CopERnIcus climate change Service Evolution

D1.2 Unified, ensemble-based global land data assimilation system

Due date of deliverable	31st December 2024
Submission date	Dec 2024
File Name	CERISE-D1-2-V1.1
Work Package /Task	WP1 Tasks 1.2 & 1.3
Organisation Responsible of Deliverable	ECMWF
Author name(s)	Pete Weston, Christoph Herbert, Ewan Pinnington, Patricia de Rosnay
Revision number	V1.1
Status	Issued
Dissemination Level	PU

Funded by the European Union

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.

1 Executive Summary

The purpose of this work is to enhance the existing land-surface data assimilation system (LDAS) in two aspects. Firstly, to move towards a more unified LDAS by assimilating more variables using more consistent data assimilation (DA) methodologies. In this report, firstly, work to incorporate the snow and soil temperature analyses into the existing simplified extended Kalman filter (SEKF), used for the soil moisture analysis, is presented. This has technical benefits by allowing the retirement of the 1-dimensional optimal interpolation (1D-OI) analysis scheme, which was previously used to analyse snow and soil temperature, thus simplifying the LDAS and reducing the maintenance overhead. It also has scientific benefits illustrated by improved near-surface temperature forecasts based on Numerical Weather Prediction (NWP).

Secondly, to enhance the use of ensemble information in the LDAS. The current background errors in the SEKF are fixed, climatological values in contrast to the hybrid background errors used in the atmospheric DA, which contain a flow dependent contribution from the ensemble of data assimilations (EDA). In this report, work to include an ensemble contribution to the SEKF background errors is presented, resulting in small but significant improvements to the short-range near-surface temperature and humidity forecasts.

Ensemble information is already used to provide the Jacobians within the SEKF, which link the observed and model variables. However, the existing EDA is under-spread near the surface and does not use the SEKF in its LDAS due to resource constraints. These issues are addressed here by adding perturbations to land-surface parameters and switching on the SEKF in the EDA thanks to infrastructure developments reducing its computational cost. The additional land-surface parameter perturbations result in significant increases in spread of many land-surface variables while switching on the SEKF in the EDA results in significant reductions in near-surface temperature and humidity forecast errors. Both changes result in an improvement to the spread-error relationship.

Table of Contents

2 Introduction

This deliverable is a report documenting the scientific configuration of the ECMWF land-surface data assimilation system to be used in the global reanalysis demonstrator ERA7-pv1 to be run in 2025 as part of CERISE WP4.

2.1 Background

The scope of CERISE is to enhance the quality of the C3S reanalysis and seasonal forecast portfolio with a focus on land-atmosphere coupling. It will support the evolution of C3S, over the project's four-year timescale and beyond, by improving the C3S climate reanalysis and the seasonal prediction systems and products towards enhanced integrity and coherence of the C3S Earth system Essential Climate Variables. CERISE will develop new and innovative ensemble-based coupled land-atmosphere DA approaches and land surface initialisation techniques to pave the way for the next generations of the C3S reanalysis and seasonal prediction systems. These developments will be combined with innovative work on observation operator developments integrating Artificial Intelligence (AI) to ensure optimal data fusion fully integrated in coupled assimilation systems. They will drastically enhance the exploitation of past, current, and future Earth system observations over land surfaces, including the Copernicus Sentinels and from the European Space Agency (ESA) Earth Explorer missions' data, moving towards an all-sky and all-surface approach. For example, land observations can simultaneously improve the representation and prediction of land and atmosphere and provide additional benefits through the coupling feedback mechanisms. Using an ensemble-based approach will improve uncertainty estimates over land and lowest atmospheric levels. By improving coupled land-atmosphere assimilation methods, land surface evolution, and satellite data exploitation, research and innovation (R&I) inputs from CERISE will improve the representation of long-term trends and regional extremes in the C3S reanalysis and seasonal prediction systems. In addition, CERISE will provide the proof of concept to demonstrate the feasibility of the integration of the developed approaches in the core C3S (operational service), with the delivery of reanalysis prototype (demonstrated in pre-operational environment), and seasonal prediction demonstrator datasets (demonstrated in relevant environment). CERISE will improve the quality and consistency of the C3S reanalysis systems and of the components of the seasonal prediction multi-system, directly addressing the evolving user needs for improved and more consistent C3S Earth system products.

2.2 Scope of this deliverable

2.2.1 Objectives

This report has two main objectives:

- To document the configuration of the land-surface data assimilation system to be used in the global reanalysis demonstrator ERA7-pv1.
- To present developments carried out in the CERISE project towards a unified and ensemble-based land DA system.

2.2.2 Work performed

The work outlined in WP1 T1.2 (Develop ensemble-based filter land DA approaches for soil moisture) and WP1 T1.3 (Unified land DA system development) is described and evaluated. The focus is the global land DA system run at ECMWF. The baseline configuration at cycle 49r1 (operational from 12th November 2024 onwards) is described briefly in Section 3. The majority of the report is devoted to describing the novel research undertaken to enhance the use of ensemble information and to move towards a unified land DA system. In particular, the analysis of snow and soil temperature variables in the SEKF is described and results are presented in Section 4.1. Initial experiments towards a unified assimilation of two-metre temperature and relative humidity are summarised in Section 4.2. The use of ensemble-based background errors is presented and assessed in Section 5.1. Enhancements to the near-surface ensemble spread are presented in Section 5.2. Finally, the use of the SEKF in the ensemble of data assimilations (EDA) is described and results presented in Section 5.3.

2.2.3 Deviations and counter measures

So far, no deviations or counter measures have been encountered.

2.2.4 Reference Documents

The reference document is the grant agreement with

[1] Project 101082139- CERISE-HORIZON-CL4-2021-SPACE-01 Grant Agreement

2.2.5 CERISE Project Partners

There are 12 project organisation partners active in the CERISE project, which are listed in the following table.

Table 1. List of the active partners, with the abbreviated and full names, in the CERISE project

ECMWF	European Centre for Medium-Range Weather Forecasts
Met Norway	Norwegian Meteorological Institute
SMHI	Swedish Meteorological and Hydrological Institute
MF	Météo-France
DWD	Deutscher Wetterdienst
CMCC	Euro-Mediterranean Center on Climate Change
BSC	Barcelona Supercomputing Centre
DMI	Danish Meteorological Institute
Estellus	Estellus
IPMA	Portuguese Institute for Sea and Atmosphere
NILU	Norwegian Institute for Air Research
MetO	Met Office

3 Baseline land data assimilation system

The full details of the ECMWF land data assimilation system (LDAS) can be found in the integrated forecast system (IFS) documentation of cycle 49r1 in Part II: Chapter 9 on land surface analysis (IFS 49r1 documentation). Here, the system is described briefly to provide a baseline on which the research in Sections 4 and 5 is based.

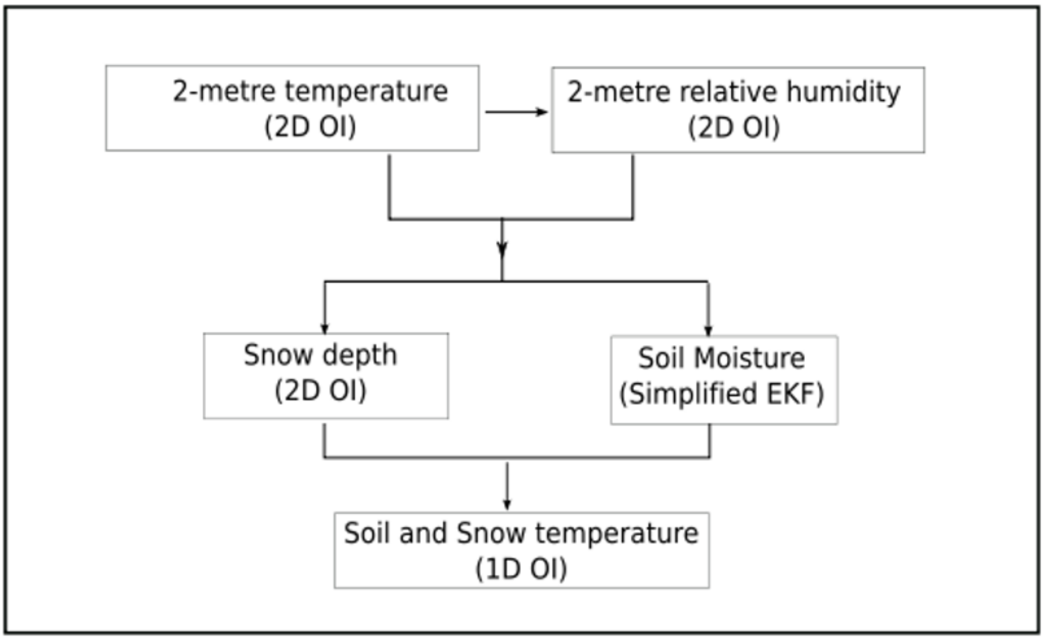

Figure 1: Schematic of ECMWF LDAS at 49r1

The LDAS comprises of several different tasks (see Figure 1) to produce an analysis of land-surface variables. The different DA methodologies and variables analysed are listed below:

- 2D-OI: 2-metre temperature, 2-metre relative humidity and snow depth are analysed in a 2-dimensional optimal interpolation (2D-OI) scheme. The assimilated observations include in-situ SYNOP reports and satellite snow cover from the Interactive Multisensor Snow and Ice Mapping System (IMS).
- 1D-OI: The top layer of soil and snow temperature are analysed in a 1-dimensional optimal interpolation scheme. The output analysis from the 2D-OI at each grid point are used as pseudo-observations.
- SEKF: The top three layers of soil moisture are analysed in a simplified extended Kalman filter. Again, the output analysis of the 2D-OI at each grid point are used as pseudo-observations as well as satellite observations from the advanced scatterometer (ASCAT) and soil moisture ocean salinity (SMOS) instruments.

The whole system provides an analysis of model prognostic variables: soil moisture at 0-7cm; 7-28cm; 28-100cm; soil temperature at 0-7cm; top layer snow temperature and multi-layer snow water equivalent (SWE) which is then used to initialise land-atmosphere coupled NWP forecasts.

4 Unified land data assimilation system

One of the aims of the CERISE project is to unify LDAS by reducing the number of separate tasks used to analyse the different land-surface variables as seen in Figure 1. The first motivation is that the SEKF is a superior DA methodology to the 1D-OI with significant benefits when introducing this method for soil moisture assimilation (de Rosnay et al, 2013; Gomez et al, 2020). The second motivation is that producing a multi-variate analysis using the same methodology should lead to enhanced consistency between the analyses of the different variables. It will also benefit from a consistent assimilation configuration including shared aspects such as quality control, observation/background error specifications, Jacobian calculations etc. The final motivation is more pragmatic. With fewer separate task running, there will be less maintenance overhead and individual methodological advances will be automatically applied to the analyses of multiple variables rather than being adapted or ported to different tasks and methodologies.

4.1 Analysing soil and snow temperature in the SEKF

This section presents the first step towards a consistent and unified LDAS using the most advanced ensemble-based SEKF. The technical changes involved the replacement of the formerly used 1D-OI scheme for first-layer snow temperature, soil temperature analysis, the integration of the first-layer snow and multi-layer soil temperature analysis into the SEKF control vector. In contrast to the 1D-OI, the main scientific benefits of the new implementation are also enabling the soil temperature analysis in deeper layers – where updates can have longer lasting effects – and the use of spread information from the EDA. In the following, the implementation of soil and snow temperature analysis in the SEKF and the involved NWP experimentation. The results are published in Herbert et. al., (2024).

4.1.1 Methodology

In the new implementation, the SEKF control vector has been extended to include the first layer snow temperature (tsn) and the soil temperatures of layers (stl1-3) as additional control variables in addition to the approach used for the current SEKF soil moisture analysis. As in the previously used 1D-OI implemented by Mahfouf et al. (2000), T2m fields provided by the screen-level analysis are assimilated as pseudo-observations for the soil and snow temperature analyses. A key difference is that the SEKF is based on the T2m analysis fields, whereas the 1D-OI approach uses the analysis increments as input. The updated SEKF control vector x_b and observation vector y are given by x^T_b = [swvl1, swvl2, swvl3, stl1, stl2, stl3, tsn] and y^T = [T2m, RH2m, SSM_{ASCAT}, SSM_{SMOS}], respectively.

The SEKF analysis was optimised in the definition of background and observation errors, to account for spurious correlations and snow-covered areas, and to ensure geophysically reasonable sensitivities between screen-level observations and control variables in the calculation of Jacobian elements. The specifications of the soil and snow temperature analysis are explained in detail in Herbert et. al. (2024) and are summarised in **Table 2.**

Table 2: Parameter settings for soil and snow temperature analysis regarding T2m observation error $\sigma_{o,T2m}$, maximum first-guess departure $o{-}b_{max}$, background error $\sigma_{b,st/tsn}$, minimum variance $var_{st/tsn,min}$, range of Jacobian elements [H_{min}, H_{max}], maximum analysis increment $\delta_{a,max}$, lower and upper limits for snow temperature analysis with respect to snow depth sd_{min} and snow water equivalent swe_{max}, respectively, and upper limit of snow depth sd_{max} for soil temperature analysis.

Parameter	Value	Parameter	Value	Parameter	Value
$\sigma_{o,T2m}$	1 K	$var_{st/tsn,min}$	10^{-3} K	sd_{min} for tsn	1 mm
$o{-}b_{max}$	5 K	[H_{min}, H_{max}]	[0, 2] K·K−1	swe_{max} for tsn	9 m
$\sigma_{b,st/tsn}$	1 K	$\delta_{a,max}$	10 K	sd_{max} for st	10 cm

In the implementation of the SEKF, various configurations were tested regarding the filtering of the Jacobian elements between the observations and the control variables. The final configuration selected for use in Cycle 49r2 for the global ERA6 reanalysis and the next NWP Cycle 50r1 has the following settings regarding the Jacobians: In case, in which a Jacobian element between a control variable and observation exceeds a threshold, the analyses of control variables related to the same geophysical variable are discarded at all soil layers, and, additionally, at both synoptic analysis times – assuring temporal consistency along the DA window. Negative covariances between T2m and the soil/snow temperature layers are set to zero, as they are considered unphysical, resulting in zero Jacobian elements. This has the advantage that more information is retained in the analysis, while consistency throughout the soil profile is maintained for large Jacobian elements. In the evaluation of the experiments, the final configuration is referred to as *SEKF(I).*

4.1.2 NWP experimentation and evaluation

A set of NWP experiments based on Cycle 49r1 was conducted to evaluate the forecast skill using the final setting of the SEKF (referred to as *SEKF(I)* in the plots*)* soil and snow temperature analysis. An implementation with the SEKF in place as well as a control setup using the former 1D-OI is presented over the period of boreal summer from 1 June to 31 August 2022. The atmospheric impact is evaluated in terms of the normalised differences in forecast error at different lead times verified against the respective analysis.

4.1.3 Results

The results presented comprise a sensitivity analysis with an assessment of the diurnal variability of the Jacobians between T2m and the control variables, a qualitative evaluation of the analysis increments, and the assessment of the impact on the atmospheric forecast skill.

Sensitivity analysis and increments

The Jacobian elements H_{ij} between the individual control variables i ∈ [stl1–3, tsn] and T2m are evaluated over a complete diurnal cycle. **Figure 2a** shows the normalised cumulative density of Jacobians over an entire day to indicate the dynamic range of values. As expected, values are typically greater than zero, implying that the air and soil temperatures are generally positively correlated. However, H(stl3,T2m) also exhibits a number of negative values, which are due to spurious non-physical negative covariances in uncoupled conditions. **Figure 2b** illustrates the diurnal cycle of the Jacobians, showing mean values and lower and upper quartiles as a function of the solar zenith angle using a binning of 30°. The diurnal cycle consistently reveals lower magnitudes for deeper layers and throughout all solar zenith angles. This is reasonable behaviour, since the coupling is supposed to be strongest at the surface and the sensitivity reduces with depth. The sensitivity calculated from the EDA spread in combination with the exclusion of negative covariances protects the analysis from an unphysical overcorrection of deeper soil layers. Thus, in contrast to the current soil moisture analysis in the SEKF, the Jacobians are not tapered with depth.

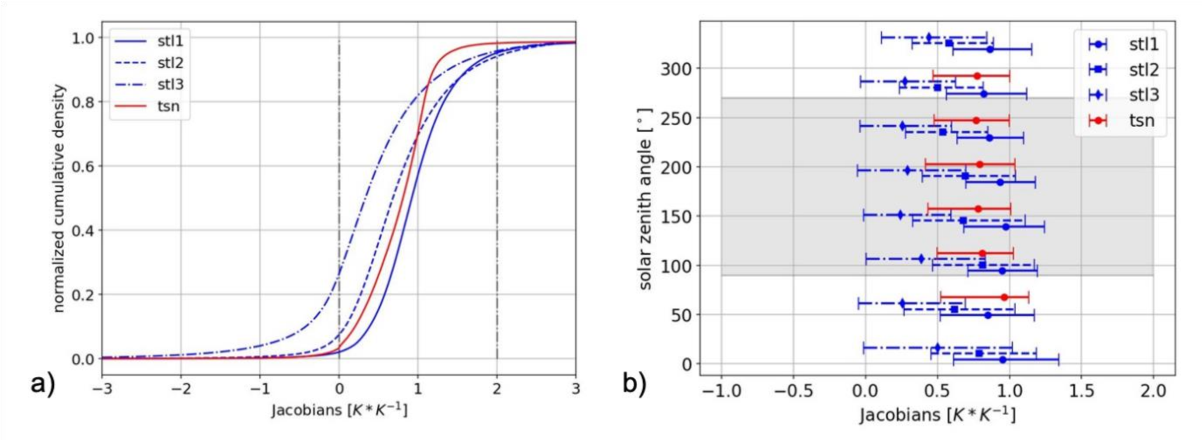

Figure 2: Dynamic range and diurnal variability of Jacobian elements between 2-m temperature and snow and multi-layer soil temperatures evaluated for an entire diurnal cycle including analysis times at 0000, 0600, 1200, and 1800 UTC on December 1, 2022. Left panel (2a): Normalized cumulative density of Jacobian elements. Right panel (2b): Means and variances of Jacobian elements as a function of mean solar zenith angle (30◦ binning) with grey shaded night-times where dusk (dawn) corresponds to 90◦ (270◦), and the sun is at its zenith at 0◦/360◦.

The soil and snow temperature analysis increments obtained for the first analysis time (valid at 00 UTC on December 1, 2022) of the NWP experiments based on the SEKF are compared with those obtained using the 1D-OI in **Figure 3**, showing those obtained with the 1D-OI (**a**,**c**) compared with those produced by SEKF (**b**,**d**). On larger scales, the 1D-OI-based increment patterns are like those obtained using the SEKF, whereas the magnitude of 1D-OI-based increments is generally larger, while SEKF-based increments are more heterogeneous on smaller scales (Herbert et al. 2024). This indicates that using the EDA to calculate the analysis results in the soil temperature increments being more indicative of the underlying local variations due to e.g. different soil and vegetation types.

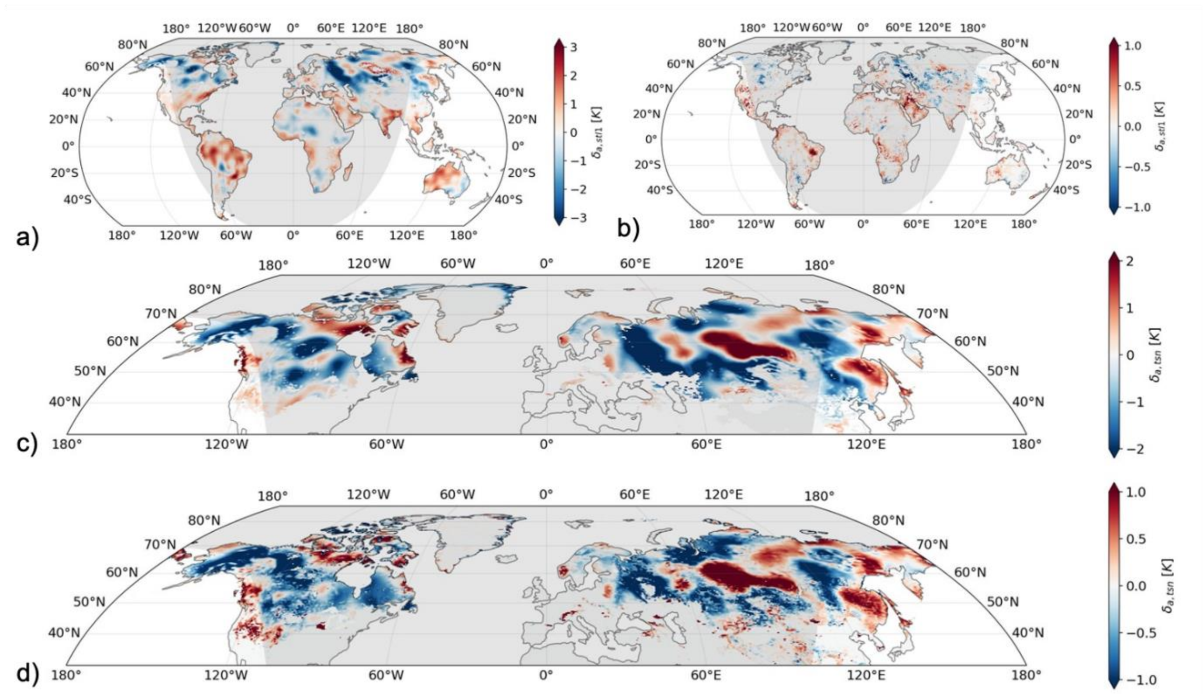

Figure 3: Comparison between maps of first-layer snow and multi-layer soil temperature analysis increments based on the SEKF and the increments at the first layer produced by the 1D-OI approach on December 1, 2022 at 0000

UTC obtained using a 6-hour short forecast. (a) Analysis increments of stl1 based on 1D-OI , (b) stl1 based on SEKF , (c) tsn based on 1D-OI and (d) tsn based on SEKF.

Impact on atmospheric forecast skill

In the following, the performance of the new SEKF soil and snow temperature analysis compared with the former setting using the 1D-OI is evaluated regarding atmospheric forecast skill. Scores are shown following the approach explain in Geer (2016).

Figure 4 shows the relative difference in forecast error in T2m evaluated for up to 10 forecast days and averaged over the three-month periods of boreal summer in the Tropics, northern and southern hemispheres (NH and SH). The red line shows the results of the discussed final SEKF configuration, and the black line corresponds to results of experiments based on an intermediate configuration. All areas indicate significant improvement in forecast skill for short lead times, with the positive impact being largest in the NH. In the 12-hour short forecast, the scores are reduced by about 4%–5% in the Tropics and the northern hemisphere, where most of the land mass is located, and by 0.5%–2% in the SH. Overall, the absolute reduction is roughly proportional to the land area over which the scores are averaged, whereas the significance of improvements depends on the season.

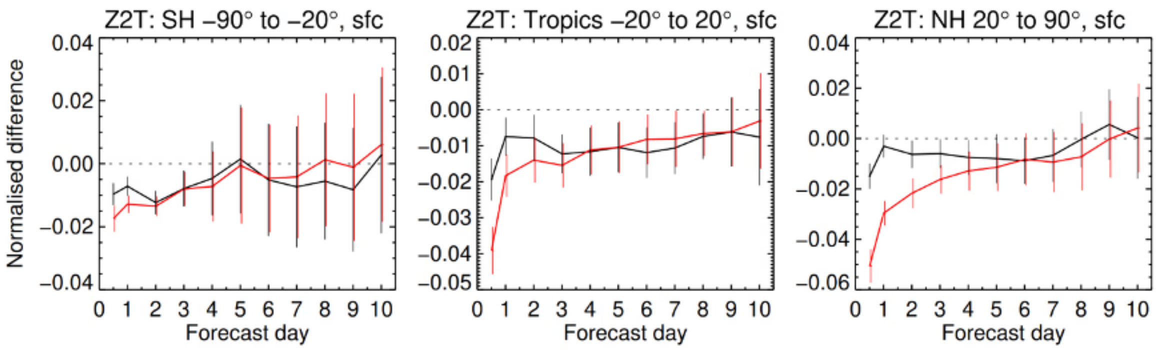

Figure 4: Normalised dRMS forecast error of T2m per forecast day of the SEKF(I) (red line) and SEKF(II) experiments (black line), compared with the 1D-OI control experiment (black dotted line at zero). The forecast of each experiment is verified against their respective own analysis, and scores are averaged over the Southern Hemisphere (SH), the Tropics, and the Northern Hemisphere (NH), and evaluated over boreal summer from June 1–August 31, 2022.

Global maps of relative differences in T2m forecast errors of the SEKF compared with the 1D-OI approach for the first 72 hours during boreal summer are presented in **Figure 5**. The improvements (indicated by blue colours) are spread over the entire snow-free land surface, where the positive impact of the multi-layer soil temperature analysis becomes evident. The effect over snow-covered areas is almost neutral.

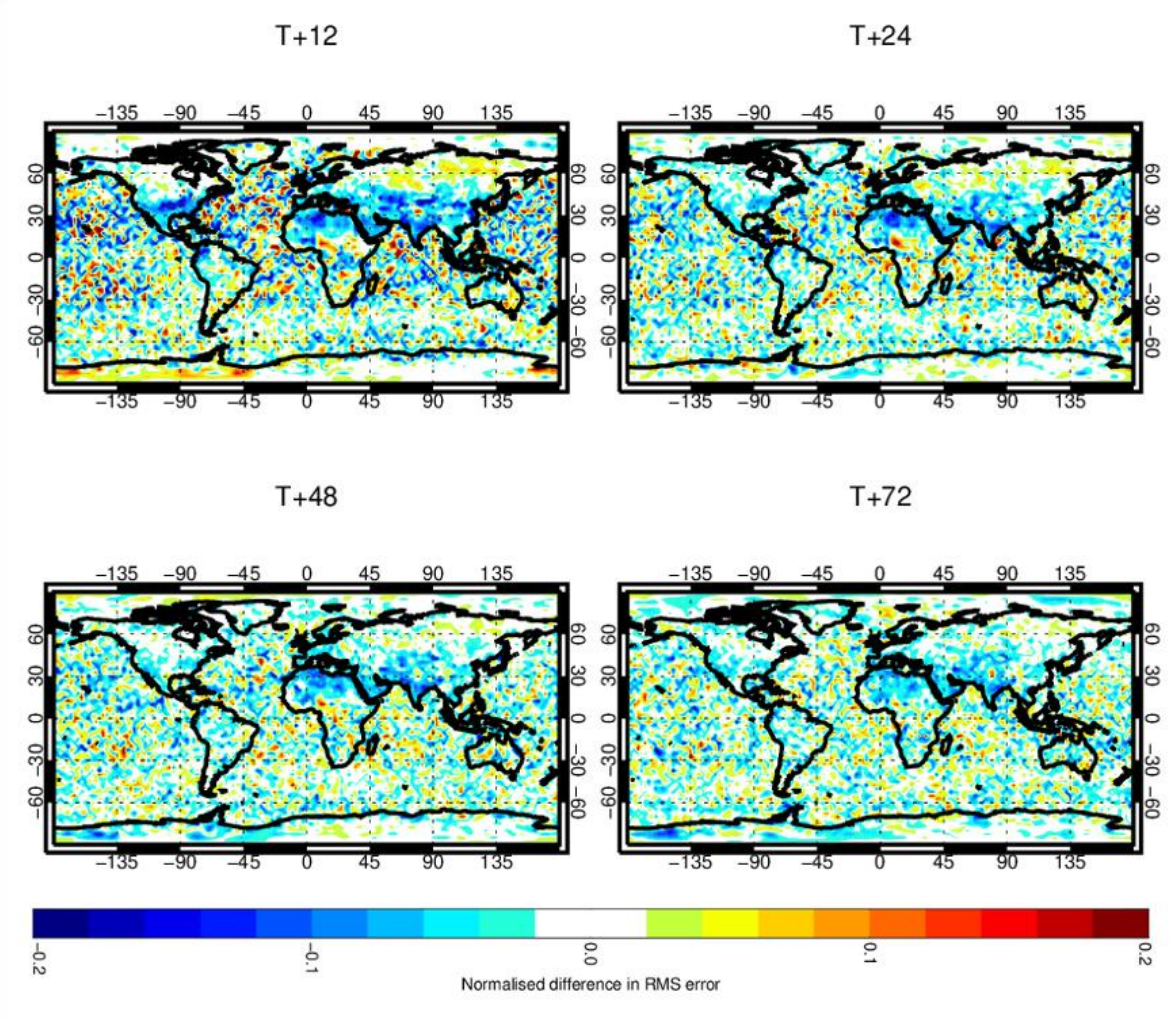

Figure 5: Global maps—gridded at a resolution of 2.5◦ × 2.5◦—of normalized dRMS forecast error of T2m for 12-, 24-, 48-, and 72-h forecasts of the SEKF(I) experiment, compared with the 1D-OI control experiment. The forecast of each experiment is verified against its respective own analysis, and scores are evaluated over three-month periods of boreal summer from June 1–August 31, 2022.

To assess the impact throughout the vertical column of the atmosphere in addition to the near-surface regions, the capability to predict the profile air temperature T is evaluated for different pressure levels. **Figure 6** illustrates the latitude–pressure diagrams of the relative forecast errors of the SEKF configuration compared with the 1D-OI, again for up to 72-h forecast lead times given over the boreal summer. The impact in forecast skill is largely consistent throughout all forecast days and widely limited to the NH, showing significant near surface improvements (blue hatched areas) in the range of 15–45°N, covering large parts of northern Africa, Europe, and North America. The effects are consistently positive near the surface, whereas at lead times of more than 24 h, some hatched areas associated with degradation and improvement can be observed.

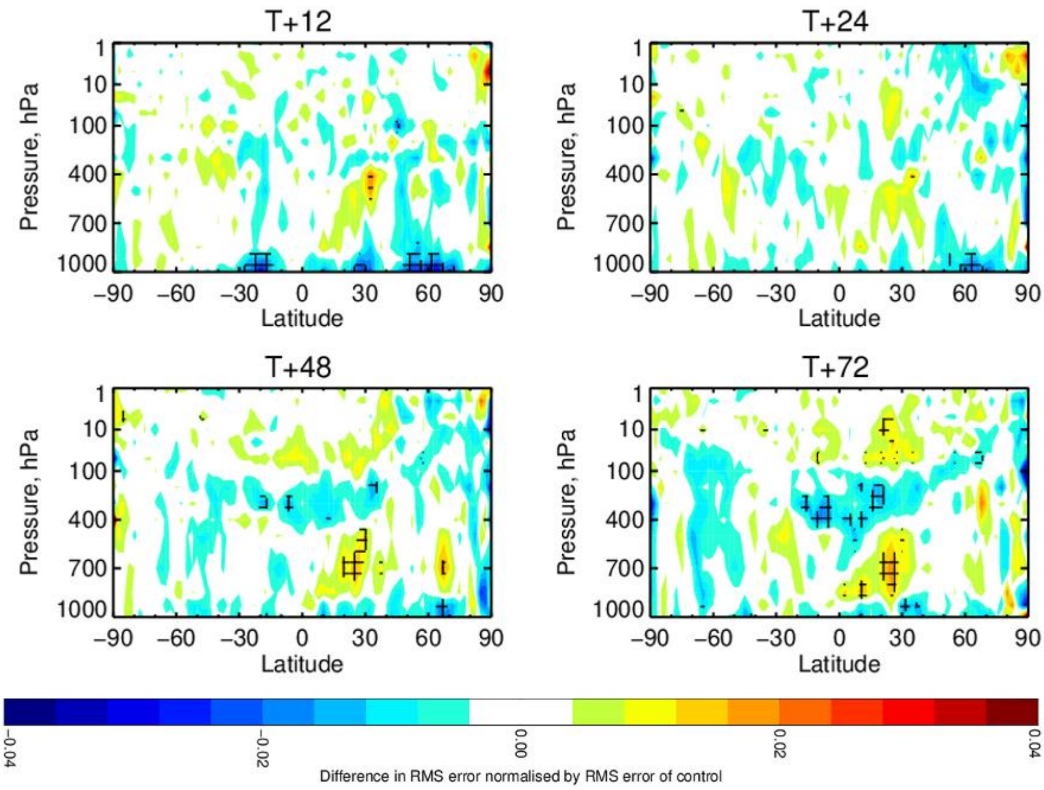

Figure 6: Latitude–pressure diagrams of normalised dRMS forecast error of T for 12-, 24-, 48-, and 72-h forecasts of the SEKF(I) experiment, compared with the 1D-OI control experiment. The forecast of each experiment is verified against its respective own analysis, and scores are evaluated over three-month periods of boreal summer from June 1–August 31, 2022.

4.2 2-metre temperature and relative humidity

As shown in Section 3, 2-metre temperature and relative humidity observations are currently assimilated in a 2D-OI, the output of which is assimilated into the SEKF. One approach to unifying the land data assimilation system which has been explored is to remove the 2D-OI completely and instead rely on the atmospheric 4D-Var to analyse the screen-level variables and provide input to the SEKF. This builds on recent work to assimilate 2-metre temperature observations into the atmospheric 4D-Var (Ingleby et al, 2024) in addition to the existing assimilation of 2-metre relative humidity observations.

Experiments have been run to replace the 2D-OI pseudo-observations used in the SEKF with pseudo-observations generated from 4D-Var. More details of these experiments can be seen in WP2 deliverable D2.1 but the results were mixed. On the one hand, there are local improvements over areas where there is sparse coverage of in-situ observations e.g. central Africa. In such areas the screen-level analysis benefits from the assimilation of surface sensitive satellite observations in 4D-Var, which are unavailable in the 2D-OI. However, there are also degradations over other areas, which could be partially due to the in-situ observations not being given as much weight in the atmospheric 4D-Var as in the 2D-OI. Another difference between these approaches is the resolution, the atmospheric 4D-Var minimisation runs at significantly lower resolution (~80km in these experiments) than the 2D-OI (~25km). Therefore, this approach is not viable for use in the operational system or demonstrators yet. However, there will be an increase in the resolution of the minimisation in 2025 and the potential use of the extended control variable framework to additionally analyse surface variables in the 4D-Var could be avenues to improve the results and eventually move away from the 2D-OI screen-level analysis.

5 Ensemble-based land data assimilation

In addition to unifying the Land Data Assimilation System (LDAS), as specified in Section 4, the CERISE project also aims to increase the use of ensemble information within the LDAS. Currently information from the ECMWF Ensemble of Data Assimilations (EDA) is used to construct the Jacobians required in the SEKF. However, background information is treated as static with a constant specification for the background error covariance matrix. Using the EDA to construct a flow-dependent ensemble-based background error covariance matrix to capture the “errors of the day” presents a significant improvement. The ability of the EDA to suitably capture spread at the surface is also an important aspect to understand.

5.1 Ensemble-based background errors

The SEKF for LDAS at ECMWF has static background errors. As part of Task 1.2 we have explored the inclusion of a flow dependent contribution to the background errors in the SEKF using the spread across the full EDA to capture "errors of the day" and build toward an ensemble-based LDAS. Last year, we prototyped, including the flow-dependent contribution only for soil moisture in the SEKF, this showed promising results in surface scores for 2-metre temperature and relative humidity. In order to capture these benefits in Cycle 49r1 at ECMWF, with a reduced development time, we included an inflation of the background error by a factor of 2. Thanks to other developments in Task 1.3 we now have soil and snow temperature in the control vector of the SEKF, this has allowed us to prototype a flow-dependent contribution in the background error for these variables in addition to soil moisture. In a set of factorial experiments, we explored adding in the ensemble-based background errors for a combination of soil moisture, soil temperature and snow temperature, we found that including the ensemble information for all three variables and all three soil layers was most beneficial when judged in a full-suite scorecard validation. The results of including the flow-dependent background errors in the SEKF are still broadly neutral when judged against the already inflated background errors but show good improvements against observations of surface 2-metre temperature and relative humidity. These improvements are shown in Figure 7 where we have improved fits of short-range forecasts to the Metar and Synop station observations of 2-metre temperature and humidity when incorporating flow-dependent ensemble information.

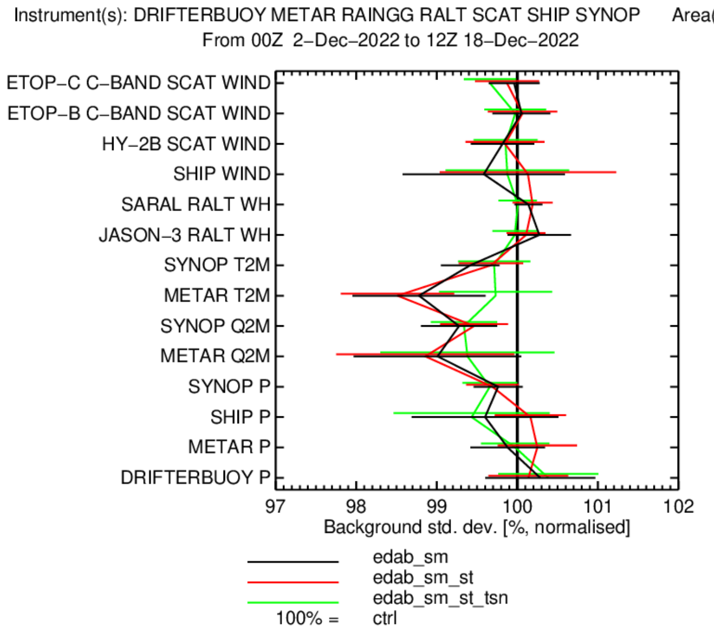

Figure 7: Normalised first-guess fits to surface observation for experiments testing the use of the EDA background errors for soil moisture (black), additionally soil temperature (red), and additionally snow temperature (green) compared to the IFS control. Scores below 100 demonstrate an improved fit to observations.

5.2 Investigate Land Surface Perturbation Methods

The EDA at ECMWF suffers from under-dispersed surface fields, this can limit the effectiveness of land surface DA. In order to increase spread at the surface across the ensemble we have experimented with the perturbation of land parameters (e.g. vegetation fraction, leaf area index) following a Stochastically Perturbed Parameter (SPP) methodology, adding spatiotemporally correlated noise. Initially these perturbation strategies were tested in an offline ensemble of land models (forced atmosphere with no feedback). The land model ensembles were run for a 2-year spin-up period and a full year's integration with and without the inclusion of perturbations in vegetation fraction and leaf area index across the ensemble. These experiments showed good increases in spread (~20%) in a variety of surface fields such as 2-metre temperature, dewpoint, soil moisture and even variables such as snow depth where the change in bare soil fraction of the grid cell allowed for more/less melting of snow. After these initial "offline" results some coupled EDA experiments were run including the most promising perturbations in the surface parameters. These full EDA experiments are slow to spin-up but show broadly similar results to the offline experiments. The results shown in Figure 8 are from the initial coupled EDA experiments with the additional land parameter perturbations run for a 3-month integration. The globally averaged spread across the ensemble is greater for all variables after inclusion of surface parameter perturbations. A set of 6-month long EDA experiments has now been completed to ensure that the spin-up in spread we see from Figure 8 saturates and does not lead to ever increasing growth in the spread for lower soil layers.

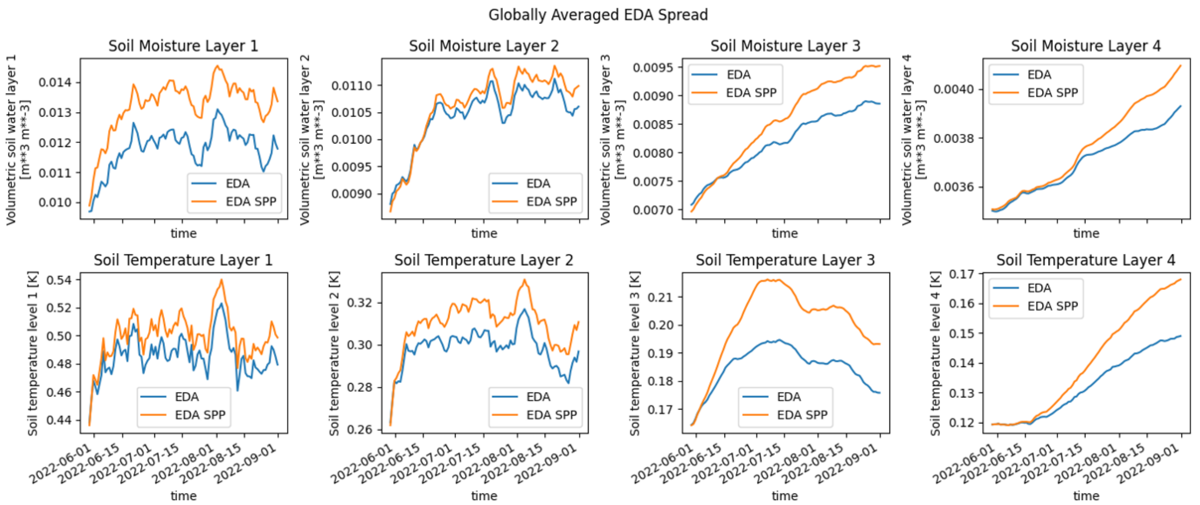

Figure 8: Globally averaged EDA spread for 8 land surface variables using the plan EDA (blue) and the EDA with SPP perturbed vegetation (orange) over a 3-month period

At ECMWF we have also started to explore more novel methods to generate ensemble/adjoint information for the surface fields using Machine Learning (ML). We have created an extensive land model database in Zarr format, comprising of model forcing, prognostic and diagnostic variables, for use in training a ML emulator of the land surface which could run at a much-reduced computational expense compared to the full ECLand model. The model is a Multi-Layer Perceptron (MLP) trained in Pytorch (after initial experimentation with gradient boosted decision trees) that allows for automatic calculations of gradients and adjoints. These preliminary results indicate that the emulator is showing promising performance, being able to accurately capture the evolution of multiple variables at long time horizons (~5 years) without much divergence from the full model fields. This presents an opportunity for land DA to generate cheap ensemble information or for explicit calculation of model Jacobians and sensitivities in the cost function. Figure 9 shows an integration of the ML emulator of the ECLand model over multiple years at a single grid-point for two differently behaving parameters (soil moisture level 1 and snow cover), demonstrating the ability of the emulator to capture the full behaviour of the ECLand model. We can see that the emulator is incredibly stable, showing no signs of divergence from the full physical model even after this 5-year integration. This behaviour bodes well for applications within land reanalysis products and long-term DA systems. The emulator can currently run a 1-month 30 km global integration in 500 milliseconds on a single GPU compared to 20 minutes for the full ECLand on 16 CPUS. Offering a potential speedup of over 2000 times over the current model runtime.

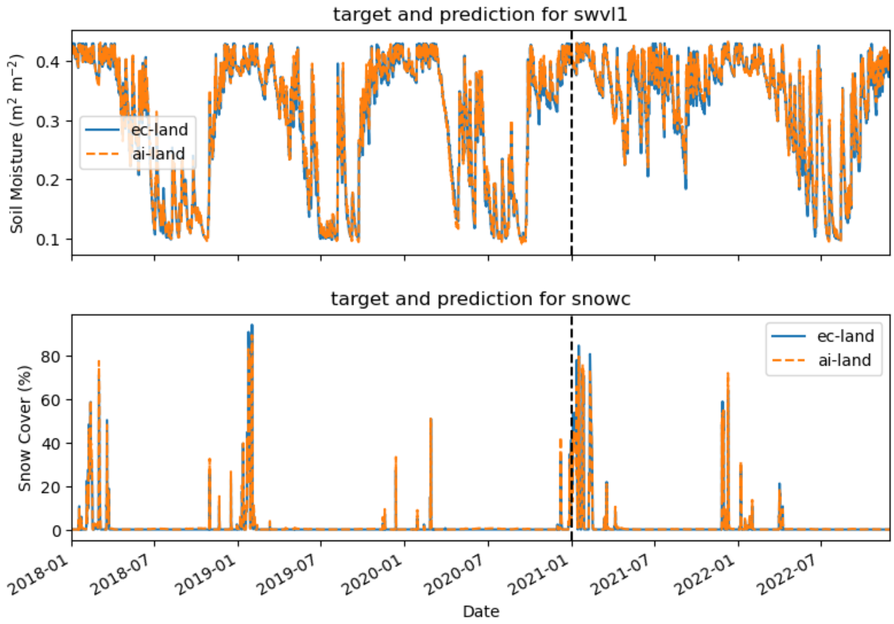

Figure 9: Machine Learning emulator and ECLand model predictions at a single grid point for soil moisture (top) and snow cover (bottom)

5.3 SEKF in the EDA

When the SEKF was first implemented at ECMWF (de Rosnay et al, 2013) it was used consistently in both the deterministic and ensemble of data assimilations (EDA) systems. This configuration was used in the operational system between 2010 and 2019, as well as in ERA5. However, in 2019 with the implementation of cycle 46r1 the number of EDA members was increased from 25 to 50, which meant that it became too computationally expensive to run the SEKF in every member of the EDA. Therefore, the SEKF was switched off in the EDA with the soil moisture analysis reverting to the previously used 1D-OI assimilation methodology.

Since then, there have been several improvements made to the SEKF to make it more computationally efficient. Firstly, rather than running multiple perturbed trajectories and using finite-difference Jacobians to link the observed variables to the model variables, in the current system the EDA spread is used to compute the Jacobians meaning a single unperturbed trajectory is run within the SEKF. Secondly, as part of WP2 task 2.1, there is now the option to run the SEKF without re-running the nonlinear trajectory. Instead, the required fields from the main trajectory are read into the SEKF. With these improvements in the efficiency of the SEKF, it is now viable to switch it back on in the EDA.

Several EDA experiments have been run to test the impact of using the SEKF in the EDA against a control using the 1D-OI. The first set of experiments were based on cycle 49r1 and the second set of experiments included the developments summarised in Section 4.1 to add snow and soil temperature in the SEKF control vector. Here, only results from the second set of experiments will be shown.

Table 3 shows the experiments which have been run including EDA controls and experiments with the SEKF switched on. In addition, deterministic experiments using the ensemble fields from the two sets of EDA experiments have been run to gauge the impact on the deterministic forecasts. All the experiments have been run for two seasons: boreal summer 2022 (1st June 2022 to 31st August 2022) and winter 2022/23 (1st December 2022 to 28th February 2023).

Table 3: Summary of EDA and HRES experiments run to test switching on the SEKF in the EDA

Name	Description	EDA experiment
EDA ctrl	EDA control at 49r2	Self
EDA SEKF	As EDA ctrl with SEKF switched on	Self
HRES ctrl	HRES control at 49r2	EDA ctrl
HRES SEKF	As HRES ctrl with ensemble information from EDA SEKF	EDA SEKF

Firstly, the impact on the EDA control member forecasts will be assessed. Figure 10 shows large improvements to the forecasts of near-surface temperature and humidity of up to ~6% at short-range which are statistically significant at the 95% confidence interval to 4 or 5 days in the NH and to 3 days in the SH. This illustrates that the SEKF is a significantly better DA methodology than the 1D-OI for soil moisture, soil temperature and snow temperature. This is consistent with the findings in deterministic experiments in Section 4.

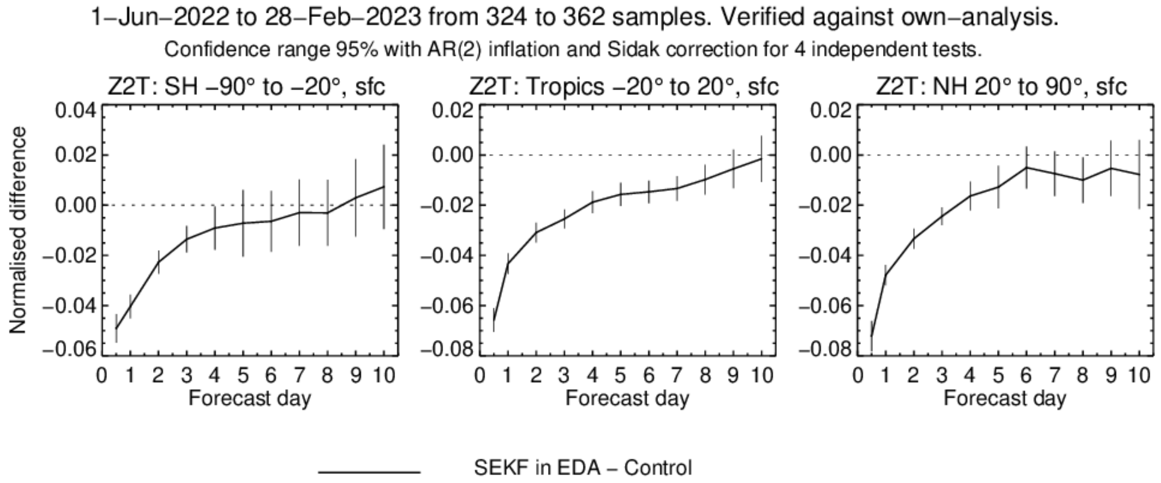

Figure 10: Normalised change in EDA control forecast RMSE for 2 metre temperature in the Southern hemisphere (left), tropics (centre) and Northern hemisphere (right). Statistics are averaged over the periods 1st June 2022 to 31st August 2022 and 1st December 2022 to 28th February 2023

As mentioned in Subsection 5.2 the existing EDA surface fields are under-spread leading to a large mismatch between spread and error in surface and lower atmospheric variables. One method for improving this relationship is to increase the spread as discussed in subsection 5.2. A second method is to reduce the error. Figure 11 shows that in the experiments where the SEKF is switched on in the EDA, the errors for near-surface temperature and especially humidity are significantly reduced, bringing the error closer to the spread.

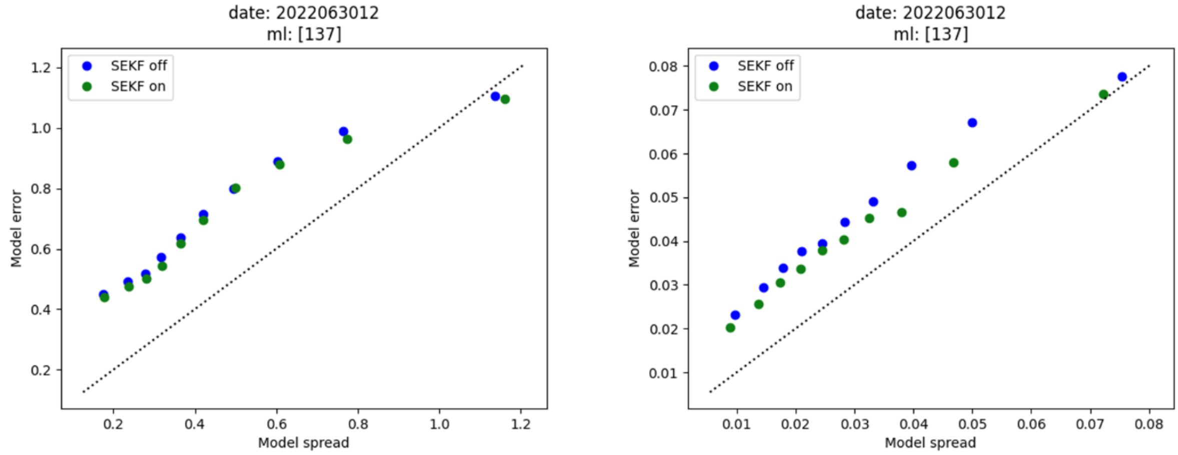

Figure 11: Scatter plots showing deciles of the spread-skill relationship with EDA spread on the x-axis and EDA error on the y-axis. Blue dots represent the values in the EDA control and green dots represent the values in the EDA with SEKF switched on. Lowest model level temperature is on the left and lowest model level relative humidity is on the right. Values are taken for the DA window between 09UTC and 21UTC on 30th June 2022 after the experiments have spun-up for 1 month

Finally, the impact of switching the SEKF on in an EDA on an accompanying deterministic (HRES) experiment can be assessed. The EDA provides flow dependent background errors to the HRES assimilation system in the atmosphere, and also from cycle 50r1 for the land-surface as detailed in Section 5.1. It also provides the Jacobians to link the observed and model variables within the SEKF. Figure 12 shows that there are small but significant reductions in forecast errors of near-surface temperature and humidity of around 1% which are significant out to 4 or 5 days in the NH. These impacts extend up to 850hPa for temperature (not shown). Figure 13 also shows that the first-guess fits against independent observations such as in-situ screen-level temperature and humidity (SYNOP & METAR T2M & Q2M), and near-surface radiosonde temperature are improved. There are also improved fits to near-surface aircraft temperature observations (not shown). This provides more evidence that the short-range near-surface temperature and humidity forecasts are improved as a result of this change.

These impacts on the deterministic forecasts are smaller than those seen directly in the EDA because the SEKF is already used in the deterministic system, it is only the impact of the improved ensemble fields which is being measured here. Still, this represents a decent improvement to the deterministic forecasts, in key variables for both NWP and reanalyses.

The results shown in this section indicate that there are significant improvements gained by switching the SEKF on in the EDA. Since 2-metre temperature is a key variable for users of reanalysis, this configuration is being tested for potential use in ERA6 as well as the ERA7-pv1 & 2 demonstrators. In addition, it is aimed to switch on the SEKF in the EDA in the ECMWF operational system with the implementation of cycle 50r1 in 2025.

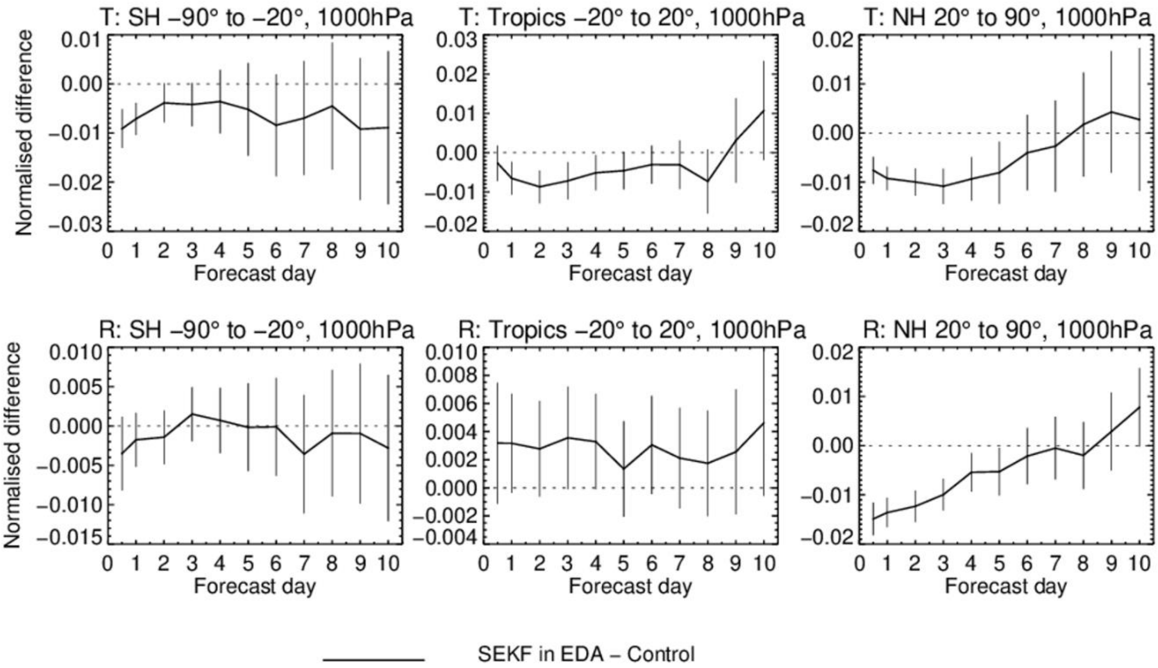

Figure 12: Normalised change in forecast RMSE for 1000hPa temperature (upper) and humidity (lower) in the Southern hemisphere (left), tropics (centre) and Northern hemisphere (right). Statistics are averaged over the periods 1st June 2022 to 31st August 2022 and 1st December 2022 to 28th February 2023

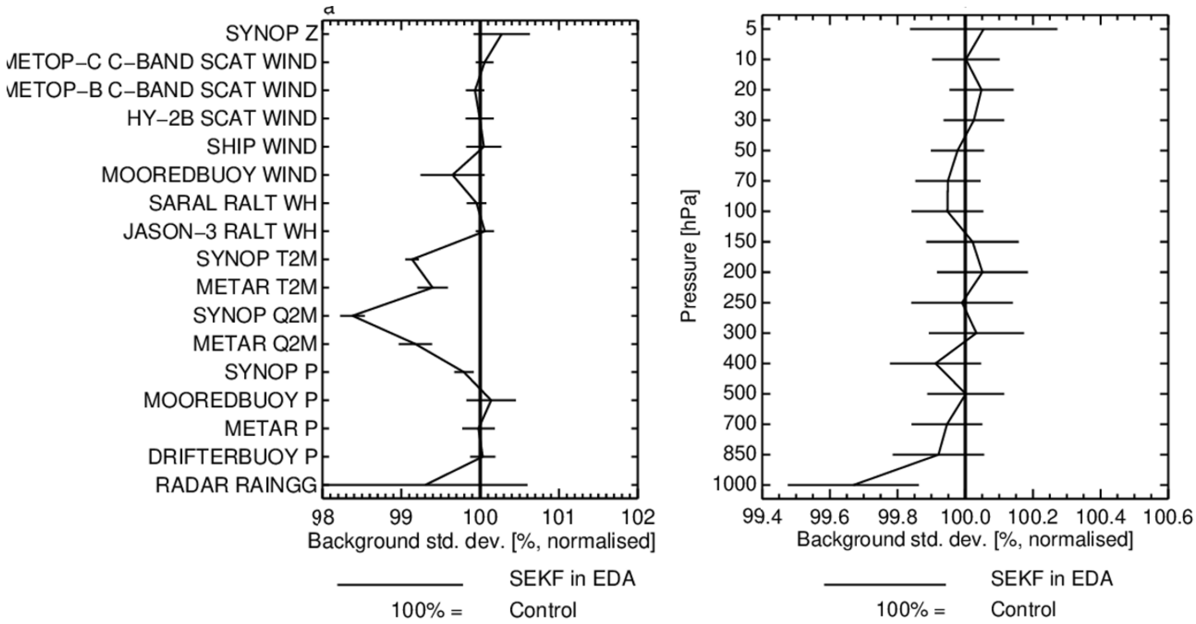

Figure 13: Normalised first-guess fits against surface (left) and radiosonde temperature (right) observations. Values less than 100% indicate an improved fit between short-range forecasts and independent observations. Statistics are averaged over the periods 1st June 2022 to 31st August 2022 and 1st December 2022 to 28th February 2023

6 Conclusions

In this report, progress towards a more unified global land data assimilation system making better use of ensemble information within on the CERISE project have been summarised. This included a considerable amount of technical work and tuning of assimilation configurations to adapt the existing system and include additional variables in the SEKF control vector. Adding soil and snow temperature to the SEKF control vector leads to significant improvements in near-surface temperature forecasts out to 3 or 4 days. In addition, it allows the previously used 1D-OI to be retired, simplifying the maintenance and further development of the LDAS. Initial testing of further unification efforts for 2-metre temperature and relative humidity have yielded mixed results which need to be understood further before moving away from the existing 2D-OI methodology. This work will continue in close collaboration with the work in WP2 on outer-loop coupling of the land and atmospheric data assimilation systems.

The global LDAS now makes better use of the ensemble information by including a contribution from the EDA errors of the day into the assumed background errors used in the SEKF. Assessment of the EDA spread has already led to significant improvements from doubling the background error in the SEKF, which has already been implemented in the operational system at cycle 49r1. In addition to this, two approaches to improving the spread-error relationship in near surface variables have been investigated. Firstly, increasing the spread by perturbing land-surface parameters such as vegetation cover and leaf area index. Secondly, by enhancing the LDAS used in the EDA by moving from the 1D-OI to the SEKF methodology. Both approaches improve the spread-error relationship, meaning a more reliable ensemble system which can be exploited more in the future. Switching the SEKF on in the EDA also yields significant forecast improvements to near-surface temperature and humidity.

Overall, there are both technical and scientific benefits from moving towards a more unified ensemble-based LDAS. The updated system will be used in the WP4 reanalysis demonstrators with further assessment and evaluation happening as part of WP6 which may lead to further enhancements and refinements of the exact LDAS configuration.

7 References

IFS 49r1 documentation (soon to be available from https://www.ecmwf.int/en/publications/ifs-documentation)

de Rosnay, P., Drusch, M., Vasiljevic, D., Balsamo, G., Albergel, C. and Isaksen, L. (2013), A simplified Extended Kalman Filter for the global operational soil moisture analysis at ECMWF. Q.J.R. Meteorol. Soc., 139: 1199-1213. https://doi.org/10.1002/qj.2023

Geer (2016) Significance of changes in medium-range forecast scores. Tellus A: Dynamic Meteorology and Oceanography, 68(1). https://doi.org/10.3402/tellusa.v68.30229

Gómez, B.; Charlton-Pérez, C.L.; Lewis, H.; Candy, B. The Met Office Operational Soil Moisture Analysis System. Remote Sens. 2020, 12, 3691. https://doi.org/10.3390/rs12223691

Herbert C., P. de Rosnay, P. Weston, and D. Fairbairn: Towards unified land data assimilation at ECMWF: Soil and snow temperature analysis in the SEKF, QJRMS 2024, DOI: https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.4808

Ingleby, B., Arduini G., Balsamo G., Boussetta S., Ochi K., Pinnington E., de Rosnay P. Improved two-metre temperature forecasts in the 2024 upgrade. ECMWF Newsletter, No. 178, Winter 2023-24, https://www.ecmwf.int/sites/default/files/elibrary/12024/81535-newsletter-no-178-winter202324.pdf

Mahfouf, J.-F., Viterbo, P., Douville, H., Beljaars, A. & Saarinen, S. (2000) A revised land-surface analysis scheme in the integrated forecasting system. ECMWF Newsletter, 88(88), 8–13.

8 Annex I:

Document History

Version	Author(s)	Date	Changes
1.0	Pete Weston, Christoph Herbert, Ewan Pinnington, Patricia de Rosnay	28/11/2024	Initial version
1.1	Pete Weston, Christoph Herbert, Ewan Pinnington, Patricia de Rosnay	16/12/2024	Revised following internal review

Internal Review History

Internal Reviewers	Date	Comments
Jelena Bojarova, Eshagh Medhi (SMHI) and Giovanni Conti (CMCC)	Dec 2024	Initial version

This publication reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained therein.