Funded by the European Union

Coordinated by ECMWF

USING ENSEMBLE LAND DATA ASSIMILATION SPREAD FOR EVALUATING THE IMPACT OF SURFACE OBSERVATIONS IN THE ISBA COUPLED HARMONIE-AROME MODEL

CopERnIcus climate change Service Evolution - CERISE

Abhishek Lodh (Ph.D.)

Postdoctoral Researcher

Swedish Meteorological and Hydrological Institute (SMHI), Norrköping, Sweden

46th EWGLAM and 31st SRNWP Meeting, Prague, 2nd October’2024

With help from:

Jelena Bojarova, Patrick Samuelsson, Magnus Lindskog, Ulf Andrae, Meto, Swapan, Åsmund Bakketun, Jostein Blyverket, Trygve Aspelien

Outline of the talk

- **Context**
- **Introduction**
- **Basics of Land Data Assimilation system**
- Theory of Extended, simplified EKF and Ensemble Kalman filter
- Results: Simulation of LDAS (sEKF & EnSRKF) test cases

2024 plans

Coupled land-atmosphere data assimilation

- **ECMWF**: (1) Outer loop land-atmosphere coupled data assimilation developments in the ECMWF IFS and evaluation for global reanalysis, (2) Coupled skin temperature assimilation developments in the IFS

MOTIVATION

- **SMHI**: Outer loop coupled DA developments in HARMONIE-AROME.
- **Met Norway**: (1) Bring the LDAS developments from WP1 into the HARMONIE-AROME coupled system, (2) coupled DA developments in HARMONIE-AROME

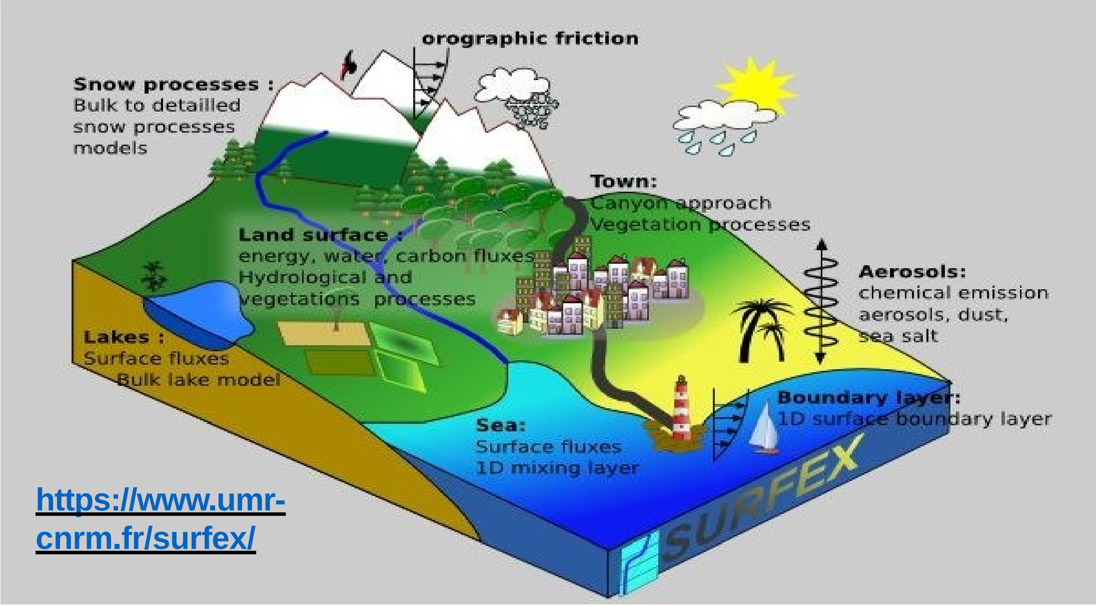
Funded by the European Union

Context

- **To Investigate the Perturbation Growth in Land Surface Variables in the Ensemble and Extended Kalman Filter- Based LDAS.**
- Conducted Land Data Assimilation system (LDAS) simulations over the Nordic (NORD_2.5km) domain on ATOS (ECMWF) with cy46h+DIF+MEB+SEKF+3DVAR and ENSRKF+3DVAR for 60 (30) days from summer priod of June-July'23 (October' 23).
- Compared perturbation growth in land surface variables like (T2m, Q2m, soil moisture, soil temperature, LHF, SHF)
-
- Signs that ENSRKF adds value to growth in perturbations of soil variables and fluxes reaching deeper soil layers, with improvements in forecats of near surface variables.

Purpose of Land Data Assimilation

Funded by the European Union



https://www.umr-cnrm.fr/surfex/

Time scales	Driving mechanism of Land-atmospheric interaction
Seconds to Hour	exchange momentum, energy, water, carbon dioxide and other chemical constituents between the land surface and the atmosphere
Day to seasons	changes in the store of soil moisture, changes in snowpack , changes in carbon allocation, and vegetation phenology
years to centuries	vegetation structure and function (e.g., disturbance, land use, stand growth) is strongly determined by climate influences

*** Land covers a substantial portion (about 30%) of the Earth's surface. The land surface consists of soil, vegetation, snow, glaciers, inland water, mountains, and human beings.**

*** "The atmosphere and the upper layers of soil (or sea) form together a united system. This is evident since the first few meters of ground has a thermal capacity comparable with 1/10 that of the entire atmospheric column standing upon it.**

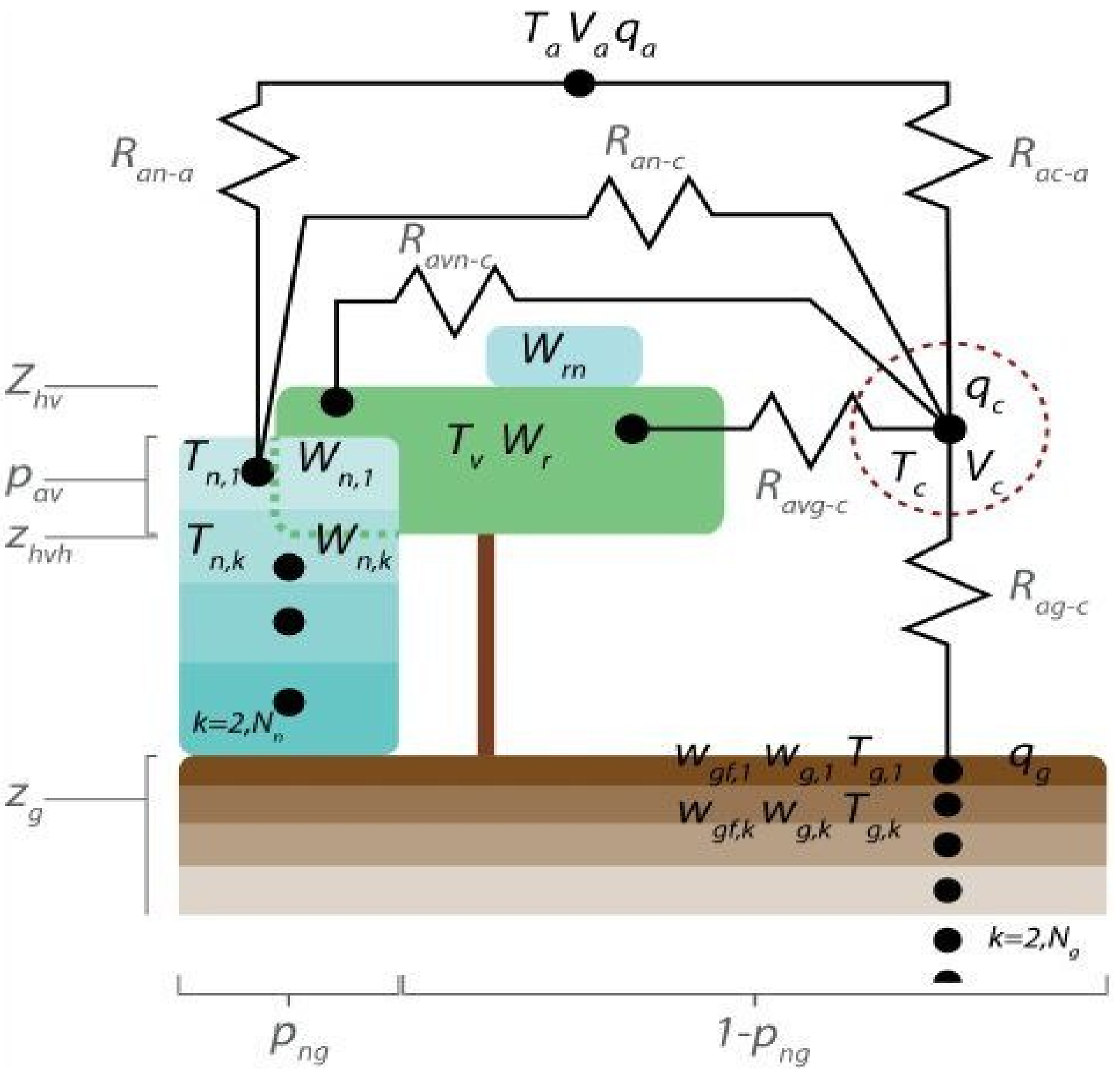
**** Improved understanding of land-atmosphere interaction and far better measurements of land-surface properties, especially soil moisture, exchanges of latent and sensible heat fluxes, would constitute a major advancement and holds the key to improvements in a number of forecasting problems.***

*** The surface variability not only determines the microclimate but also affects the mesoscale atmospheric circulation**

Hence, proper representation of initial state (boundary condition) of Land Surface in regional NWP (Climate) models is important for medium range and S2S forecasts.

Funded by the European Union

SURFEX



Without Land DA the state of the soil easily drifts away. Thus, we correct bad physics with heavy LDAS.

Explicit snow (12 layers), Decharme et al. (2016, doi:10.5194/tc-10-853-2016)

MEB and forest snow, Napoly et al. (2020, doi:10.5194/gmd-13-6523-2020)

Diffusion soil (14 layers), Decharme et al. (2011, doi:10.1029/2011JD016002)

Explicit canopy: MEB (Multi-Energy Balance), Boone et al. (2017, doi:10.5194/gmd-10-843-2017)

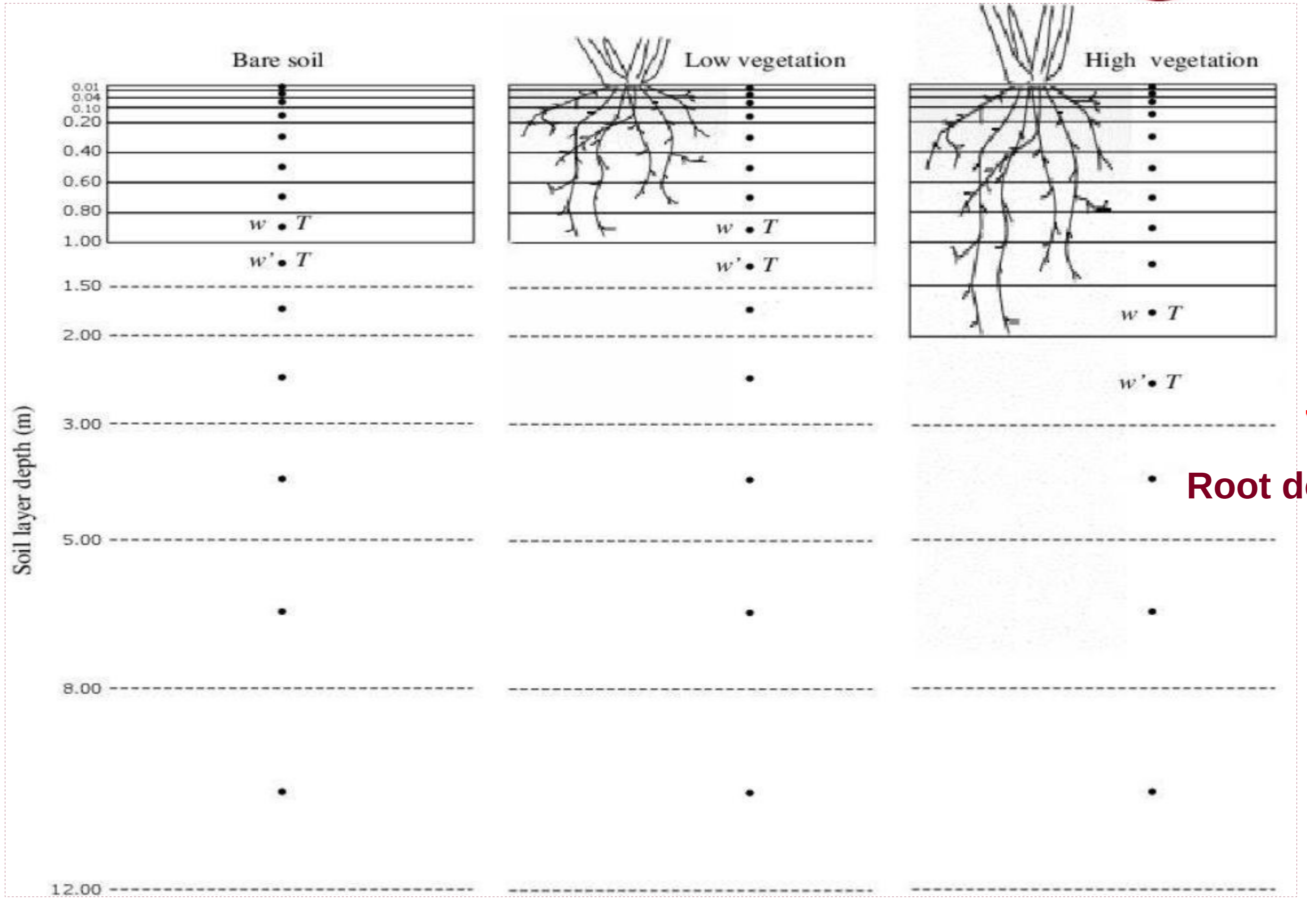
Litter layer in forest Napoly et al. (2017, doi:10.5194/gmd-10-1621-2017) Low heat capacity. Stores energy and water/ice.

The LSM in SURFEX is 2nd generation in HARMONIE-AROME NWP system (Research mode)

Funded by the European Union

ISBA (Inter-action between the Soil, Biosphere and Atmosphere) LSM: Soil Diffusion

- The heat and soil moisture transfers within the soil are computed using 14 layers up to a 12 m depth.
- The depth of the 14 layers (see figure) have been chosen to minimize numerical errors in solving the finite-differenced diffusive equations, especially in the uppermost meter of the soil. The same default grid thicknesses are used everywhere.
- Hydrological grids, enclosed by the solid black lines in the figure, are defined by root depth for vegetated surfaces. Thus the soil water prognostic equations do not extend as deeply as the thermal computations.
- The root depth is essential for the transpiration estimates.

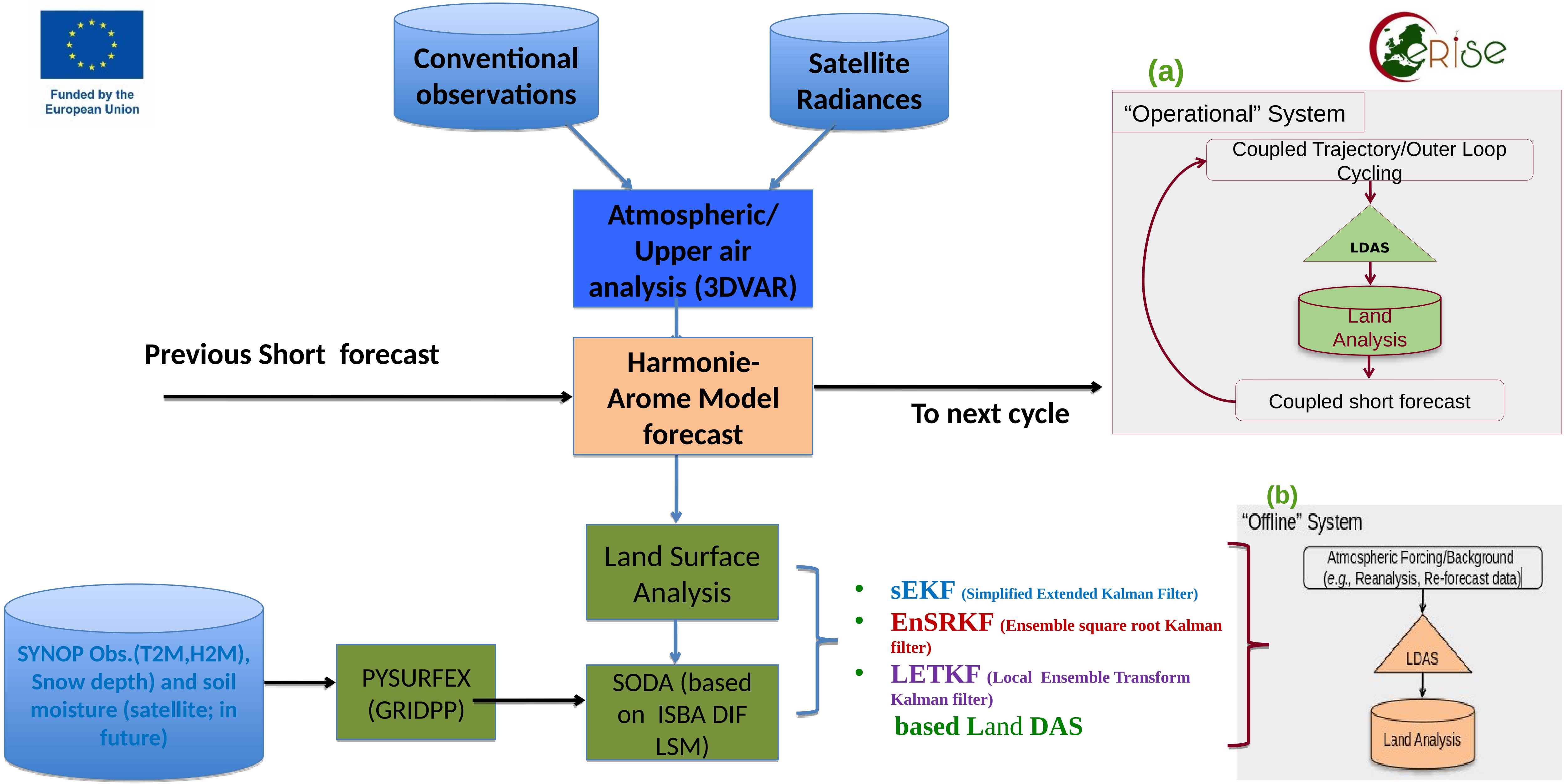


Root depth

Figure: in SURFEX Scientific documentation for v8.1, P. Le Moigne, February 23, 2018. http://www.umr-cnrm.fr/surfex/IMG/pdf/surfex_scidoc_v8.1.pdf

Noilhan and Planton, 1989; Noilhan and Mah-Fouf, 1996, Decharme et al. 2011, doi:10.1029/2011JD016002

Funded by the European Union



Basic Schematic of Land data Assimilation

Funded by the European Union

Extended Kalman Filter (EKF)

EKF equations:

$$\mathbf{X}_B = M(\mathbf{X}_\mathbf{A})$$

$$\mathbf{B} = \mathbf{M} \cdot \mathbf{A} \cdot \mathbf{M}^\mathbf{T} + \mathbf{Q} \quad \text{- evolving B-matrix}$$

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

$$\mathbf{X}_\mathbf{A} = \mathbf{X}_\mathbf{B} + \mathbf{K}(\mathbf{Y} - H(\mathbf{X}_\mathbf{B}))$$

$$\mathbf{A} = (\mathbf{I} - \mathbf{K} \cdot \mathbf{H}) \cdot \mathbf{B}$$

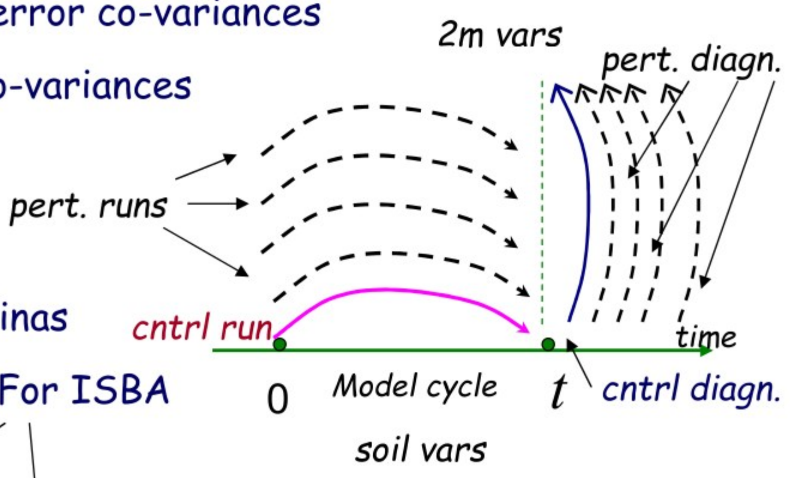
Q - matrix of the internal model error co-variances

A - matrix of the analysis error co-variances

I - unit matrix

M - matrix of model Jacobinas

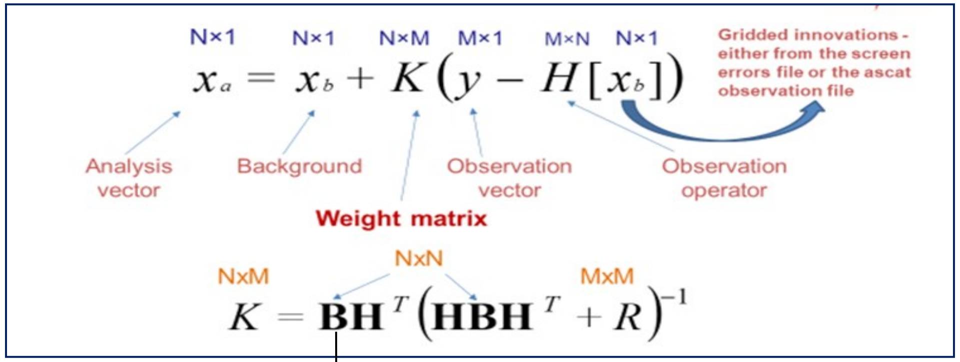
H - matrix of obs operator Jacobinas



$$\mathbf{M} = \begin{Bmatrix} \frac{\partial T_S^t}{\partial T_S^0} & \frac{\partial T_S^t}{\partial T_G^0} & \frac{\partial T_S^t}{\partial W_S^0} & \frac{\partial T_S^t}{\partial W_G^0} \\ \frac{\partial T_G^t}{\partial T_S^0} & \frac{\partial T_G^t}{\partial T_G^0} & \frac{\partial T_G^t}{\partial W_S^0} & \frac{\partial T_G^t}{\partial W_G^0} \\ \frac{\partial W_S^t}{\partial T_S^0} & \frac{\partial W_S^t}{\partial T_G^0} & \frac{\partial W_S^t}{\partial W_S^0} & \frac{\partial W_S^t}{\partial W_G^0} \\ \frac{\partial W_G^t}{\partial T_S^0} & \frac{\partial W_G^t}{\partial T_G^0} & \frac{\partial W_G^t}{\partial W_S^0} & \frac{\partial W_G^t}{\partial W_G^0} \end{Bmatrix}$$

For ISBA

$$\mathbf{H} = \begin{Bmatrix} \frac{\partial T2m^t}{\partial T_S^t} & \frac{\partial T2m^t}{\partial T_G^t} & \frac{\partial T2m^t}{\partial W_S^t} & \frac{\partial T2m^t}{\partial W_G^t} \\ \frac{\partial RH2m^t}{\partial T_S^t} & \frac{\partial RH2m^t}{\partial T_G^t} & \frac{\partial RH2m^t}{\partial W_S^t} & \frac{\partial RH2m^t}{\partial W_G^t} \end{Bmatrix}$$



Background error covariance

Using the flow-dependent and evolving B in an ideal EKF, analysis A is expected to be improve. However, it is computationally expensive as $n \sim 10^7$

Source: Kalnay & Ekaterina Kurzeneva

Simplified Extended Kalman filter (EKF)….

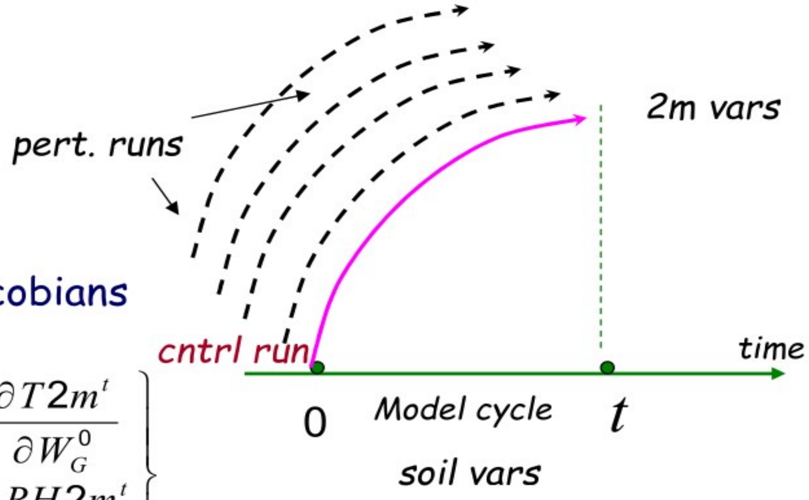
$$\mathbf{X_B} = M(\mathbf{X_A})$$

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

$$\mathbf{X_A} = \mathbf{X_B} + \mathbf{K}(\mathbf{Y} - H(\mathbf{X_B}))$$

B - matrix does not evolve, but H-Jacobians are calculates as follows:

$$\mathbf{H} = \begin{Bmatrix} \frac{\partial T2m^t}{\partial T_S^0} & \frac{\partial T2m^t}{\partial T_G^0} & \frac{\partial T2m^t}{\partial W_S^0} & \frac{\partial T2m^t}{\partial W_G^0} \\ \frac{\partial RH2m^t}{\partial T_S^0} & \frac{\partial RH2m^t}{\partial T_G^0} & \frac{\partial RH2m^t}{\partial W_S^0} & \frac{\partial RH2m^t}{\partial W_G^0} \end{Bmatrix}$$



SEKF is not much cheaper than EKF: it contains the same number of additional model runs. The only economy is in matrix calculations.

The main reason of SEKF is that B-matrix might be unstable.

- **sEKF Requires**

$|\nabla_x M_i(x + \delta x) - \nabla_x M_i(x)| < |\nabla_x M_i(x)|$, (M is tangent linear of model)

$|\nabla_x H_i(x + \delta x) - \nabla_x H_i(x)| < |\nabla_x H_i(x)|$, (H is linearised observation operator)

* Also, for sEKF to work, the state must be “linearly” constrained - that is, constrained to a degree when linearised operators can be applied within the limits or the characteristic uncertainty range.

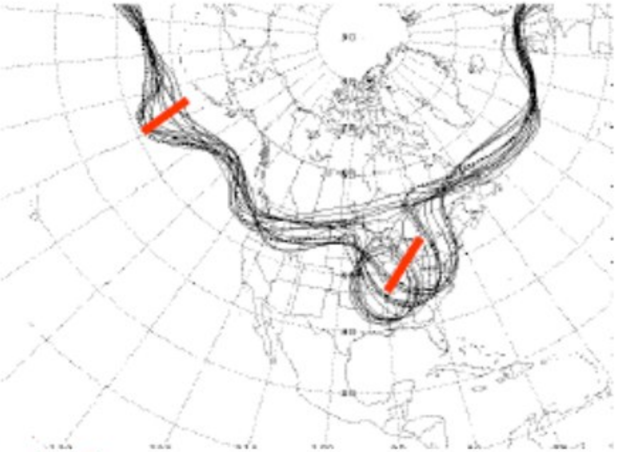
Cost of simplification: wrong representation of physics. There is almost no response of $T2m^t$ to T_s^0 (due to short memory), but strong response of $T2m^t$ to T_s^t.

Source: Ekaterina Kurzeneva

Funded by the European Union

Ensemble Kalman Filter (EnKF)

$$\mathbf{P}_i^b = \mathbf{L}_{i-1}\mathbf{P}_{i-1}^a\mathbf{L}_{i-1}{}^T + \mathbf{Q} \quad *$$



Physically,

▪“errors of day” are the instabilities of the background flow. Strong instabilities have a few dominant shapes (perturbations lie in a low-dimensional subspace).

▪ It makes sense to assume that large errors are in similarly low-dimensional spaces that can be represented by a low order EnKF.

❖ Although the dimension of $\mathbf{P}_i^f$ is huge, the rank ($\mathbf{P}_i^f$) << n (dominated by the errors of the day)

$$\mathbf{P}_i^b \approx \frac{1}{m}\sum_{k=1}^{m}(x_k^f - x^t)(x_k^f - x^t)^T$$

Ideally $\quad m \to \infty$

❖ Using ensemble method to estimate *

$$\boxed{\begin{aligned}\mathbf{P}_i^b &\approx \frac{1}{K-1}\sum_{k=1}^{K}(x_k^f - \overline{x^f})(x_k^f - \overline{x^f})^T \\ &= \frac{1}{K-1}\mathbf{X}^b \bullet \mathbf{X}^{bT}\end{aligned}}$$

K ensemble members, K<<n

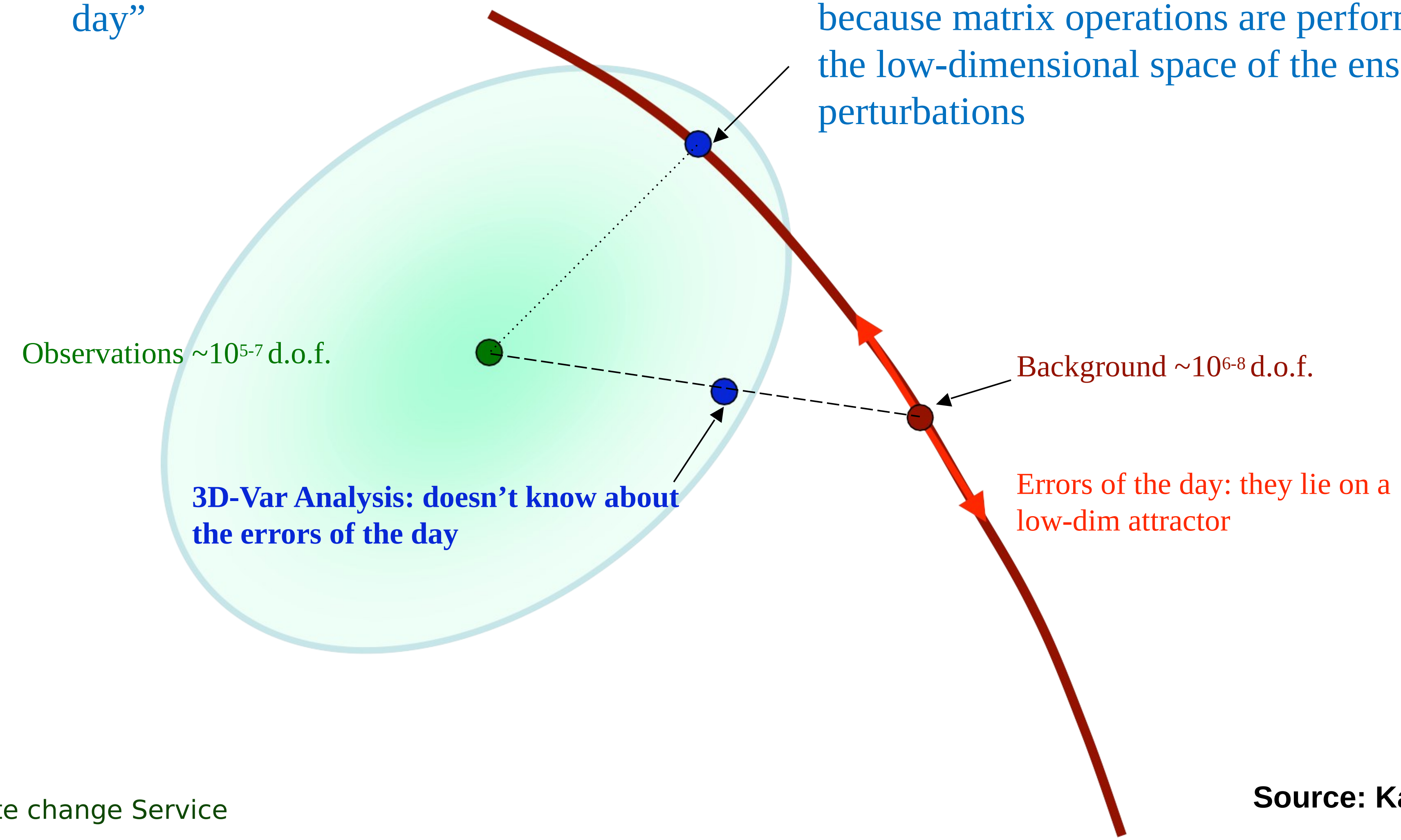
❖ Problem left: How to update ensemble ? i.e.: How to get $\mathbf{x}_i^a$ for each ensemble member?

Source: Kalnay

Funded by the European Union

With **Ensemble Kalman Filter** based LDAS the perturbations pointing to the directions of the “errors of the day”

Ensemble Kalman Filtering is efficient because matrix operations are performed in the low-dimensional space of the ensemble perturbations



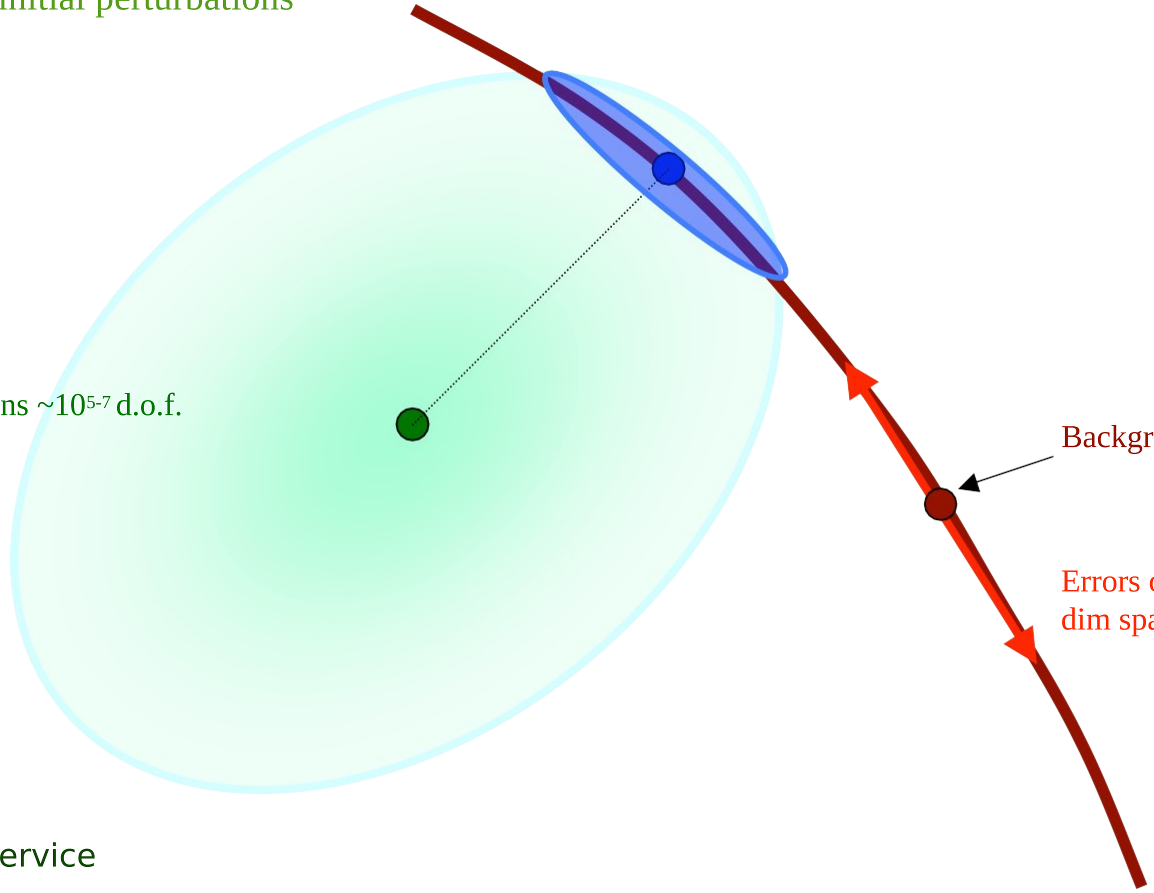
Source: Kalnay

After the EnKF computes the analysis and the analysis error covariance **A**, the new ensemble initial perturbations are computed

$$\sum_{i=1}^{k+1} \delta \mathbf{a}_i \delta \mathbf{a}_i^T = \mathbf{A}$$

initial perturbations

These perturbations represent the analysis error covariance and are used as initial perturbations for the next ensemble forecast



Source: Kalnay

Funded by the European Union

Ensemble Square root filter approach

- Observations are assimilated to update only the ensemble mean.

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

- Assume analysis ensemble perturbations can be formed by transforming the forecast ensemble perturbations through a transform matrix

$$\mathbf{x}_i^b = M\mathbf{x}_{i-1}^a$$

$$\mathbf{P}_i^b \approx \frac{1}{K-1}\sum_{k=1}^{K}(x_k^b - \overline{x^b})(x_k^b - \overline{x^b})^T$$

$$\mathbf{K}_i = \mathbf{P}_i^b\mathbf{H}^T[\mathbf{H}\mathbf{P}_i^b\mathbf{H}^T + \mathbf{R}]^{-1}$$

$$\boxed{\overline{\mathbf{x}_i^a} = \overline{\mathbf{x}_i^b} + \mathbf{K}_i(\mathbf{y}_i^o - H\overline{\mathbf{x}_i^b})}$$

$$\mathbf{X}_i^a = \mathbf{T}_i\mathbf{X}_i^b$$

$$\mathbf{x}_i^a = \overline{\mathbf{x}_i^a} + \mathbf{X}^a$$

$$\frac{1}{k-1}\mathbf{X}^a\mathbf{X}^{aT} = \mathbf{P}^a_{i\ n\times n} = [\mathbf{I} - \mathbf{K}_i\mathbf{H}]\mathbf{P}_i^b = [\mathbf{I} - \mathbf{K}_i\mathbf{H}]\frac{1}{k-1}\mathbf{X}^b\mathbf{X}^{bT} \Rightarrow \mathbf{X}_i^a = \mathbf{T}_i\mathbf{X}_i^b$$

Source: Kalnay

Advantages of EnKF vs sEKF

- Some major problems associated with using the EKF in connection with (larger) nonlinear models:
  - Inaccuracy in the evolution of the model error covariance matrix and huge computational requirements associated with the storage and forward integration of this matrix
  - Use of the central forecast as the estimate of the state. For non-linear dynamics the central forecast is not equal to the mean or expected value

- The EnKF was designed to resolve the points above, and has gained in popularity due to its simple conceptual framework and relative ease of implementation
  - No derivation of a tangent linear operator
  - Model error covariance implicitly defined through maintaining a set of model states in the form of an ensemble
  - The mean of the ensemble representing the estimated state
- In EnKFs each ensemble member is run forward in time through the model
- Uncertainty (or spread in the ensemble) is introduced by stochastic model dynamics (stochastic physics) when integrating each ensemble member forward in time
- In the EKF uncertainty in the estimated state is introduced in the update of the B-matrix (background error covariance) and in the added Q-matrix
- However, both algorithms are optimal and correct only when the underlying PDFs (prior and posterior to the observations) are Gaussian

Source: Sam-Erik Walker (NILU)

Harmonie-Arome Model Configuration Used in the Study

Code: https://github.com/josteinblyverket/Harmonie/tree/EnKF_CY46h1; **Multi-layer physics:** ISBA-DIF , 3-L for Snow scheme , Soil heat capacity = 2.0E-5

Upper Air: 3DVAR; Surface analysis: (a) ENSRKF LDA and (b) sEKF for Land Data Assimilation
Experiment: cold start at (a) 2023-10-01, 3hr cycling for 4 weeks. (b) 2023-06-01, 3hr cycling for 8 weeks; Local settings like upper air DA common to both the LDAS runs.

Multi-Layer surface physics

Force-restore

- **ISBA-3L** 3 layer soil (top, root, deep)
- **D95** bulk snow scheme
- **OI** surface analysis

Multi-layer physics

- **ISBA-DIF** 14 layer soil (0.01m, ..., 12m)
- **MEB** Multi Energy Balance for vegetation
- **SEKF** Simplified Extended Kalman Filter for surface analysis (constant **B**)
- Ensemble Square Root Kalman Filter for surface analysis (for Soil Moisture)
- LETKF Filter for surface analysis (For Soil Mositure)

EnKF experiment settings:

- T2m and RH2m gridded observations (gridPP)
- **Observation error: 1K and 0.4**
- 16 ensemble members
- **One control run (using the deterministic forecast)**
- **Control variables**: soil temperature layer 1 to 2 and soil moisture layer 2 to 6 & 7.
- 3 hour cycles

MOTIVATION

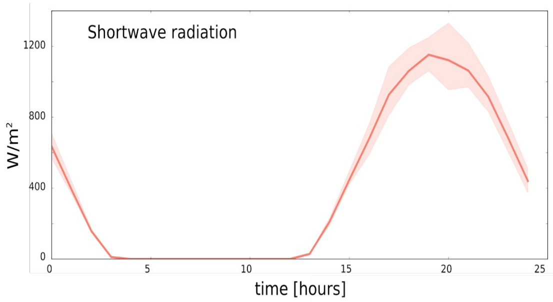
Task 1.2 (Lead - SMHI): Develop ensemble-based filter LDAS approaches for soil moisture (M3-18)

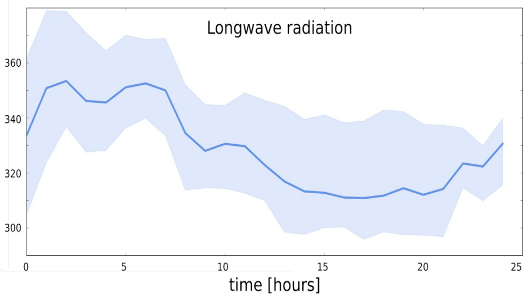
Funded by the European Union

Funded by the European Union

Results

1) Perturbations are applied following Charrois et al 2016, Blyverket et al. (2019); Fields are perturbed using the spatial-temporal perturbation methodology are precipitation, shortwave downward radiation and long wave downward radiation (additive). For surface, soil moisture perturbations are multiplicative while the soil temperature perturbations are additive.





Cross-correlated AR(1) perturbations:

- Multiplicative: shortwave radiation and precipitation
- Additive: Longwave radiation

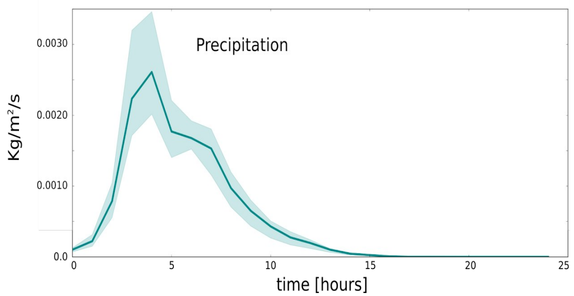


Figure: Ensemble mean (solid line) and spread (shaded region) for forcing of atmospheric variables. (Source: Jostein Blyverket (MET-Norway)

2) Ensemble Kalman filter based land data assimilation system for Harmonie-Arome system tested for three domains : SOR_TEST (smaller), METCOOP25D (bigger) and NORD_2.5km (intermediate domain). As of now only the SYNOP observations are assimilated. Experiments are run for to test the impact of domains and initial conditions on the growth of perturbation of the land surface variables (TG1, TG7, TG14, WG1, WG7 and WG14) and land surface fluxes (LHF, SHF) .

Funded by the European Union

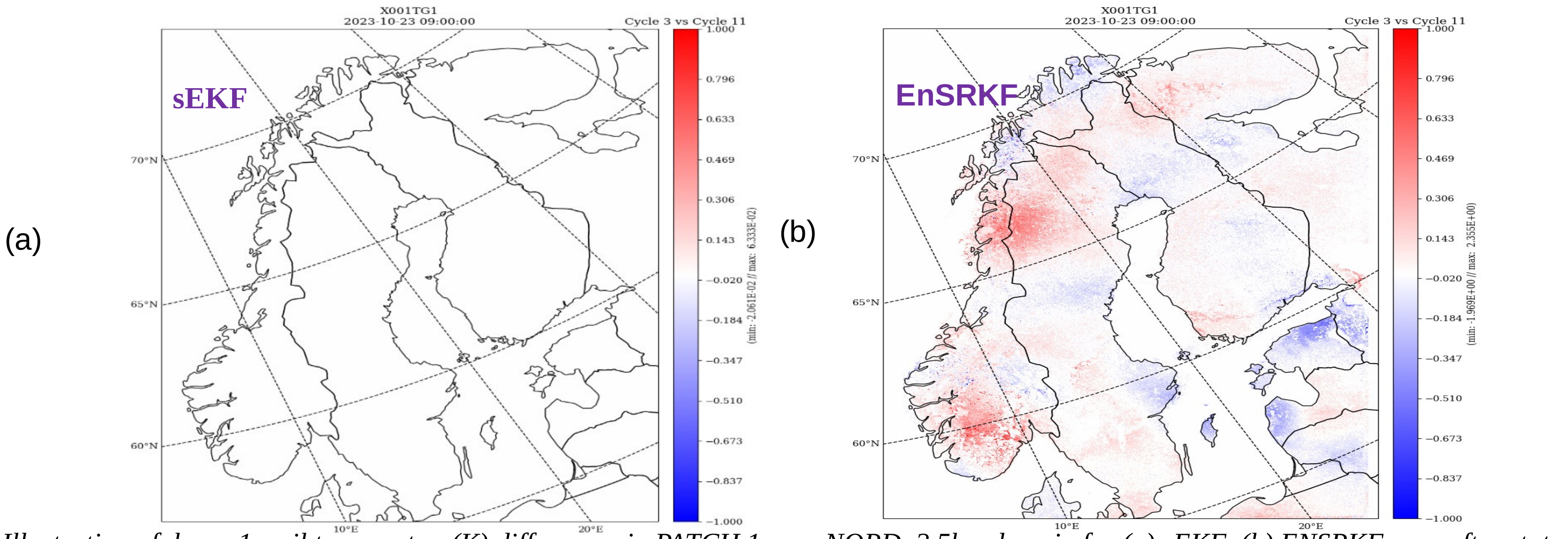


Figure : *Illustration of layer 1 soil temperature(K) differences in PATCH 1 over NORD_2.5km domain for (a) sEKF (b) ENSRKF runs after state surface perturbations*

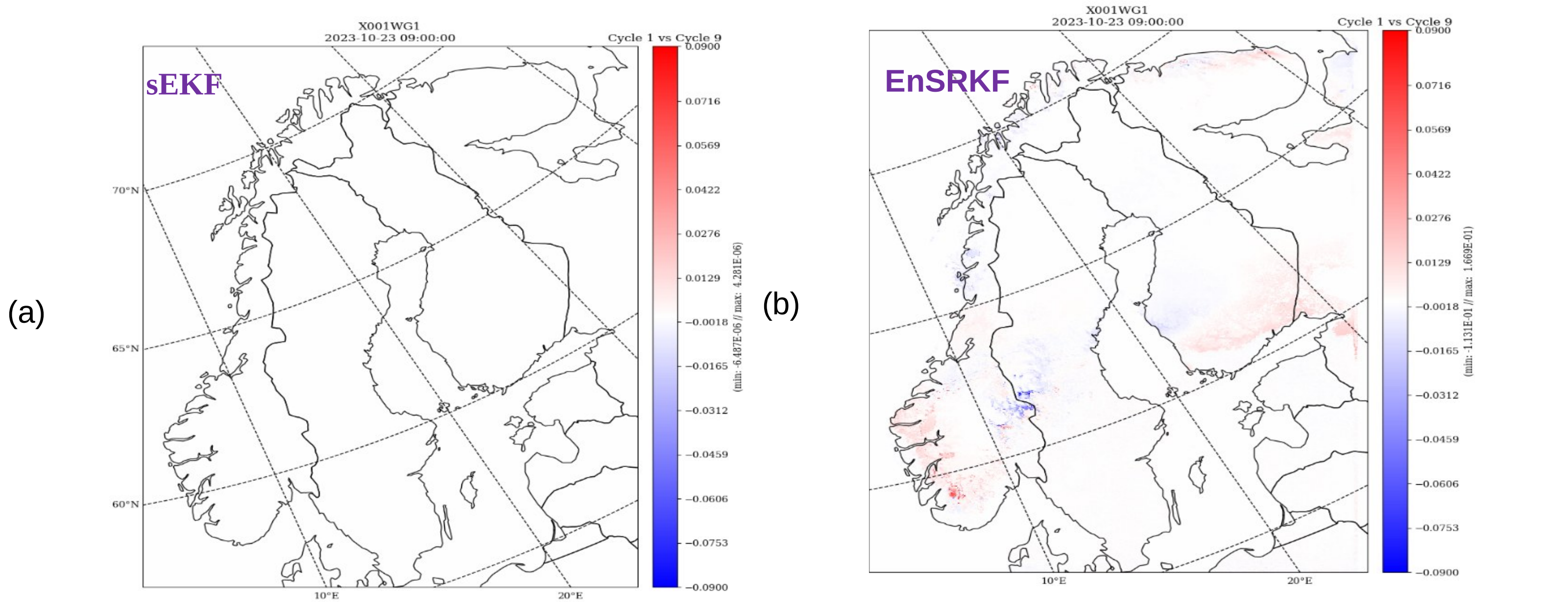


*Figure : Illustration of layer 1 soil moisture differences (kg m**-3) in PATCH 1 over NORD_2.5km domain for (a) sEKF (b) ENSRKF runs after state surface perturbations*

Funded by the European Union

(a) (b)

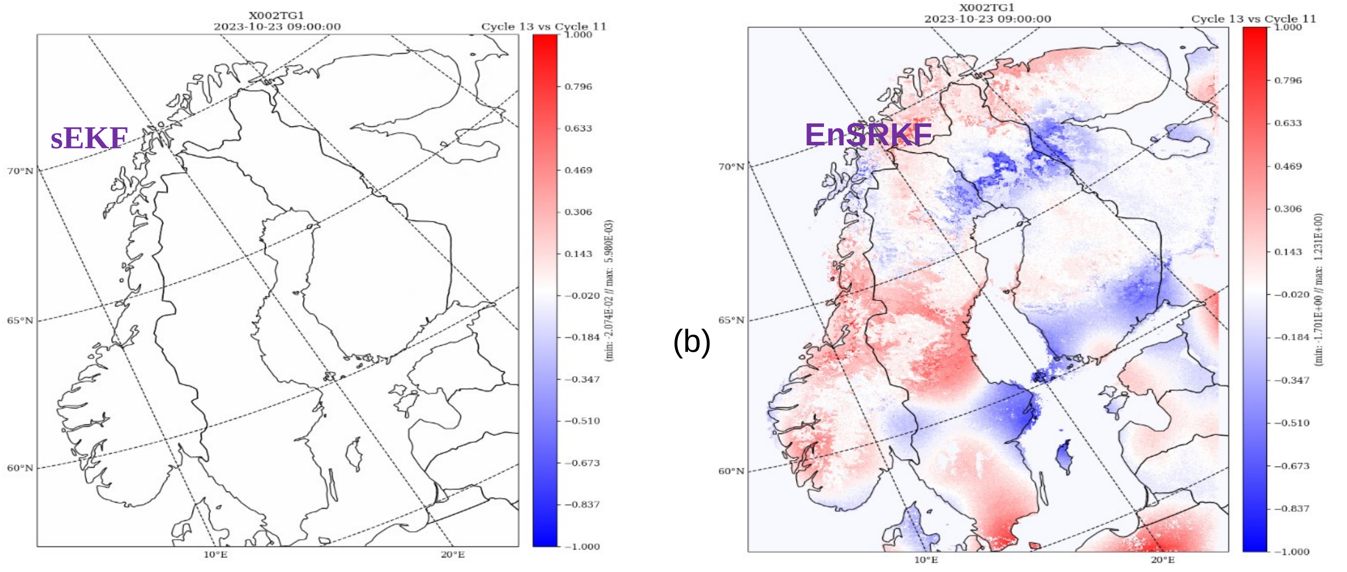


Figure : *Illustration of layer 1 differences in soil temperature (K) over PATCH2 over the NORD_2.5km domain for (a) sEKF (b) ENSRKF , after state surface perturbations*

(a) (b)

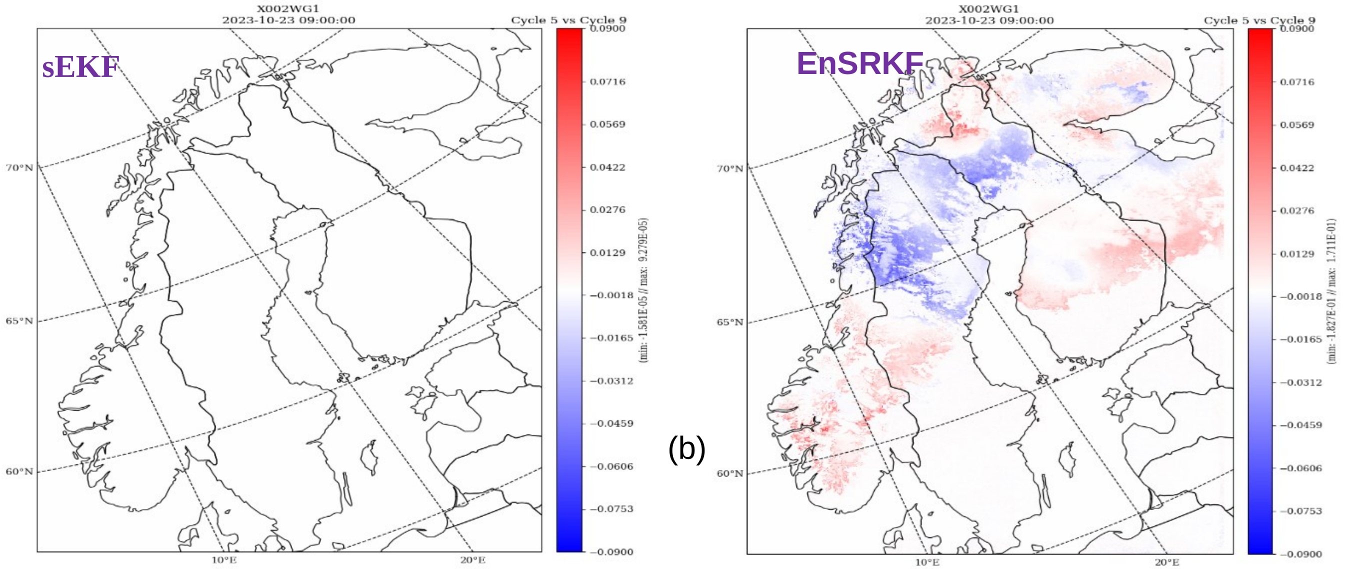


Figure : *Illustration of layer 1 differences in soil moisture (kg m**-3) over PATCH2 over the NORD_2.5km domain for (a) sEKF (b) ENSRKF, after state surface perturbations.*

Funded by the European Union

(a)

(b)

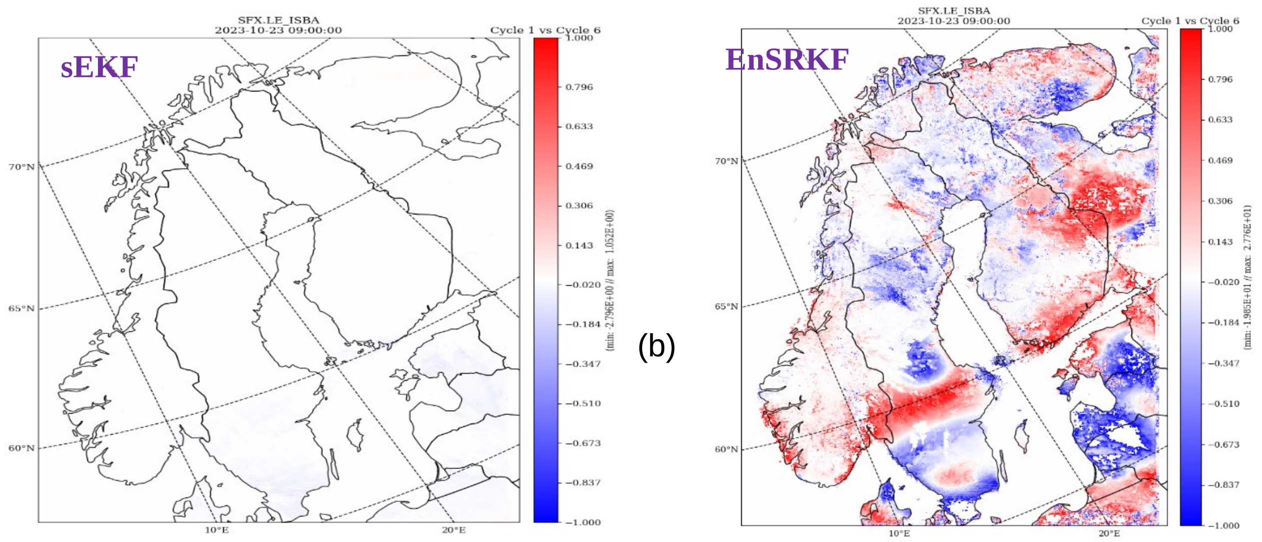


Figure : *Illustration of differences in perturbations of latent heat flux (W/m2) over the NORD_2.5km domain for (a) EKF (b) ENSRKF, after state surface perturbations.*

(a)

(b)

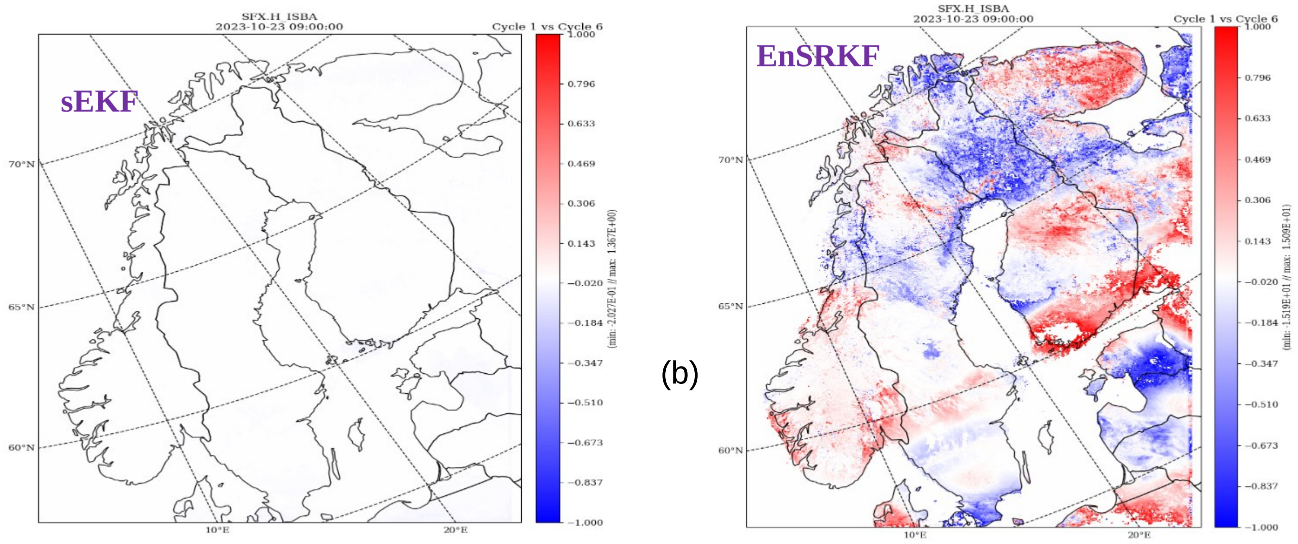


Figure : *Illustration of differences in perturbations of sensible heat flux (W/m2) over the NORD_2.5km domain for (a) EKF (b) ENSRKF, after state surface perturbations.*

Funded by the European Union

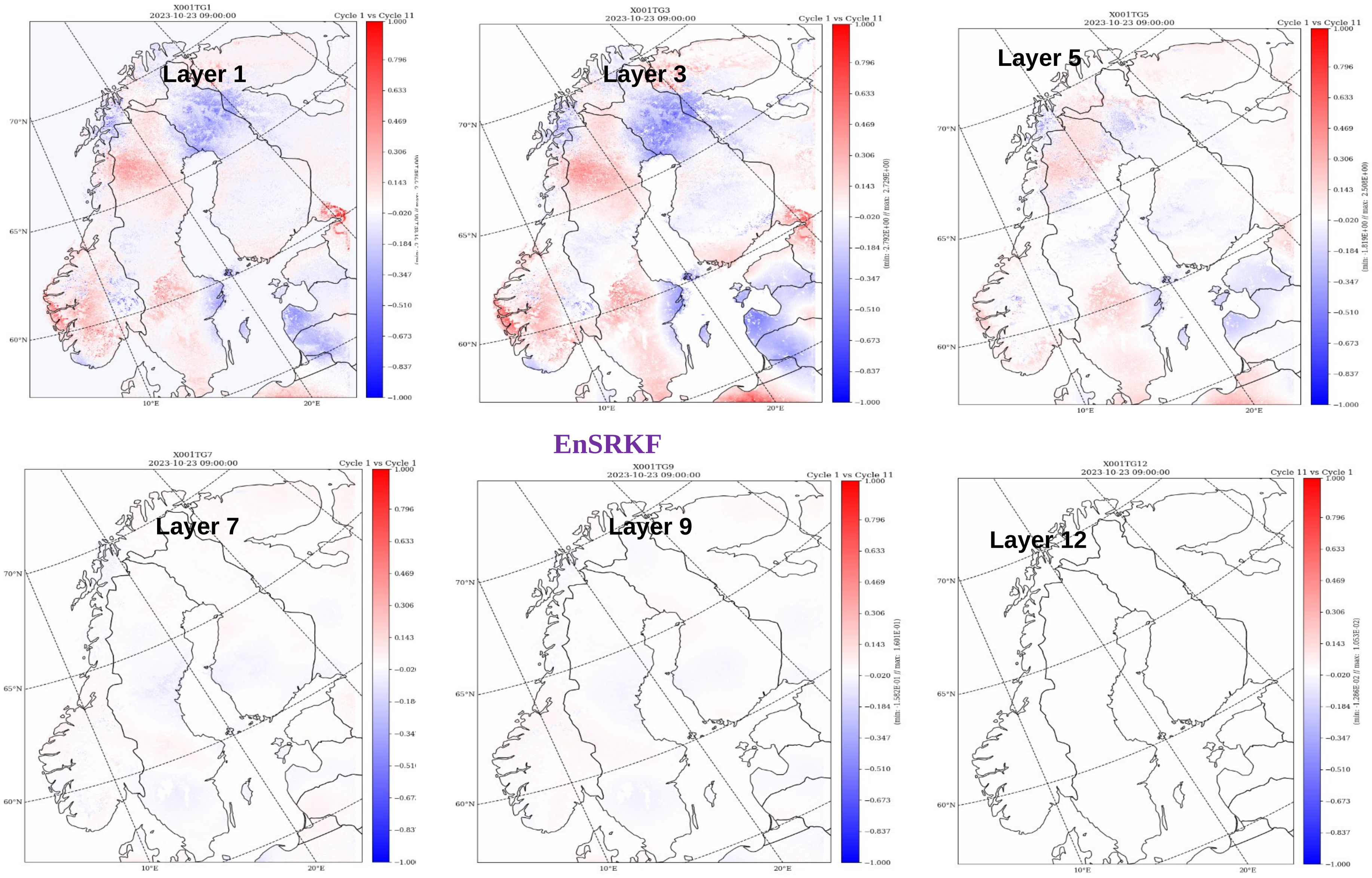


Figure: Illustration of spread of soil temperature differences (K) for PATCH 1 (EnSRKF), in layer 1 to layer 12 after state surface perturbations over NORD_2.5km domain valid at day-23 of the run from 1st October'23.

Funded by the European Union

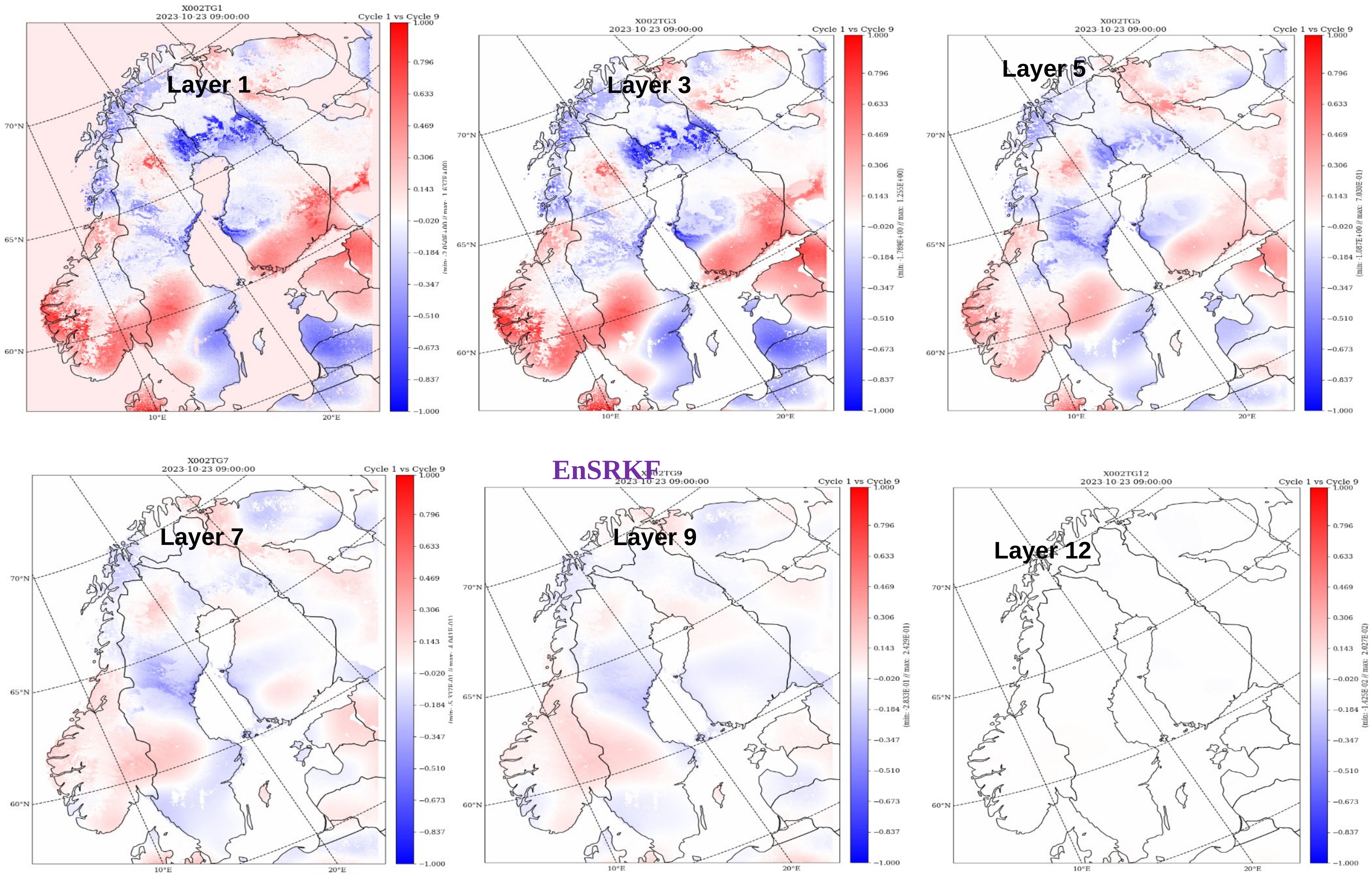


Figure: Illustration of spread of soil temperature differences (K) over PATCH 2 (EnSRKF), in layer 1 to layer 12 after state surface perturbations over NORD_2.5km domain valid at day-23 of the run from 1st October’23.

Funded by the European Union

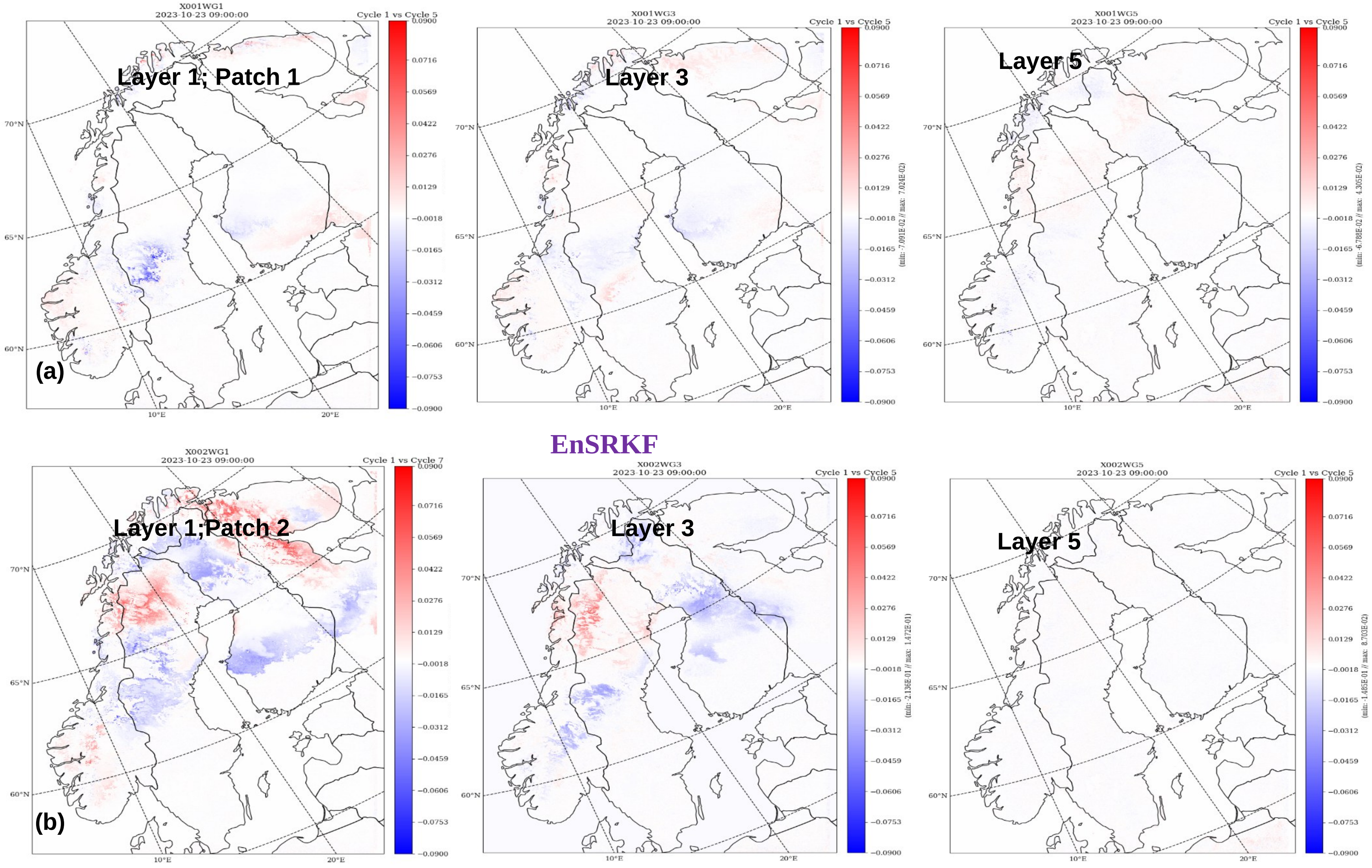


*Figure: Illustration of spread of soil moisture differences (kg m**-3) for (a) PATCH 1 (b) PATCH 2, in layer 1 to layer 5, after state surface perturbations in EnSRKF based LDAS (over NORD_2.5km domain, valid at day-23 of the run from 1st October'23.*

Funded by the
European Union

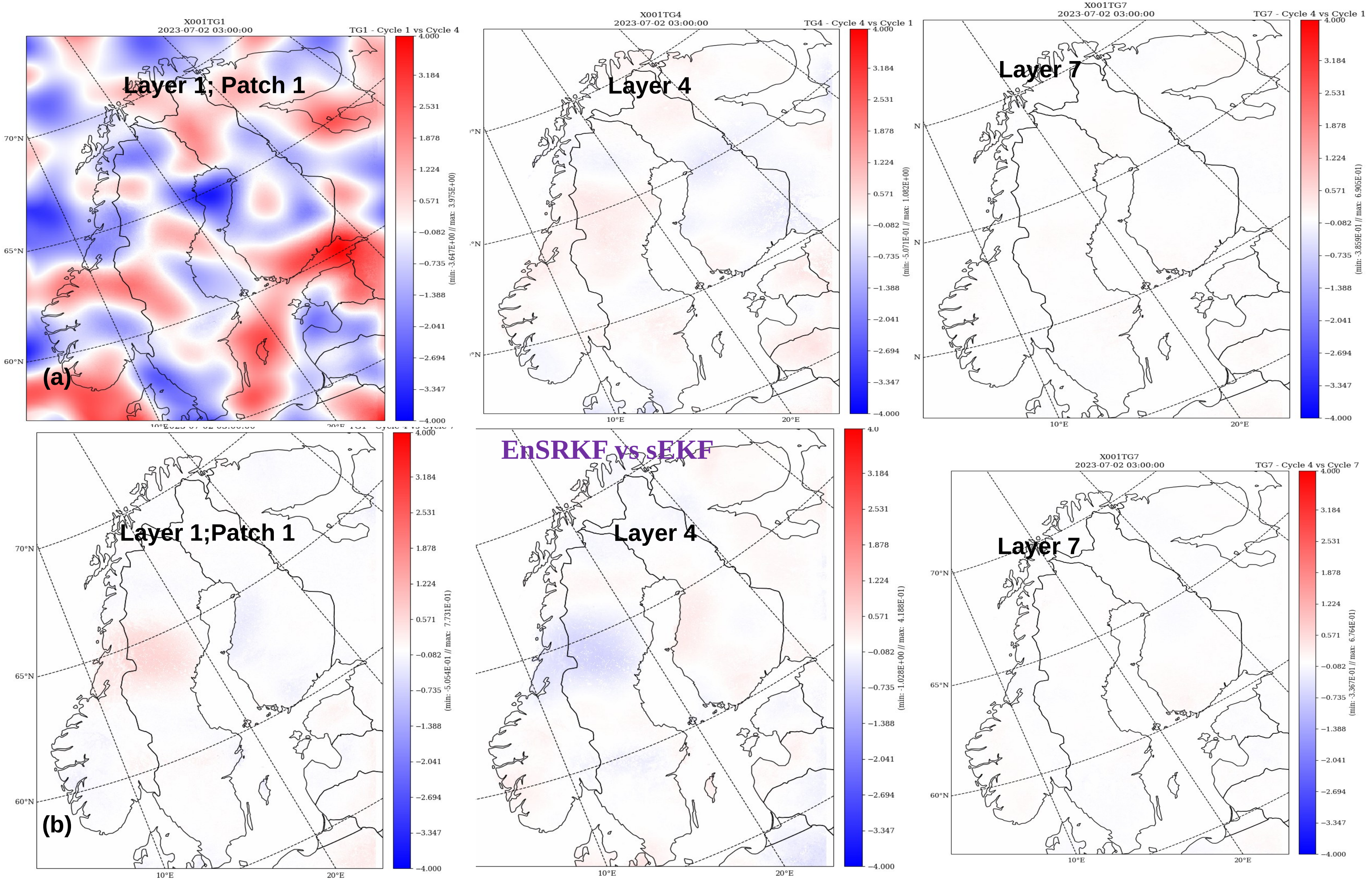
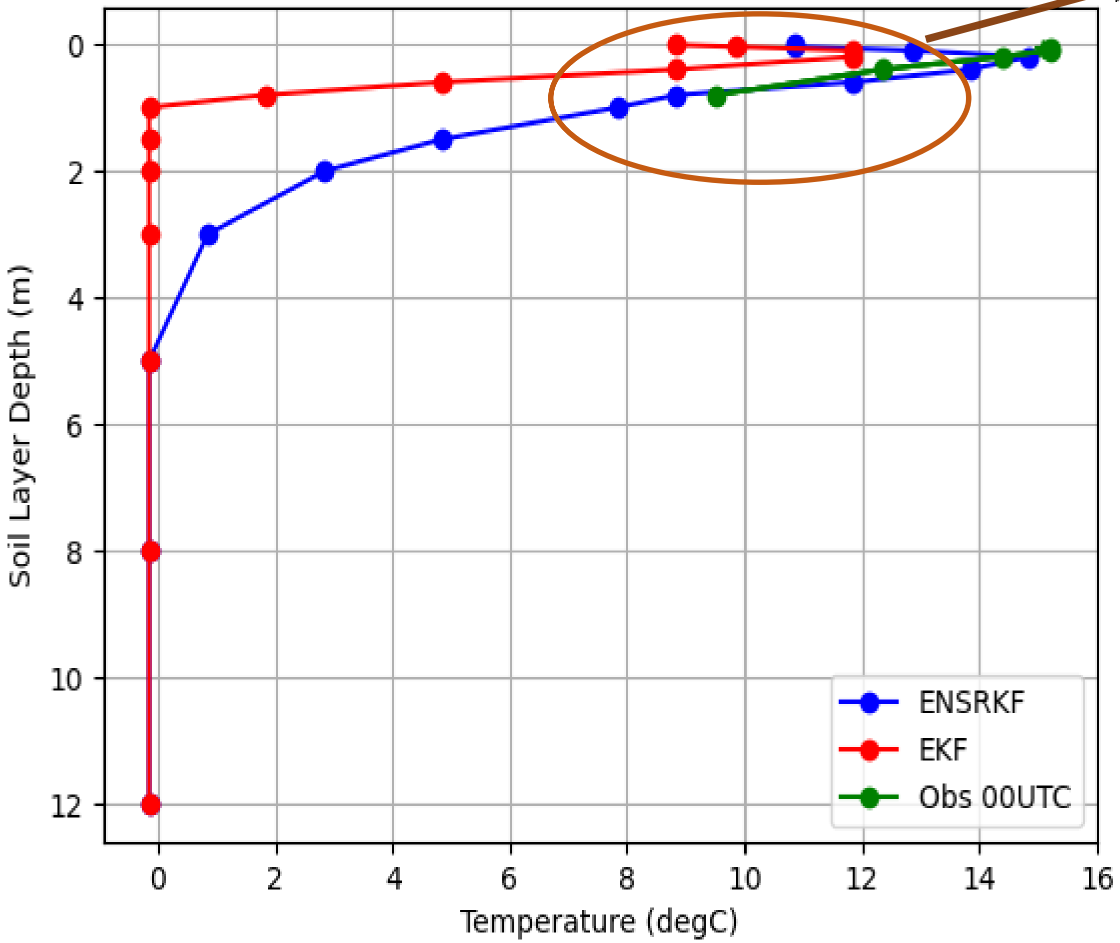
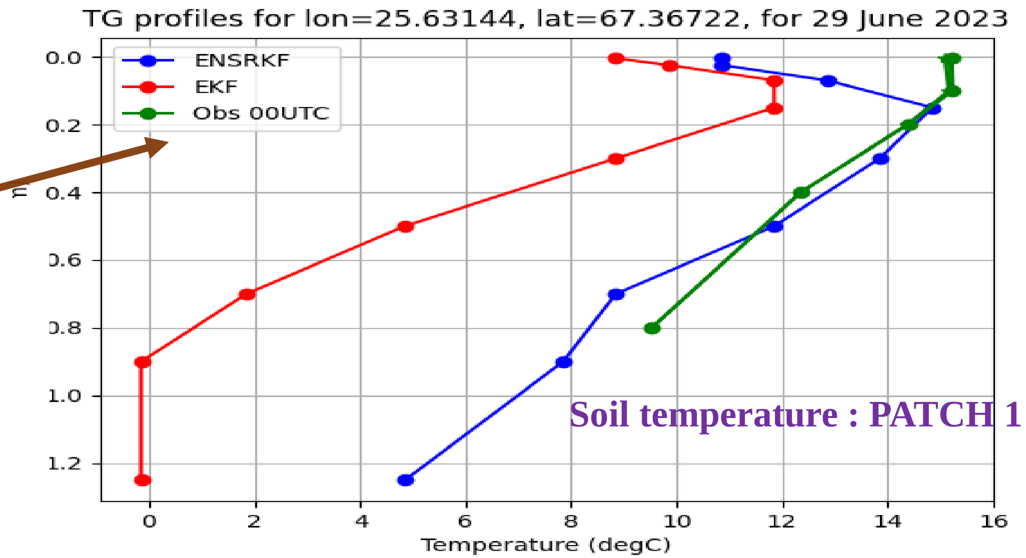


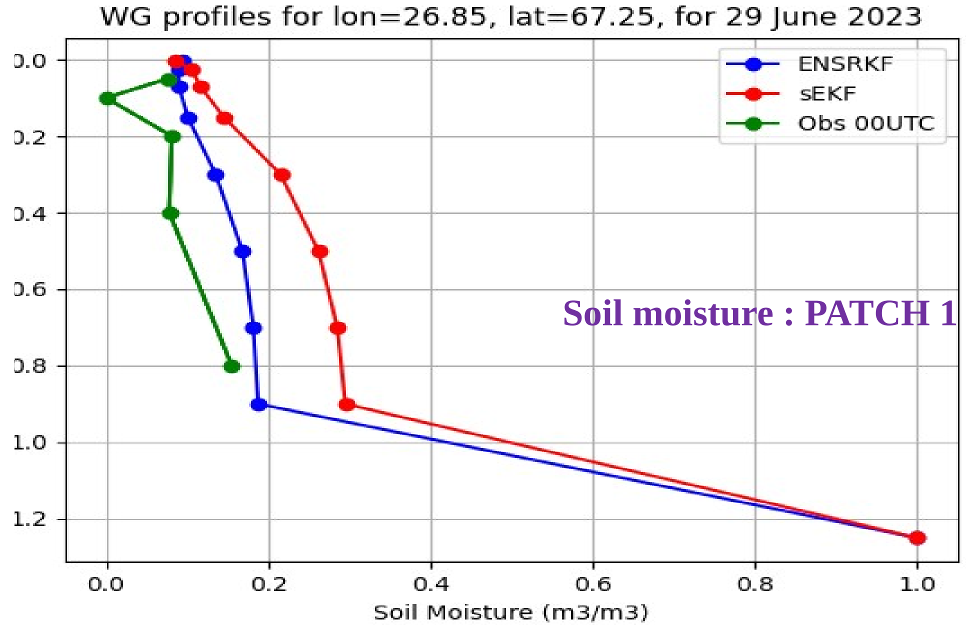
Figure: Illustration of spread of soil temperature differences (K) for PATCH 1, in layer 1, layer 4 and layer 7, after state surface perturbations in (a) EnSRKF (b) sEKF based LDAS (over NORD_2.5km domain, valid at day-32 of the run from 1st June'23.

Funded by the
European Union

Vertical Soil temperature and soil moisture profile at SODANKYLÄ (Station: LUO0009 (ST); LUO0009 (SM))

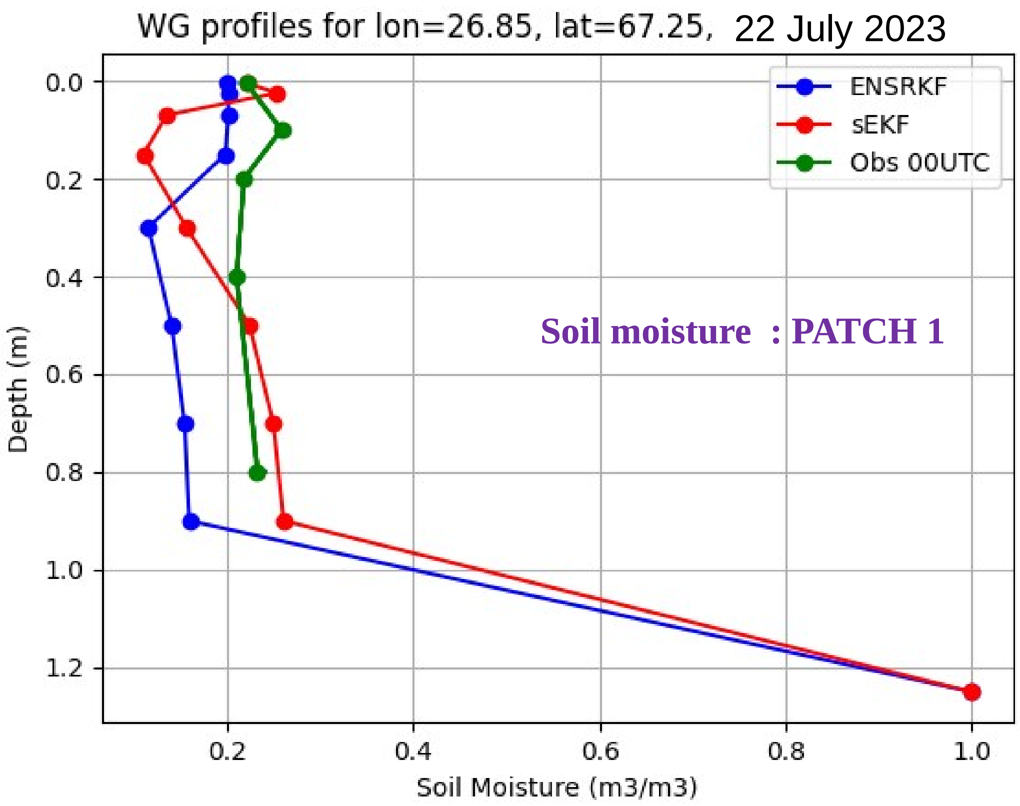


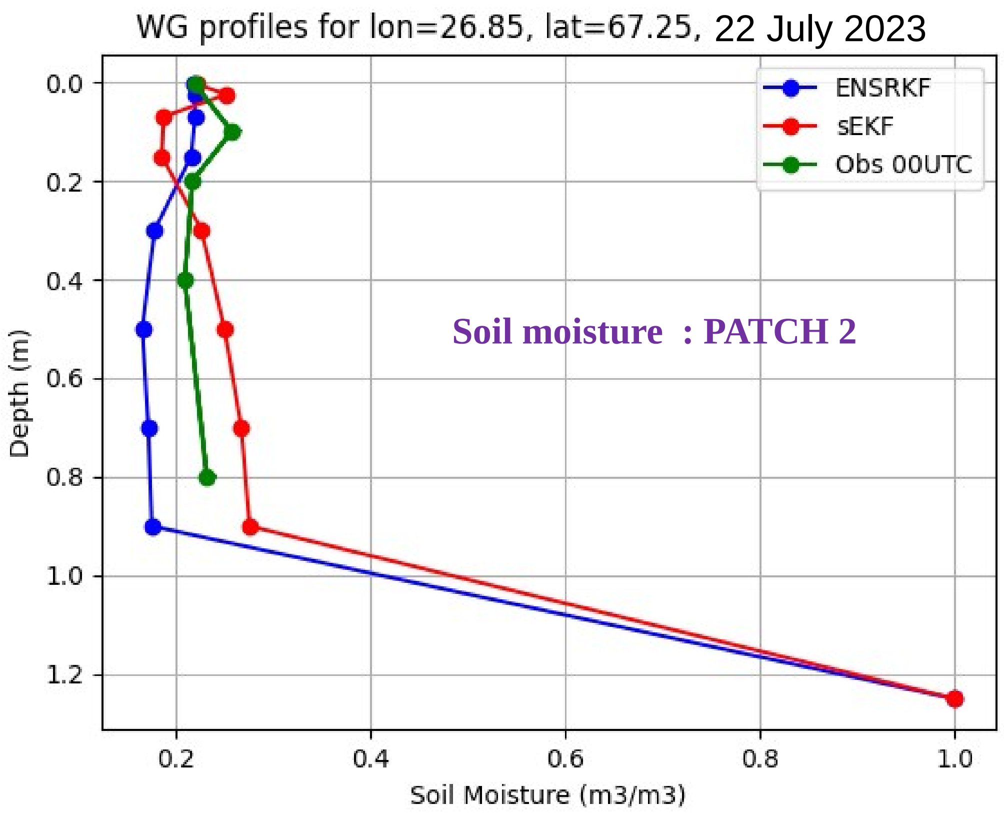




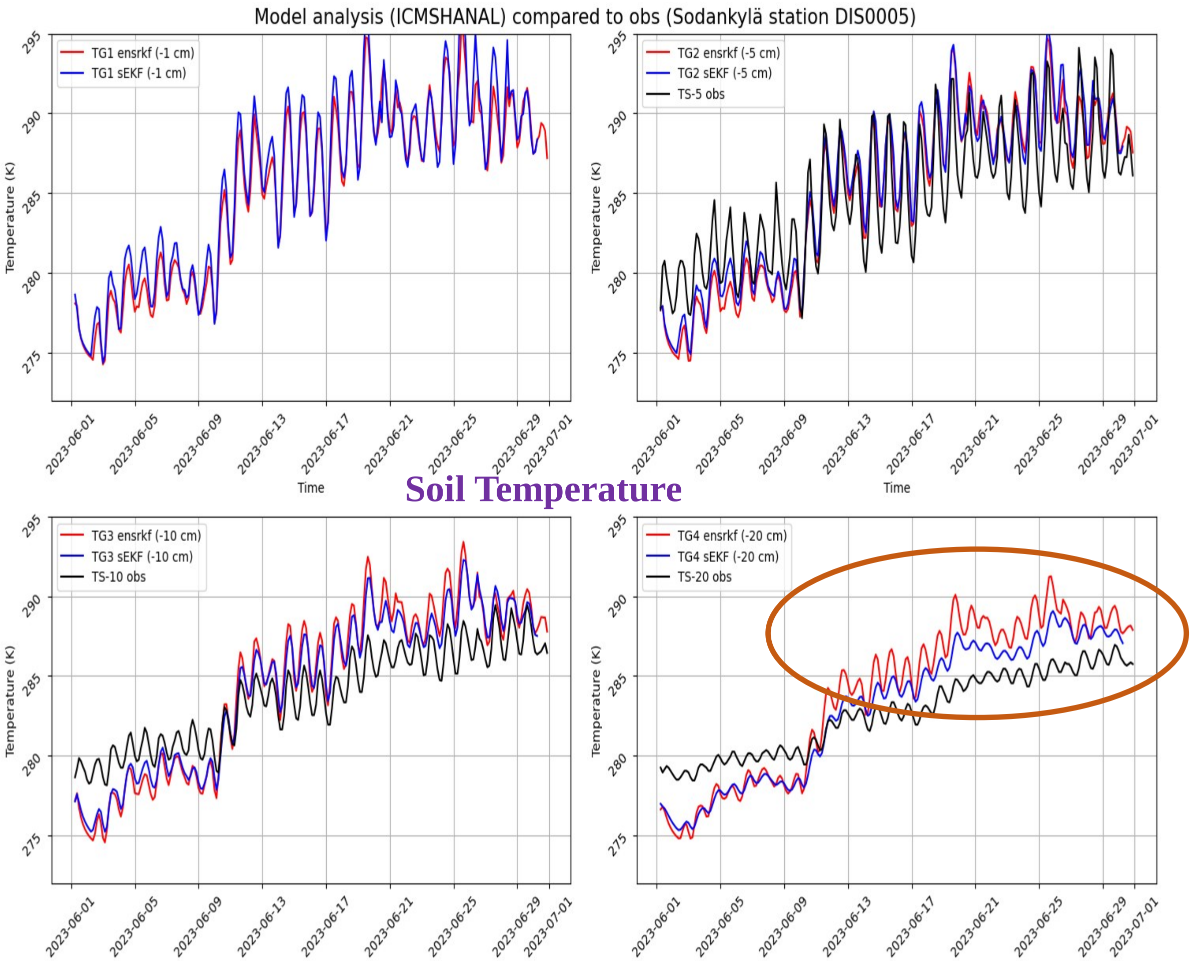
SODANKYLÄ (Station - DIS0005)

Experiment 2 : cold start at (a) 2023-06-01, 3hr cycling for 8 weeks; (a) ENSRKF LDA and (b) sEKF for Land Data Assimilation; but the length of control vectors increased to include WG6 and WG7

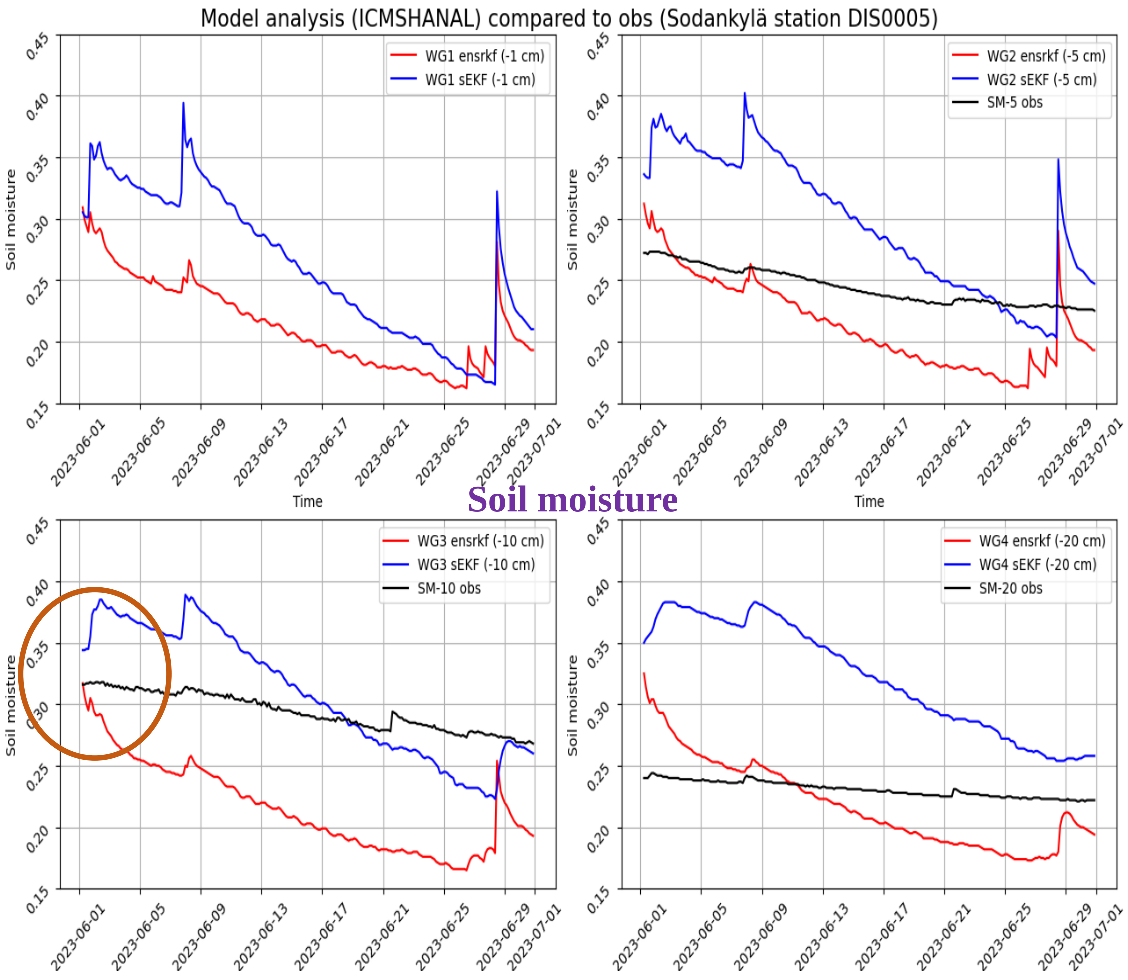




Funded by the European Union



Soil Temperature

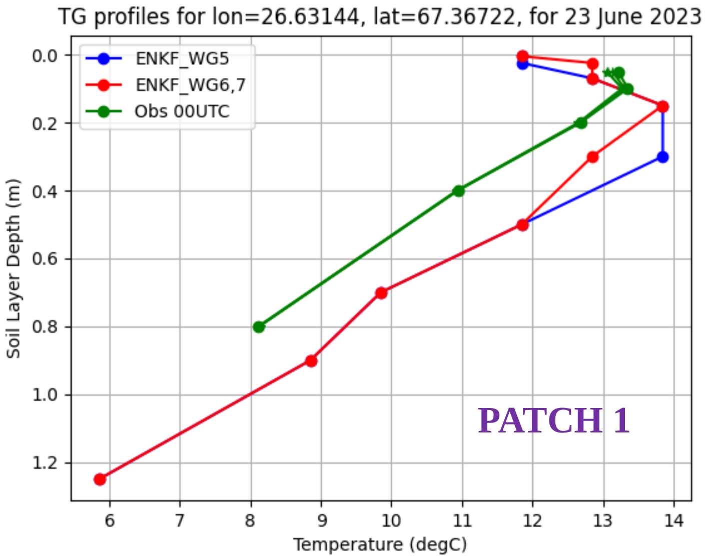


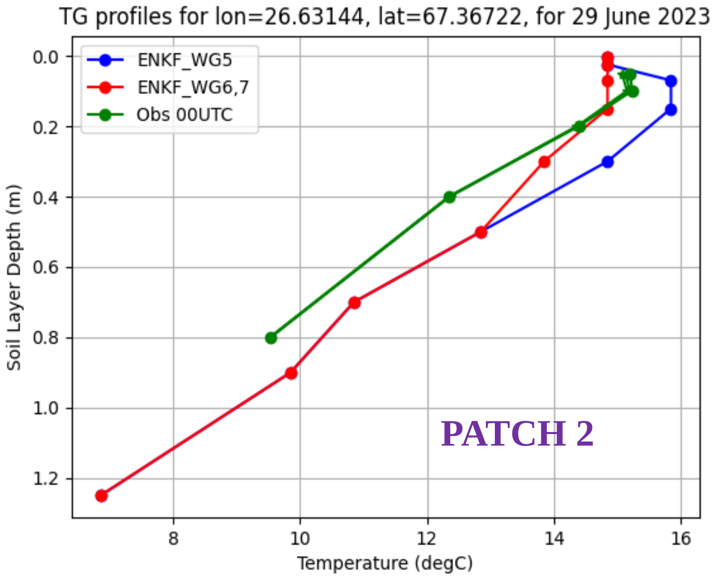
Soil moisture

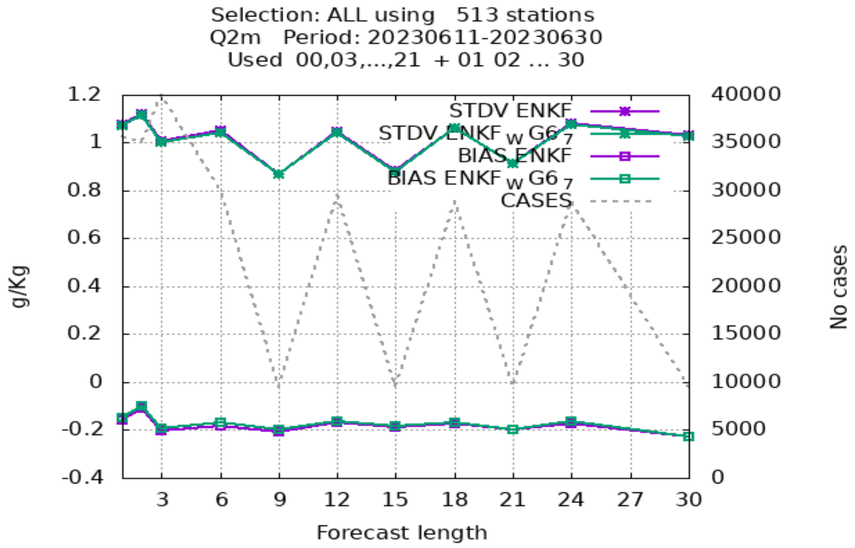
Thanks :Mikael Hasu

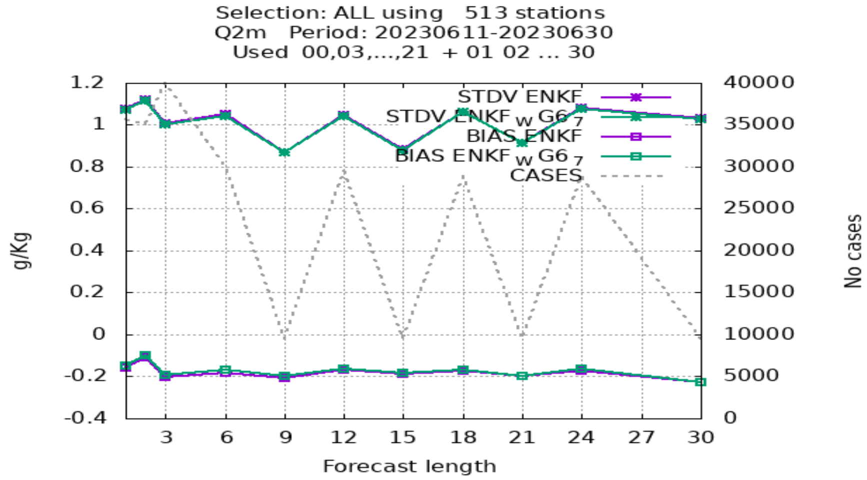
SODANKYLÄ (Station - DIS0005)

Experiment 3 : cold start at (a) 2023-06-01, 3hr cycling for 8 weeks; (a) ENSRKF (TG1,TG2, WG1 to 5) and (b) ENSRKF (TG1,TG2, WG1 to 7) for Land Data Assimilation



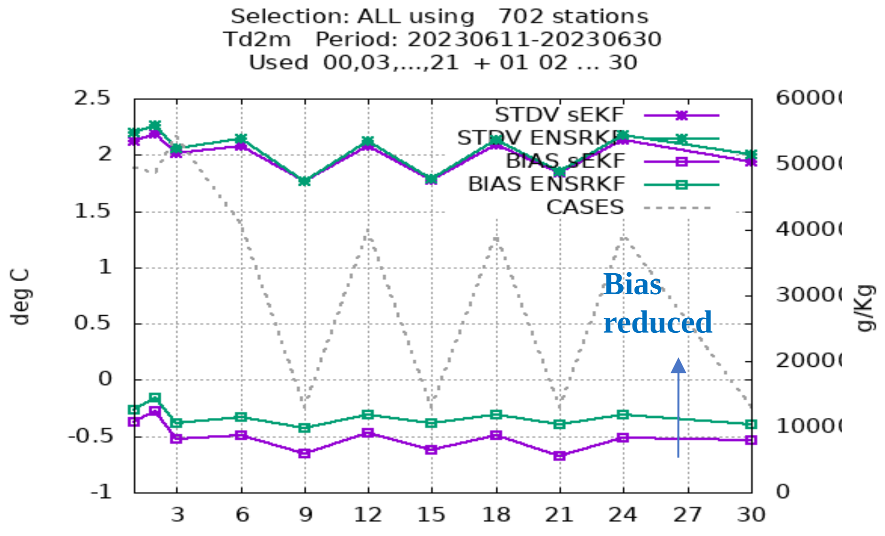


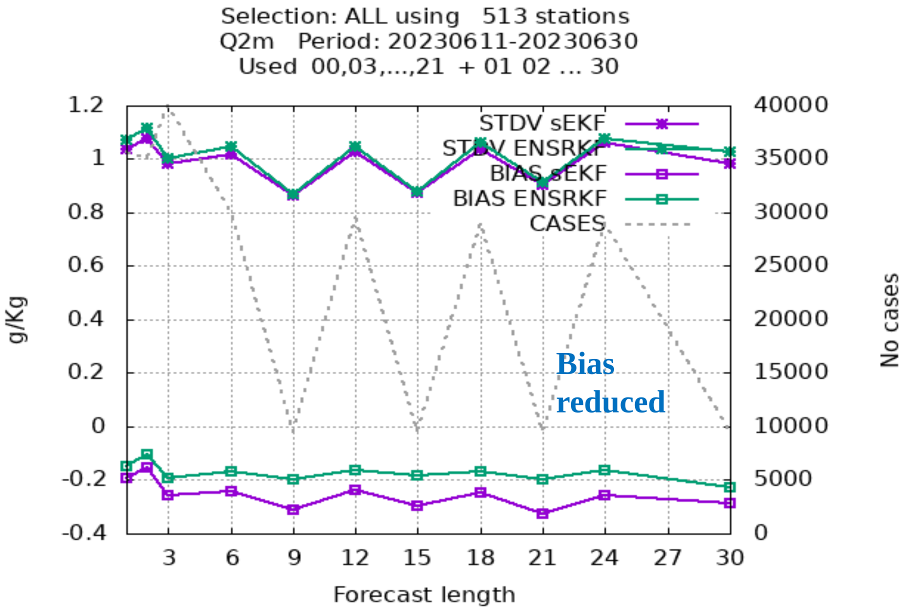


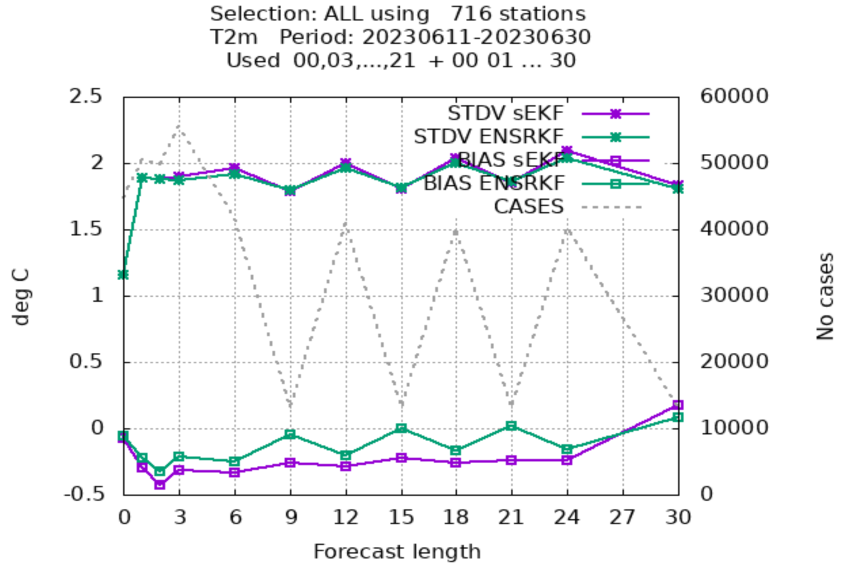


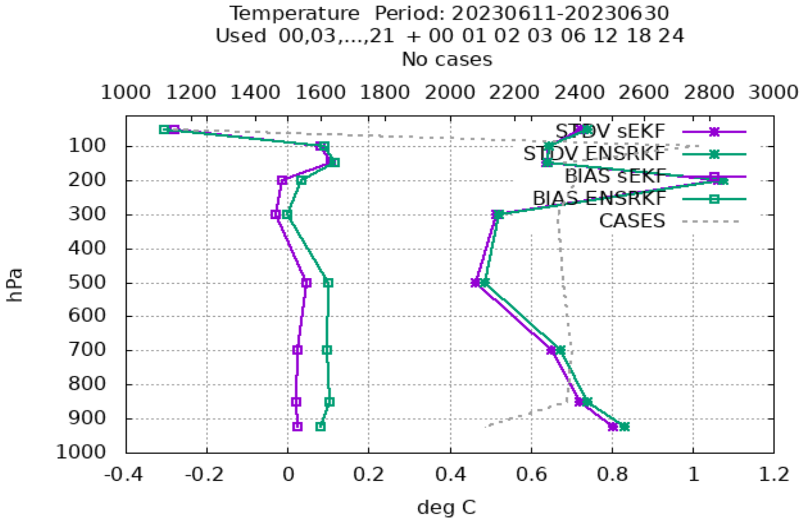
Funded by the European Union

June









Major improvements with ENSRKF LDAS in reducing the bias in forecasts of dew-point temperature, specific humidity and 2m-temperature

Funded by the European Union

Conclusions

1. An investigation of ENSRKF vs sEKF LDA setup are compared on their performance for surface forecasts over NORDIC domain, in the HARMONIE-AROME SODA environment (offline). ENSRKF adds value to growth in perturbations of soil variables and land surface fluxes reaching deeper soil layers, with improvements in forecasts of near surface variables. The LDA assimilation experiments are carried out using SCREEN level SYNOP observations.

2. Further refinement of control vector set-up shows that adding two more control vectors in soil temperature resolves the state of the near surface variables better. Results shows that what levels are to be updated depends on the surface characteristics and root-zone depth.

3. The statistical scores suggests that the improvements in forecasts of near surface varaibles of T2m, TD2m, Q2m and U10m with the ENSRKF based LDAS vs sEKF during the June-July 2023 period. Thus, ENSRKF gives better scores concerning reduction of the systematic errors.

4. ENSRKF scheme is intended to be further investigated and used for the coupled DA developments. The local ensemble Kalman filter (LETKF) to be further tested for sequential land data assimilation to improve the land surface model performance.

Thank you!

CopERnIcus climate change Service Evolution - CERISE

The CERISE project (grant agreement No 101082139) is funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the Commission. Neither the European Union nor the granting authority can be held responsible for them.