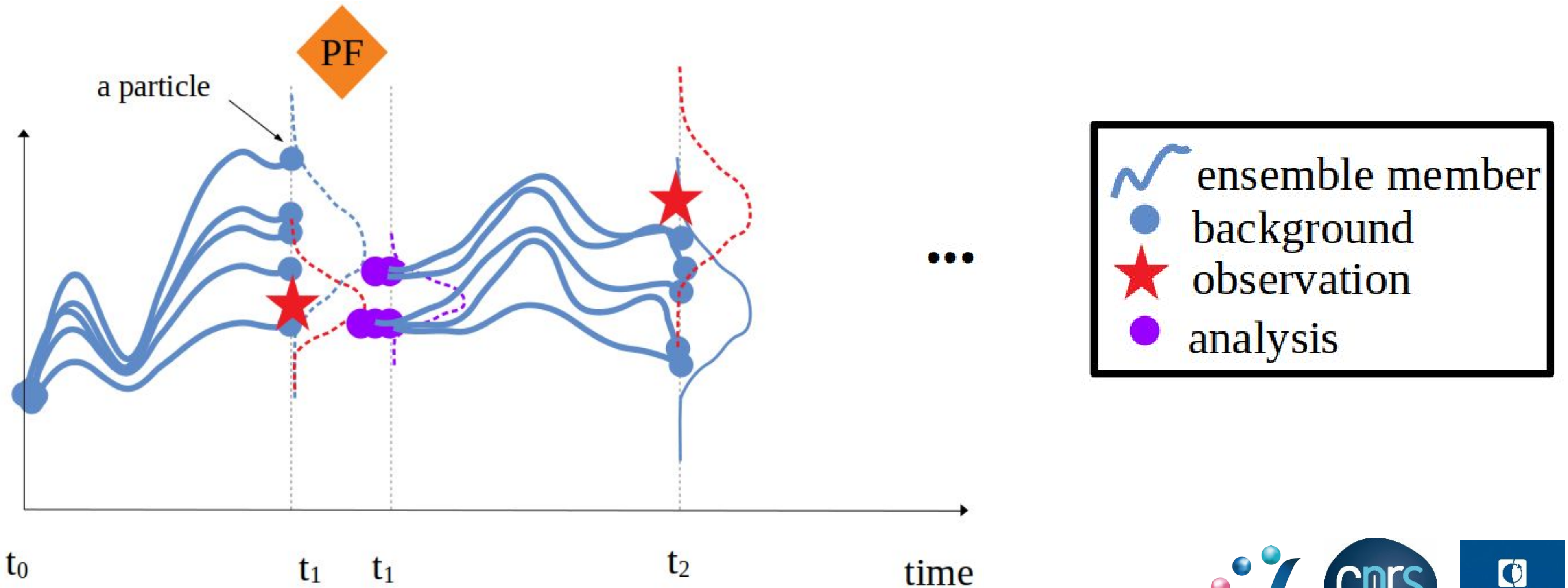




Massifs modelled in October 2022

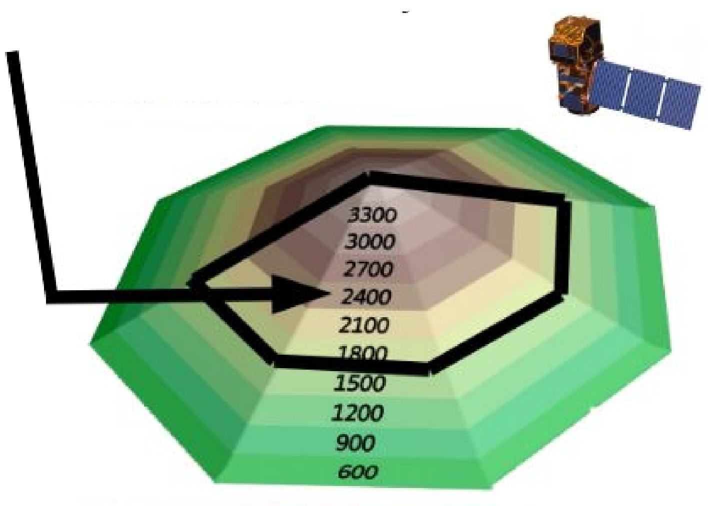
Data Assimilation in Crocus: particle filter

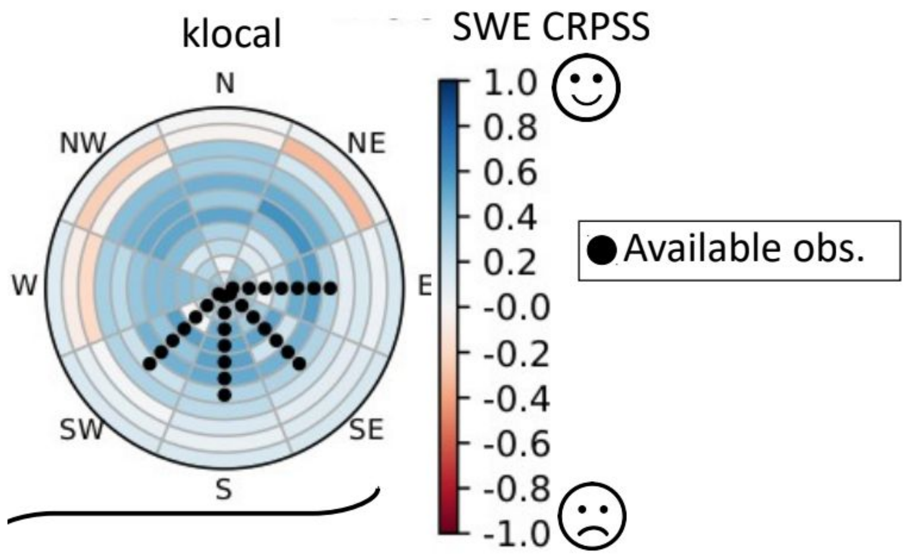
- **Sequential Particle Filter** (Gordon *et al.*, 1993 ; Charrois *et al.*, 2016)
  - Resampling of most reliable members to reduce ensemble spread
  - Well suited for highly non-linear models ***with a varying number of variables such as varying number of layers and parameters in Crocus***



An example of assimilating reflectances

- **Based on** Cluzet *et al.* (2021)
  - Assimilation of optical reflectances available once a week only on sunny slopes and above the tree line
  - Propagate information to non-observed areas using spatial correlations of background
  - Evaluation using Snow Water Equivalent (SWE)





Outline

- Bases of data assimilation in SURFEX
- Optimal Interpolation for continental surfaces in NWP
- Simplified Extended Kalman Filter for land surface monitoring
- Particle filter for snow model Crocus
- Further topics

Advanced Data Assimilation in SURFEX

- OI and SEKF are the main data assimilation algorithms proposed in **SURFEX v8.1** for tile **NATURE**

- Nevertheless, more advanced data assimilation approaches are currently developed within the SURFEX framework:

  - An Ensemble Kalman Filter (Ensemble Square Root Filter) has been developed for the **ISBA** land surface model by the CNRM/VEGEO team (Bonan *et al.*, 2020). Current version in SURFEX v8.1 works only for one observation. Updated version of code will be put in a dev. branch of the SURFEX v9 git

  - Another Ensemble Square Root Filter was developed by MetNorway for NWP purposes within SURFEX. This work is based on Blyverket *et al.* (2019)

- For the Crocus snow model, a particle filter is also available

Technical points

- It can be difficult to write scripts to run SURFEX with assimilation including preparing forcing files, observations files, settings in OPTIONS.nam, plotting results … Fortunately toolboxes have been developed to simplify this task:

  - **LDAS-Monde**: python and shell scripts developed by CNRM/GMME/VEGEO in the context of vegetation and soil moisture monitoring.
    More information can be found here:
    https://www.umr-cnrm.fr/spip.php?article1022&lang=en
    https://opensource.umr-cnrm.fr/projects/openldasmonde/files

  - **pysurfex**: python API developed by MET Norway for surface assimilation in NWP.
    More information can be found here:
    https://metno.github.io/pysurfex

  - **CroCO**: contact crocus@meteo.fr if you plan using either Crocus or CroCO, they will provide you all the help you need to run it.

- A complete documentation of the namelists used in SURFEX v8.1 can be found here:
  http://www.umr-cnrm.fr/surfex/IMG/pdf/surf-v8-1d.pdf

Thank you for your attention. Time for questions!

Bibliography

- Albergel, C. *et al.* (2017), Sequential assimilation of satellite-derived vegetation and soil moisture products using SURFEX_v8.0: LDAS-Monde assessment over the Euro-Mediterranean area, *Geosci. Model Dev.*, 10, 3889 – 3912, doi: 10.5194/gmd-10-3889-2017

- Barbu, A. L. *et al.* (2014), Integrating ASCAT surface soil moisture and GEOV1 leaf area index into the SURFEX modelling platform: a land data assimilation application over France, *Hydrol. Earth Syst. Sci.*, 18, 173 – 192, doi: 10.5194/hess-18-173-2014

- Beljaars, A. C. M. *et al.* (1996), The anomalous rainfall over the United States during July 1993: Sensitivity to land surface parameterization and soil moisture anomalies, *Mon. Weather Rev.*, 124, 362 – 382, doi: 10.1175/1520-0493(1996)124<0362:TAROTU>2.0.CO;2

- Bonan, B. *et al.* (2020), An ensemble square root filter for the joint assimilation of surface soil moisture and leaf area index within the Land Data Assimilation System LDAS-Monde: application over the Euro-Mediterranean region, *Hydrol. Earth Syst. Sci.*, 24, 325–347, doi: 10.5194/hess-24-325-202

Bibliography

- Brun, E. *et al.* (1989), An energy and mass model of snow cover suitable for operational avalanche forecasting, *J. Glaciol.*, 35(121), 333–342, doi: 10.3189/S0022143000009254

- Brun, E. *et al.* (1992), A numerical model to simulate snowcover stratigraphy for operational avalanche forecasting, *J. Glaciol.*, 38(128), 13–22, doi: 10.3189/S0022143000009552

- Blyverket, J. *et al.* (2019), An Evaluation of the EnKF vs. EnOI and the assimilation of SMAP, SMOS and ESA CCI soil moisture data over the contiguous US, *Remote Sensing*, 11, 478, doi: 10.3390/rs11050478

- Cluzet, B. *et al.* (2021), CrocO_v1.0 : a Particle Filter to assimilate snowpack observations in a spatialised framework, *Geosci. Model Dev.*, 14, 1595–1614, doi: 10.5194/gmd-14-1595-2021

- Coiffier, J. *et al.* (1987), The operational hemispheric model at the French meteorological service, *Collection of Papers presented at the WMO/IUGG Symposium, Tokyo, 4-6 Aug. 1986. Tokyo, Meteorological Society of Japan*, 337-345 https://www.jstage.jst.go.jp/article/jmsj1965/64A/0/64A_337/_pdf/-char/ja

Bibliography

- Draper, C. *et al.* (2009), An EKF assimilation of AMSR-E soil moisture into the ISBA land surface scheme, *J. Geophys. Res.*, 114, D020104, doi: 10.1029/2008JD011650

- Drusch, M. *et al.* (2005), Observations operators for the direct assimilation of TRMM microwave imager retrieved soil moisture, *Geophys. Res. Lett.*, 32, L15403, doi: 10.1029/2005GL023623

- Durand, Y. *et al.* (1999), A computer-based system simulating snowpack structures as a tool for regional avalanche forecasting, *J. Glaciol.*, 45, 469–484, doi: 10.3189/S0022143000001337

- Giard, D., and Bazile, E. (2000), Implementation of a new assimilation scheme for soil and surface variables in a global NWP model, *Mon. Weather Rev.*, 128, 997–1015, doi: 10.1175/1520-0493(2000)128<0997:IOANAS>2.0.CO;2

- Jiang, C. *et al.* (2016), Inconsistencies of interannual variability and trends in long-term satellite leaf area index products, *Glob. Change Biol.*, 23(10), 4133–4146, doi: 10.1111/gcb.13787

Bibliography

- Mahfouf, J.-F. (1991), Analysis of soil moisture from near-surface parameters: A feasibility study, *J. Appl. Meteorol.*, 30, 506–526, doi: 10.1175/1520-0450(1991)030%3C1534:AOSMFN%3E2.0.CO;2

- Mahfouf, J.-F. *et al.* (2009), A comparison of two off-line soil analysis schemes for assimilation of screen level observations, *J. Geophys. Res.*, 114, D08105, doi: 10.1029/2008JD011077

- Nichols, N. K. (2010), Mathematical concepts of data assimilation, in: Data assimilation: making sense of observations, Springer-Verlag, Berlin, Germany, 13–40, doi: 10.1007/978-3-540-74703-1_2

- Reichle, R. H. and Koster, R. D. (2004), Bias reduction in short records of satellite soil moisture, Geophys. Res. Lett., 31, L19501, doi: 10.1029/2004GL020938

- Scipal, K. *et al.* (2008), Assimilation of a ERS scatterometer derived soil moisture index in the ECMWF numerical weather prediction system, *Adv. Water Resour.*, 31, 1101–1112, doi: 10.1016/j.advwatres.2008.04.013

Bibliography

- Vernay, M. *et al.* (2022), The S2M meteorological and snow cover reanalysis over the French mountainous areas: description and evaluation (1958–2021), *Earth Syst. Sci. Data*, 14, 1707–1733, doi: 10.5194/essd-14-1707-2022

- Vionnet, V. *et al.* (2012), The detailed snowpack scheme Crocus and its implementation in SURFEX v7.2, *Geosci. Model Dev.*, 5, 773–791, doi: 10.5194/gmd-5-773-2012