REQUEST FOR A SPECIAL PROJECT 2025–2027

MEMBER STATE: Italy

Principal Investigator: Dr Carla Cardinali

Affiliation: CMCC

Address:
Viale C. Berti Pichat 6/2
40127 Bologna, Italy

Other researchers: Giovanni Conti, Gustavo de Gonçalves, Marcelo Guatura, Alvaro Vasconcellos, Andrea Cipollone
Collaborations: Aeronautica Militare e Meteorologia, ESYDA-CMCC Division

Project Title: Towards a land/atmosphere/ocean/ice coupled data assimilation system: 20-year coupled reanalysis with an ensemble Kalman filter to initialise longer time-range forecasts, specifically seasonal forecasts (towards C3S multi-model project)

To make changes to an existing project please submit an amended version of the original form.)

If this is a continuation of an existing project, please state the computer project account assigned previously.	**SP** ………….	
Starting year: (A project can have a duration of up to 3 years, agreed at the beginning of the project.)	2025	
Would you accept support for 1 year only, if necessary?	**YES**	

Computer resources required for project year:	**2025**	**2026**	**2027**
High Performance Computing Facility [SBU]	300M	300M	300M
Accumulated data storage (total archive volume)[2] [GB]	150T	150T	150T

EWC resources required for project year:	**2025**	**2026**	**2027**
Number of vCPUs [#]			
Total memory [GB]			
Storage [GB]			
Number of vGPUs[3] [#]			

Continue overleaf.

Principal Investigator: Dr Carla Cardinali

Project Title: Towards a land/atmosphere/ocean/ice coupled data assimilation system: 20-year coupled reanalysis with an ensemble Kalman filter to initialise longer time-range forecasts, specifically seasonal forecasts (towards C3S multi-model project)

Extended abstract

All Special Project requests should provide an abstract/project description including a scientific plan, a justification of the computer resources requested and the technical characteristics of the code to be used. The completed form should be submitted/uploaded at https://www.ecmwf.int/en/research/special-projects/special-project-application/special-project-request-submission.

Following submission by the relevant Member State the Special Project requests will be published on the ECMWF website and evaluated by ECMWF and its Scientific Advisory Committee. The requests are evaluated based on their scientific and technical quality, and the justification of the resources requested. Previous Special Project reports and the use of ECMWF software and data infrastructure will also be considered in the evaluation process.

Requests exceeding 10,000,000 SBU should be more detailed (3-5 pages).

Introduction

Over the last 20 years, the field of prediction has seen significant advancements. Using observations and theoretical frameworks, models have developed and they are capable of forecasting various weather and climate phenomena with remarkable accuracy. These models range from those that predict the evolution of weather systems and local extreme events, with lead times from hours to days, to those that offer global weather predictions up to weeks in advance.

All these predictive models are grounded in and are continually refined by observations of the natural world, ensuring that they remain as accurate and relevant as possible. This continuous feedback loop between observation and modelling is crucial for improving the reliability of different-range forecasts and projections (forecasting future conditions based on hypothetical scenarios), helping the society to better prepare for and respond to environmental changes.

The observations play a critical role in ensuring that the initial conditions for these predictions accurately reflect the current state of the Earth system. They are also essential for the construction, calibration, and validation of the prediction models.

A key component of this process is the initialisation of predictions through data assimilation (DA). DA involves the incorporation of hundreds of millions of observations from around the world daily, achieving a physically consistent and precise starting point for forecasts. This advanced technique enhances the accuracy of weather predictions by continuously updating the model state with new observational data. Weather extremes are more and more necessary to be predicted reliably. To obtain a sufficient level of detail, the integration of models and observations is essential, embedding a multilayered fusion, which characterises the varied levels of information across multiple weather scales.

The ambition for Earth system prediction and forecasting is expanding, exemplified by the concept of a digital twin of the Earth (Bauer et al. 2021). A digital twin would be a highly detailed and dynamic representation of the Earth's systems, created through an advanced blending of models and observational data, a process that could be referred to as model-data fusion (Gettelman et al., 2022). Digital twins also introduce new levels of interactivity fir users with data and information. This approach aims to enhance the precision and utility of predictions, ultimately supporting more informed and effective responses to environmental challenges.

The future focus of prediction needs to extend beyond the traditional surface weather parameters such as temperature, wind speed, rain rate, and runoff, and will encompass the full components of the Earth system, including land surfaces and atmospheric chemistry, but also ocean and cryosphere. This holistic approach is essential for capturing the complex interactions and feedback mechanisms that drive Earth's climate and environmental conditions.

The integration of land surface processes and atmospheric chemistry is crucial for understanding and predicting phenomena such as air quality, carbon cycling, and ecosystem dynamics. Land models will include vegetation growth, soil processes, and land use changes, while atmospheric chemistry models will represent pollutants, greenhouse gases, and other chemical constituents.

The ocean plays a vital role for regulating climate by absorbing heat and carbon dioxide. Ocean models will simulate currents, temperature, salinity, and biogeochemical processes. These models help predict phenomena like El Niño and La Niña, ocean acidification, and the impacts of climate change on marine ecosystems and coastal communities. The cryosphere, which includes glaciers, ice sheets, sea ice, and permafrost, significantly influences global climate and sea level rise. Models will account for ice dynamics, melting rates, and interactions with the atmosphere and ocean. Accurate cryosphere modelling is essential for predicting sea-level change, freshwater availability, and feedback effects on global temperatures.

By integrating these components, prediction models will provide a comprehensive understanding of the Earth system. This integration allows for better predictions of weather, climate, and environmental changes, enabling more effective management and mitigation strategies. The inclusion of land, chemistry, ocean, and cryosphere models ensures that we consider all critical factors influencing our planet, leading to more accurate and actionable forecasts.

Improvements in DA techniques have significantly contributed to the advancements in numerical weather prediction (NWP) over recent decades. These techniques have enabled the extraction of much more information from the same set of observations, enhancing the accuracy and reliability of the predictions. For the initialization of the Earth system components, a diverse range of techniques is employed, which include optimal interpolation, (simplified) ensemble Kalman filters, and variational methods. Operational methods for land, ocean, and ice components tend to be less complex than those used for the atmosphere. However, aligning these methods across different Earth system components could yield significant benefits. Converging on a unified approach across Earth-system components would enhance the integration and consistency of predictions across the entire Earth system. For extended predictions, such as sub-seasonal to seasonal and initialized decadal forecasts, improvements in predictability are largely expected to come from the slower-varying components of the Earth system: land, ice, and ocean. These components have longer memory and influence over extended periods, making them crucial for accurate long-term predictions. Consequently, there is a strong motivation to advance assimilation methods for these components to improve the overall predictive capability.

Several active EU projects such as CERISE and CONCERTO are addressing this matter by including coupled data assimilation developments, primarily focusing on the land and atmosphere coupled system. These projects represent significant progress, but they need to be expanded to also include the crucial ocean component, which is vital for a comprehensive understanding of the Earth system. The ocean's ability to store and release heat over long periods affects the system’s memory and therefore drives weather patterns from seasons to years. Also changes in salinity, which impacts ocean circulation and ocean currents that distribute heat and nutrients around the globe provides important information for seasonal prediction and ultimately for the global climate system.

By addressing these gaps and expanding the focus to also include the ocean component in the initialisation process, the scientific community can significantly improve the accuracy and reliability of Earth system predictions, providing more robust tools for managing and mitigating the impacts of environmental changes.

DA and coupled DA activities at CMCC

In 2020 the Euro-Mediterranean Center on Climate Change (CMCC, Italy) approved a new strategy on longer forecast range predictions (e.g. the seasonal forecast) strongly supporting the use of a proper initialisation of these predictions by a weakly coupled data assimilation system (WCDA). The CMCC strategy foresees therefore the development of a coupled atmosphere, land , ocean and cryosphere data assimilation system initialising seasonal predictions.

Open-source modelling plays an important role for creating transparency and reproducibility of scientific results, fostering collaboration and community engagement that encourages knowledge sharing, exchange of ideas, and collective problem-solving, leading to the development of more robust and comprehensive models. CMCC has therefore also decided to make the development of WCDA an open-source data assimilation system for operational use.

Starting with the open-source DART (NCAR, https://dart.ucar.edu/) system, **SPREADS** (Scalable PaRallelised EArth Data assimilation System; Cardinali et al. paper to be submitted) has been developed. It thrived by modifying and implementing new features essential for its operational use. CMCC's development of SPREADS as an open-source data assimilation system builds upon the strengths of the DART framework while customizing and extending it to meet the specific requirements of operational forecasting and decision support.

Atmospheric DA

The Ensemble Adjustment Kalman Filter (EAKF) is a DA technique developed by Jeff Anderson and his colleagues (Anderson 2001; Anderson 2003; Anderson and Collins 2007; Reader et al. 2012) in DART. It is a variant of the traditional Ensemble Kalman Filter (EnKF) designed to address some of the limitations of the standard EnKF, particularly in situations where the ensemble size is small or where the model and observation error statistics are poorly known. The EAKF adjusts the ensemble state using a least squares solver approach, aiming to find the best estimate of the true state that is consistent with both the model dynamics and the available observations. Unlike the traditional EnKF, which directly updates the ensemble members based on their covariances with observations, the EAKF adjusts the ensemble mean and spread to better fit the observations.

Observations and ensemble members are assimilated within localized regions, which helps to mitigate the impact of spurious long-range correlations and reduces computational costs. A least squares approach can be used within each local region to estimate the analysis increments based on observations and ensemble forecasts. The EAKF uses a different algorithm for updating the ensemble once the observations become available. One of the advantages of the EAKF is its robustness to uncertainties in model dynamics and observation error statistics. By incorporating a least squares adjustment step, the EAKF can effectively assimilate observations even when the error characteristics of model and observations are poorly known or inadequately represented. Very importantly, the EAKF is designed to perform effectively with relatively small ensemble sizes.

One of the first upgrades developed in SPREADS is the FGAT (First Guess at Appropriate Time) approach to improve the efficiency and accuracy of the assimilation process. In FGAT, the model's first guess (also known as the background or forecast) is adjusted based on observations to produce an analysis of the current state of the system matching the time of observations with model time step. Observational data are interpolated to the time of the model's first guess, which in SPREADS has been designed splitting the commonly used 6-hour window into 11 time slots of 30’ and 2 time slots (at the beginning and the end of the window) of 15’ each. This ensures that the observations and model output are aligned in time, allowing for a meaningful comparison between them. Several advantages are obtained with this approach: a) by adjusting the model's first guess directly to the observation time, the FGAT approach reduces the need for interpolating observations to the model's time grid, leading to computational efficiency. b) It ensures consistency between the model's initial conditions and observational data, improving the accuracy of the assimilation process, and, c) it can be implemented with various data assimilation techniques, allowing for flexibility in adapting to different modelling and observational setups. For reference, during the development of the variational data assimilation system at ECMWF, the largest assimilation improvement was obtained when the system moved from 6-hour window 3DVar to FGAT 3DVar (Andersson et al. 1998).

To ensure a flexible and fast observation handling throughout the entire assimilation process, a query language observation database (Database for Observations, D4O) has been developed. D4O manages and controls the observational data flow through detailed definitions of the observations and a query language that is organised in a hierarchical tree-like structure from which is easy to select the desired information and place it in a data matrix for further examination. This relational-like process is particularly efficient for MPI-parallel data access and query-coordination for data shuffling between MPI-tasks. High vectorization efficiency for storing and retrieving observations is therefore achieved, enabling a fast, flexible and configurable I/O management. D4O can efficiently handle large amounts of satellite observations necessary to produce a the state of art analysis.

SPREADS is a modular and parallelised (full parallelisation of the system is a very recent CMCC Earth System Data Assimilation Division achievement) system assimilating 15 million atmospheric observations per 6-hour window, namely satellite infrared (IR) and microwave (MW) brightness temperatures (BT), Global Positioning System - Radio Occultations (GPS-RO), Atmospheric Motion Vectors (AMV), scatterometer and conventional observations (wind profilers, radiosondes, aircraft land and marine surface pressure). Methods for observational bias correction and cloud detection are being developed using machine learning techniques. At the moment, the CMCC atmospheric reanalysis is under production.

SPREADS is fully integrated in the Community Earth System Model (CESM; https://www.cesm.ucar.edu/) infrastructure developed in a joint collaboration between CMCC and NCAR. CESM is a software framework for setting up and running a combination of models, each designed to represent a part of the Earth system. In particular, SPREADS is coupled with the atmosphere model CAM, the land model CLM, the cryosphere model CICE and the ocean model NEMO.

Atmosphere and Land Weakly Coupled DA

In the framework of the CERISE EU project, CMCC is engaged in the development of a weakly coupled land-atmosphere assimilation system using SPREADS, CAM and CLM. The first prototype of the system is going to be evaluated by the end of 2024. Land observations being assimilated are Leaf Area Index (LAI), Soil Moisture (SM) and Snow Cover Fraction and Snow Water Equivalent (SWE), all satellite derived products. The coupling between land and atmosphere is obtained throughout the CLM and CAM models by updating the combined fluxes between the two components. In particular, observations native to each Earth system component only directly impact the state vector for that component. Information is passed between components indirectly through the short-term coupled model forecasts that provide the EAKF background ensemble. In this configuration, CLM and CAM couple every 30'. The SPREADS assimilation updates the CAM state with atmospheric observations every 6 hours using the FGAT approach. Snow and SM land observations are assimilated using a separate SPREADS system every 24 hours (at 00 UTC), whilst LAI is assimilated every 8 days (at 00 UTC).

Scientific Plan

Atmosphere Land and Ocean Weakly Coupled DA

Motivation

Uncertainties in the ocean analysis state estimates result from accumulated errors from all system components (ocean model, boundary condition forcing, observations and DA method). Part of the discrepancy in the atmospheric (uncoupled) analyses is due to the treatment of the prescribed boundary conditions (e.g., sea ice and surface temperature). Also, both coupled climate models and ocean-ice models, driven by prescribed atmospheric forcing, cannot adequately represent, for example, the observed polar trends whereas ocean analyses (and reanalyses) have proven quite adequate to capture these trends when observations are available to constrain the system.

Current state-of-the-art coupled forecasting systems face challenges in analysing interface states in a self-consistent manner:

- Sea Surface Temperature (SST) retrieval products: many atmospheric and oceanic DA systems nudge towards SST retrieval products (Storto and Masina, 2016). However, empirical methods converting satellite-measured radiances into SST retrievals contain significant caveats. Errors in model calibration, particularly at high latitudes, and the use of skin SST estimates to constrain bulk temperature present substantial issues.
- Diurnal SST variations: processes like diurnal variations of SST and near-surface cooling in the microlayer are well observed but not well represented in coupled atmosphere-ocean general circulation models. Reproducing SST variability accurately remains a challenge.

Coupled Earth system models offer the potential for forecasts across multiple timescales: at prediction timescales, the diurnal cycle impacts coupled processes in the atmosphere and ocean boundary layers. Mesoscale interactions between SST, fronts and near-surface winds have broader implications for atmospheric dynamics throughout the troposphere.

Satellite-based measurements of SST are inherently coupled due to influences from not only the sea surface but also the full atmospheric column above it. The derived SST is highly influenced by both atmosphere and ocean boundary layers as well as the strength of upward longwave radiation and turbulent heat flux exchanges. To avoid dealing with the complex calibration issues associated with satellite radiances, current prototype WCDA systems typically rely on SST data products produced by specialists and assimilate gridded SST such as OSTIA from UK Met Office.

A recommendation from an ECMWF workshop (Balmaseda et al., 2018) has been to directly assimilate satellite radiances to constrain SST and sea ice, just as done in data assimilation for atmospheric quantities. WCDA therefore offers a great opportunity to treat the interfaces within the coupled model in a more robust and consistent manner.

The integrated ocean observing system is advancing rapidly with the growing constellation of satellite observing platforms. These platforms provide a comprehensive view of the ocean surface. Sparse in situ measurements and largely unobserved small-scale processes critical to air-sea interactions and the deep ocean are notable challenges. Recent and upcoming satellite missions promise increasingly high-resolution datasets of the ocean surface. Instrumentation progress, starting with the Advanced Very High Resolution Radiometer (AVHRR), the Visible Infrared Imaging Radiometer Suite (VIIRS) and towards the Sea and Land Surface Temperature Radiometer (SLSTR) on board Sentinel-3 (A and B), along with missions like the Surface Water and Ocean Topography (SWOT), are expected to provide large volumes of enhanced data suitable for assimilation into ocean models.

Also, information on the ocean surface can be extracted directly from a large network of satellite radiance measurements already present in the atmospheric data assimilation system. For example, infrared sensors such as the Infrared Atmospheric Sounding Interferometer (IASI) and the Cross-track Infrared Sounder (CrIS) have many channels, which are sensitive to radiation emitted from the top few microns of the ocean surface. These spaceborne IR can detect changes in the surface skin temperature of the ocean, which can differ from the bulk SST of the water just below.

The air-sea interface is a key focus area for WCDA explorations. One approach is to use radiative transfer models to assimilate observed brightness temperature directly, instead of relying on SST retrieval products. This new capability enables a more accurate and consistent analysis of interface states such as SST, sea surface salinity (SSS) and sea ice.

Recently, Akella et al. (2017), have modified the skin temperature by modelling and assimilating near sea surface sensitive satellite IR observations. The atmosphere-ocean interface layer of their atmospheric general circulation model is updated to include near surface diurnal warming and cool-skin effects and the analysis system also directly assimilate SST-relevant AVHRR observations. The capability to assimilate satellite radiances in coupled systems has improved the predictability of NASA's Global Modeling and Assimilation Office (GMAO) system, most notably near the surface.

Moreover, the monitoring of SSS is crucial for understanding and forecasting the ocean circulation, water cycle and the changing of the climate. Recent satellite missions like ESA's Soil Moisture Ocean Salinity (SMOS), NASA's Aquarius and Soil Moisture Active Passive (SMAP) missions have made it possible for the first time to measure SSS from space.

The SPREADS assimilation system is also coupled with the ocean model NEMO (using the CESM coupler), therefore offering a great possibility to analyse atmosphere, land and ocean all together. In particular, for the ocean the assimilation of temperature and salinity profiles from of in situ observations and SST and SSS from satellite radiances observations is proposed.

The high modularity of SPREADS, the FGAT approach and the use of RTTOV enable the direct assimilation of radiances also for the ocean component. As for the atmosphere, the ocean assimilation is then performed every 6 hours whilst the NEMO ocean model is coupled with the CAM atmospheric model every 90 minutes.

It is important to notice that the time dependent background error covariance estimates are a fundamentally beneficial ingredient of the analysis configuration proposed (Penny et al., 2015; Penny, 2017). The full introduction of flow-dependent background errors involves estimating the ocean background error covariances from the ensemble, accounting therefore for anisotropic and inhomogeneous correlations that are difficult to estimate with other analysis methods.

Experimentation plan

This project prepares a WCDA atmosphere-land-ocean prototype based on the EnKF methodology that promises substantial advantages over other methods as outlined above. It is based on a well established research system at NCAR that is extended to a fully coupled system with an entirely revised observational data handling focusing on computational efficiency and scalability to large observational data volumes and instrument diversity.

The ultimate goal is to produce a 20-year coupled reanalysis that will allow sufficient sampling of seasonal weather system and teleconnection variability to perform both scientific analyses of forecasts and performance assessment of the WCDA while maintaining a fairly stable observational data record throughout the period.

<u>Since the available computing resources for Special Projects may not be sufficient to complete the entire reanalysis production, we propose to invest the first two years in deriving the ultimate production configuration, and then aim to produce as much of the prototype reanalysis in years two and three. The completion of the entire 20-year reanalysis can then be carried out according to further resource availability.</u>

The configuration of the proposed WCDA prototype is summarised in the table below.

	Atmosphere assimilation	**Ocean assimilation**	**Land assimilation**
DA method	SPREADS-CAM 6H window	SPREADS-NEMO 6H window	SPREADS-CLM 24H window
Members	80	80	80

	Atmosphere assimilation	Ocean assimilation	Land assimilation
Resolution	0.5°	0.5°	0.5°
Covariance inflation	Adaptive multiplicative	Adaptive multiplicative	Adaptive multiplicative
Localization	yes	yes	yes
CESM model forecast frequency		90’	30’
In situ observations	Land&Marine surface PS, Wind profiler, Radiosondes, Aircrafts	Temperature and salinity profilers	
Satellite observations	AMV, Satterometer, GPS-RO, Amsu-A, Airs, IASI	SST and SSS	
Satellite products			LAI, Soil moisture, Snow dept and SWE

The project is divided into three stages:

First year

Scientific consolidation and optimisation of the WCDA infrastructure:

- **Observations:** Observations quality control, observations bias correction and cloud detection (for IR radiances) procedure will be assessed and eventually revised. In the atmospheric analysis, these last two processes are using machine learning methodology whereas training is performed by using only different and complementary satellite observations.
- **Parametrisations:** The introduction and use of the skin SST in the control vector may induce changes in air-sea interaction parametrisation to better represent mesoscale interactions between SST and fronts, near-surface wind and the ocean bulk temperature (cool skin effect). Similarly, parametrisations for unresolved turbulent processes, subsurface flow fields generating a range of dynamical behaviours may need to be retuned to respect global momentum conservation. Tuning of atmospheric boundary layer parametrisation could be required. Inclusion of altimeter observations may be necessary as well (coupled wave model).
- **WCDA**. Adding the ice component to the ocean is a strong possibility depending on the scientific assessment of the first year results (SPREADS-CICE already has a working configuration).

Second year

Further consolidation of the WCAD prototype in preparation for the ultimate reanalysis configuration. Focus of the second year will be on the establishment of the final configuration of observation datasets to be assimilated in the reanalysis. If possible, the reanalysis will already be launched in year-2 to relax the overall production schedule and allow for iterations in case of unforeseen scientific or technical issues.

Third year

The third year will be fully dedicated to the production of the 20-year coupled reanalysis, the assessment of observational data monitoring, analysis increments in the coupled system, and forecast performance evaluation compared to existing reanalyses.

Computational cost and data storage estimate

It is assumed that the first year will comprise shorter-term experiments, while years two and three invest more computing time on longer reanalysis-type experiments. Apart from shorter-period testing, sensible experimentation periods should be several months to represent sufficient variability. Whether configurations with the full ensemble or with fewer members can be trialled remains to be seen during the project.

The cost estimates are based on:

- The present performance of a single ensemble member of the atmospheric component of WCAD on CMCC's Juno and ECMWF's Atos machines. The CMCC Juno CPU - ECMWF Atos CPU equivalence for time to solution (t0) is about: 2 Juno nodes with 36 cores ~ 1.2 Atos node with 128 cores. For running an 80-member ensemble, 80 Atos nodes would be desirable, but 160 nodes would be ideal.
- The fact that the atmospheric analysis is by far more costly than the coupled analysis in the EnKF configuration, so that the trials with the atmospheric analysis on CMCC's Juno system represent a realistic estimate of total expected cost.

The SBU calculation follows the prescribed formula: SBU = P * N * T with P = 424,074,479 / (8090 * 128 * 86,400), N = number of cores and T = time in seconds.

The WCDA analysis consists of three phases, namely initial forecast, interpolation to observation locations, and regression. From using the above formula and our trials we computed per analysis cycle: 80-ensemble member forecast task using 80 Atos nodes (1 node/member) = 60,000 SBU + interpolation task = 7,000 SBU + regression task = 95,000 SBU. The 80-member forecast task cost could be reduced to 40,000 SBU if allocated on 160 nodes.

The total cost per cycle is therefore 162,000 SBU. 4 cycles per day result in 650,000 SBU, 1 month (1 year) of analyses is 19.5 million (235 million) SBU. The entire 20-year reanalysis would be 4.7 billion SBU.

As described above, the reanalysis production is already expected to start in year-2, so that our expectation is to complete at least 2-3 years of the reanalysis in years 2 and 3, which amounts to 470-700 million SBU. The preparatory experimentation comprises at east one year worth of analyses. We also accounted for some overhead for experiment reruns in case of scientific or technical issues. This leads to the overall 300 million SBU/year estimate.

The data storage estimate is based on the present WCDA data output per cycle: 190 Gbyte = 800 Gbyte/day. Our estimate is based on the assumption that we can retain several months worth of data in the active archive without the need for permanent storage. 6 months of output storage results in about 150 Tbyte of active storage per year.

References

Akella, S., Todling, R. and Suarez, M. (2017), Assimilation for skin SST in the NASA GEOS atmospheric data assimilation system. Quart. J. Roy. Meteor. Soc., 143, 1032-1046. Available at https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.2988

Andersson, E., J. Haseler, P. Undén, P. Courtier, G. Kelly, D. Vasiljevic, C. Brankovic, C. Cardinali, C. Gaffard, A. Hollingsworth, C. Jakob, P.A.E.M. Janssen, E. Klinker, A. Lanzinger, M.J. Miller, F. Rabier, A.Simmons, B. Strauss, J-N Thépaut, and P. Viterbo, 1998: The ECMWF implementation of three-dimensional variational assimilation (3D-Var). Part III: Experimental results. Quart. J. Roy. Meteor. Soc., 124, 1831-1860

Anderson J. L., 2001: An Ensemble Adjustment Kalman Filter for data assimilation. Mon. Wea. Rev., 129: 2884–2903. doi:10.1175/1520-0493(2001)129<2884:AEAKFF>2.0.CO;2

Anderson, J. L., 2003: A local least squares framework for ensemble filtering. Mon. Wea. Rev., 131 , 634–642.

Anderson, J.L. and Collins, N. (2007) Scalable implementations of ensemble filter algorithms for data assimilation. Journal of Atmospheric and Oceanic Technology, 24, 1452–1463

Bauer P., Stevens B., Hazeleger W., A digital twin of earth for the green transition. Nat. Clim. Chang. 11, 80–83 (2021)

Gettelman A, Geer AJ, Forbes RM, Carmichael GR, Feingold G, Posselt DJ, Stephens GL, van den Heever SC, Varble AC, Zuidema P. The future of Earth system prediction: Advances in model-data fusion. Sci Adv. 2022 Apr 8;8(14):eabn3488. doi: 10.1126/sciadv.abn3488. Epub 2022 Apr 6. PMID: 35385304; PMCID: PMC8985915

Penny, S. G., D. Behringer, J. Carton, and E. Kalnay, 2015: A hybrid global ocean data assimilation system at NCEP. Mon. Wea. Rev., 143, 4660–4677, https://doi.org/10.1175/MWR-D-14-00376.1

Penny S. G., Hamill T. M., Coupled data assimilation for integrated earth system analysis and prediction. Bull. Am. Meteorol. Soc. 98, ES169–ES172 (2017)

Raeder, K., Anderson, J.L., Collins, N., Hoar, T.J., Kay, J.E., Lauritzen, P.H. and Pincus, R. (2012) DART/CAM: an ensemble data assimilation system for CESM atmospheric models. Journal of Climate, 25, 6304–6317. https://doi. org/10.1175/JCLI- - D- - 11- - 00395.1

Storto, A. and Masina, S.: C-GLORSv5: an improved multipurpose global ocean eddy-permitting physical reanalysis, Earth Syst. Sci. Data, 8, 679–696, https://doi.org/10.5194/essd-8-679-2016, 2016.