**CopERnIcus climate change Service Evolution**

# D2.1 Documentation of coupled assimilation infrastructure and methodology and preliminary assessment towards optimal degrees of coupling for coupled global reanalysis

| | |
|---|---|
| Due date of deliverable | 31/12/2024 |
| Submission date | Dec 2024 |
| File Name | CERISE-D2-1-V2.0 |
| Work Package /Task | WP2 Tasks 2.1 and 2.1 |
| Organisation Responsible of Deliverable | ECMWF |
| Author name(s) | Patricia de Rosnay, Christoph Herbert, David Fairbairn, Ewan Pinnington, and Peter Weston |
| Revision number | 2.0 |
| 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 develop the next generation of C3S (Copernicus Climate Change Service) global scale reanalysis codes infrastructure to support modular coupled assimilation and monitoring. In the ECMWF Integrated Forecasting System (IFS), which is used to product the C3S global reanalysis, we developed and tested innovative coupled surface-atmosphere assimilation. This document presents the infrastructure and the coupled data assimilation methodology that was developed in the IFS, and it presents preliminary results. More balanced initial conditions between the different outer loops showed an overall positive impact on atmospheric near surface forecasts and improved fit to independent atmospheric data. However, verification against Land Surface Temperature data shows degradations in terms of skin temperature forecast skills This will be further investigated with Extended Control Variable soil temperature coupled data assimilation to constrain skin temperature at the land-atmosphere interface.

# Table of Contents

## Contents

# 2 Introduction

## 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 4-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 data assimilation 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 from the Copernicus Sentinels and from the European Space Agency (ESA) Earth Explorer missions, 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, 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 datasets (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 of this deliverables

This deliverable describes the IFS coupled assimilation infrastructure and methodology, and it presents preliminary assessment towards optimal degrees of coupling for coupled global reanalysis.

### 2.2.2 Work performed in this deliverable

In this deliverable the work outlined in WP2 T2.1: *Infrastructure developments to support coupled DA developments, and code efficiency and modularity,* and in WP2 T2.2 *Development of outer loop land-atmosphere coupling* is described and evaluated.

The current land-atmosphere coupling approach used for the C3S global reanalysis, consists of a weakly coupled DA system, where results from atmospheric and land surface analyses are fed back into the next 12-hour data assimilation window by the short connecting forecast. In the context of WP2 Task 2.1, we conducted infrastructure developments in the ECMWF IFS to improve the system efficiency and to prepare for coupled data assimilation workflow developments based on an outer coupling approach as described in de Rosnay et al (2022). As part of WP2 Task 2.2, we developed a quasi-strongly coupled land-atmosphere DA system based on outer loop coupling. These developments give the capability of activating the land data assimilation as part of several 4D-Var outer loops and to initialize the atmosphere and the land components of subsequent outer loops with updated land analyses within the same assimilation window. The methodology and preliminary results are presented here.

### 2.2.3 Deviations and counter measures

No deviations have been encountered.

### 2.2.4 Reference Documents

[1] Project 101082139- CERISE-HORIZON-CL4-2021-SPACE-01 Grant Agreement
[2] Project 101082139- CERISE-HORIZON-CL4-2021-SPACE-01 D7.2 Albedo, vegetation and LST satellite datasets in the CERISE verification database
[3] Project 101082139- CERISE-HORIZON-CL4-2021-SPACE-01 D1.2 Unified, ensemble-based global land data assimilation system and documentation

### 2.2.5 CERISE Project Partners:

| | |
|---|---|
| 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 Infrastructure developments for code efficiency and modularity

Infrastructure developments were conducted in the ECMWF IFS to improve the system efficiency and to prepare for outer loop coupling data assimilation workflow developments.

## 3.1 Code efficiency

In the existing weakly-coupled data assimilation system the atmospheric 4D-Var and land-surface data assimilation systems (LDAS) run separately providing analyses of the atmospheric and land-surface variables which are then used to initialise coupled forecasts (de Rosnay et al, 2022). Despite running separately, the most advanced part of the LDAS, the simplified extended Kalman filter (SEKF), shares much of its code with the atmospheric 4D-Var. In particular, the SEKF runs its own full non-linear trajectory to provide the background and first-guess fields for the land-surface analysis.

There are two approaches to making the SEKF more efficient. Both approaches require significantly fewer resources than the current operational configuration because they both avoid re-running the full non-linear trajectory which is the most computationally expensive part of the SEKF. Firstly, it can be run as part of the main 4D-Var task, using the existing fields which are in memory from running the original first trajectory. This approach effectively merges the SEKF and the main trajectory. Secondly, the main 4D-Var trajectory can write out the fields required but the SEKF itself remains a separate task.

The first approach was developed and implemented in CERISE as part of milestone M3 and was used as the baseline configuration for the developments carried out in section 4. However, this had a few drawbacks. Firstly, it didn't work under the object-oriented prediction system (OOPS, English et al, 2017) which is now used to run the operational assimilation system at ECMWF. This meant that all of the experiments documented in section 4 were run with OOPS switched off. Secondly, there were concerns over the robustness of the merged system for future operational applications, for example a problem in the SEKF could lead to a delay in the production of the entire atmospheric analysis. Thirdly, the SEKF requires input from the 2D-OI (two-dimensional Optimal Interpolation) 2 metre temperature and relative humidity analysis which would, in the current configuration, go into the time critical path leading to delays in the assimilation system as a whole.

For these reasons, as part of the CERISE Task 2.1 work, the second approach has now also been developed and implemented within the OOPS framework. The main 4D-Var trajectories now output the required fields to provide the background, first-guess and quality control for the SEKF. This should allow a more flexible coupling configuration in addition to addressing concerns about the robustness of the system whilst not affecting the time critical path.

## 3.2 Code modularity

The capability to assimilate screen level observations along with soil moisture and snow cover data in the offline Land Data Assimilation System (LDAS) was also developed in line with the coupled approach. These developments ensure seamless coupled reanalysis streams initialisation from a single stream land reanalysis. They also enable full consistency of reforecast and real time streams of seasonal prediction systems.

# 4 Outer loop land-atmosphere coupling

## 4.1 Overview

Developments focused on outer loop coupled assimilation methodology for consistent land-atmosphere assimilation for global reanalysis purposes. In the ECMWF IFS system, we implemented the unified LDAS from WP1 (Herbert et al., 2024) under the outer loop level of 4D-Var, following the approach developed for ocean-atmosphere coupling (Laloyaux et al., 2016) and following the strategy defined in de Rosnay et al., 2022. Several configurations were tested to identify the optimal number of outer loops required for optimal land-atmosphere coupled assimilation by running short term numerical experiments with the global IFS.

## 4.2 Methodology

The current operational analysis relies on 12-hour data assimilation (DA) windows, within which land and atmospheric DA systems are weakly-coupled. The atmospheric data assimilation system uses a 4D-Var (four-dimensional variational) approach (Rabier et al., 2000), and the land surface data assimilation includes a 2D-OI screen analysis and a SEKF soil analysis (de Rosnay et al., 2022). The latter requires to run a separate coupled model trajectory initialized from the atmospheric and land background fields. The control configuration and starting point of the outer loop coupling developments have the SEKF running as part of the first outer loop (*uptraj0*) in a shared trajectory with atmospheric 4D-Var **(Figure 1).** The SEKF can be flexibly activated in any outer loop whereas the respective trajectory – used in both 4D-Var and SEKF – is initialized by the updated atmospheric fields, while still relying on the land background fields (one-way coupling). Two-way coupling is achieved by also feeding back updated land conditions to subsequent outer loop(s) after the first SEKF has been run. Outer loop coupling falls in the category of 'quasi-strong' coupling (Penny et al., 2017). The trajectory of an outer loop relies on both the most updated atmospheric and land surface conditions that are fed back from previous outer loop(s) within the same DA window.

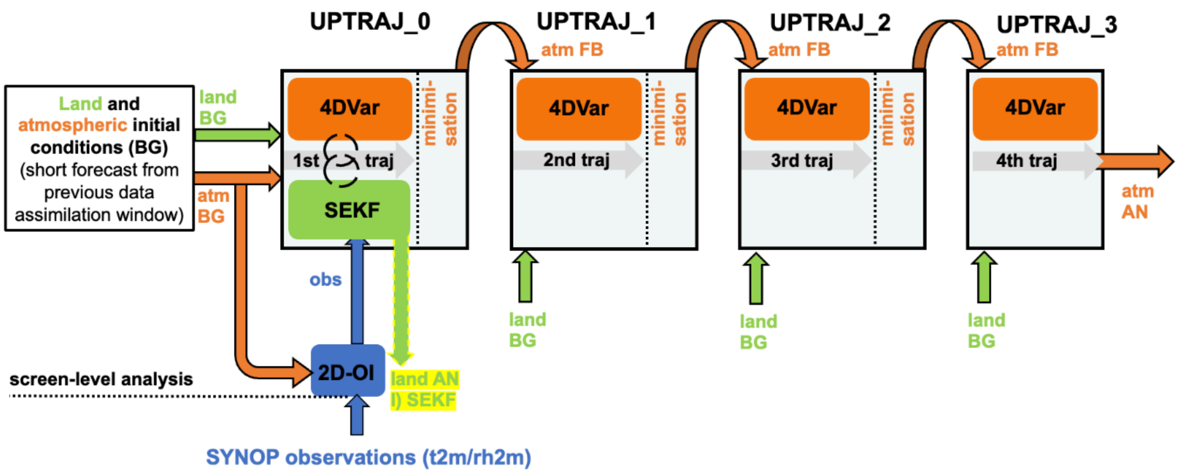


*Figure 1: Control configuration used as baseline, with the SEKF activated once per 12-hour assimilation window, in the first outer loop (uptraj0) without land feedback. Gridded T2m/RH2m analysis fields based on the 2D-OI are assimilated and land analysis fields are directly stored in the SEKF.*

## Outer loop coupling developments

We have carried out the technical development and scientific evaluation of different outer loop coupling configurations to find the optimal degree of land-atmosphere coupling. Both the developments regarding merged SEKF/4D-Var coupled trajectories and outer loop coupling were all based on IFS Cycle 49R1 outside the operationally used OOPS environment in research mode, where model trajectory and minimization in an outer loop are run in separate tasks.

The current SEKF control vector contains three-layer soil moisture and soil temperature, respectively, and first-layer snow temperature (see WP1 deliverable D1.2 and Herbert et al., 2024). For each of the seven control variables an analysis increment is obtained at each grid point that is representative for a 12-hour DA window. The analysis increments are added to the initial background fields to produce updated first-guess fields that are picked up by the next outer loop(s). The updated fields are provided as initial conditions for the non-linear trajectories at the same spatial resolution. They are interpolated to lower-resolution grids required for the inner loops in the remaining atmospheric minimization(s). An example of an outer loop coupling configuration is given in **Figure 2**, where the SEKF is activated in *uptraj0* and updated land feedback is provided to initialise the next outer loop (green arrows at the bottom).

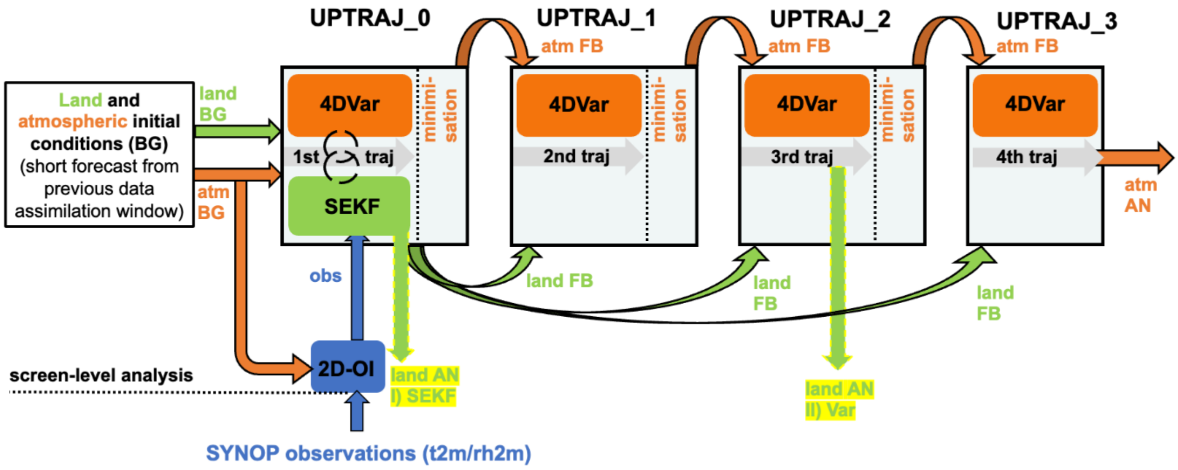


*Figure 2: Outer loop coupling configuration where the SEKF is activated in the first outer loop (uptraj0) and land feedback is passed on to the next outer loops. Gridded T2m/RH2m analysis fields based on the 2D-OI are assimilated and land analysis fields can be either I) directly stored in the SEKF or II) written out in the penultimate trajectory (uptraj2).*

Having the land feedback in place enables the final land analysis fields to be created in two different ways (see **Figure 2,** output marked in yellow): Firstly, directly in the SEKF at the respective analysis times and secondly, written out in the trajectory of a later outer loop. With the latter approach, the SEKF becomes identical to a simplified 2D-Var method (Mahfouf et al. 2009), where land variables evolve using a trajectory, accounting for the increment information used to initialise the trajectory. In the outer loop coupling framework, this setup has the advantage that land variables provided in later outer loops are based on more updated atmospheric conditions, and the land fields – stored and used in a short forecast to generate initial conditions at the start of the next DA window – are more consistent with those the final outer loop relies on. Our preliminary investigations (not shown) demonstrated that writing out the land variables in a later trajectory in a 2D-Var-like approach provides consistent improvement over storing the land analysis fields directly in the SEKF of a previously coupled trajectory. Therefore, this approach is used in the rest of the document.

The different configurations investigated involve activating the SEKF in one or several, earlier or later outer loops. We also compared assimilating in the SEKF the screen-level pseudo-observations from either the 2D-OI, or using updated fields produced in 4D-Var – with recent developments having the capability of assimilating SYNOP observations into 4D-Var (Ingleby et al., 2024). One of the key lessons learnt is the importance of properly incorporating the SYNOP observations to constrain the land variables to result the system in having a positive impact on the atmospheric forecast without negatively affecting the land surface variables. As will be presented in Section 4.3, most promising outer loop coupling setup relies on two-way coupling where the SEKF is activated in the first two outer loops and pseudo-observations are produced from multiple 2D-OI which use the updated T2m/RH2m first guess fields that are available in the respective trajectory (see **Figure 3**).

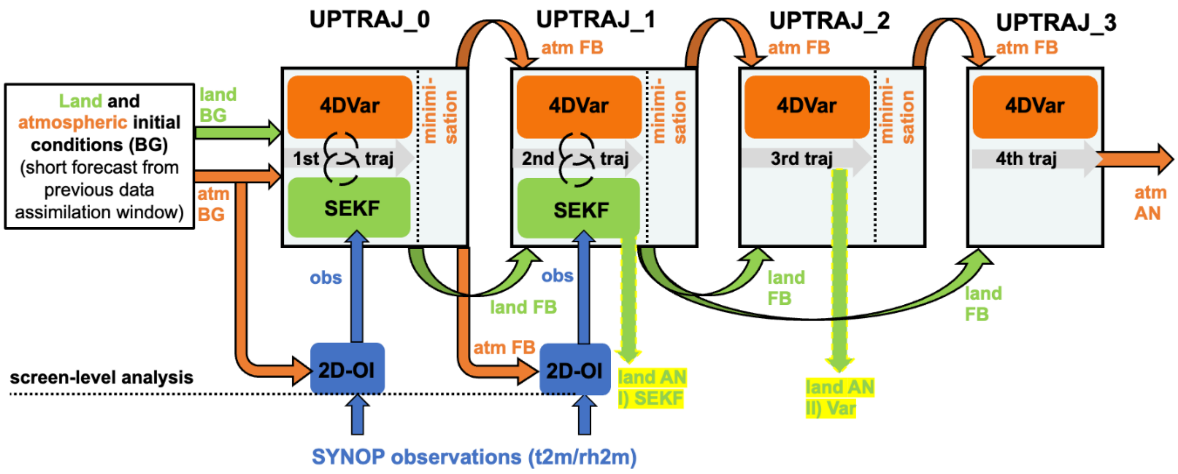


*Figure 3: Outer loop coupling configuration where the SEKF is activated in the first and second outer loops (uptraj0 and uptraj1) and land feedback is passed on to the next outer loops, respectively. Gridded T2m/RH2m analysis fields based on the 2D-OI (using updated atmospheric T2m/RH2m first guess) are assimilated and land analysis fields can be either I) directly stored in the second SEKF or II) written out in the penultimate trajectory (uptraj2).*

## Using 4D-Var based pseudo-observations

The land surface analysis strongly relies on screen-level observations, and the current operational system uses the 2D-OI to produce the gridded T2m/RH2m analysis to be assimilated in the SEKF. From IFS Cycle 49R1, SYNOP T2m/RH2m observations are assimilated into 4D-Var (Ingleby et al., 2024) and hence updated fields are produced in the trajectory of each outer loop. The outer loop coupling framework enables using these fields as pseudo-observations instead of those produced by the 2D-OI. So far in Cycle 49R1, T2m observations assimilated in 4D-Var are considered only in the first 6 hours due to increased temperature bias as observations are assimilated further into the DA window. Current developments include the activation of weak-constraint 4D-Var in the boundary layer, which comprises model-bias correction with a diurnal cycle that results in reduced bias allowing to assimilate SYNOP screen-level observations throughout the entire DA window. Besides having a technical advantage to potentially replace the current 2D-OI, the motivation behind testing the impact of using 4D-Var-based screen-level fields is that T2m/RH2m fields analyzed in 4D-Var not only contain the essential information from SYNOP observations, but are also constrained by satellite data, which can provide added value for the land surface analysis in areas where SYNOP observations are sparse.

The infrastructural benefit of using the 2D-OI is that it is performed in a separate task, at the same spatial resolution as the non-linear trajectories, and both SEKF and 2D-OI can be initialized consistently from the same first guess. In contrast in 4D-Var, the first analysis increment is produced in the first minimization, at lower spatial resolution, in *uptraj0* and the analysis increment is then added to the first guess in the beginning of the DA window of the next outer loop. This results in the analyzed fields to be only available for use in the SEKF in the following trajectory (*uptraj1*). Therefore, when using 4D-Var-based pseudo-observations, a staggered setup is required that is less consistent with respect to the outer loop in which the analysis fields are generated and in which they are assimilated by the SEKF. Several configurations were implemented to compare the impact of assimilated screen-level fields produced by the 2D-OI and generated in different 4D-Var outer loops. The 'cleanest' comparison between the two setups was obtained when activating the SEKF in *uptraj1* with using i) the fields from 2D-OI and ii) the fields from 4D-Var produced in the minimization in *uptraj0* and stored in the trajectory in *uptraj1*, with both approaches using the initial T2m/RH2m background. An example of a staggered setup with assimilating 4D-Var based fields in *uptraj1* is given in **Figure 4**.

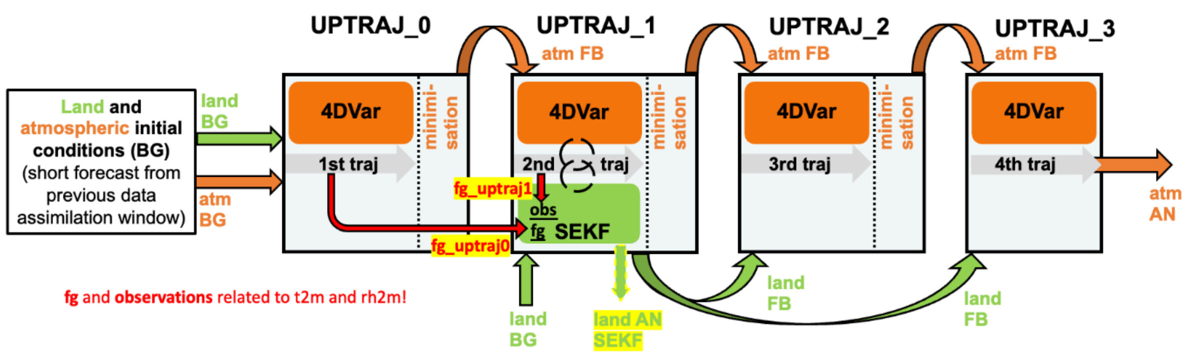


*Figure 4: Outer loop coupling configuration using 4D-Var based T2m/RH2m analysis as pseudo-observations in the SEKF activated in uptraj1.*

## 4.3 Experiments

In the framework of outer loop coupling, the impact of NWP experiments along boreal summer JJA 2022 and winter DJF 2022/23 are evaluated. The atmospheric forecast of experiments based on different configurations are compared to the control setup (see **Figure 1**) verified against ECMWF operational analysis and against near-surface and surface-sensitive observations. The following configurations are discussed.

I) Comparison between 4D-Var-based and 2D-OI-based pseudo-observations used in the SEKF that is activated in *uptraj1*, and
II) Performance of a different number of coupled outer loops using pseudo-observations based on multiple 2D-OI in the SEKF.

## 4.4. Results

Initial findings (not shown) revealed that activating the SEKF in *uptraj0* has the major advantage for the atmospheric forecast. The omission of the SEKF in *uptraj0* degrades scores no matter how many later outer loops are coupled. This implies that having the first outer loop coupled is beneficial and serves a good baseline towards the optimal degree of coupling.

In the configuration where the SEKF is activated within *uptraj0*, and updated land conditions are fed back to initialize subsequent outer loops, storing land analysis fields in a later trajectory (following a variational-type approach as explained in Section 4.2) results show improvement over storing field directly in the SEKF. In case the SEKF is also activated in later outer loops, the effect is largely neutral for most atmospheric forecast fields (see black vs red and green vs blue lines in **Figure 5** – T2m forecast error as a function of lead times for different configurations). This can be explained by the fact that the SEKF has already been initialized with updated, more accurate atmospheric conditions (two-way coupling).

Providing land feedback to subsequent outer loops without storing land variables as variational-type approach can slightly degrade scores, in particular if the SEKF is only activated in the first outer loop (not shown). This might be due to inconsistency of the land analysis fields stored in the SEKF which are used for the short forecast to provide initial conditions for the next cycle and those used in the atmosphere in subsequent outer loops of the same cycle.

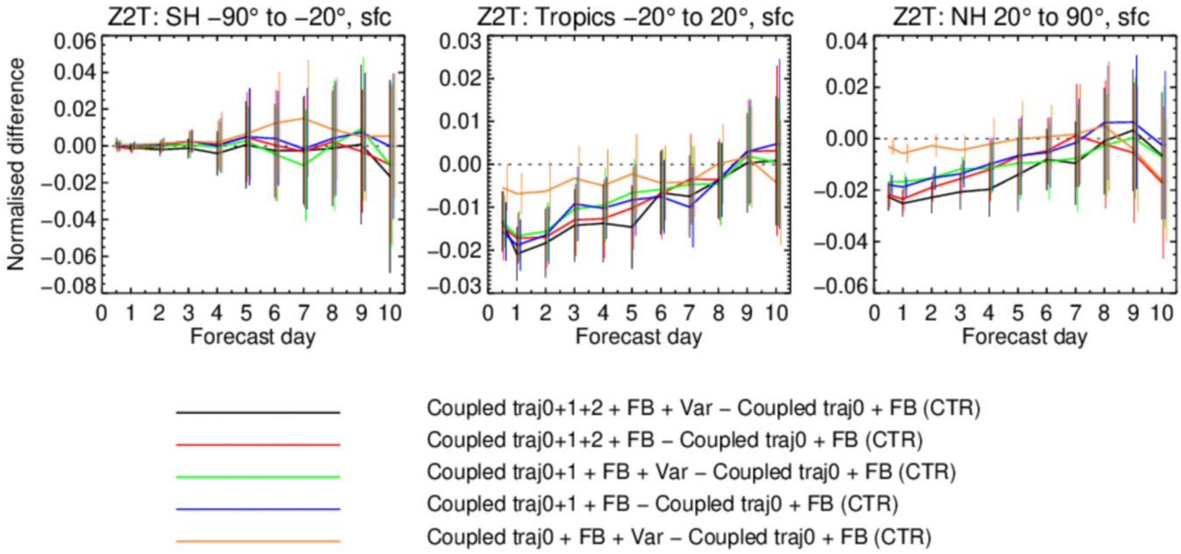


*Figure 5: Normalised RMS forecast error difference of T2m comparing configurations regarding different number of coupled outer loops, using land feedback only while storing land analysis variables in the SEKF (blue, red land black dashed line at zero), or variational-type methods (brown, green and black lines). Experiments are run over boreal summer period and results are compared to the control experiment with the SEKF in the first outer loop with no land feedback as per Figure 1 (black dashed line at zero).*

## 4D-Var vs. 2D-OI-based pseudo-observations

As aforementioned, T2m/RH2m analyses from the atmospheric 4D-Var can have a benefit of being reconstructed based on predicted variables that are constrained by satellite data. In contrast, if using 2D-OI-based analyses, the land is – apart from the background containing some forward-carried satellite information – in principle constrained by SYNOP observations. The impact of assimilating 4D-Var is compared to that of assimilating 2D-OI based T2m/RH2m fields as pseudo-observations in the SEKF running in coupled *uptraj1*. Hereby, the relative impact of the two configurations is tested on top of CY49R1 – where T2m assimilated in 4D-Var in first half of DA window – and on top of recent developments that include weak-constraint

4D-Var in the boundary layer, where screen-level variables are assimilated along the entire DA window.
Using 4D-Var fields results in a higher forecast error against observations compared to using those from the 2D-OI (**Figure 6**). In particular, the verification of the first-guess fields against near-surface in-situ measurements shows increased departures for METAR and SYNOP screen-level observations (red dashed box in **Figure 6c**). **Figure 7** shows the relative differences in forecast error of temperature at 1000hPa and T2m. Using 4D-Var-based pseudo-observations shows local improvement in forecast skill in the tropics in areas with sparse SYNOP observations but with overall no clear advantage over the 2D-OI in higher latitudes. Similar results were obtained when the two configurations were tested on top of enabling the weak-constraint 4D-Var in the boundary layer, showing only small but non-significant positive relative impact (not shown here).

Besides the inconsistency of the 'staggered setup' when using 4D-Var-based fields, two aspects could explain the negative impact. First, the T2m/RH2m increments obtained in 4D-Var are much smaller than those generated by 2D-OI and not constraining the land sufficiently, and second, for computing cost reasons the inner resolution used in the minimization of the first trajectory to obtain these increments is lower than the full experiment resolution used in 2D-OI. Thus, the use of 4D-Var based instead of 2D-OI-based fields requires further investigation.

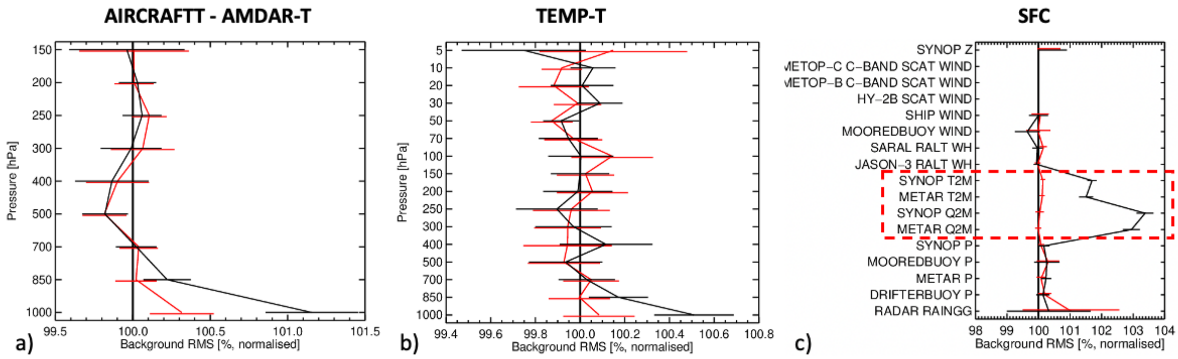


*Figure 6: Globally averaged relative differences in RMS first-guess departures regarding a) aircraft temperature, b) temperature from radiosondes and c) further near-surface-sensitive observations along the boreal summer, comparing configuration using 4D-Var-based (black line, see setup in Figure 4) and 2D-OI-based (red line) pseudo-observations in the SEKF outer loop coupled uptraj1 to the control with the SEKF activated in uptraj0.*

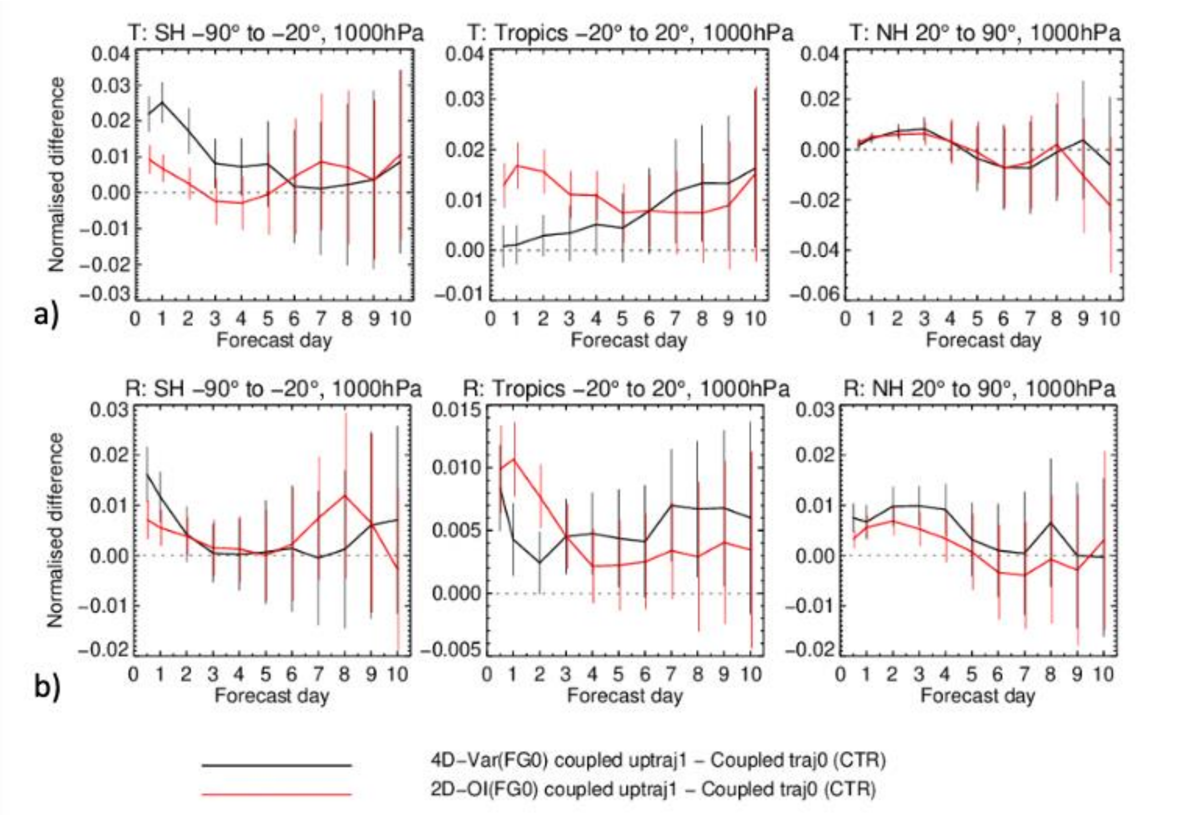


*Figure 7: Relative differences in RMS forecast error verified against operational analysis of a) Temperature at 1000hPa and b) T2m, averaged over the NH, the tropics and the SH along the boreal summer, comparing configurations – using 4D-Var-based (black line, see setup in Figure 4) and 2D-OI-based (red line) pseudo-observations in the SEKF outer loop coupled uptraj1 – to the control with the SEKF activated in uptraj0.*

## Different number of coupled outer loops with multiple 2D-OI

The results of testing different degrees of coupling using several 2D-OI with activated SEKF in *uptraj0* and later outer loops are presented below evaluating experiments during boreal winter 2022/23. All configurations store the analysis fields as in a variational approach.

**Figure 8** shows the relative differences in forecast error of T2m and temperature at 850hPa and 1000h, respectively. Having the SEKF activated in coupled *uptraj1* uses two-way coupling with updated atmospheric fields leads to improvement. In the tropics the configuration of coupled *uptraj0-2* outperforms those with less coupling showing significantly reduced forecast error up to five days. Higher degree of coupling results in improvement apart from 1000 hPa temperature in the Northern Hemisphere (NH), where coupled *uptraj0-1* leads to the best scores and coupled *uptraj0-2* slightly is degrading.

In Central Africa there is a prominent strong warm temperature model bias at 850hPa geopotential height in the current IFS cycle. Maps of relative differences in forecast error for temperature at 850hPa (**Figure 9**) for the two configurations coupled *uptraj0-1* and coupled *uptraj0-2* show that increased coupling alleviates the bias leading to significant reduction of forecast error over 3 days that is attributed to substantial cooling in the region.

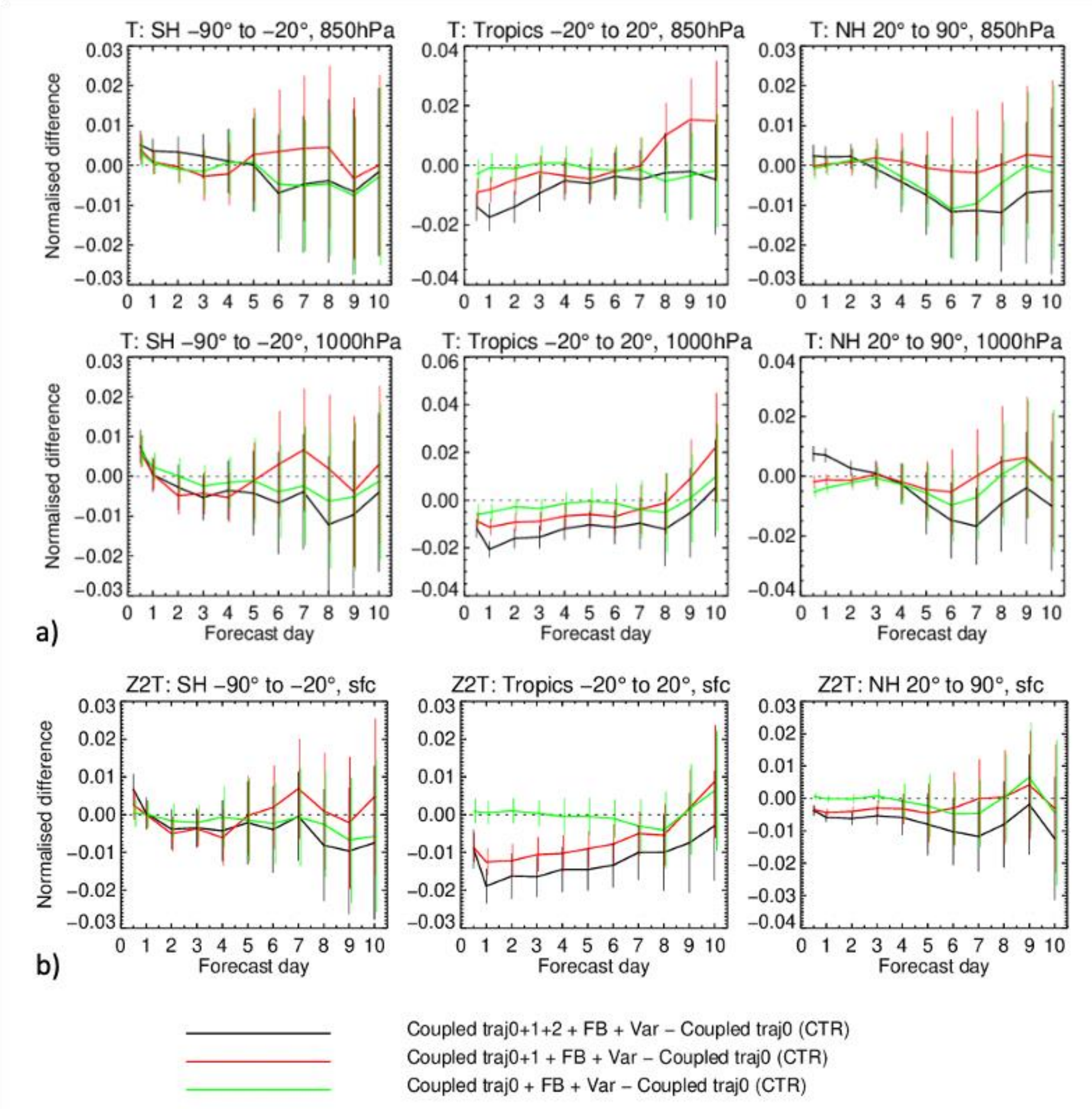


*Figure 8: Relative differences in RMS forecast error verified against operational analysis of a) temperature at 850hPa and 1000hPa, and b) T2m – averaged over the NH, the tropics and the SH along the boreal winter, comparing configurations using 2D-OI-based pseudo-observations in the SEKF outer loop coupled uptraj0 (green line), uptraj0-1 (red line) and uptraj0-2 (black line) to the control with the SEKF activated in uptraj0.*

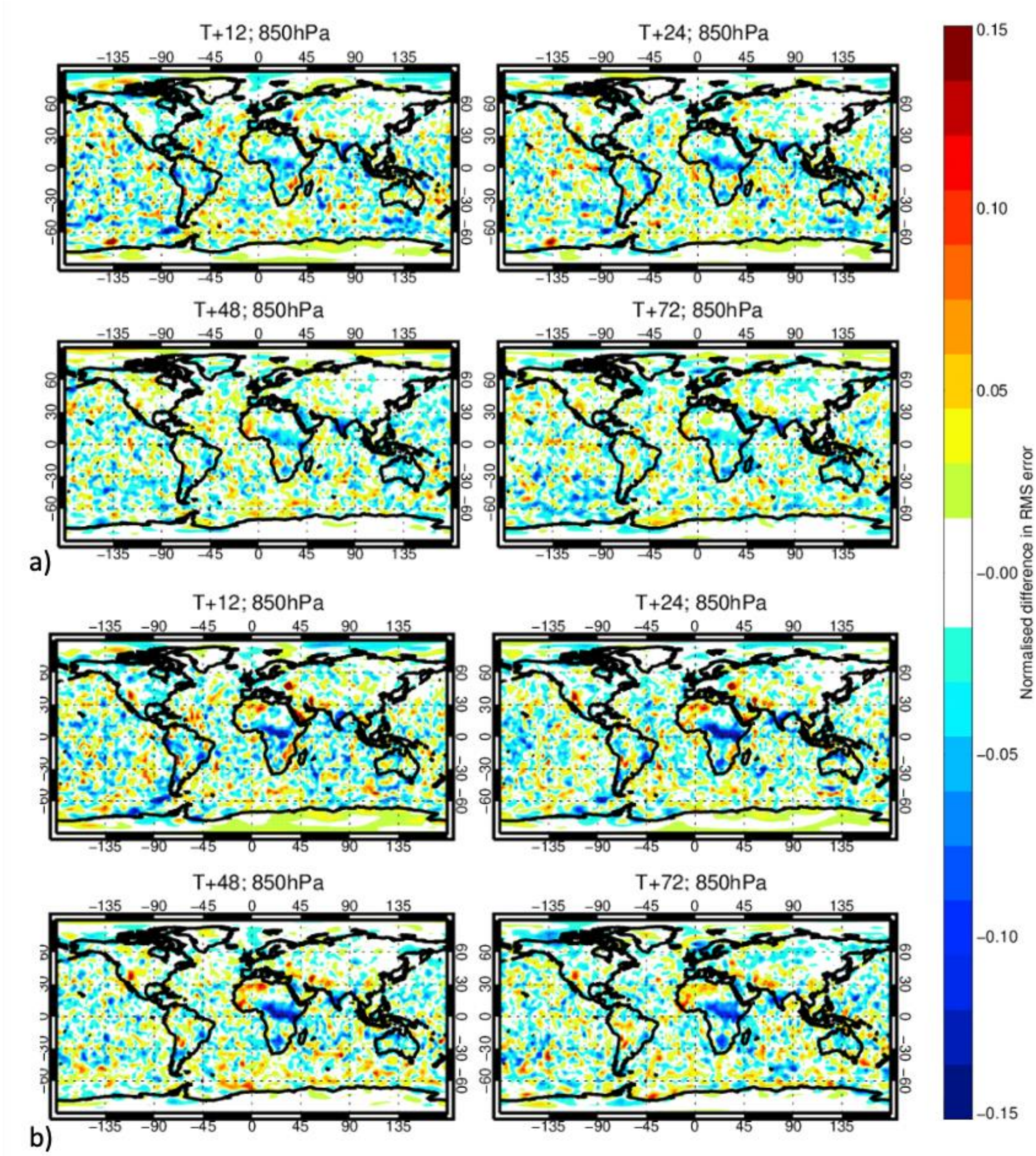


*Figure 9: Relative differences in RMS forecast error for different lead times (12 hour to 3 days) verified against operational analysis of temperature at 850hPa averaged over the NH, the tropics and the SH along the boreal winter, comparing configurations using 2D-OI-based pseudo-observations in the SEKF outer loop coupled a) uptraj0 and b) uptraj0-1, to the control with the SEKF activated in uptraj0.*

**Figure 10** shows the globally averaged relative differences in first-guess departures against observations between the three configurations in comparison to the uncoupled control. There is a significant reduction in first-guess departures regarding in-situ aircraft and radiosonde data at near-surface pressure levels and for screen-level observations (T2m and RH2m, marked in red dashed box). Most of the positive impact is already obtained for coupled *uptraj0-1*, while coupling *uptraj2* adding about a slight improvement. For configuration coupled *uptraj0-1*, the spatial distribution of first-guess departures against screen-level T2m and RH2m observations (corresponding to **Figure 10c**) is shown in **Figure 11**. The departures are smaller overall and well-distributed across the globe, while a noticeable degradation for all variables is limited to the Sahara Desert which is under investigation.

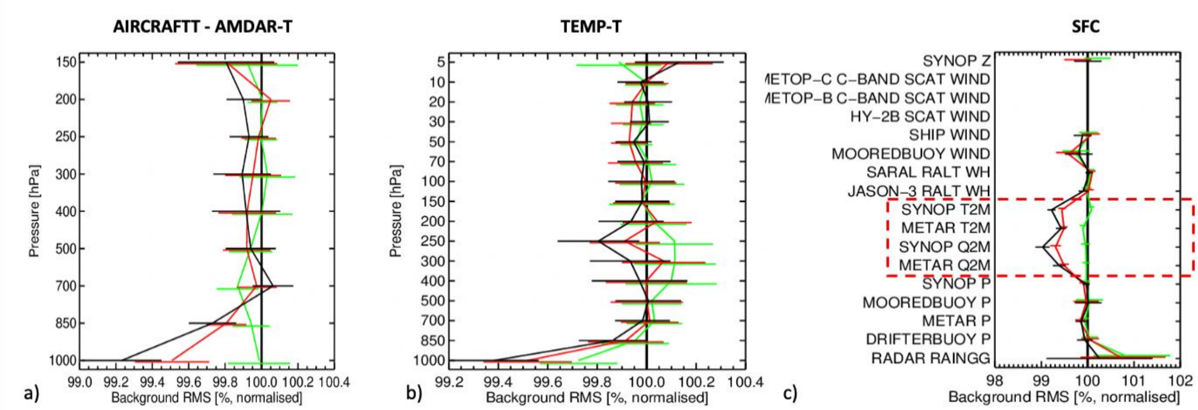


*Figure 10: Globally averaged relative differences in RMS first-guess departures regarding a) aircraft temperature, b) temperature from radiosondes and c) further near-surface-sensitive observations along the boreal winter, comparing configurations using 2D-OI-based pseudo-observations in the SEKF outer loop coupled uptraj0 (green line), uptraj0-1 (red line) and uptraj0-2 (black line) to the control with the SEKF activated in uptraj0.*

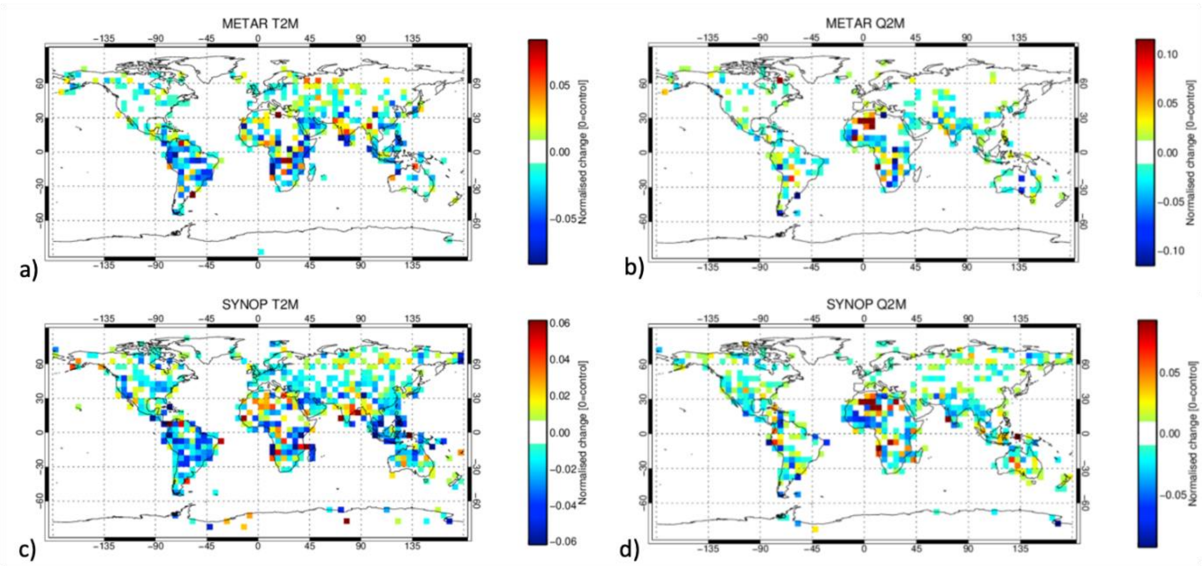


*Figure 11: Relative differences in RMS first-guess departures regarding screen-level observations including a) METAR T2m, b) METAR RH2m, c) SYNOP T2m and d) SYNOP RH2m along the boreal winter, comparing configuration using 2D-OI-based pseudo-observations in the SEKF outer loop coupled uptraj0-1 (spatial maps of the average value of the red line in the red-dashed box in Figure 9c) to the control with the SEKF activated in uptraj0.*

To assess the impact of stronger coupling at the land-atmosphere interface, the forecast results are evaluated in terms of skin temperature (SKT) forecast, which are verified against both the operational analysis and SEVIRI-based land-surface temperatures (LST) retrievals (provided as part of WP7, see D7.2 “Albedo, vegetation and LST satellite datasets in the CERISE verification database”). **Figure 12** shows the relative differences in SKT forecast error for different lead times for the three configurations, verified against ECMWF operational analysis. Especially at short lead times, scores show significant degradations for configurations coupled *uptraj0-1* and *uptraj0-2* while being mostly neutral for coupled *uptraj0.*

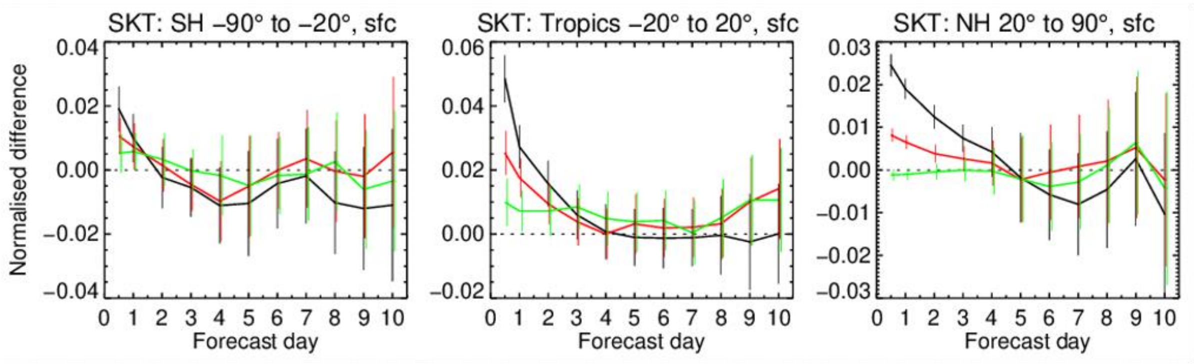


*Figure 12: Relative differences in RMS forecast error of skin temperature for different lead times verified against operational analysis averaged over the NH, the tropics and the SH along the boreal winter, comparing configurations using 2D-OI-based pseudo-observations in the SEKF outer-loop coupled uptraj0 (green line), uptraj0-1 (red line) and uptraj0-2 (black line) to the control with the SEKF activated in uptraj0.*

The spatial distribution of T2m and SKT forecast errors at 12-hour lead time for the three configurations are presented in **Figure 13**. For the smallest degree of coupling with coupled *uptraj0*, the scores over the Sahara Desert are improved for both SKT and T2m. For stronger coupling, as observed similarly as in the verification against screen-level observations, degradations of T2m are most important in the Sahara Desert while for SKT they are much stronger and widespread across Africa and extending to other continents such as South Asia.

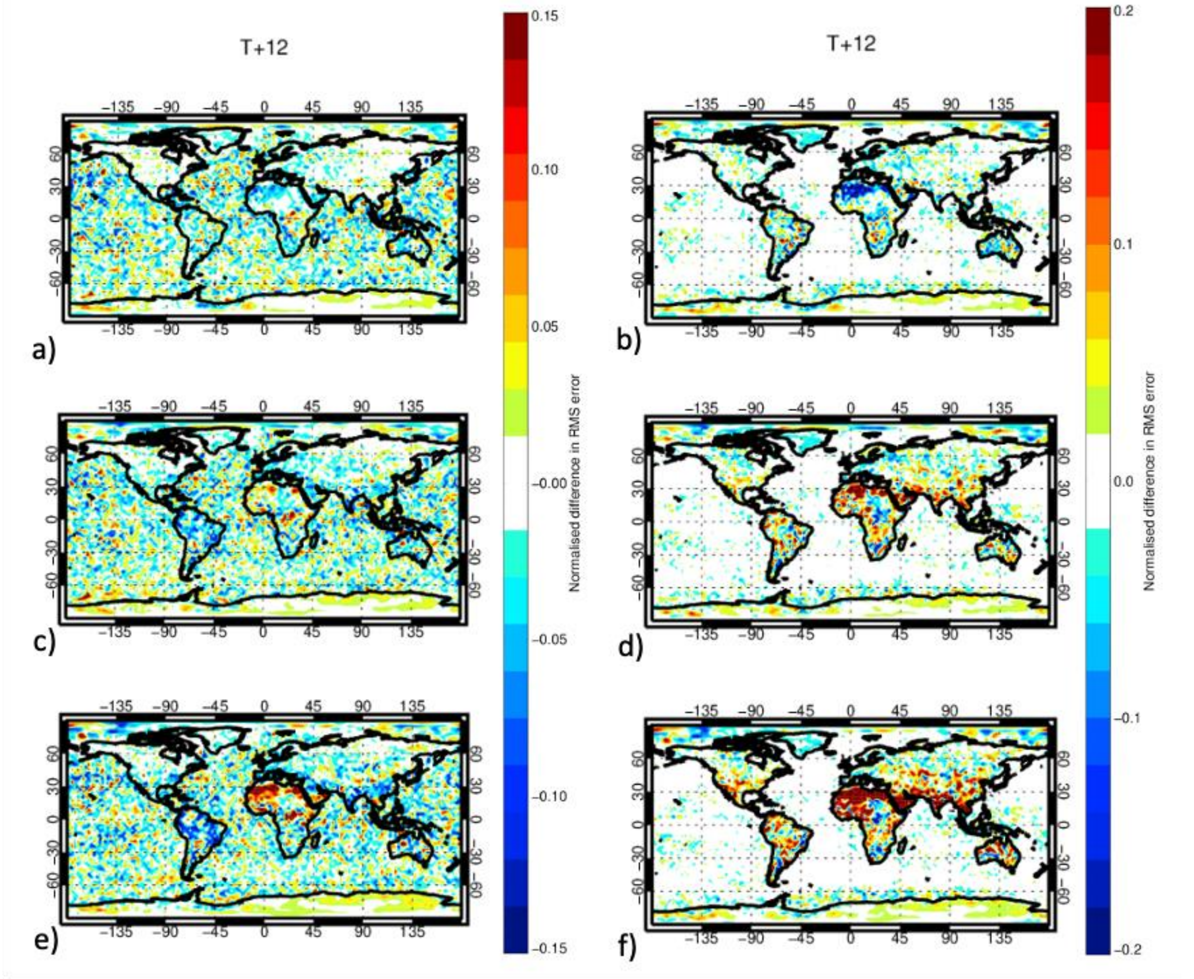


*Figure 13: Relative differences in RMS forecast error for 12 hour forecast of T2m (a, c and e), and skin temperature (b, d and f) along the boreal winter, comparing configurations using 2D-OI-based pseudo-observations in the SEKF with outer loop coupled uptraj0 (a and b), uptraj0-1 (c and d) and uptraj0-2 (e and f) to the control with the SEKF activated in uptraj0.*

**Figure 14** shows the relative differences in SKT forecast error of the three configurations against SEVIRI LSTdisc. The smaller differences between SKT and LST for coupled *uptraj0* over the Sahara in **Figure 14a** are in agreement with the reduced forecast error in **Figure 13b**. Similarly, the larger differences between SKT and LST for configuration coupled *uptraj0-2* in **Figure 14c** can be clearly associated with increased forecast error in **Figure 13f**. This confirms the reduced SKT forecast skill obtained against operational analysis in Figures 12 and 13.

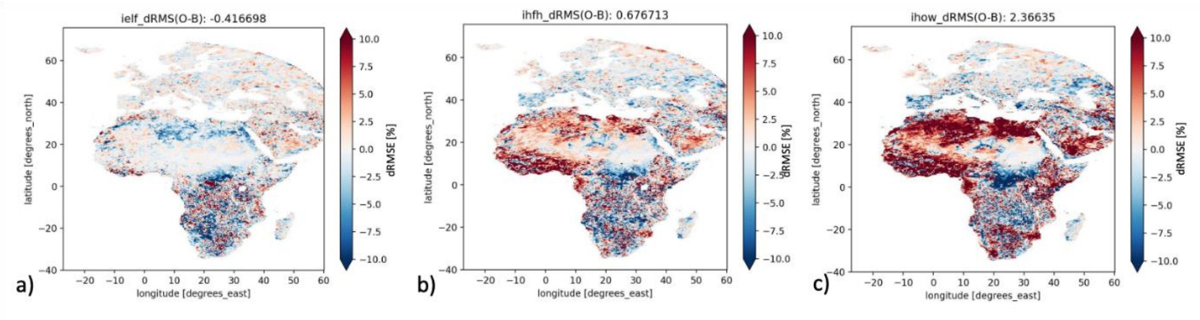


*Figure 14: Relative differences in RMSE between SKT forecast and SEVIRI LST along the boreal winter, comparing configurations using 2D-OI-based pseudo-observations in the SEKF outer loop coupled uptraj0 (a), uptraj0-1 (b) and uptraj0-2 (c) to the control with the SEKF activated in uptraj0.*

The differences between the SKT forecast and the LST are evaluated for the diurnal cycle on the basis of the range of values between 12UTC and 00UTC. Hereby, SKT is considered at the corresponding validity times of the 12-hour forecast. **Figure 15a** and **b** illustrate the diurnal cycle for LST as well as the SKT forecast of the control experiment, respectively, exhibiting high values for the Sahara Desert. Large parts of Southern Africa have low diurnal cycle that corresponds to the relatively noisy areas of RMS differences between SKT and LST. **Figure15c** shows the differences in diurnal cycle between LST and SKT (control), revealing a smaller temperature range by around 10K for the control experiment. The differences between SKT of the three configurations and SKT of the control are shown **Figure 15d-f**. While differences for coupled *uptraj0* are relatively small, for coupled *uptraj0-1* and *uptraj0-2* the diurnal cycle over the Sahara Desert is even more reduced on top of the already reduced temperature range in the control experiment. Thus, apart from the positive impact for atmospheric temperature forecast, too strong coupling can lead to degraded SKT in regions with seasonally strong diurnal cycle, showing increased forecast error accompanied with reduced diurnal cycle compared to LST retrievals from the SEVIRI satellite.

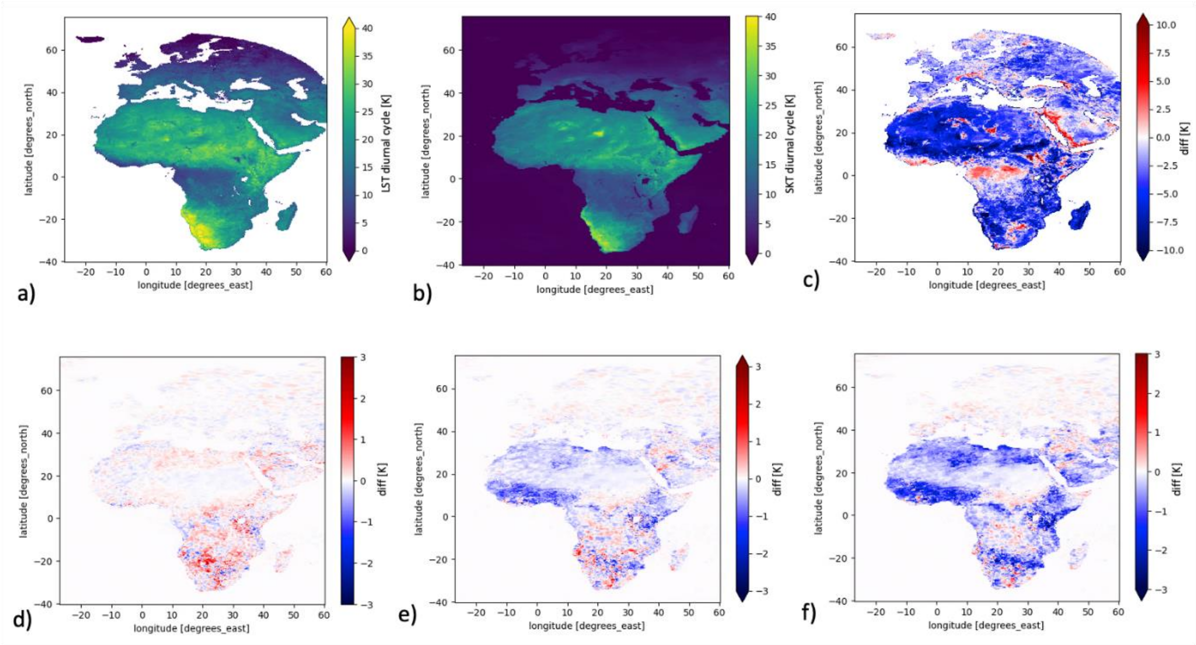


*Figure 15: Evaluation of the SKT diurnal cycle in terms of differences between 12UTC and 00UTC averaged over the boral winter; a) LST from SEVIRI, b) SKT forecast of the control (SEKF activated in uptraj0), c) SKT (control) minus LST, d) SKT (coupled uptraj0) minus SKT (control), e) SKT (coupled uptraj0-1) minus SKT (control) and f) SKT (coupled uptraj0-2) minus SKT (control).*

The reason for the reduced diurnal cycle with stronger coupling could be due to the currently estimated skin conductivity in the coupled model, which determines the heat exchange between soil and skin temperature. Further research needs to be conducted to better understand the impact on land variables and sensible heat flux.

# 5 Summary and Conclusions

This report summarises major infrastructure developments that were conducted in the ECMWF IFS system in preparation of the next generations of the C3S for global reanalysis and seasonal prediction. They include the capability to run the soil analysis within the main 4D-Var trajectory, saving the cost of a trajectory in each 12-hour data assimilation window. The unified LDAS of WP1 was integrated in the coupled infrastructure and the consistency between the offline and coupled land DA systems was enhanced by implementing consistent observation interface in the offline land DA system. The capability to assimilate screen level observations in the offline LDAS was also developed in line with the coupled approach. These developments are key to ensure seamless coupled reanalysis streams initialisation from a single stream land reanalysis, as well as to ensure consistency of reforecast and real time streams of seasonal prediction systems.

Outer loop land-atmosphere coupled data assimilation was developed in the IFS and different coupling configurations were investigated to identify the optimal number of coupled outer loops and strength of coupling. More balanced initial conditions between the different outer loops showed an overall positive impact on atmospheric near surface forecasts and improved fit to independent atmospheric data. Having the capability of running the unified LDAS of WP1 in combination with enhanced land-atmosphere coupling of WP2 allows to constrain land surface variables (soil and snow temperature) through coupled assimilation. However, verification against LST data showed degradations in terms of skin temperature forecast skills related to diurnal cycle representation issues with strongest coupling configurations. This will be further investigated and the possibility of better interface observations assimilation through an Extended Control Variable approach to constrain skin temperature will be explored in WP2 T2.3.

# 6 References

de Rosnay, P., Browne, P., de Boisséson, E., Fairbairn, D., Hirahara, Y., Ochi, K., et al. (2022) Coupled data assimilation at ECMWF: current status, challenges and future developments. Quarterly Journal of the Royal Meteorological Society, 148(747), 2672–2702. Available from: https://doi.org/10.1002/qj.4330

English, S., D. Salmond, M. Chrust, O. Marsden, A. Geer, E. Holm et al., 2017: Progress with running IFS 4D-Var under OOPS, ECMWF Newsletter No. 153, pp. 13–14. https://www.ecmwf.int/en/newsletter/153/news/progress-running-ifs-4d-var-under-oops

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., G. Arduini, G. Balsamo, S. Boussetta, K. Ochi, E. Pinnington, and P. de Rosnay: "Improved two-metre temperature forecasts in the 2024 upgrade", ECMWF Winter 2024 Newsletter, doi: https://www.ecmwf.int/en/newsletter/178/earth-system-science/improved-two-metre-temperature-forecasts-2024-upgrade

Mahfouf, J.-F., K. Bergaoui, C. Draper, F. Bouyssel, F. Taillefer, and L. Taseva (2009), A comparison of two off-line soil analysis schemes for assimilation of screen level observations, J. Geophys. Res., 114, D08105, doi:10.1029/2008JD011077.

Penny S.G., Akella S., Alves O., Bishop C., Buehner M., Chevallier M., Counillon F., Draper C., Frolov S., Fujii Y., Karspeck A., Kumar A., Laloyaux P., Mahfouf J.-F., Martin M., Peña M., de Rosnay P., Subramanian A., Tardif R., Wang Y., Wu X.: Coupled Data Assimilation for Integrated Earth System Analysis and Prediction: Goals, Challenges and Recommendations. World Meteorol. Org. (WMO), WWRP 2017-3, 2017. https://library.wmo.int/doc_num.php?explnum_id=10830

Rabier, F., Järvinen, H., Klinker, E., Mahfouf, J.-F. and Simons, A. (2000) The ECMWF operational implementation of four-dimensional variational assimilation. I: Experimental results with simplified physics. Quarterly Journal of the Royal Meteorological Society, 126(564), 1143–1170. https://doi.org/10.1002/qj.49712656415

# 7 Annex I:

## Document History

| Version | Author(s) | Date | Changes |
|---|---|---|---|
| 1.0 | Patricia de Rosnay | 25 October 2024 | Initial version |
| 1.1 | Patricia de Rosnay, Christoph Herbert, David Fairbairn, Ewan Pinnington, Pete Weston | 25 November 2024 | First full draft |
| 2.0 | Patricia de Rosnay | 11 December 2024 | Issued |
| | | | |
| | | | |

## Internal Review History

| Internal Reviewers | Date | Comments |
|---|---|---|
| Etienne Tourigny (BSC) and Harald Schyberg (Met Norway) | December 2024 | First full draft |
| | | |
| | | |
| | | |
| | | |