Funded by the European Union

Coordinated by ECMWF

ENSEMBLE VERSUS EXTENDED KALMAN FILTER BASED LAND DATA ASSIMILATION FOR THE SOIL DIFFUSION BASED ISBA MODEL

CopERnIcus climate change Service Evolution - CERISE

Abhishek Lodh (Ph.D.)

Postdoctoral Researcher

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

Presented at :

S2S Community Workshop
Toward Minimizing Early Model Biases and Errors in S2S Predictions,
Jun. 5 to Jun. 7, 2024 ; Boulder, CO, USA

With thanks to
Jelena Bojarova, Patrick Samuelsson, Magnus Lindskog, Ulf Andrae, Meto & Åsmund Bakketun, Jostein Blyverket, Trygve Aspelien

ECMWF · Norwegian Meteorological Institute · SMHI · Deutscher Wetterdienst (Wetter und Klima aus einer Hand) · cmcc · Barcelona Supercomputing Center (Centro Nacional de Supercomputación) · METEO FRANCE · The Danish Meteorological Institute · Estellus · IPMA · NILU · Met Office

Funded by the European Union

Outline of the talk

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

Funded by the European Union

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

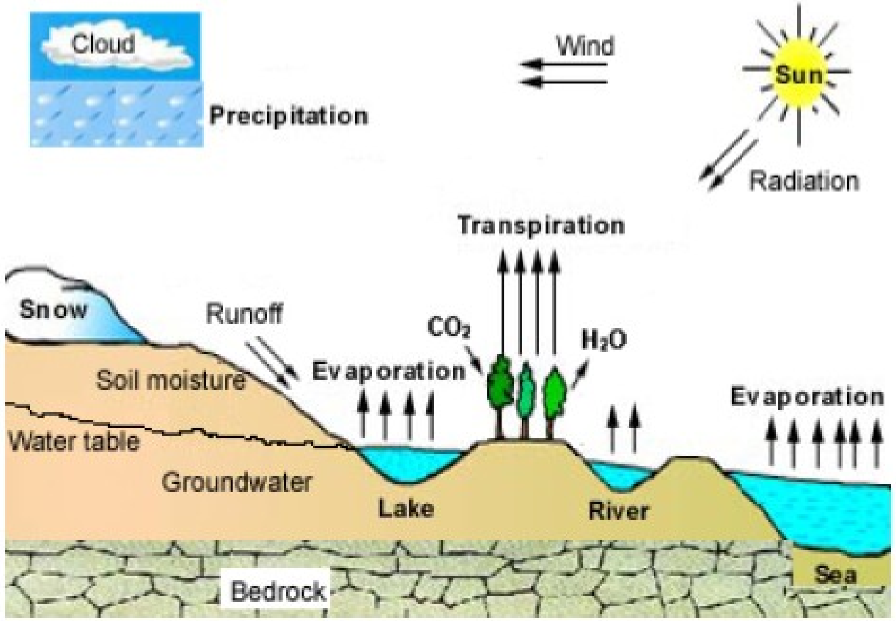
- Conducted Land Data Assimilation system (LDAS) simulations over the Nordic (NORD_2.5km) domain on ATOS (ECMWF) with cy46h+DIF+MEB+SEKF+3DVAR and ENSRKF+3DVAR for 23 days and more (simulations ongoing).

- 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.

Land surface-atmosphere interaction

Funded by the European Union



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, animals, human beings, their shelters, and much more.

Land surface processes, in principal, refer to the exchanges of heat, water, CO2, and other trace constituents among these components.

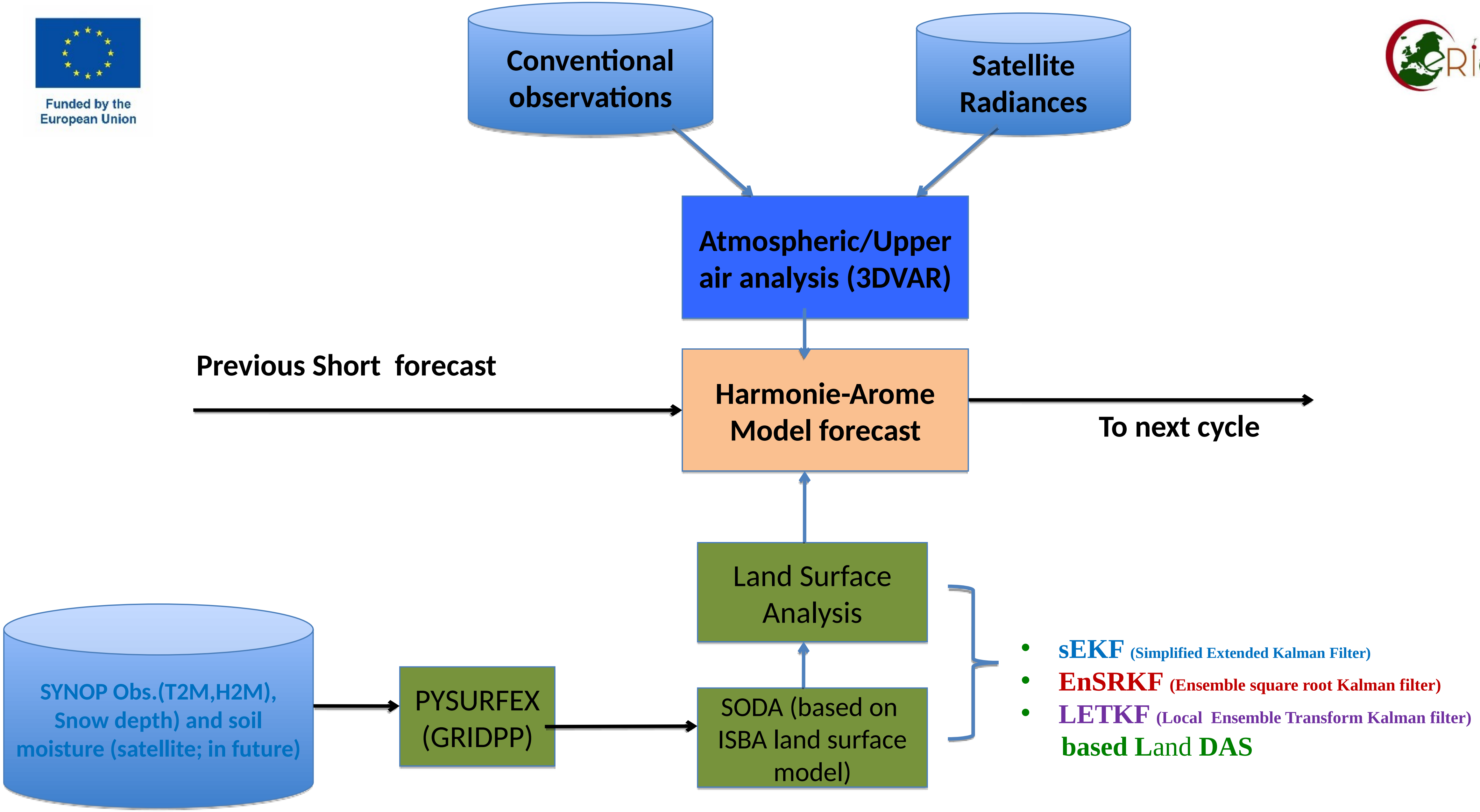
The surface variability not only determines the microclimate but also affects the mesoscale atmospheric circulation

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

Purpose of Land Data Assimilation

- Soil Moisture strongly influences the partitioning of available energy into sensible and latent heat flux and hence the evolution of the lower atmospheric conditions.
- Imperfect parameterisations of land surface and soil processes and failures in simulating precipitation and cloud cover can lead to considerable drifts of soil moisture; assimilation is needed to control forecast drifts.
- The use of in-situ land surface observations is unfeasible, because no extensive observation network exists.
- Conventional data, e.g. screen-level parameters (T2m and RH2m), and satellite data (eg. ASCAT), can be used to adjust soil moisture in an assimilation framework.
- Soil Moisture Observations **– In-situ** (limited) & **Satellite** (latest addition-SMAP, SMOS, ASCAT)
- In NWP – proper land surface state is required for initialize the model forecast (soil moisture, snow, soil temperature, LST controls the partitioning of the energy at soil-atmosphere interface)– Requirement of land surface Analyses



Basic Schematic of Land data Assimilation

Kalman Filter and its different flavors: Overview

Funded by the European Union

- **Kalman filter (Kalman, 1960) :** Propagation and update of state error covariance and mean for a linear stochastic system
- **Extended Kalman Filter (Smith et al., 1962) :** Propagation of state error covariance with linearised version of the model
- **Ensemble Kalman filter (Evensen, 1994; Burgers et al., 1998) :** Monte-Carlo approximation of state error covariance and its update; propagation of state error covariance and mean by ensemble integration
- **Ensemble square root filter (Anderson 2001; Bishop et al. 2001; Whitaker and Hamill 2002; also Pham 2001) :** Deterministic representation and update of state error covariance in ensemble form

- The Kalman Filter provides a recursive solution of the least squares minimization problem in the linear case.
- The Kalman Filter provides optimal solution for the current state of the system given past observations.
- The state of the DA system at any stage is given by (i) state estimate **x** and (ii) state error covariance estimate **P**.
- The assimilation cycle breaks into two stages: propagation and analysis.

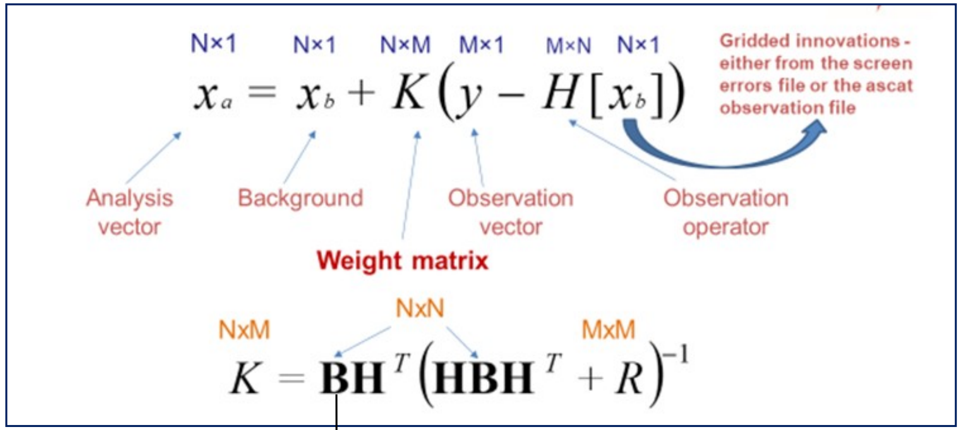
$$\mathbf{x}^f(t_i) = M_{i-1}\left[\mathbf{x}^a(t_{i-1})\right]$$

- The sensitivity of analysis to innovation = Kalman gain,

$$\mathbf{x}^a - \mathbf{x}^f = \mathbf{K}\,[\mathbf{y} - H(\mathbf{x}^f)].$$

Funded by the European Union

Extended Kalman filter (EKF)….



Background error covariance

- EKF Requires

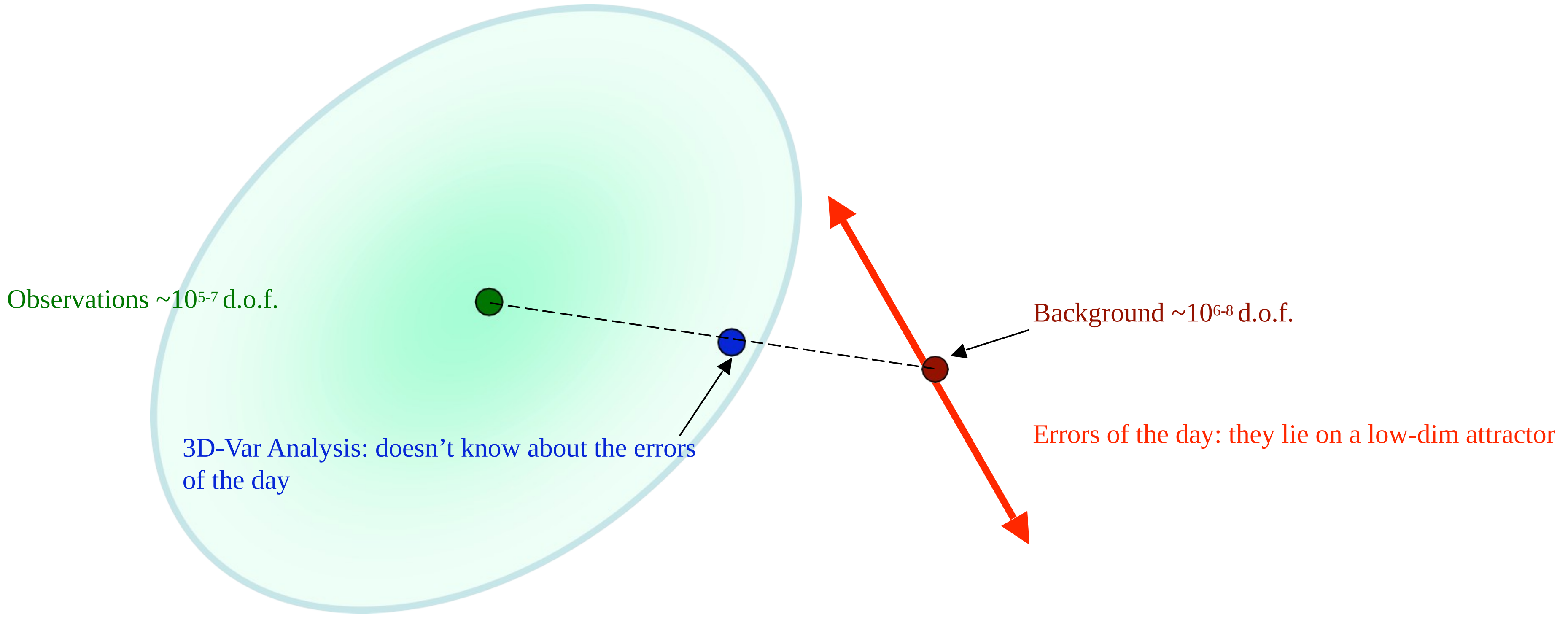
$|\nabla_x M_i(x + \delta x) - \nabla_x M_i(x)| < |\nabla_x M_i(x)|$, (M is tangent linear of model)

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

- Therefore, for EKF 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.

Funded by the
European Union

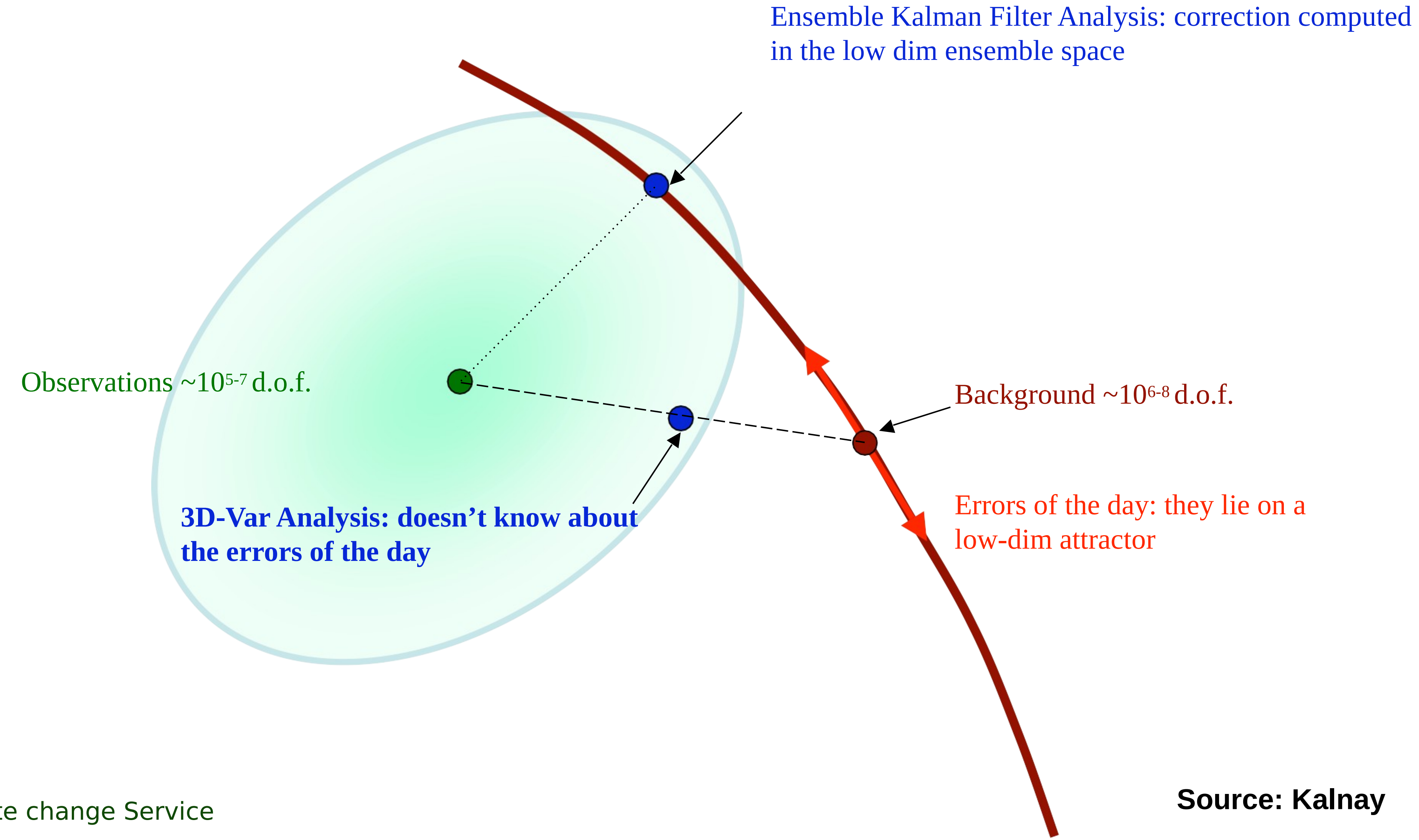
With Ensemble Kalman Filter based LDAS we will get perturbations pointing to the directions of the "errors of the day"



Source: Kalnay

Funded by the European Union

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

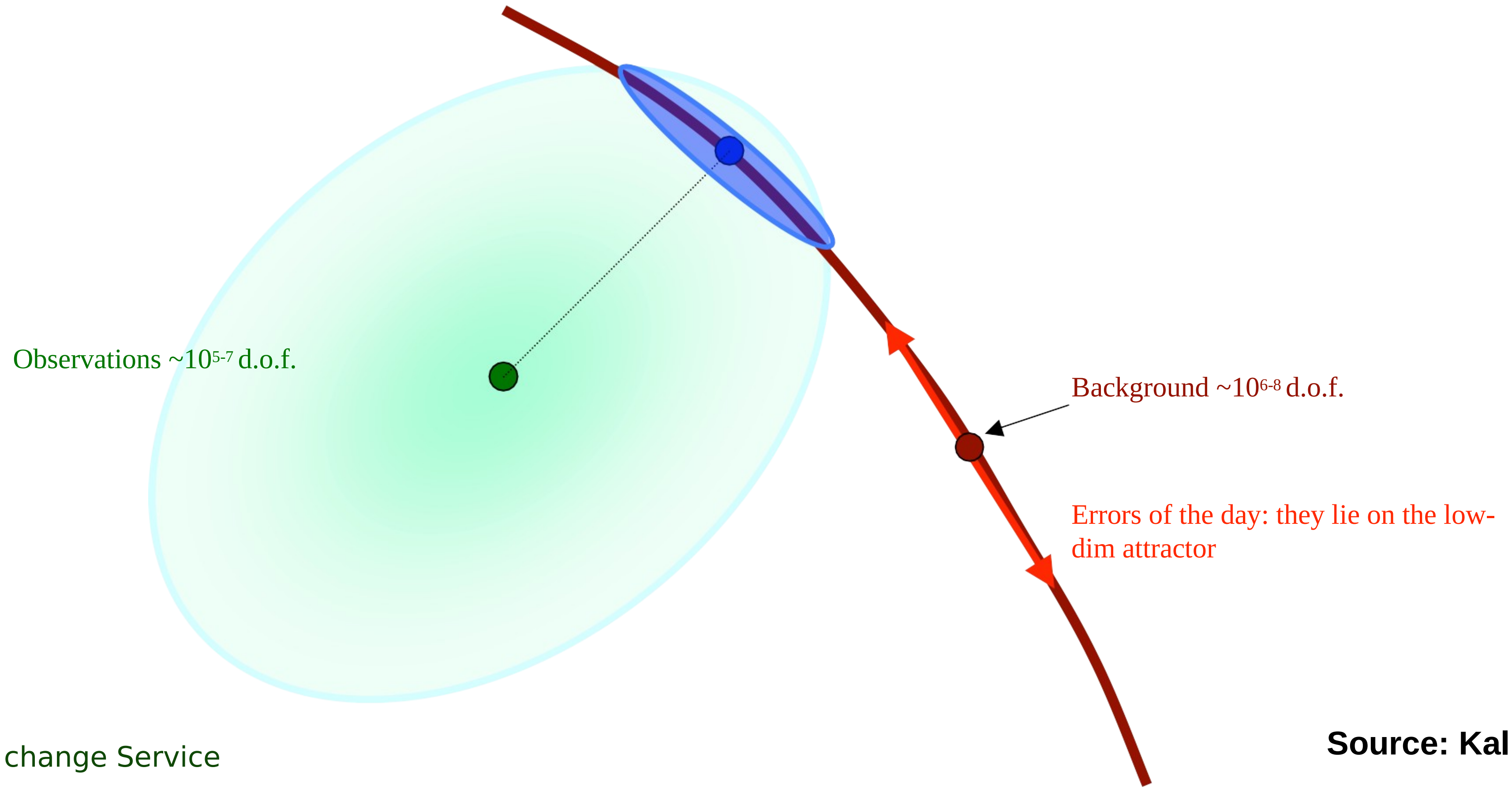


Source: Kalnay

After the EnKF computes the analysis and the analysis error covariance **A**, the new ensemble initial perturbations $\delta \mathbf{a}_i$ are computed:

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

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

Extended Kalman Filter (EKF)

Funded by the European Union

- Forecast step

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

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

- Analysis step

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

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

$$\mathbf{P}_{i\ n\times n}^a = [\mathbf{I} - \mathbf{K}_i\mathbf{H}]\mathbf{P}_i^b = [(\mathbf{P}_i^b)^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}]^{-1}$$

- Using the flow-dependent $\mathbf{P}_i^b$, analysis is expected to be improved significantly

However, it is computational hugely expensive. $\mathbf{P}_i^b$, $\mathbf{L}_i$ n*n matrix, n~10^7

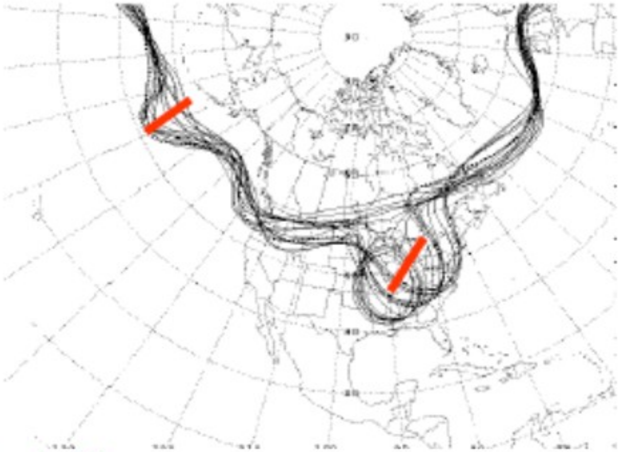
computing equation * directly is impossible

Source: Kalnay

Ensemble Kalman Filter (EnKF)

Funded by the European Union

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



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 *

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

$$= \frac{1}{K-1}\mathbf{X}^b \bullet \mathbf{X}^{bT}$$

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

Ensemble Update: two approaches

Funded by the European Union

1. Perturbed Observations method:

An "ensemble of data assimilations"

- It has been proven that an observational ensemble is required (e.g., Burgers et al. 1998). Otherwise $\mathbf{P}^a_{i\ n\times n} = [\mathbf{I} - \mathbf{K}_i\mathbf{H}]\mathbf{P}^b_i$ is not satisfied.

- Random perturbations are added to the observations to obtain observations for each independent cycle

$$\mathbf{y}^o_{i\,(k)} = \mathbf{y}^o_i + noise$$

- However, perturbing observations introduces a source of sampling errors (Whitaker and Hamill, 2002).

$$\mathbf{x}^b_{i\,(k)} = M\mathbf{x}^a_{i-1(k)}$$

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

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

$$\boxed{\mathbf{x}^a_{i\,(k)} = \mathbf{x}^b_{i\,(k)} + \mathbf{K}_i(\mathbf{y}^o_{i\,(k)} - H\mathbf{x}^b_{i\,(k)})}$$

Source: Kalnay

Ensemble Update: two approaches

Funded by the European Union

2. Ensemble square root filter (EnSRF)

- 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

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

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

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

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

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

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

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

Source: Kalnay

Harmonie-Arome Model Configuration Used in the Study

Funded by the European Union

code: https://github.com/josteinblyverket/Harmonie/tree/EnKF_CY46h1multi-layer physics: ISBA-DIF , 3-L for Snow scheme , Soil heat capacity = 2.0E-5

Surface analysis: (a) ENSRKF and (b) sEKF for Land Data Assimilation
Experiment: cold start at 2023-10-01, 3h cycling for 3 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)

MOTIVATION

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

Funded by the European Union

ISBA: 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.

Decharme et al. 2011, doi:10.1029/2011JD016002

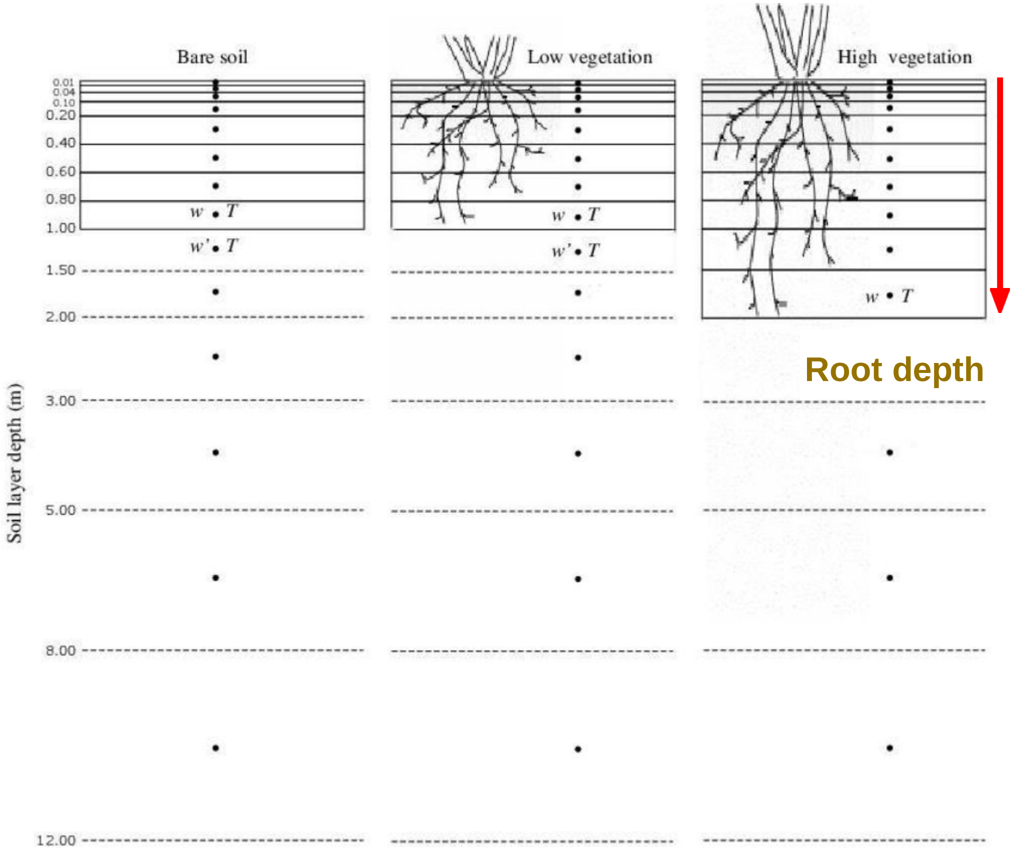


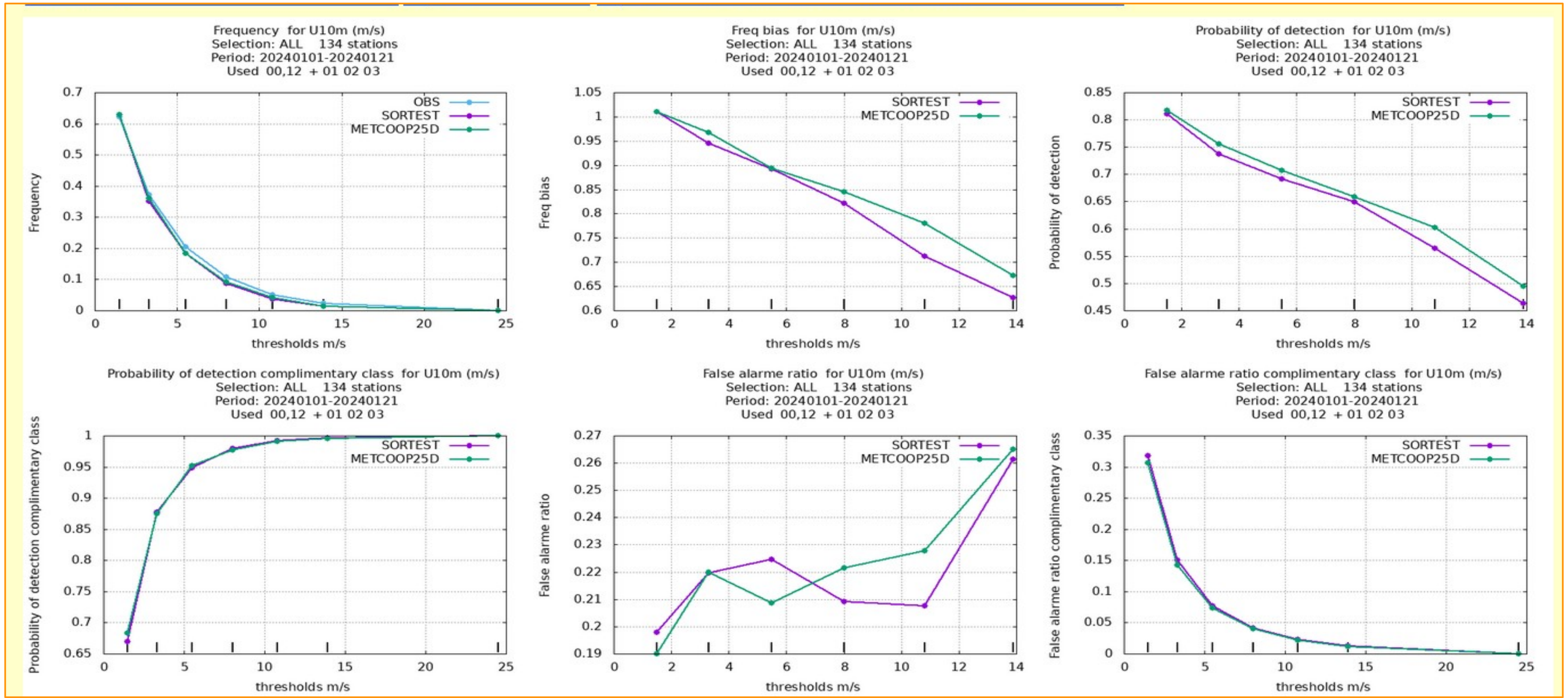
Figure 4.7 in SURFEX Scientific documentation for v8.1, P. Le Moigne, February 23, 2018.
http://www.umr-cnrm.fr/surfex/IMG/pdf/surfex_scidoc_v8.1.pdf

Source: Patrick Samuelsson

Results

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

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

(a)

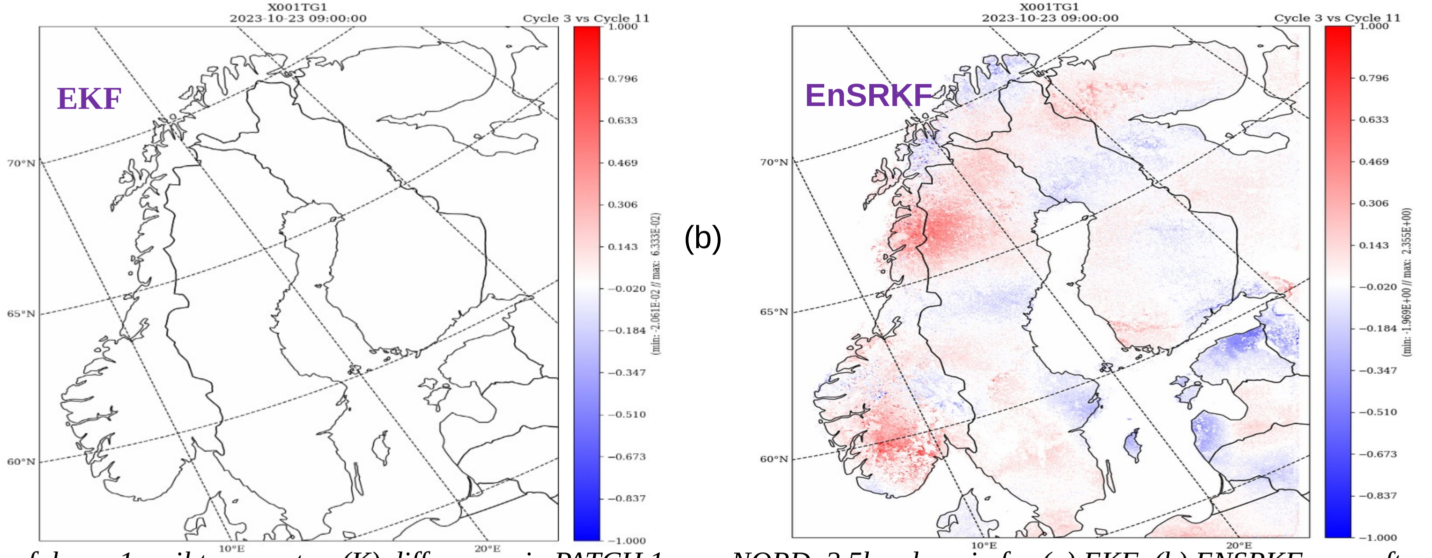


Figure : *Illustration of layer 1 soil temperature(K) differences in PATCH 1 over NORD_2.5km domain for (a) EKF (b) ENSRKF runs after state surface perturbations*

(a)

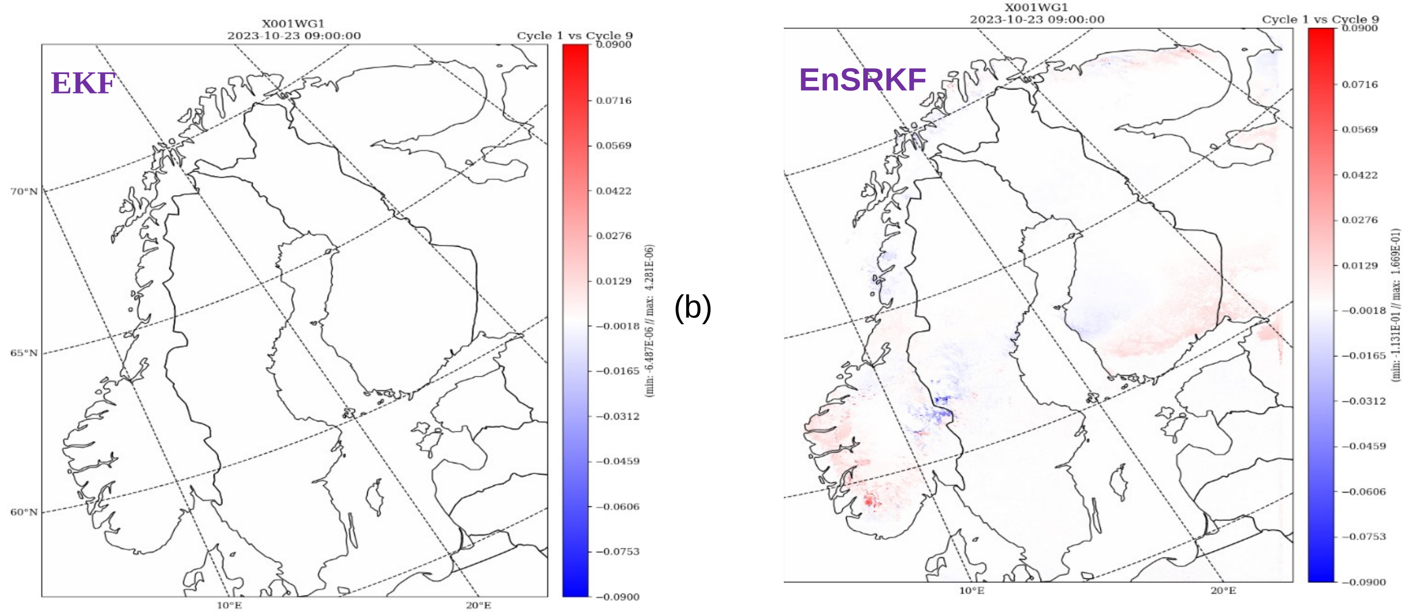


*Figure : Illustration of layer 1 soil moisture differences (kg m**-3) in PATCH 1 over NORD_2.5km domain for (a) EKF (b) ENSRKF runs after state surface perturbations*

(a) (b)

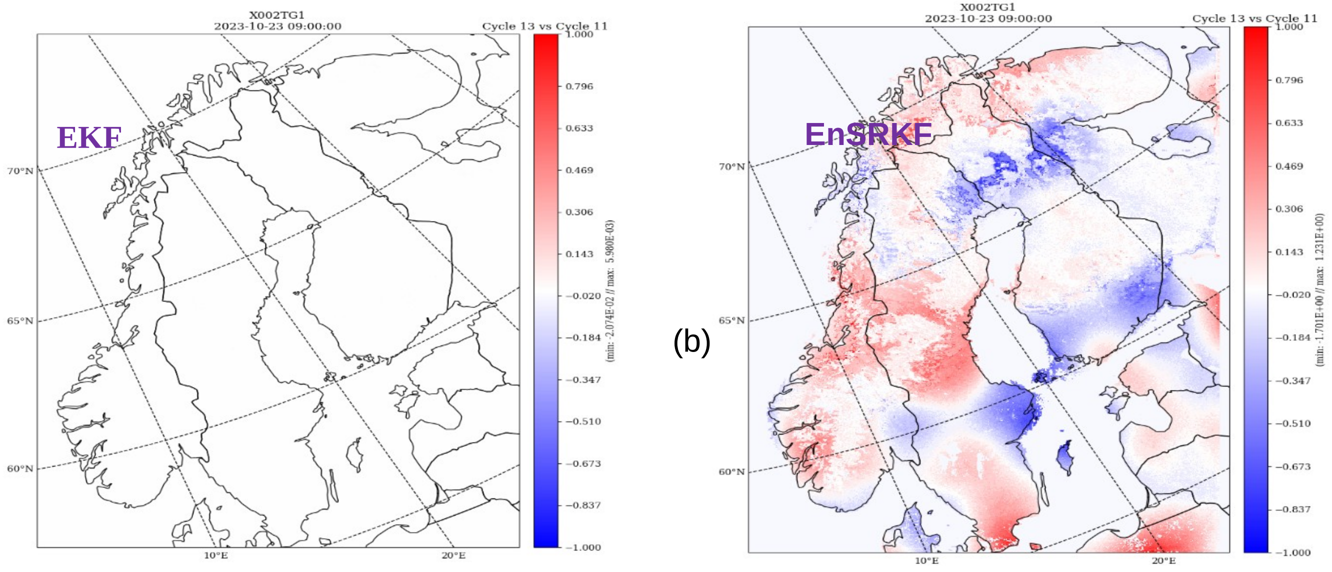


Figure : *Illustration of layer 1 differences in soil temperature (K) over PATCH2 over the NORD_2.5km domain for (a) EKF (b) ENSRKF , after state surface perturbations*

(a) (b)

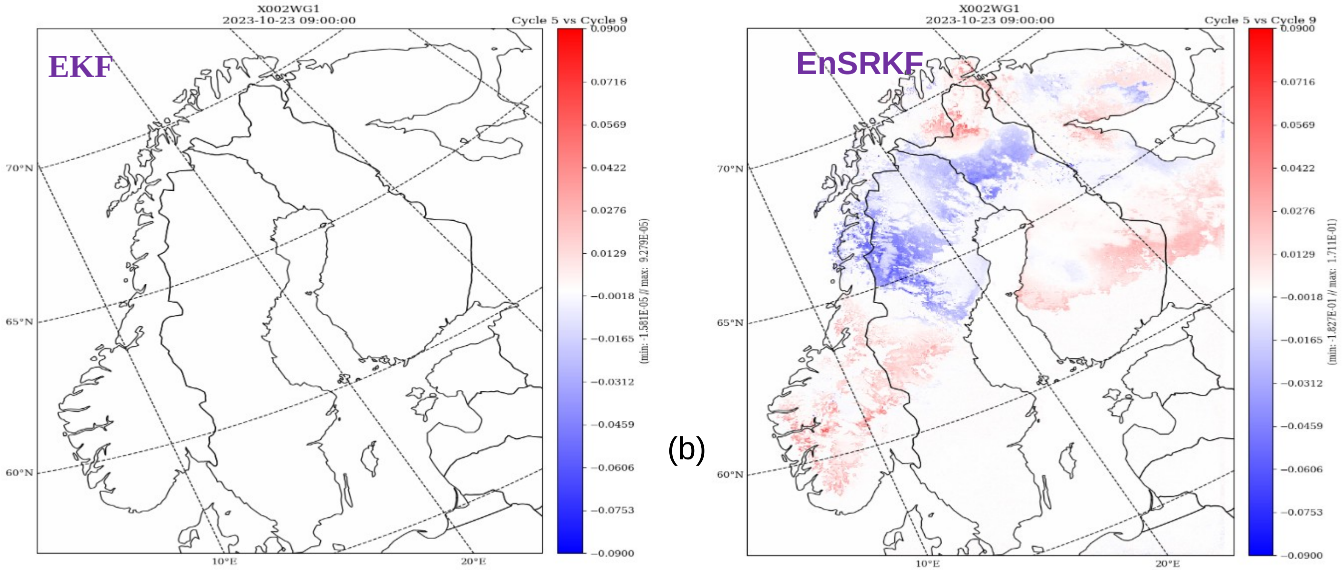


Figure : *Illustration of layer 1 differences in soil moisture (kg m**-3) over PATCH2 over the NORD_2.5km domain for (a) EKF (b) ENSRKF, after state surface perturbations.*

Funded by the European Union

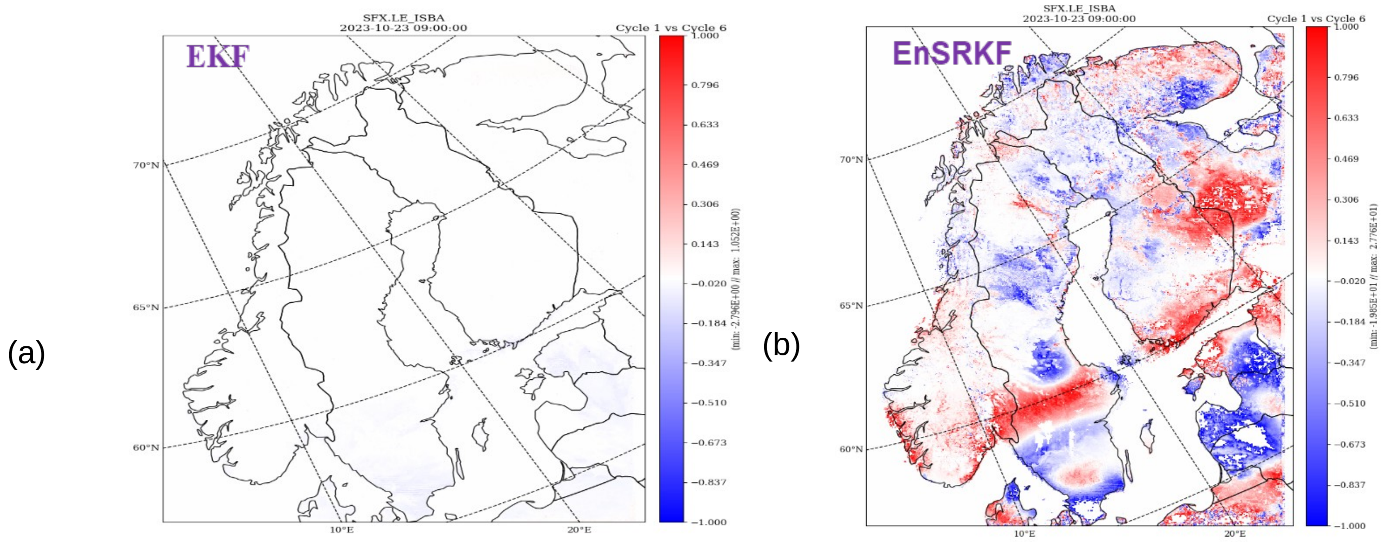


Figure : *Illustration of differences in perturbations of latent heat flux (W/m2) over the NORD_2.5km domain for (a) EKF (b) ENSRKF, after state surface perturbations.*

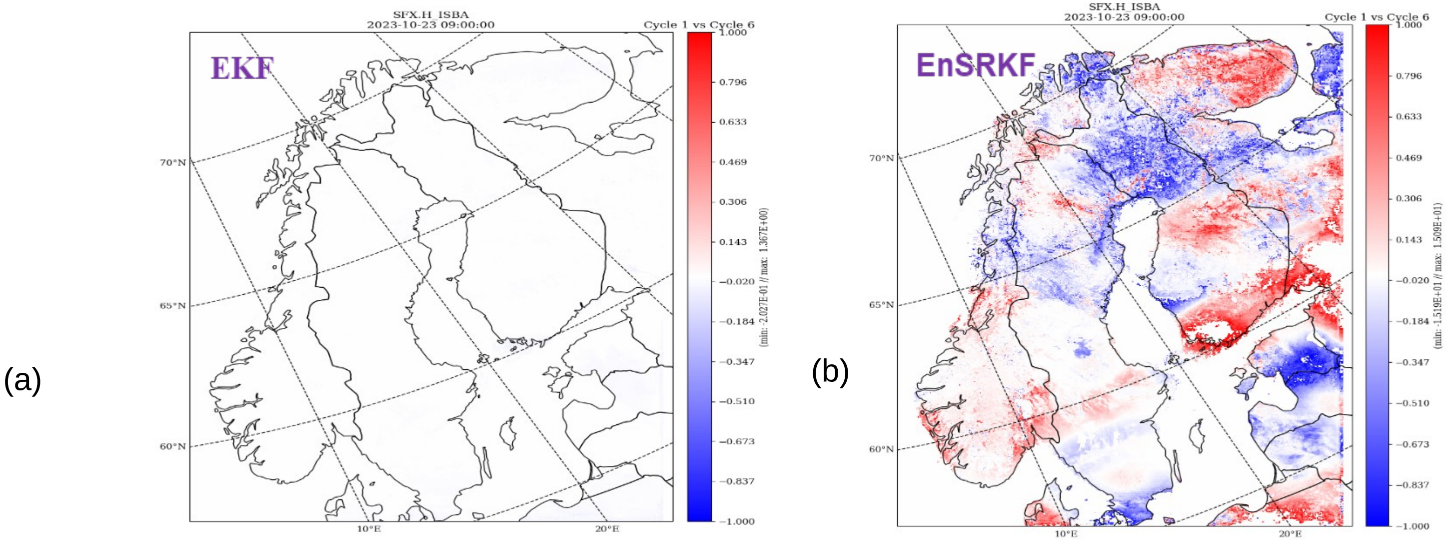


Figure : *Illustration of differences in perturbations of sensible heat flux (W/m2) over the NORD_2.5km domain for (a) EKF (b) ENSRKF, after state surface perturbations.*

Funded by the European Union

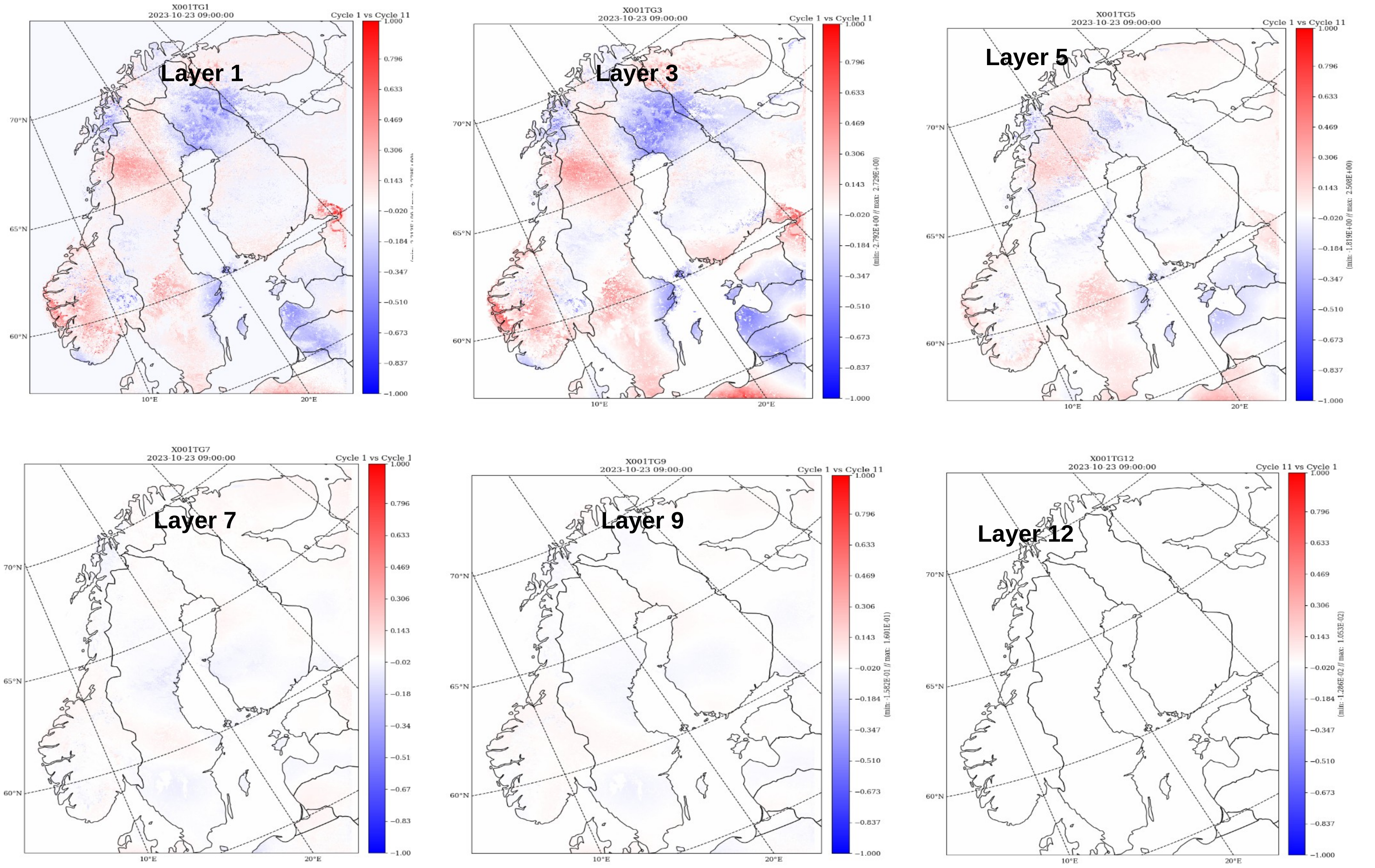


Figure: Illustration of spread of soil temperature differences (K) for PATCH 1, in layer 1 to layer 12 after state surface perturbations over NORD_2.5km domain valid at day-23 of the run from 1st October’23.

Funded by the European Union

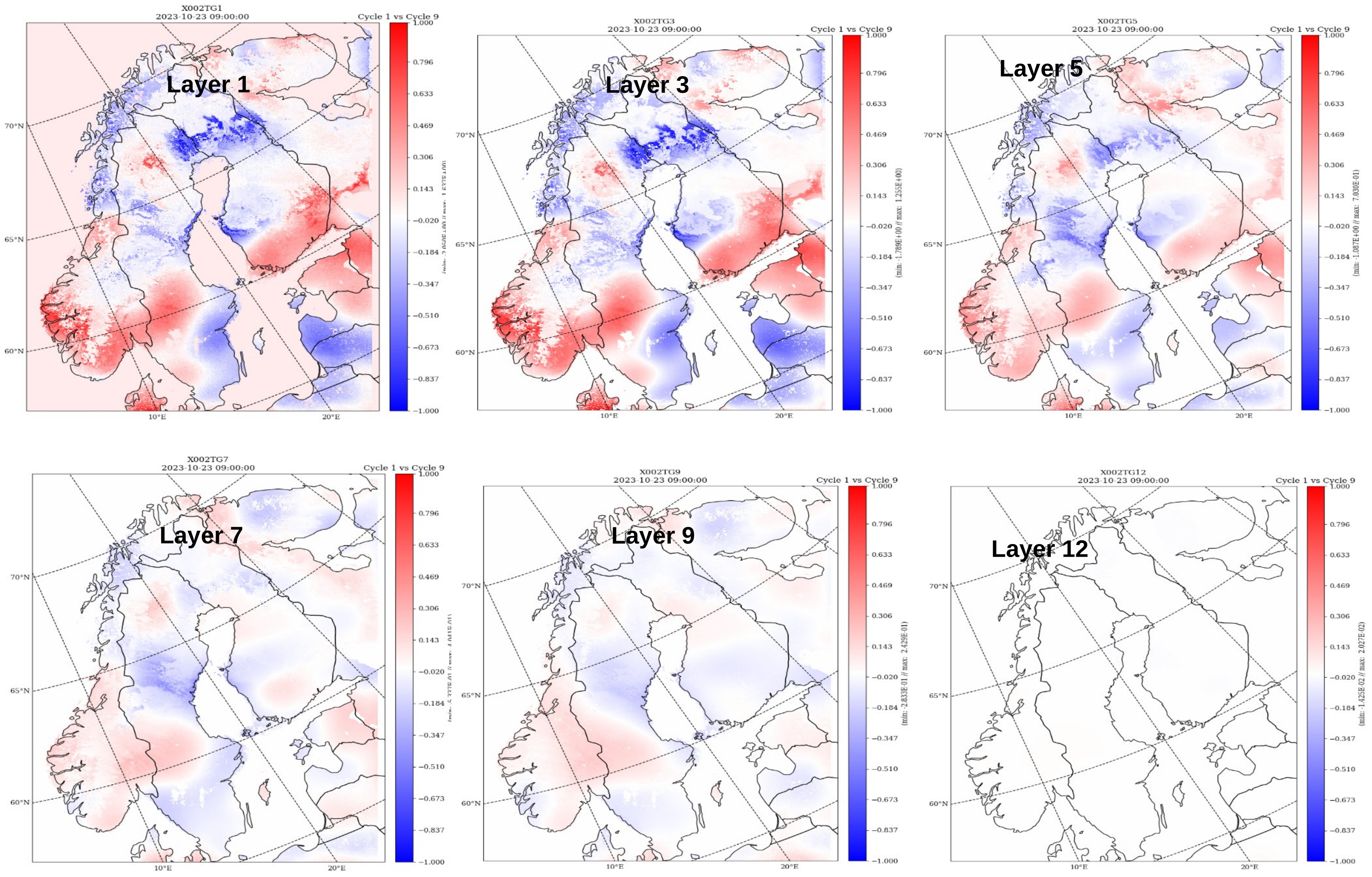


Figure: Illustration of spread of soil temperature differences (K) over PATCH 2, in layer 1 to layer 12 after state surface perturbations over NORD_2.5km domain valid at day-23 of the run from 1st October'23.

Funded by the European Union

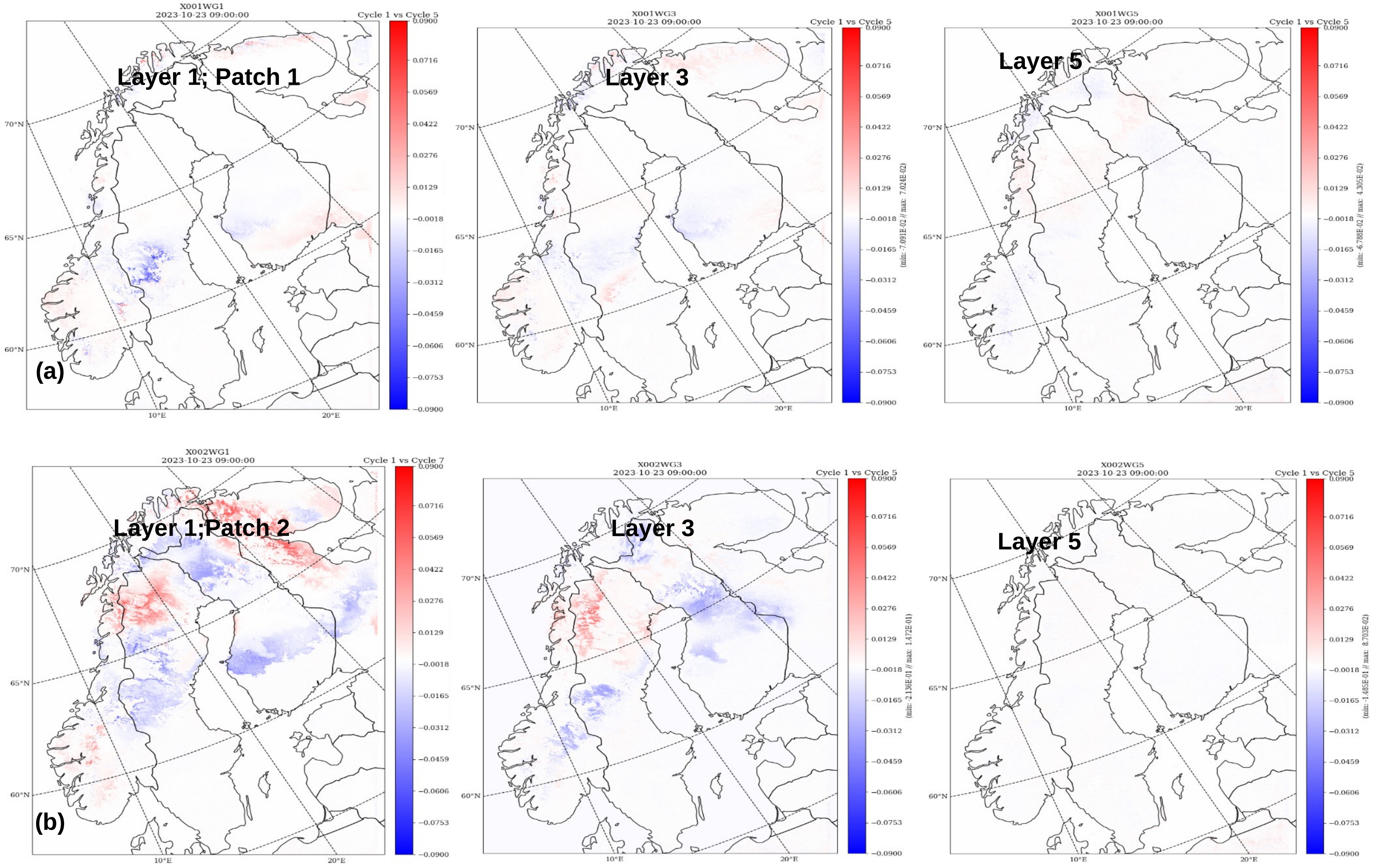
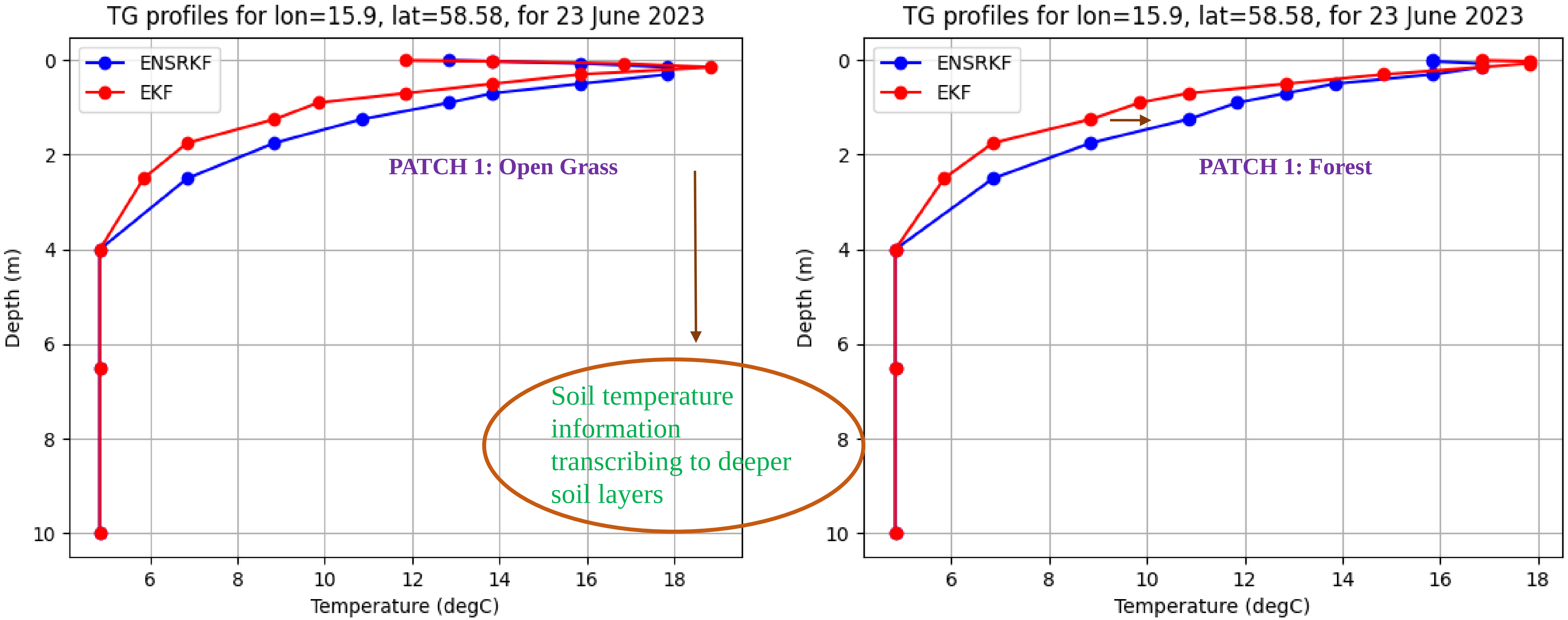


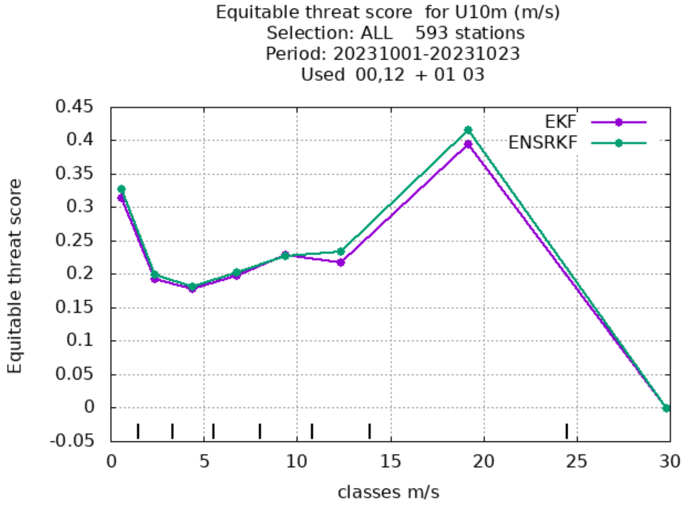
*Figure: Illustration of spread of soil moisture differences (kg m**-3) for (a) PATCH 1 (b) PATH 2, in layer 1 to layer 5, after state surface perturbations over NORD_2.5km domain, valid at day-23 of the run from 1st October'23.*

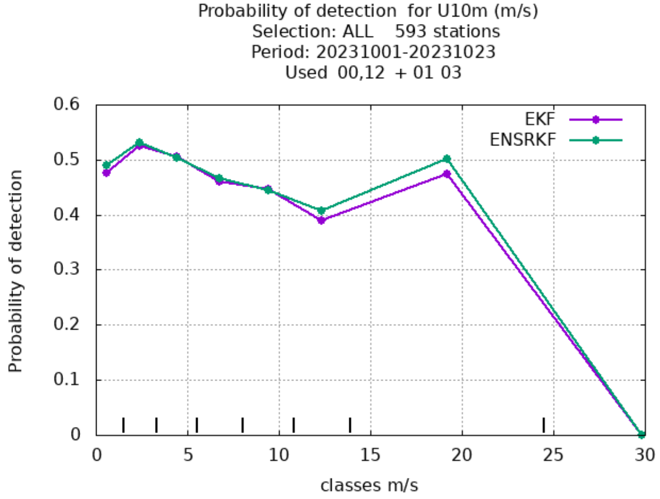
Funded by the
European Union

Vertical Profiles of Soil Temperature

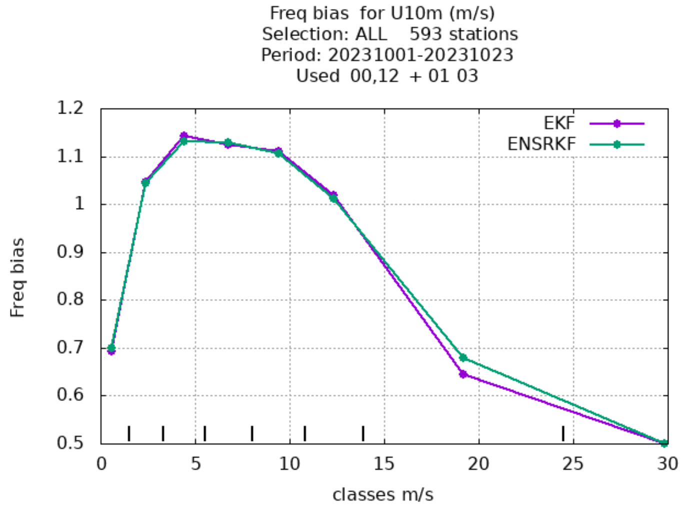


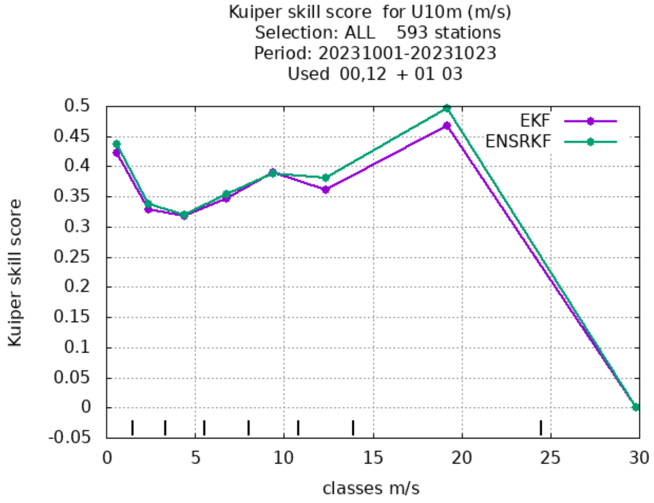
U10m





Minor improvements in predicting U10m with ENSRKF LDAS with SYNOP obs.



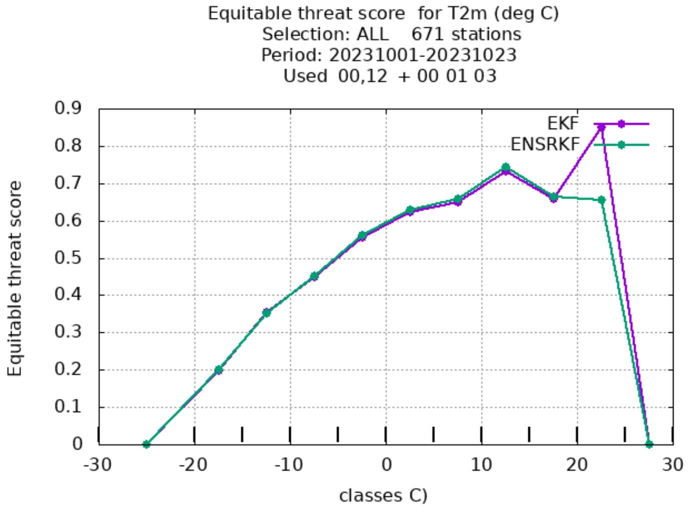


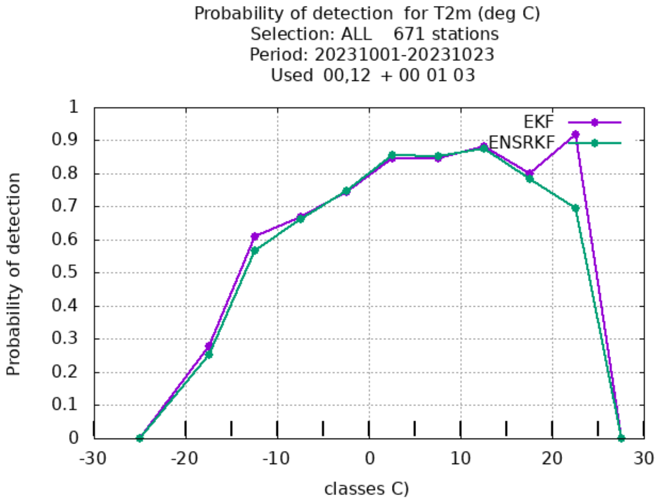
Funded by the European Union

Funded by the European Union

T2m

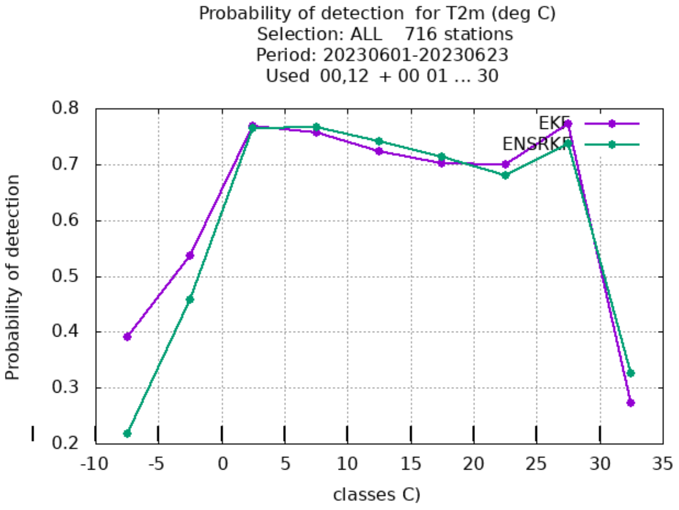
October

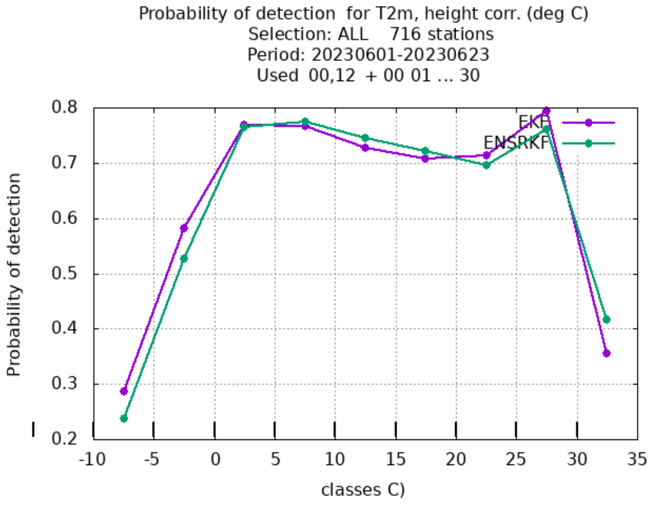




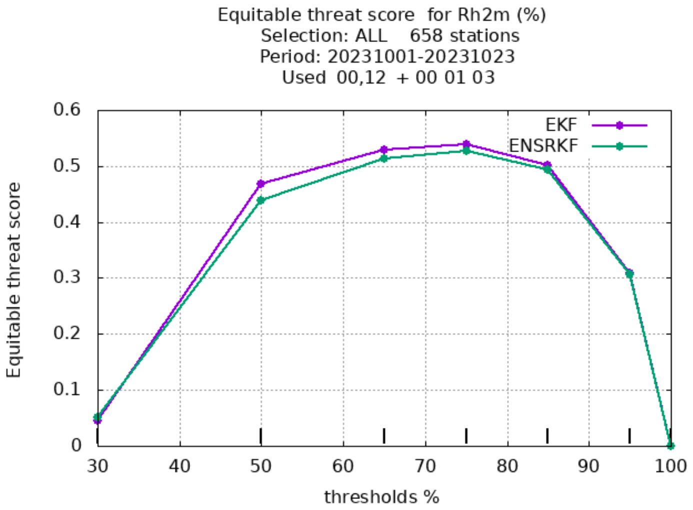
For T2m not much of improvement; EKF and EnSRKF gives same scores for October and June

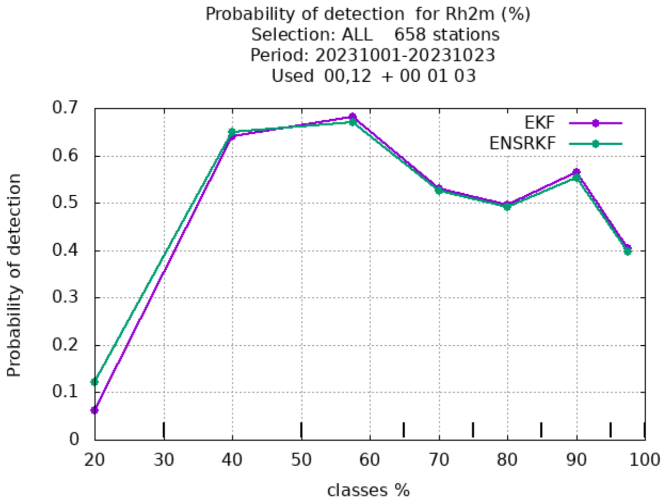
June



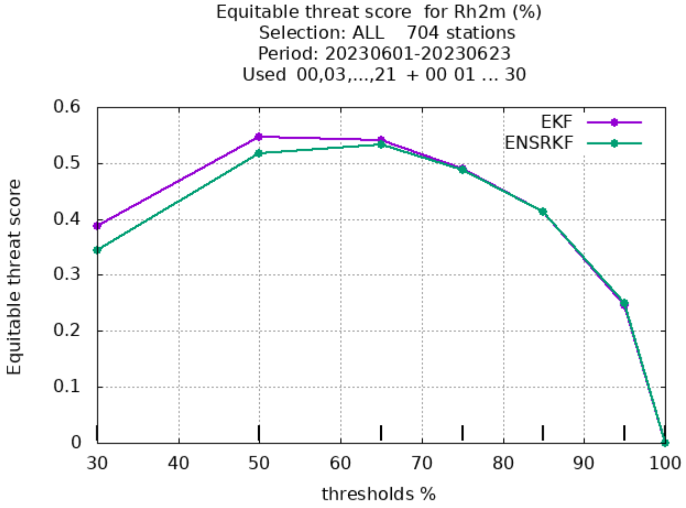


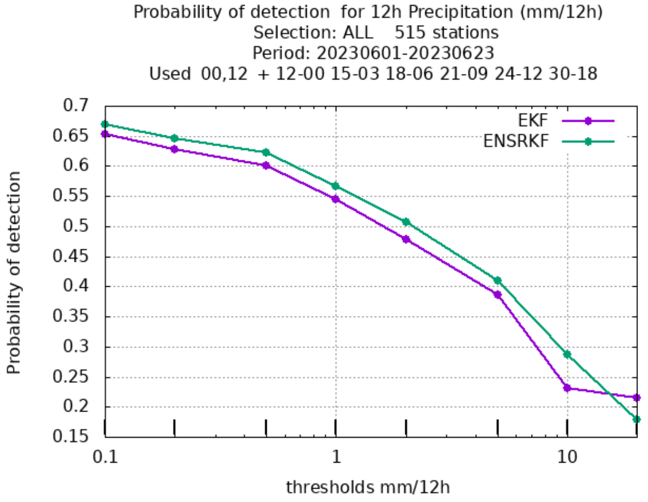
RH2m: October

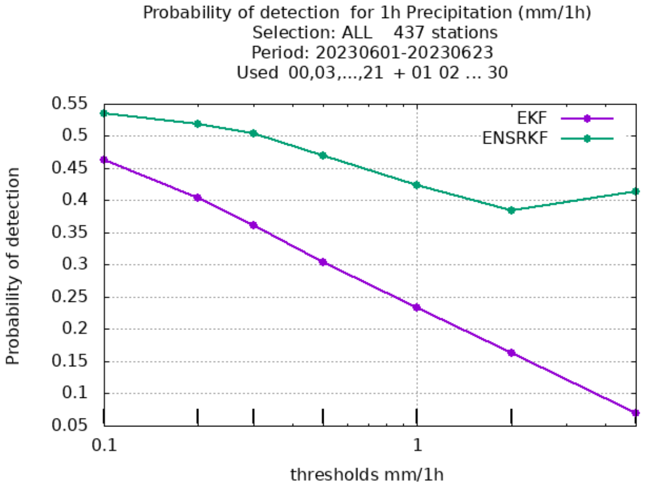




RH2m and Precipitation: June



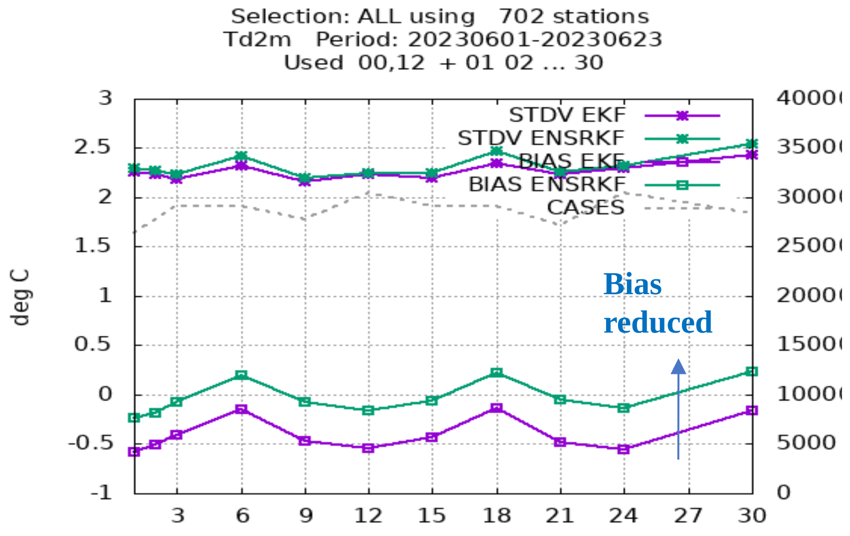


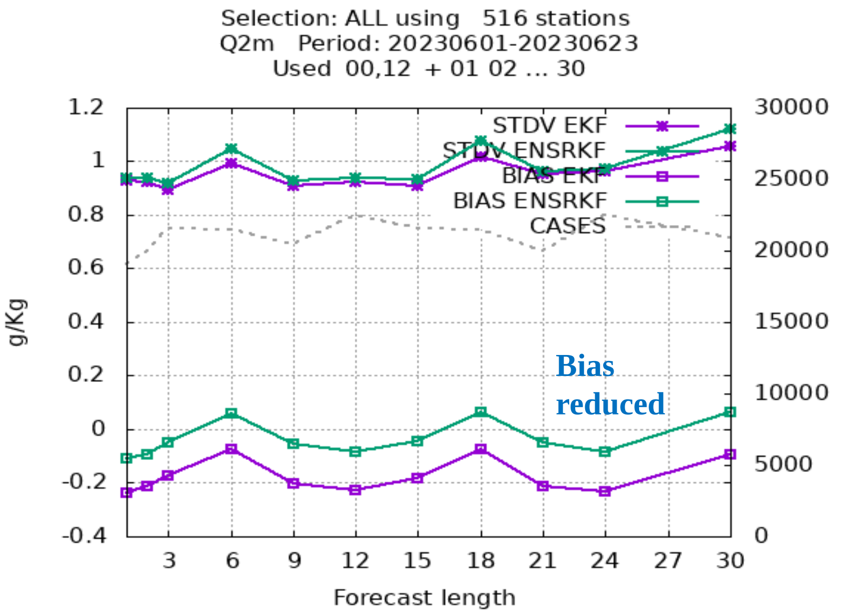


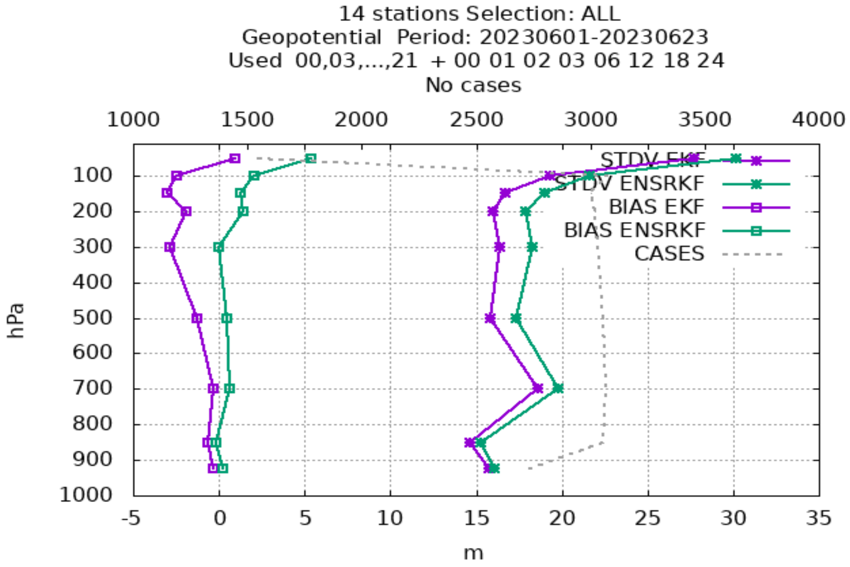
Major improvements with ENSRKF based LDAS in improving the precipitation POD scores (only SYNOP observations assimilated)

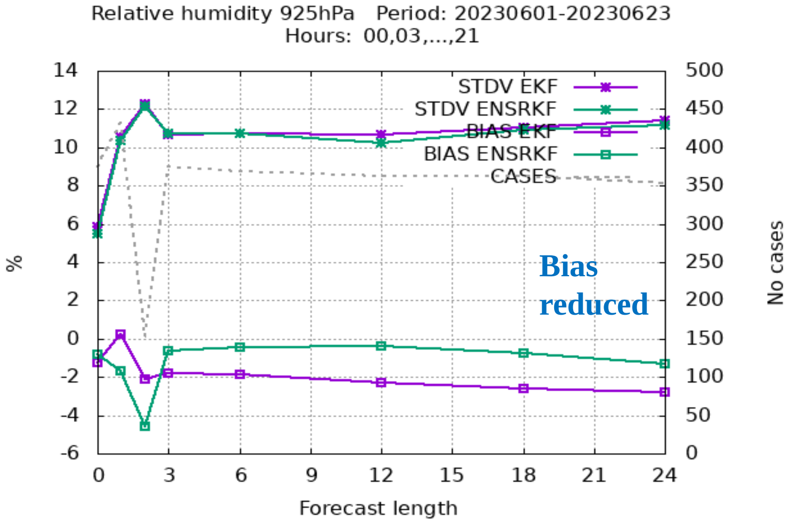
Funded by the European Union

June









Major improvements with ENSRKF LDAS in reducing the bias in forecasts of dew-point temperature, specific humidity,RH at 925 hPa and vertical profile of Geo-potential.

Conclusions

1. Ensemble Kalman filter based LDAS simulations tested over SOR_TEST (smaller) and METCOOP25D (bigger) domain in the HARMONIE-AROME system. The statistical scores like POD for U10m is better for SOR_TEST domain than METCOOP25D domain because during the time-period of the simulations, the weather features are better resolved in smaller domain. But smaller domain not advised for Data assimilation work.

2. Over the MetCoop domain (NORD_2.5km) the spread in perturbations of land surface variables and fluxes like soil temperature, soil moisture (surface to root zone), latent heat flux and sensible heat flux is more pronounced for the Ensemble kalman filter based LDAS in comparison to the simplified extended Kalman filter based LDAS.

3. The statistical scores suggests that the improvements in forecasts of T2m is limited whereas forecasts of TD2m, RH2m, Geopotential and U10m with the ENSRKF based LDAS is pronounced with better statistical verification scores. Thus, ENSRKF gives better scores concerning reduction of the systematic errors. Simulations of 6 months are required to confirm the results, which looks promising with ENSRKF LDAS.

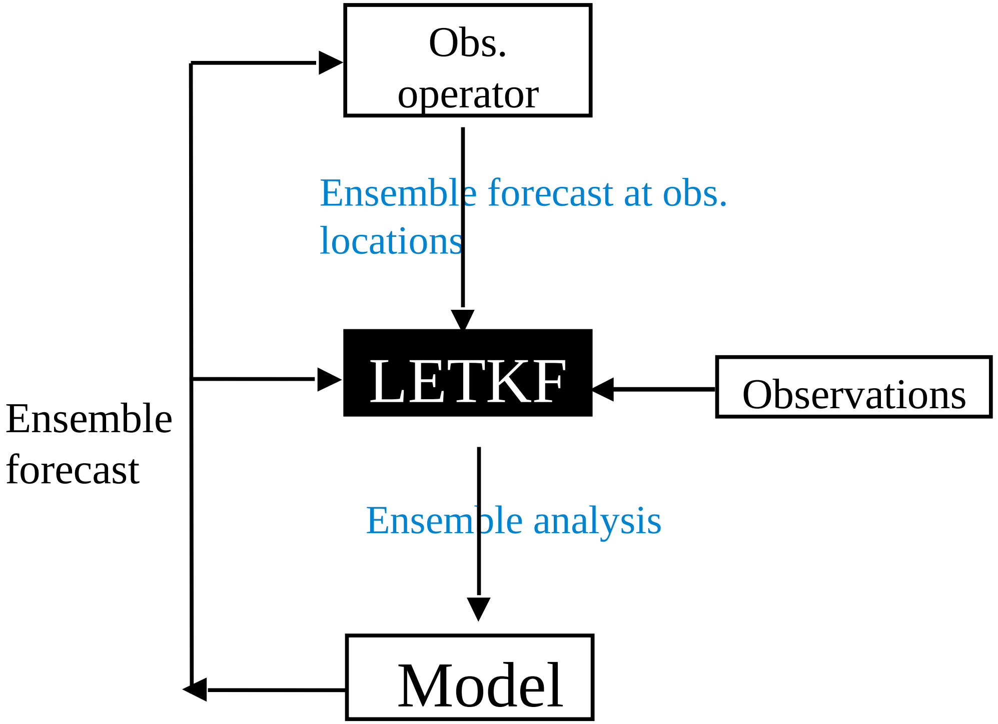
Funded by the European Union

Future work with Offline LDAS : LETKF

- EnKF is simple and model independent, while 4D-Var requires the development and maintenance of the adjoint model (model dependent)
- Variational DA can assimilate asynchronous observations, while EnKF assimilate observations at the synoptic time.
- Using the weights w^a at any time 4D-LETKF can assimilate asynchronous observations and move them forward or backward to the analysis time

Disadvantage of EnKF:

- Low dimensionality of the ensemble in EnKF introduces sampling errors in the estimation of P^b.
- Covariance localization can solve this problem.



Source: Kalnay

Funded by the European Union

Coordinated by ECMWF

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.